Chapter 12

Physiological Monitoring with Factorial Switching Linear Dynamical Systems

John A. Quinn[1] *and Christopher K.I. Williams*[2]

12.1 Introduction

A common way to handle nonlinearity in complex time series data is to try splitting the data up into a number of simpler segments. Sometimes we have domain knowledge to support this piecewise modelling approach, for example in condition monitoring applications. In such problems, the evolution of some observed data is governed by a number of hidden factors that switch between different modes of operation. In real-world data, e.g. from medicine, robotic control or finance, we might be interested in factors which represent pathologies, mechanical failure modes, or economic conditions respectively. Given just the monitoring data, we are interested in recovering the state of the factors that gave rise to it.

A good model for this type of problem is the Switching Linear Dynamical System (SLDS), which has been discussed in previous chapters. A latent "switch" variable in this type of model selects between different linear-Gaussian state spaces. In this chapter we consider a generalisation, the Factorial Switching Linear Dynamical System (FSLDS), where instead of a single switch setting there are multiple discrete factors that collectively determine the dynamics. In practice there may be a very large number of possible factors, and we may only have explicit knowledge of commonly occurring ones.

We illustrate how the FSLDS can be used in the physiological monitoring of premature babies in intensive care. This application is a useful introduction because it has complex observed data, a diverse range of factors affecting the observations, and the challenge of many "unknown" factors. It also provides an opportunity to demonstrate the ways in which domain knowledge can be incorporated into the FSLDS model. Many of the specific modelling details here are also directly applicable to physiological monitoring of adults, in intensive care and other settings.

Observations and factors

Babies born three or four months prematurely are kept, in their first days or weeks *post partum*, in a closely regulated environment, with a number of probes continuously collecting physiological data such as heart rate, blood pressure, temperature

[1]Department of Computer Science, Faculty of Computing and IT, Makerere University, Uganda.
[2]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, United Kingdom.

Channel name	Label
Core body temperature (°C)	Core temp.
Diastolic blood pressure (mmHg)	Dia. Bp
Heart rate (bpm)	HR
Peripheral body temperature (°C)	Periph. temp.
Saturation of oxygen in pulse (%)	SpO_2
Systolic blood pressure (mmHg)	Sys. Bp
Transcutaneous partial pressure of CO_2 (kPa)	$TcPCO_2$
Transcutaneous partial pressure of O_2 (kPa)	$TcPO_2$
Incubator temperature (°C)	Incu Temp.
Incubator humidity (%)	Incu Humidity.

Table 12.1: Physiological (upper) and environmental (lower) measurement channels in this application, with labels used to denote them later in the chapter.

and concentrations of gases in the blood. These vital signs (literally "signs of life") are used in neonatal intensive care to help diagnose the condition of a baby in a critical state. The state of health of a baby cannot be observed directly, but different states of health are associated with particular patterns of measurements. Given observations of the heart rate, body temperature and so on, inferences can therefore be made about the operation of the underlying physiological systems – e.g. whether they are functioning normally, or whether there seems to be evidence of some pathology. This task is complicated by the fact that the observations depend not just on the state of a baby's physiology but also on the operation of the monitoring equipment. There is observation noise due to inaccuracies in the probes, and some operations can cause the measurements to become corrupted with artifact. The specific data channels we consider here are listed in Table 13.1, each sampled once per second.

Types of known, common factors which affect these observations fall into two categories: artifactual and physiological. The known factors we concentrate on here are as follows:

- *Bradycardia* – temporary decrease in heart rate (physiological),
- *Blood sample* – artifactual rise in systolic and diastolic blood pressure measurements while a sample is taken (artifactual),
- *Temperature probe disconnection* – the core temperature probe cools to ambient temperature (artifactual),
- *Handling* – opening of the incubator, leading to a drop in incubator humidity with increased physiological variation (artifactual).

In addition to these common factors there are many examples of physiological variation due to rare factors, or for which no explanation is available.

Outline of the chapter

We describe the model in section 13.2 and compare it to other models for condition monitoring in section 13.2.1. In section 13.3 we describe how to handle novel dynamics, i.e. the presence of unknown factors. In section 13.4 we describe parameter estimation of the model, showing how domain knowledge can be incorporated. Inference is discussed in section 13.5 and we demonstrate the operation of the system

in section 13.6. The description in this chapter gives an overview of the application and model; for more specific implementation details, see Quinn, Williams and McIntosh (2009) and [8]. Demonstration code is also available [9].

12.2 Model

We first recap on the SLDS before generalising to the factorial case. In such models the observations $y_{1:T}, y_t \in \mathbb{R}^{d_y}$, are generated by hidden dynamics $x_{1:T}, x_t \in \mathbb{R}^{d_x}$, according to:

$$x_t \sim \mathcal{N}\left(A^{(s_t)} x_{t-1} + d^{(s_t)}, Q^{(s_t)}\right), \quad y_t \sim \mathcal{N}\left(C^{(s_t)} x_t, R^{(s_t)}\right) \tag{12.1}$$

where $s_t \in 1, \ldots, K$ is a discrete variable defining which of a set of K different dynamics is active at time t. Here $A^{(s_t)}$ is a square system matrix, $d^{(s_t)}$ is a drift vector, $C^{(s_t)}$ is the state-observations matrix, and $Q^{(s_t)}$ and $R^{(s_t)}$ are noise covariance matrices. In this formulation, all dynamical parameters can be switched between regimes. Similar models referred to in the above literature sometimes switch only the state dynamics $\{A, Q\}$, or the observation dynamics $\{C, R\}$. Conditioned on a setting of s_t, the model is equivalent to a linear Gaussian state-space model (Kalman filter).

It is possible to factorise the switch variable, so that M factors $f_t^{(1)}, \ldots, f_t^{(M)}$ affect the observations y_t. The factor $f^{(m)}$ can take on $L^{(m)}$ different values. The state space is the cross product of the factor variables,

$$s_t = f_t^{(1)} \otimes f_t^{(2)} \otimes \cdots \otimes f_t^{(M)} \tag{12.2}$$

with $K = \prod_{m=1}^{M} L^{(m)}$ being the number of settings that s_t can take on. The value of $f_t^{(m)}$ depends on $f_{t-1}^{(m)}$, and the factors are a priori independent, so that

$$p(s_t \,|\, s_{t-1}) = \prod_{m=1}^{M} p\left(f_t^{(m)} \,|\, f_{t-1}^{(m)}\right) \ . \tag{12.3}$$

The joint distribution of the model is

$$p(s_{1:T}, x_{1:T}, y_{1:T}) = p(s_1)p(x_1)p(y_1 \,|\, x_1, s_1) \prod_{t=2}^{T} p(s_t \,|\, s_{t-1})p(x_t \,|\, x_{t-1}, s_t)p(y_t \,|\, x_t, s_t)$$

where $s_{1:T}$ denotes the sequence $s_1, s_2, \ldots, s_T$ and similarly for $x_{1:T}$ and $y_{1:T}$. $p(x_t \,|\, x_{t-1}, s_t)$ and $p(y_t \,|\, x_t, s_t)$ are defined in Eq. (13.1), and $p(s_t \,|\, s_{t-1})$ in Eq. (13.3).

By considering the factored nature of the switch setting, we have an observation term of the form $p(y_t \,|\, x_t, f_t^{(1)}, \ldots, f_t^{(M)})$. This can be parameterised in different ways. In this work, we specify conditional independencies between particular components of the observation y_t given the factor settings. This is explained further in section 13.4.5. Although we make use of prior factored dynamics in Eq. (13.3) in this work, it is very simple to generalize the model so that this no longer holds. The inference algorithms described in section 13.5 can still be applied. However, the separate factors are crucial in structuring the system dynamics and observations model.

In the physiological monitoring application, factor settings $f_t^{(1)}, \ldots, f_t^{(M)}$ represent different conditions (e.g. whether a probe has fallen off or not, whether there is a specific problem with the circulatory system or not). The state x_t can contain estimates of the "true" values of physiological properties, based on noisy, artifact-prone observations y_t.

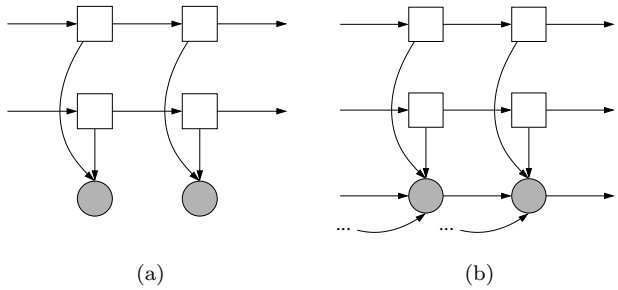


Figure 12.1: Graphical representations of different factorial models, with $M = 2$ factors. Squares are discrete values, circles are continuous and shaded nodes are observed. (a) The Factorial HMM, (b) the Factorial AR-HMM, in which each observation depends on previous values.

12.2.1 Comparison with other switching models for condition monitoring

We have assumed the existence of a discrete switch variable which indexes different modes of operation. In our formulation, the problem of condition monitoring is essentially to infer the value of this switch variable over time from new data. We are particularly interested in the class of models in which there are first-order Markovian transitions between the switch settings at consecutive time steps. Given the switch setting it is possible to characterise the different dynamic regimes on other ways, yielding alternative models for condition monitoring. In this section, we first review the hidden Markov model (HMM) and autoregressive hidden Markov model (AR-HMM), and then discuss their advantages and disadvantages for condition monitoring with respect to the (F)SLDS.

A simple model for a single regime is the Gaussian distribution on y_t. When this is conditioned on a discrete, first-order Markovian switching variable, we obtain an instance of a HMM. This model can therefore be used for condition monitoring when the levels and variability of different measurement channels are significant (though note that in general the HMM can use any reasonable distribution on y_t).

Autoregressive (AR) models are a common choice for modelling stationary time series. Conditioning an AR model on a Markovian switching variable we obtain an autoregressive hidden Markov model (AR-HMM), also known as a switching AR model – see e.g. [13]. This provides a model for conditions in which observations might be expected to oscillate or decay, for example. During inference, the model can only confidently switch into a regime if the last p observations have been generated under that regime; there will be a loss of accuracy if any of the measurement channels have dropped out in that period, for example, or another artifactual process has affected any of the readings.

The general condition monitoring problem involves independent factors which affect a system. In both of these models the switch variable can be factorised, giving the factorial HMM [4] and the factorial AR-HMM respectively. The graphical models for these two constructions are shown in Figure 13.1.

By characterising each regime as a linear Gaussian state-space model we obtain the (F)SLDS. The SLDS can be thought of as a "hybrid" model, having both discrete switch settings as in the HMM and continuous hidden state as in a linear dynamical system. The FSLDS is similar, though with the discrete switch setting structure of the factorial HMM. Note, however, that observations in the FHMM [4] are generated through an additive process in which each factor makes a contribution.

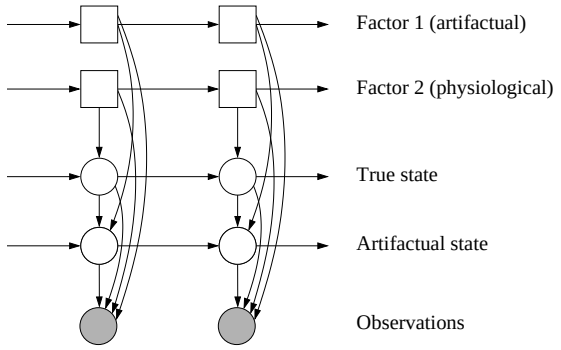


Figure 12.2: Factorial switching linear dynamical system for physiological condition monitoring, with $M = 2$ factors as an example. The state is split up into two sets of variables, containing estimates of the 'true' physiology and of the levels of artifactual processes.

The mechanisms used to generate observations under different factor settings can in general be more complex and nonlinear than this, as in the the overwriting mechanism explained in section 13.4.5.

(F)SLDS models have a number of representational advantages for condition monitoring. First, we can have many dimensions of hidden state for each observed dimension. This allows us to deal with situations in which different elements affect the observations, as we will demonstrate in the examples later in this chapter.

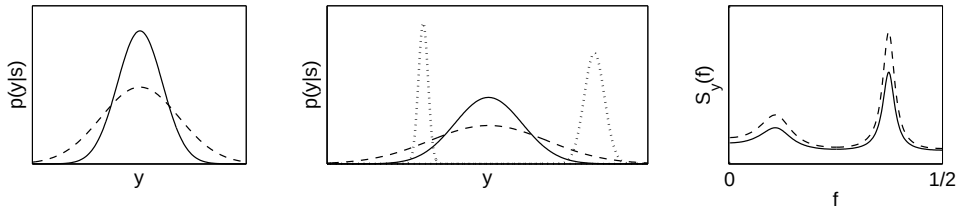
In the physiological monitoring case we can construct detailed representations of the causes underlying observations. For instance, the state can be split into two groups of continuous latent variables, those representing the "true" physiology and those representing the levels associated with different artifactual processes, as we demonstrate in sections 13.4.1 and 13.4.2. In this application, factors can be physiological or artifactual processes. Physiological factors can affect any state variable, whereas artifactual processes affect only artifactual state. This formulation of the model for physiological condition monitoring is illustrated in Figure 13.2.

The (F)SLDS also gives us the ability to represent different sources of uncertainty in the system. We can explicitly specify the intra-class variability in the dynamics using the parameter Q and the measurement noise using the parameter R. There is no way to make this distinction in either of the other models, which have only one noise term per regime. However, this flexibility in the FSLDS is obtained at the cost of greater complexity, particularly in terms of computing inferences, as we examine in section 13.5.

12.3 Novel conditions

So far we have assumed that the monitoring data contains a limited number of regimes, for which labelled training data is available. In real-world monitoring applications, however, there is often such a great number of potential dynamical regimes that it might be impractical to model them all, or we might never have comprehensive knowledge of them. It can therefore be useful to include a factor in the condition monitoring model which represents all "unusual cases".

In this section we present a method for modelling previously unseen dynamics as an extra factor in the model, referred to as the "X-factor". This represents all dynamics which are not normal and which also do not correspond to any of the known regimes. A sequence of data can only be said to have novelty relative to



(a) (b) (c)

Figure 12.3: (a) Class conditional likelihoods in a static 1D model, for the normal class (solid) and the X-factor (dashed). (b) Likelihoods of the normal class and X-factor in conjunction with other known, abnormal regimes (shown dotted). (c) The power spectral density of a latent AR(5) process with white observation noise (solid), and that of a corresponding X-factor process (dashed).

some reference, so the model is learnt taking into account the parameters of the normal regime. The inclusion of this factor in the model has two potential benefits. First, it is useful to know when novel regimes are encountered, e.g. in order to raise an alarm. Second, the X-factor provides a measure of confidence for the system. That is, when a regime is confidently classified as "none of the above", we know that there is some structure in the data which is lacking in the model.

12.3.1 The X-factor

First consider a case in which we have independent, one-dimensional observations which normally follow a Gaussian distribution. If we expect that there will also occasionally be spurious observations which come from a different distribution, then a natural way to model them is by using a wider Gaussian with the same mean. Observations close to the mean retain a high likelihood under the original Gaussian distribution, while outliers are claimed by the new model.

The same principle can be applied when there are a number of known distributions, so that the model is conditionally Gaussian, $y \,|\, s \sim \mathcal{N}\left(\mu^{(s)}, \Sigma^{(s)}\right)$. For condition monitoring we are interested in problems where we assume that the possible settings of s represent a "normal" mode and a number of known additional modes. In physiological monitoring, for example, the normal mode corresponds to the times when the physiology is stable and there is no artifactual corruption of the observed data. Additional modes correspond e.g. to known problems with the monitoring equipment or specific pathologies. We assume here that the normal regime is indexed by $s = 1$, and the additional known modes by $s = 2, \ldots, K$. In this static case, we can construct a new model, indexed by $s = *$, for unexpected data points by inflating the covariance of the normal mode, so that

$$\Sigma^{(*)} = \xi\Sigma^{(1)}, \qquad \mu^{(*)} = \mu^{(1)} \,, \tag{12.4}$$

where normally $\xi > 1$. We refer to this type of construction for unexpected observations as an "X-factor". The parameter ξ determines how far outside the normal range new data points have to fall before they are considered "not normal".

The likelihood functions for a normal class and a corresponding X-factor are shown in Figure 13.3(a). Clearly, data points that are far away from the normal range are more likely to be classified as belonging to the X-factor. For condition monitoring this can be used in conjunction with a number of known classes, as shown in 13.3(b). Here, the X-factor has the highest likelihood for regions which are far away from any known modes, as well as far away from normality.

We can generalise this approach to dynamic novelty detection by adding a new factor to a trained factorial switching linear dynamical model, by inflating the system noise covariance of the normal dynamics

$$Q^{(*)} = \xi Q^{(1)} , \tag{12.5}$$

$$\left\{A^{(*)}, C^{(*)}, R^{(*)}, d^{(*)}\right\} = \left\{A^{(1)}, C^{(1)}, R^{(1)}, d^{(1)}\right\} \tag{12.6}$$

To help understand why (13.5) and (13.6) are a dynamic generalisation of (13.4), consider the specific case of a hidden scalar AR(p) process,

$$x_t \sim \mathcal{N}\left(\sum_{k=1}^{p} \alpha_k x_{t-k}, \sigma_q^2\right), \qquad y_t \sim \mathcal{N}(x_t, \sigma_r^2) . \tag{12.7}$$

The power spectral density for the hidden process x_t at frequency f is given by

$$S_x(f) = \frac{\sigma_q^2}{\left|1 - \sum_{k=1}^{p} \alpha_k e^{-2\pi i f k}\right|^2} , \tag{12.8}$$

where $-\frac{1}{2} \leq f \leq \frac{1}{2}$, assuming one observed value per unit of time. By inflating σ_q^2 (as specified in (13.5)) we observe that the power is increased at each frequency. The observed process has the spectrum $S_y(f) = S_x(f) + \sigma_r^2$. As the scale of $S_y(f)$ is determined by the magnitudes of the two noise variances, inflating σ_q^2 will have the effect of increasing the power at every frequency, as illustrated in Figure 13.3(c).

In the LDS, any sequence of x's is jointly Gaussian. Consider the case where the state is a scalar variable; the eigenfunctions are sinusoids and the eigenvalues are given by the power spectrum. As increasing the system noise has the effect of increasing the power at all frequencies in the state sequence, we have a dynamical analogue of the static construction given above.

A similar model for changes in dynamics is mentioned by [12, p. 458 and §12.4], who suggest it as the parameterisation of an extra state in the unfactorised SLDS for modelling large jumps in the x-process, and suggest setting $\xi = 100$. Their analysis in §12.4.4 shows that this is used to model single-time-step level changes, and not (as we are doing) sustained periods of abnormality. We find a much smaller value $\xi = 1.2$ to be effective for our task (larger values of ξ mean that an observation sequence must deviate further from normal dynamics to be claimed by the X-factor). A different generative model for the X-factor in principle would be white noise, but we find in practice that this model is too dissimilar to the real signal and is not effective.

The nature of the measurement noise, and hence the value of the parameter $R^{(s)}$, is assumed to be the same for both the normal regime and for the X-factor. Care needs to be taken that the known factor dynamics do not have a very high variance compared to the normal dynamics. It is clear from Figure 13.3(b) that the X-factor will not be effective if any of the factors are wider than normality. This can be ascertained by examining the spectra of the different model dynamics.

12.4 Parameter estimation

In a condition monitoring problem, it is assumed that we are able to interpret at least some of the regimes in the data – otherwise we would be less likely to have an interest in monitoring them. We can therefore usually expect to obtain some labelled training data $\{y_{1:T}, s_{1:T}\}$. When available, this data greatly simplifies the learning process, because determining the switch setting in the (F)SLDS makes the

model equivalent to a linear dynamical system, therefore making the process of parameter estimation a standard system identification problem.

Given training data with known switch settings, the learning process is therefore broken down into the training of a set of LDS models – one per switch setting. We might choose a particular parameterisation, such as an autoregressive (AR) model of order p hidden by observation noise and fit parameters accordingly. Expectation maximisation can be useful in this setting to improve parameter settings given an initialisation [2]. We describe particular methods used for parameter estimation in the physiological monitoring application which incorporate both of these ideas.

When labelled training data is available, estimates of the factor transition probabilities are given by $P(f_t^{(m)} = j \mid f_{t-1}^{(m)} = i) = \frac{n_{ij}+\zeta}{\sum_{k=1}^{M}(n_{ik}+\zeta)}$, where n_{ij} is the number of transitions from factor setting i to setting j in the training data. The constant terms ζ (set to $\zeta = 1$ in the experiments described later in the chapter) are added to stop any of the transition probabilities being zero or very small.

Some verification of the learned model is possible by clamping the switch setting to a certain value and studying the resulting LDS. One simple but effective test is to draw a sample sequence and check by eye whether it resembles the dynamics of training data which is known to follow the same regime [**?**, §5.7]. Some insight into the quality of the parameter settings can also be gained by considering estimation of the hidden state x in the LDS. The Kalman filter equations yield both an *innovation sequence*, $\tilde{y}_{1:T}$ (the difference between the predicted and actual observations), and a specification of the covariance of the innovations under ideal conditions. An illuminating test is therefore to compare the actual and ideal properties of the innovation sequence when applied to training data. In particular, the innovations $\tilde{y}_t$ should come from a Gaussian distribution with zero mean and a specific covariance, and should be uncorrelated in time. We find in practice that such tests are useful when training (F)SLDS models for condition monitoring. For more details about verification in linear dynamical systems, see [**?**, §5.5].

We now show examples of learning different aspects of the FSLDS in the physiological monitoring setting. We begin with system identification for the "normal" physiological dynamics (i.e. the dynamics which are observed when the baby is stable and there is no artifactual influence) in section 13.4.1. We then show how to model an artifactual process, the drawing of a blood sample, in section 13.4.2. In section 13.4.3, we introduce the training of a physiological factor associated with bradycardia, a specific heart problem. Learning the X-factor parameter ξ is covered in section 13.4.4, and in section 13.4.5 we demonstrate how to combine dynamical models into the overall factorised SLDS.

12.4.1 Learning normal dynamics: heart rate

Looking at examples of normal heart rate dynamics as in the top left and right panels of Figure 13.4, it can be observed first of all that the measurements tend to fluctuate around a slowly drifting baseline. This motivates the use of a model with two hidden components: the signal x_t, and the baseline b_t. These components are therefore used to represent the true heart rate, without observation noise. The dynamics can be formulated using autoregressive (AR) processes, such that an AR(p_1) signal varies around an AR(p_2) baseline, as given by the following equations:

$$x_t - b_t \sim \mathcal{N}\left(\sum_{k=1}^{p_1} \alpha_k (x_{t-k} - b_{t-k}), \eta_1\right) \quad , \quad b_t \sim \mathcal{N}\left(\sum_{k=1}^{p_2} \beta_k b_{t-k}, \eta_2\right) \qquad (12.9)$$

where η_1, η_2 are noise variances. For example, an AR(2) signal with AR(2) baseline has the following state-space representation:

$$x_t = \begin{bmatrix} x_t \\ x_{t-1} \\ b_t \\ b_{t-1} \end{bmatrix}, \quad A = \begin{bmatrix} \alpha_1 & \alpha_2 & \beta_1 - \alpha_1 & \beta_2 - \alpha_2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \beta_1 & \beta_2 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$Q = \begin{bmatrix} \eta_1 + \eta_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \eta_2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad C = [1\ 0\ 0\ 0]\ . \tag{12.10}$$

It is straightforward to adjust this construction for different values of p_1 and p_2. The measurements are therefore generally taken to be made up of a baseline with low frequency components and a signal with high frequency components. We begin training this model with a heuristic initialisation, in which we take sequences of training data and remove high frequency components by applying a symmetric 300-point moving average filter. The resulting signal is taken to be the low frequency baseline. The residual between the original sequences and the moving-averaged sequences are taken to contain both stationary high frequency hemodynamics as well as measurement noise. These two signals can be analysed according to standard methods and modelled as AR or integrated AR processes (specific cases of autoregressive integrated moving average (ARIMA) processes [11]) of arbitrary order. Heart rate sequences were found to be well modelled by an AR(2) signal varying around an ARIMA(1,1,0) baseline. An ARIMA model is a compelling choice for the baseline, because with a low noise term it produces a smooth drift[3]. Having found this initial setting of the model parameters, EM updates are then applied. These are similar the updates given in [2], though constrained so that we retain the structure in Eq. (13.10). This has been found to be particularly useful for refining the estimates of the noise terms Q and R.

Examples of the heart rate model being applied using a Kalman filter to sequences of heart rate data are shown in Figure 13.4, which plots the noisy observations y_t (upper panels) and estimates of the baseline b_t and high frequency components $x_t - b_t$ (middle and lower panels respectively).

12.4.2 Learning artifactual dynamics: blood sampling

An arterial blood sample might be taken every few hours from each baby. This involves diverting blood from the arterial line containing the pressure sensor, so that measurements are entirely unrelated from the baby's physiology. Throughout the operation a saline pump acts against the sensor, causing an artifactual ramp in the blood pressure measurements. The slope of the ramp is not always the same, as the rate at which saline is pumped can vary. See Figure 13.6(a) for an example. During this process, the systolic and diastolic blood pressures of the baby evolve as normal but are unobserved.

[3]The ARIMA(1,1,0) model has the form $(X_t - \beta X_{t-1}) = \alpha_1(X_{t-1} - \beta X_{t-2}) + Z_t$ where $\beta = 1$ and $Z_t \sim N(0, \sigma_Z^2)$. This can be expressed in un-differenced form as a non-stationary AR(2) model. In our implementation we set $\beta = 0.999$ and with $|\alpha_1| < 1$ we obtain a stable AR(2) process, which helps to avoid problems with numerical instability. This slight damping makes the baseline mean-reverting, so that the resulting signal is stationary. This has desirable convergence properties for dropout modelling.

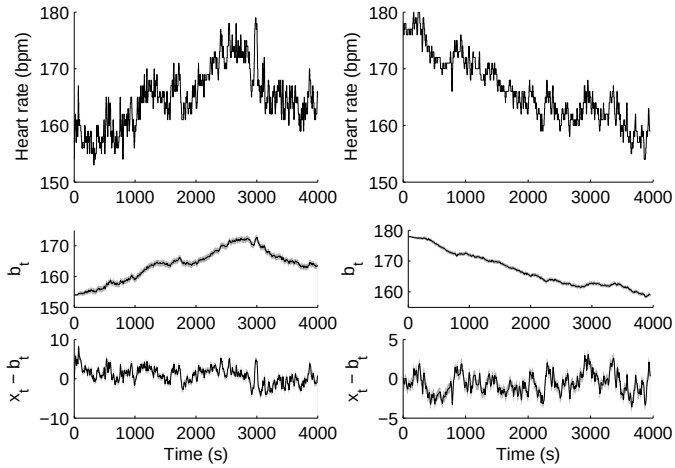


Figure 12.4: In these two examples, HR measurements (in the top left and top right panels) are varying quickly within normal ranges. The estimated distributions of the underlying signal (bottom left and bottom right panels) are split into a smooth baseline process and zero-mean high frequency component, given by the applying the Kalman filter equations with parameters learnt as in section 13.4.1.

We specify a structural model for this artifactual rise, in which the artifactual measurements a_t evolve according to a gradient which is subject to a random walk:

$$a_t \sim \mathcal{N}\left(a_{t-1} + d_a + c_{t-1}, \sigma_a^2\right), \quad c_t \sim \mathcal{N}\left(c_{t-1}, \sigma_c^2\right) \ . \tag{12.11}$$

Each of these terms are scalars. Parameter d_a is a positive constant specifying the average drift, which is modified by the random walk term c_t. The Gaussian noise on a_t with variance σ_a^2 models the differences in slope of blood samples taken at different times, while the noise on c_t with variance σ_c^2 models the change in slope within a single blood sample operation. During the procedure, both blood pressure readings (systolic and diastolic) are generated by the same underlying value a_t.

We now represent these dynamics in state space form. The hidden state dynamics for a_t, c_t are given by

$$x_t = \begin{bmatrix} a_t \\ c_t \end{bmatrix}, \ d_{\text{BS}} = \begin{bmatrix} d_a \\ 0 \end{bmatrix}, \ A_{\text{BS}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ Q_{\text{BS}} = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_c^2 \end{bmatrix}.$$

The observations y_t are two-dimensional, where $y_{t,1}$ is the systolic blood pressure and $y_{t,2}$ is the diastolic blood pressure. The observation model is then given by

$$C_{\text{BS}} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \ R_{\text{BS}} = \begin{bmatrix} r_{\text{SysBP}} & 0 \\ 0 & r_{\text{DiaBP}} \end{bmatrix}.$$

The parameters are straightforward to learn from training data. Let T denote a 2-dimensional time series spanning n time steps and containing an example of the observed systolic and diastolic blood pressures during a blood sample operation, such that $T = \{y_{t,1:2} \,|\, 1 \le t \le n\}$. If we have a set of such training sequences $T_1, \ldots, T_N$ then parameter estimation proceeds as follows:

$$d_a = \frac{1}{N}\sum_{i=1}^{N} \text{slope}(T_i), \quad \sigma_a^2 = \mathsf{V}\left(\{\text{slope}(T_i) \,|\, 1 \le i \le N\}\right) \tag{12.12}$$

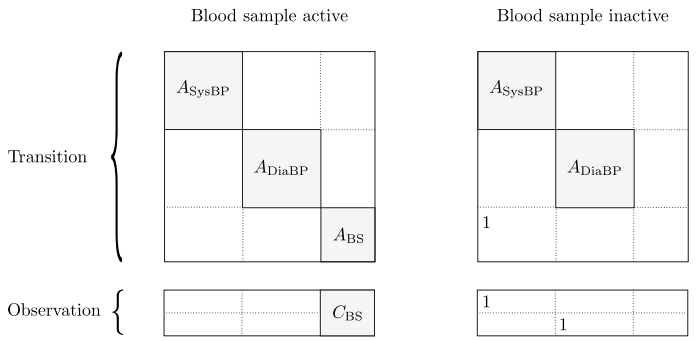


Figure 12.5: State spaces for normal dynamics and for blood sampling.

where the slope of the measurements can be found either by simply calculating the gradient $\text{slope}(T) = \frac{1}{2(n-1)}\left[(y_{1,1} + y_{1,2}) - (y_{n,1} + y_{n,2})\right]$ or, better, with linear regression. The variance of the noise on c_t is found using $\sigma_c^2 = \frac{1}{N}\sum_{i=1}^{N} \mathsf{V}\left(\text{diff}(T_i)\right)$, where $\text{diff}(T) = \left\{\frac{1}{2}\left[(y_{i,1} + y_{i,2}) - (y_{i-1,1} + y_{i-1,2})\right] \mid 2 \leq i \leq n\right\}$. We assume that the monitoring equipment has the same observation noise whether a blood sample is being taken or not. Therefore R_{BS} is the same as for the normal dynamics, learnt as described in the previous section.

Having set the values of each of the parameters, these dynamics can be incorporated into the full switching model. Figure 13.5 shows the structure of the full transition matrix A and the observation matrix C, for blood sample dynamics ($f^{(\text{BS})} = 1$) and for normal dynamics ($f^{(\text{BS})} = 0$). During a blood sample, observations are generated only by the artifactual state variables. State variables representing the baby's actual systolic and diastolic blood pressure are evolved according to the normal dynamics, but not observed (see Figure 13.13(a) for an illustration of hidden state inferences in this case). During normal dynamics, the physiological state variables are observed, and a_t is tied to the estimate of systolic blood pressure.

One way to check that such a model has been adequately fitted is to sample from it and verify that the results have similar characteristics to the training data. Figure 13.6(a) shows a genuine example of measurements made while a blood sample being taken. We then draw a sample of the same length from the switching model, Figure 13.6(b), using the same initial state at $t = 1$ and with the switch settings clamped to be the same as for the genuine sequence. Note that the observation noise appears greater on the simulated version partly because i.i.d. Gaussian noise is being used to model quantisation effects in the training observations.

12.4.3 Learning physiological dynamics: bradycardia

Bradycardia is a (usually temporary) reduction in the heart rate, and brief episodes are common for premature infants. It can have many causes, some benign and some serious. Examples of this phenomenon are shown in Figure 13.7. Because there is no simple physical process governing this pattern of observations, we have less opportunity to parameterise the model based on domain knowledge. Therefore we turn to statistical principles to learn the dynamics and valildate our chosen model.

There are several ways to approach modelling this pattern. For example, we could learn the dynamics of the whole sequence as a single regime, or split it up into falling and rising dynamics. If the former, we could specify some positive second

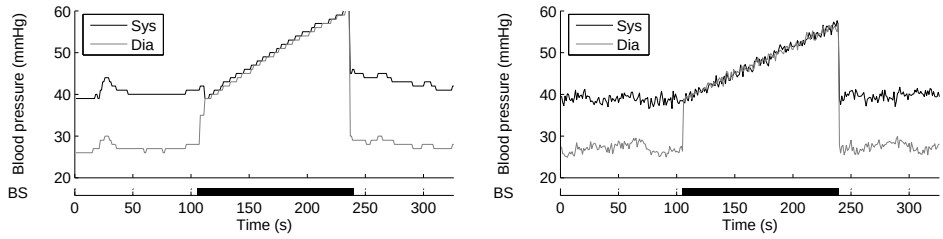


Figure 12.6: Genuine blood sample, and a sample drawn from the model with switch settings clamped to be the same. BS indicates the factor switch setting, where black denotes $f_t^{(\mathrm{BS})} = 1$ (blood sample dynamics) and white denotes $f_t^{(\mathrm{BS})} = 0$ (normal dynamics).

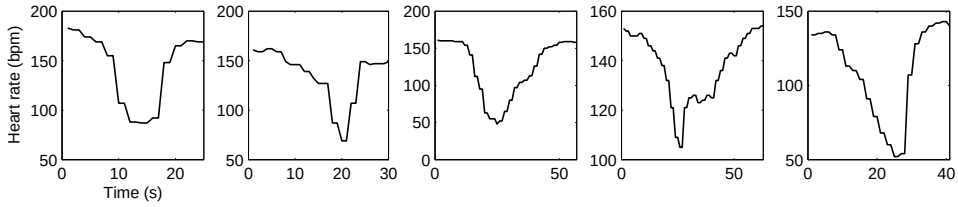


Figure 12.7: Examples of periods of bradycardia. Heart rate is measured in beats per minute (bpm).

derivative, or look at the variability of the observations. If the latter then we could use drift terms to model the fall and rise, or use exponential decays towards different mean values.

Bradycardic drops and subsequent rises in heart rate were found to be adequately modelled by retraining the ARIMA(1,1,0) model for baseline heart rate dynamics. The high frequency heart rate dynamics are kept the same as for the stable heart rate regime. As for the normal regime, this model learnt in terms of hidden ARIMA processes was used as an initial setting and updated with three iterations of EM.

12.4.4 Learning the novelty threshold

Unlike the factors for which we have an interpretation, we do not assume that labelled training data is available for learning X-factor dynamics. We therefore consider a partial labelling of the training data $y_{1:T}$, comprising of annotations for known factors and for some representative quantity of normal dynamics. The remainder of the training data is unlabelled, giving us a semi-supervised learning problem.

To apply the expectation-maximisation algorithm to the X-factor within a SLDS (non-factorised switch setting), the M-step update to ξ is given by

$$\tilde{\xi} = \frac{1}{\sum_{t=2}^{T} p(s_t = * \,|\, y_{1:T}, \theta_{\mathrm{old}})} \times \sum_{t=2}^{T} \left\langle (x_t - A^{(1)} x_{t-1})^\top {Q^{(1)}}^{-1} (x_t - A^{(1)} x_{t-1}) \right\rangle_{x_t, x_{t-1} \,|\, y_{1:T}, \theta_{\mathrm{old}}} p(s_t = * \,|\, y_{1:T}, \theta_{\mathrm{old}}) \tag{12.13}$$

where $s_t = *$ indexes the X-factor switch setting at time t. We describe strategies for

calculating the expectation term in sections 13.5.1 and 13.5.2. It can be convenient to use the filtering estimate $\langle\cdot\rangle_{x_t,x_{t-1}\,|\,y_{1:t},\theta_{\text{old}}}$ as an approximation.

The parameters $A^{(1)}$ and $Q^{(1)}$ are the system matrix and system noise covariance matrix respectively for the normal dynamical regime. Intuitively, this update expression calculates a Z-score, considering the covariance of novel points and the covariance of the normal regime. Every point is considered, and is weighted by the probability of having been generated by the X-factor regime. Eq. (13.13) does not explicitly constrain $\tilde{\xi}$ to be greater than 1, but with appropriate initialisation it is unlikely to violate this condition.

The factorial case is a little more complicated due to the possibility that different combinations of factors can overwrite different channels. For example, if a bradycardia is occurring in conjunction with some other, unknown regime, then the heart rate dynamics are already well explained and should not be taken into account when re-estimating the X-factor parameter ξ.

An alternative to (13.13) and an extension to the factorial case is given in [8, §C.4].

12.4.5 Learning the factorial model

The previous discussion assumes that we train the model conditioned on each switch setting independently, and then combine parameters. Where there are many factors this implies a great quantity of training data is needed. In practice, however, this requirement can be mitigated.

Where there are several measurement channels it may be found that some factors "overwrite" others. For example, if we are monitoring the physiological condition of a patient, we might have two factors: *bradycardia* and *incubator open*. If there is a period of bradycardia while the incubator is open, then we would see the same measurements as though there was only the bradycardia. It is often possible to specify an ordering of factors such that some overwrite measurement channels of others in this way. This ordering specifies conditional independencies in the factors and observations, such that

$$f^{(i)} \text{ overwrites } f^{(j)} \text{ on measurement channel } d \implies y_{t,d} \perp\!\!\!\perp f_t^{(j)} \,|\, f_t^{(i)} > 0 \ ,$$

assuming that the switch setting $f^{(j)} = 0$ means that factor j is "inactive" and positive integers index the active dynamics of that factor. The significance of this is that examples of every combination of factors do not need to be found in order to train the factorial model. The factors can be trained independently, and then combined together by reasoning about which channels are overwritten for each combination.

12.5 Inference

Exact inference in the switching linear dynamical model is intractable, so we need to make approximations regardless of whether we are doing filtering or smoothing. We describe each case in turn in sections 13.5.1 and 13.5.2, including its relevance to this application and the techniques which can be used to make it tractable. We then discuss modifications of the standard inference procedures to suit this application in sections 13.5.3 and 13.5.4.

12.5.1 Filtering

During deployment of the neonatal monitoring system we are primarily interested

in filtering, though fixed-lag smoothing with a small time delay would also be conceivable. We receive measurements second by second and are required to make an immediate diagnosis based only on the history.

The time taken to calculate the exact filtering distribution $p(s_t, x_t \mid y_{1:t})$ in the switching linear Gaussian state-space model scales exponentially with t, making it intractable. This is because the probabilities of having moved between every possible combination of switch settings in times $t-1$ and t are needed to calculate the posterior at time t. Hence the number of Gaussians needed to represent the posterior exactly at each time step increases by a factor of K, the number of cross-product switch settings. The intractability of inference in this model is proved in [6], which also concentrates on a fault diagnosis setting.

Gaussian Sum approximations have been reviewed previously. While performing the forward pass to calculate the filtering distribution, at each time step we maintain an approximation of $p(x_t \mid s_t, y_{1:t})$ as a mixture of I Gaussians. Calculating the Kalman updates and likelihoods for every possible setting of s_{t+1} will result in the posterior $p(x_{t+1} \mid s_{t+1}, y_{1:t+1})$ having KI mixture components, which can be collapsed back into I components by matching means and variances of the distribution for each setting of s_t.

Rao-Blackwellised particle filtering (RBPF) [7] is another technique for approximate inference, which exploits the conditionally linear dynamical structure of the model to try to select particles close to the modes of the true filtering distribution. A number of particles are propagated through each time step, each with a switch state s_t and an estimate of the mean and variance of x_t. A value for the switch state s_{t+1} is obtained for each particle by sampling from the transition probabilities, after which Kalman updates are performed and a likelihood value can be calculated. Based on this likelihood, particles can be either discarded or multiplied. Because Kalman updates are not necessarily calculated for every possible setting of s_{t+1}, this method can give an increase in speed when there are many factors. The fewer particles used, the greater the trade-off of speed against accuracy, as it becomes less likely that the particles can collectively track all modes of the true posterior distribution. RBPF has been shown to be successful in condition monitoring problems with switching linear dynamics, for example in fault detection in mobile robots [1].

In the factorised model, the number of switch settings may be high. There may also be some transitions with low probability. In this case, sampling from the discrete transition prior $p(s_t \mid \hat{s}_{t-1}^{(i)})$ might be problematic, as there is a chance that no particles are sampled from certain switch settings. We modify the prior so that we sample from $\hat{s}_t^{(i)} \sim q(s_t \mid \hat{s}_{t-1}^{(i)})$ where

$$q(s_t = k \mid s_{t-1} = j) = \frac{p(s_t = k \mid s_{t-1} = j) + \zeta}{\sum_{l=1}^{M} \left(p(s_t = l \mid s_{t-1} = j) + \zeta \right)} , \tag{12.14}$$

which makes it more likely that particles are sampled from different switch settings. In the following experiments we use $\zeta = 0.1$. If a large setting of ζ was thought to be distorting the results, then it would be straightforward to adjust the importance weight of each particle i by a factor of $\frac{p\left(\hat{s}_t^{(i)} \mid \hat{s}_{t-1}^{(i)}\right)}{q\left(\hat{s}_t^{(i)} \mid \hat{s}_{t-1}^{(i)}\right)}$ at each time step to compensate.

12.5.2 Smoothing

While carrying out parameter estimation, we are interested in smoothing. Given offline training data, we can afford to exploit backwards information in order to refine estimates of the system parameters. The 'E' steps of the EM procedures

cited in section 13.4 therefore use a backwards pass of the data – though in this case we already know the switch settings s_t for all t, so we do smoothing during parameter estimation on linear dynamical systems, not on the full switching model. It can also be interesting to see how the performance on testing data changes when calculating smoothed inferences, even though this is not realistic for deployment.

Approximate smoothed inference in the full switching model is possible with various schemes. In the previous section we described the simplifying assumption that the true filtering distribution is a mixture of Gaussians (the Gaussian sum approximation). We can make the same assumption about the smoothing distribution. Gaussian sum smoothing is possible using Expectation Correction (see the chapter by Barber in this book). Smoothed inference can also be done with Expectation Propagation (see chapter by Zoeter and Heskes in this book), or by variational inference [3].

We now discuss some adaptations to the standard inference routines, for handling missing observations and constraining the number of possible switch transitions.

12.5.3 Handling missing observations

Zero observations in this application denote missing values, for example where a probe has been disconnected. It would be possible in principle to add an extra factor for each channel to indicate a missing observation. However each extra factor with two states slows inference down by a factor of two. To add a dropout channel for each of 10 observed channels would slow down inference by a factor of 2^{10}.

Instead we can test at each time whether there are any dimensions in y_t which are zero. Within the inference routines, for each switch setting we then use the updated observation matrix given by

$$H^*_{ij} = \begin{cases} H_{ij} & \text{where } y_j \neq 0 \\ 0 & \text{otherwise .} \end{cases} \tag{12.15}$$

The Kalman filter updates, which are incorporated in all the inference schemes we consider, have two stages: prediction (in which the variance of the estimates increases) and correction (in which the variance decreases). The effect of this modification to the observation matrix is to cancel the correction step whenever observations are unavailable. Typical inference results are shown in Figure 13.8, where a period of heart rate measurements are missing and we show the estimated distribution of "true" heart rate. We made the normal dynamics stable and mean-reverting in section 13.4.3, so that the estimates reach an equilibrium in this situation.

12.5.4 Constraining switch transitions

In this application, factors change switch settings slowly relatively to the sampling rate. It is therefore unlikely that more than one factor changes its setting in any one time step. We can use this to constrain the transitions.

The discrete transition probabilities are defined by a matrix Z, where $Z_{ij} = p(s_t = j \,|\, s_{t-1} = i)$, given by Eq. (13.3). We can use an updated transition matrix Z^* such that

$$Z^*_{ij} = \begin{cases} Z_{ij} & \text{where } 0 \leq \sum_{s=1}^{M} I\left[f^{(i)}[s] = f^{(j)}[s]\right] \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{12.16}$$

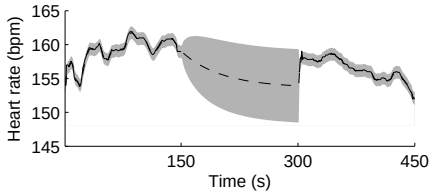


Figure 12.8: Inference of true heart rate under a dropout, with automatic handling of missing observations. The solid black line shows HR observations, where available, the dashed line shows the mean of the estimated distribution of true HR, and the shaded area shows two standard deviations of the estimated distribution.

where $I[\cdot]$ is the indicator function and $f^{(i)}[s]$ denotes the switch setting for factor i in the cross product switch setting s. This reduces the inference time from $O(K^2)$ to $O(K \log K)$.

12.6 Experiments

This section describes experiments used to evaluate the model for condition monitoring. Other than the X-factor, we consider here the incubator open/handling of baby factor (denoted 'IO'), the blood sample factor (denoted 'BS'), the bradycardia factor (denoted 'BR') and the temperature probe disconnection factor (denoted 'TD'). We demonstrate the operation of the transcutaneous probe recalibration factor (denoted 'TR'), but do not evaluate it quantitatively due to a scarcity of training data.

Some conventions in plotting the results of these experiments are adopted throughout this section. Horizontal bars below time-series plots indicate the posterior probability of a particular factor being active, with other factors in the model marginalised out. White and black indicate probabilities of zero and one respectively[4]. In general the plots show a subset of the observation channels and posteriors from a particular model – this is indicated in the text.

24-hour periods of monitoring data were obtained from fifteen premature infants in the intensive care unit at Edinburgh Royal Infirmary. The babies were between 24 and 29 weeks gestation (around 3-4 months premature), and all in around their first week *post partum*.

Each of the fifteen 24-hour periods was annotated by two clinical experts. At or near the start of each period, a 30 minute section of normality was marked, indicating an example of that baby's current baseline dynamics. Each of the known common physiological and artifactual patterns were also marked up.

Finally, it was noted where there were any periods of data in which there were clinically significant changes from the baseline dynamics not caused by any of the known patterns. While the previous annotations were made collaboratively, the two annotators marked up this 'Abnormal (other)' category independently. The software package TSNet [5] was used to record these annotations, and the recorded intervals were then exported into Matlab. The number of intervals for each category, as well as the total and average durations, are shown in Table 13.2. The figures for the 'Abnormal' category were obtained by combining the two annotations, so that

[4] A convenient property of the models evaluated here, from the perspective of visualisation, is that the factor posteriors tend be close to zero or one. This is partly due to the fact that the discrete transition prior $p(s_t|s_{t-1})$ is usually heavily weighted towards staying in the same switch setting (long dwell times).

Factor	Incidences	Total duration	Average duration
Incubator open	690	41 hours	3.5 mins
Abnormal (other)	605	32 hours	3.2 mins
Bradycardia	272	161 mins	35 secs
Blood sample	91	253 mins	2.8 mins
Temp. disconnection	87	572 mins	6.6 mins
TCP recalibration	11	69 mins	6.3 mins

Table 12.2: Number of incidences of different factors, and total time for which each factor was annotated as being active in the training data (total duration of training data $15 \times 24 = 360$ hours).

the total duration is the number of points which either annotator thought to be in this category, and the number of incidences was calculated by merging overlapping intervals in the two annotations (two overlapping intervals are counted as a single incidence).

The rest of this section shows the results of performing inference on this data and comparing it to the gold standard annotations provided by the clinical experts.

12.6.1 Evaluation of known factors

The dataset for evaluation consisted of fifteen 24-hour periods of monitoring data (one day of monitoring data for each of fifteen babies). Evaluation was done using leave-one-out cross validation, so that the 24-hour period from each baby was used for testing in turn, using data from the other 14 babies for training.

From each 24-hour period, a 30 minute section near the start containing only normal dynamics was reserved for calibration (learning normal dynamics according to section 13.4.3). Testing was therefore conducted on the remaining $23\frac{1}{2}$ hour periods.

The quality of the inferences made were evaluated using receiver operating characteristic (ROC) curves for the different inference methods. The ROC curve plots the rate of true positives against the rate of false negatives, such that the area under the curve (AUC) gives a balanced measure of performance even when the class sizes in the testing data are unequal. Another useful statistic which can be obtained from the ROC curve is the Equal Error Rate (EER), which is the error rate for the threshold setting at which the false positive rate is equal to the false negative rate. We give error rates, so smaller numbers are better (some authors give 1 - EER).

Fig. 13.9 shows a plot of AUC against processing time for Gaussian sum filtering, Gaussian sum smoothing using Expectation Correction, and RBPF with varying numbers of particles. Fig. 13.10 shows a corresponding plot for EER against processing time.

Gaussian sum filtering (forward mixture size $I = 1$) had good performance on all four factors. Gaussian sum smoothing (with a forward mixture size $I = 1$ and backward mixture size $J = 1$) with Expectation Correction gave improved performance in the inference of bradycardia, and similar or slightly improved performance for the other three factors.

RBPF results were not so good, even with high numbers of particles, though it is interesting to note that a certain level of accuracy can be achieved with much faster inference than the other methods.

Specific examples of the operation of these models are now given, each using filtered inference. Figures 13.11-13.13 show inferences of switch settings made with the FSKF with Gaussian sum approximation. In each case the switch settings have

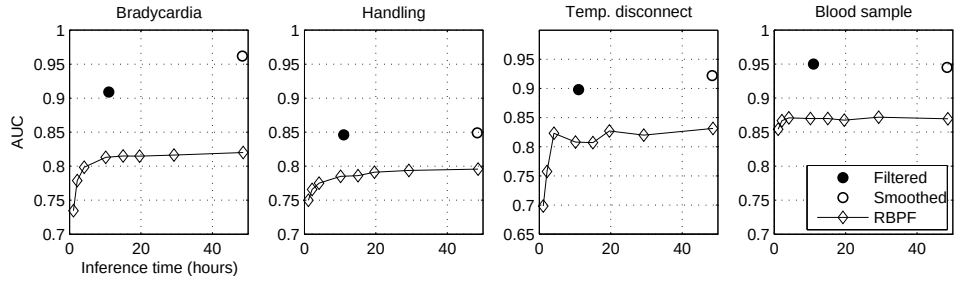


Figure 12.9: AUC for different inference methods on four known factors, for 360 hours of monitoring data. 'Filtered' and 'Smoothed' denote the performance of Gaussian sum filtering and EC smoothing respectively. RBPF inference was done with 5, 10, 20, 50, 75, 100, 150 and 250 particles.

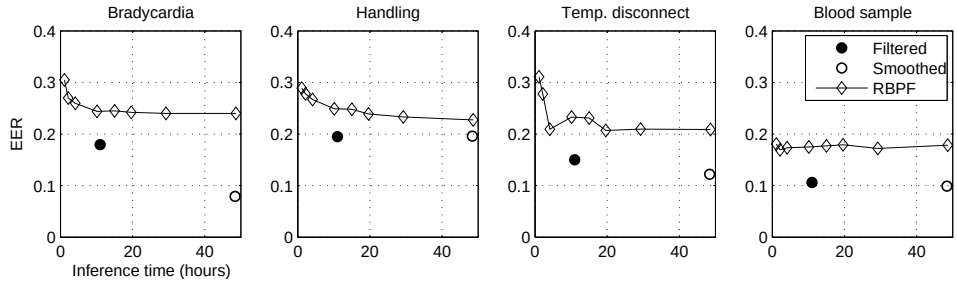
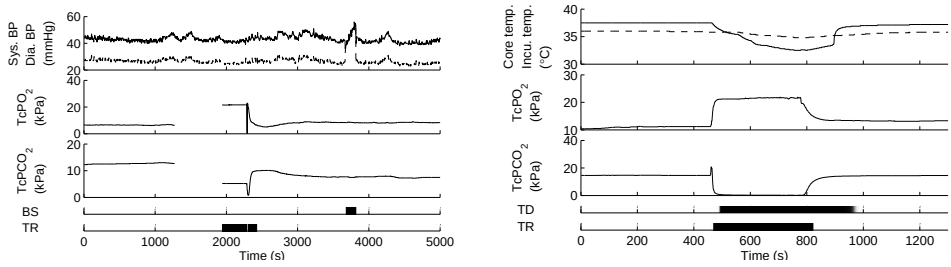
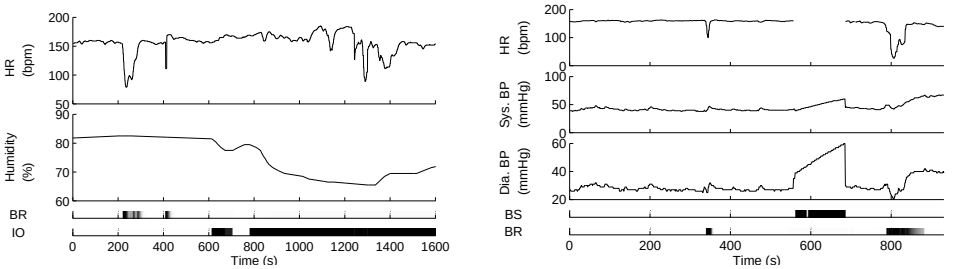


Figure 12.10: EER for different inference methods on four known factors, with data and inference methods as in Figure 13.9.



(a) (b)

Figure 12.11: Inferred filtering distributions of switch settings for two situations involving recalibration of the transcutaneous probe. BS denotes a blood sample, TR denotes a recalibration, and TD denotes a core temperature probe disconnection. In panel (a) the recalibration is preceeded by a dropout, followed by a blood sample. Diastolic BP is shown as a dashed line which lies below the systolic BP plot. Transcutaneous readings drop out at around $t = 1200$ before the recalibration. In panel (b), the solid line shows the core temperature and the dashed line shows incubator temperature. A core temperature probe disconnection is identified correctly, as well as the recalibration. Temperature measurements can occasionally drop below the incubator temperature if the probe is near to the portals; this is accounted for in the model by the system noise term Q.



(a) (b)

Figure 12.12: Inferred filtering distributions of switch settings for two further situations in which there are effects due to multiple known factors. In panel (a) there are incidences of bradycardia, after which the incubator is entered. There is disturbance of heart rate during the period of handling, which is correctly taken to be associated with the handling and not an example of spontaneous bradycardia. In panel (b), bradycardia and blood samples are correctly inferred. During the blood sample, heart rate measurements (supplied by the blood pressure sensor) are interrupted.

been accurately inferred. Figure 13.11 shows examples of transcutaneous probe recalibration, correctly classified in conjunction with a blood sample and a core temperature probe disconnection. In 13.11(b) the recalibration and disconnection begin at around the same time, as a nurse has handled the baby in order to access the transcutaneous probe, causing the temperature probe to become detached.

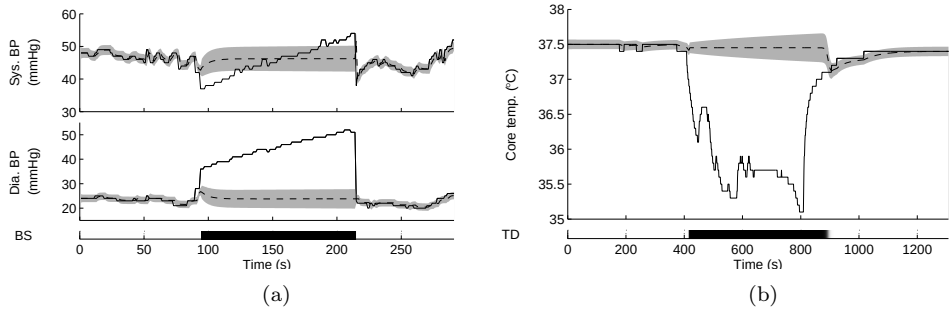
Figure 13.12 shows inference of bradycardia, blood sampling, and handling of the baby. In 13.12(a) that it has been possible to recognise the disturbance of heart rate at $t = 800$ as being caused by handling of the baby, distinguished from the bradycardia earlier where there is no evidence of the incubator having been entered.

For the blood sample and temperature probe disconnection factors, the measurement data bears no relation to the actual physiology, and the model should update the estimated distribution of the true physiology in these situations accordingly. Figure 13.13 contains examples of the inferred distribution of true physiology in data periods in which these two artifacts occur. In each case, once the artifactual pattern has been detected, the physiological estimates remain constant or decay towards a mean. As time passes since the last reliable observation, the variance of the estimates increases towards a steady state.

12.6.2 Inference of novel dynamics

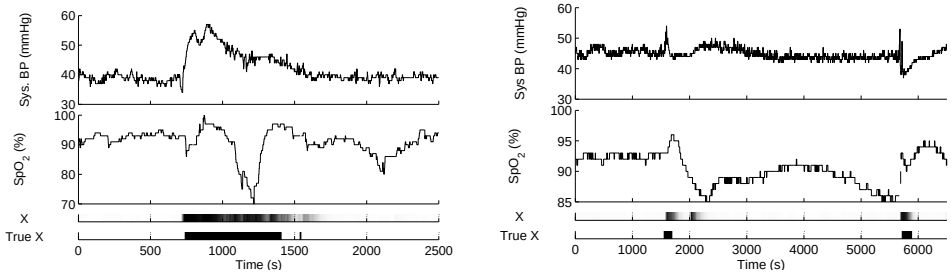
Examples of the operation of the X-factor are shown in Figures 13.14-13.16. We employ two models with different sets of factors. The label '(1)' on the plots denotes the FSLDS with only one factor, the X-factor. The label '(5)' denotes the FSLDS which has five factors – the four known factors and the X-factor. Figure 13.14 shows two examples of inferred switch settings under model (5) for periods in which there are isolated physiological disturbances. Both the posteriors for the X-factor and the gold standard intervals for the 'Abnormal (other)' category are shown. The physiological disturbances in both panels are cardiovascular and have clearly observable effects on the blood pressure and oxygen saturation measurements.

In Figure 13.14(a), the X-factor is triggered by a sudden, prolonged increase in blood pressure and a desaturation, in broad agreement with the ground truth annotation. In Fig. 13.14(a) there are two spikes in BP and shifts in saturation which are picked up by the X-factor, also mainly in agreement with the annotation. A minor turning point in the two channels was also picked up at around $t = 2000$,



(a) (b)

Figure 12.13: Inferred filtering distributions of the true physiological state during artifactual corruption of measurements. Panel (a) shows correct inference of the duration of a blood sample, and panel (b) shows correct inference of a temperature probe disconnection. Measurements are plotted as a solid line, and estimates $\hat{x}_t$ relating to true physiology are plotted as a dashed line with the gray shading indicating two standard deviations. In each case, during the period in which measurements are corrupted the estimates of the true physiology are propagated with increased uncertainty.



(a) (b)

Figure 12.14: Filtering inferences of X-factor switch settings during periods of cardiovascular disturbance, compared to the gold standard annotations.

which was not considered significant in the gold standard (a false positive).

Effects of introducing known factors to model (1) are shown in Figure 13.15. In panel (a), there are two occurrences of spontaneous bradycardia, HR making a transient drop to around 100bpm. The X-factor alone in model (1) picks up this variation. Looking at the inferences from model (5) for the same period, it can be seen that the bradycardia factor provides a better match for the variation, and probability mass shifts correctly: the X-factor is now inactive. In panel (b), a similar effect occurs for a period in which a blood sample occurs. The X-factor picks up the change in dynamics when on its own, and when all factors are present in model (5) the probability mass shifts correctly to the blood sample factor. The blood sample factor is a superior description of the variation, incorporating the knowledge that the true physiology is not being observed, and so able to handle the discontinuity at $t = 900$ effectively.

Figure 13.16 shows examples of inferred switch settings from model (5) in which there are occurrences of both known and unknown types of variation. In Fig. 13.16(a) a bradycardia occurs in the middle of a period of elevated blood pressure and a deep drop in saturation. The bradycardia factor is active for a period which corresponds closely to the ground truth. The X-factor picks up the presence of a change dynamics at about the right time, but its onset is delayed when compared

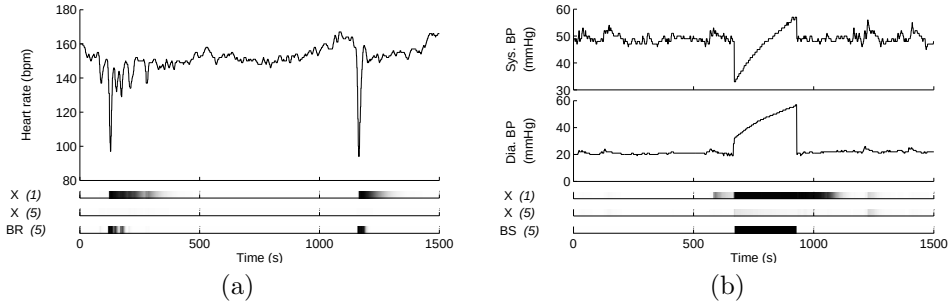


Figure 12.15: Filtering inferences of switch settings for the X-factor, and for known patterns for models (1) and (5). Model (1) contains the X-factor only, whereas model (5) includes the X-factor and all known factors. Panel (a) shows two instances of bradycardia, (b) shows a blood sample.

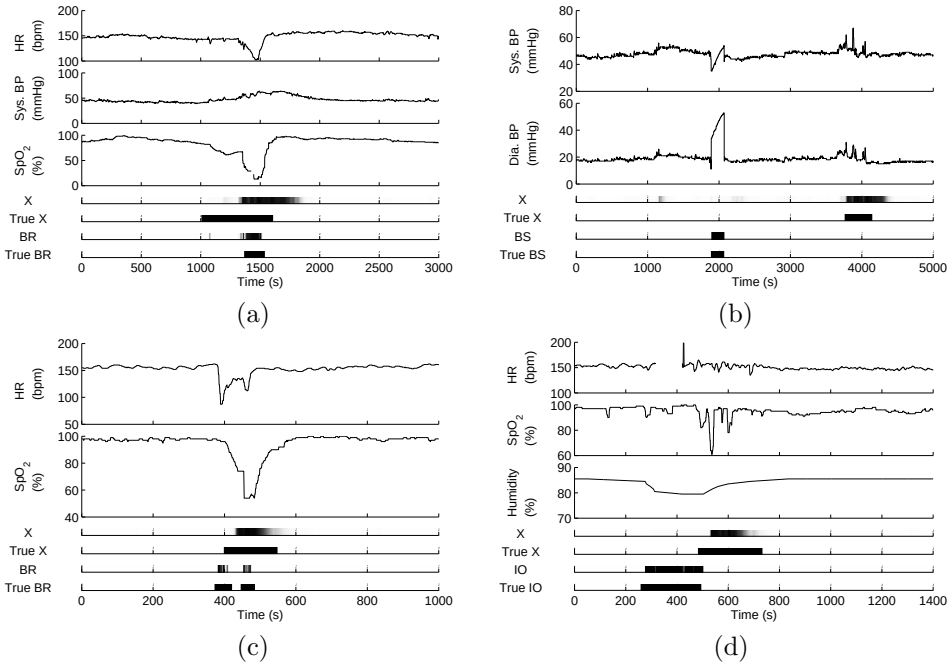


Figure 12.16: Filtering inferences of switch settings for the X-factor, in regions where other factors are active. In panel (a) a bradycardia occurs in conjunction with a rise in blood pressure and deep desaturation. The X-factor is triggered around the right region but is late compared to ground truth. In panel (b), unusual BP variation is correctly classified as being due to a blood sample, followed by variation of unknown cause. Panel (c) shows bradycardia with a desaturation picked up by the X-factor, and (d) shows the X-factor picking up disturbance after the incubator has been entered.

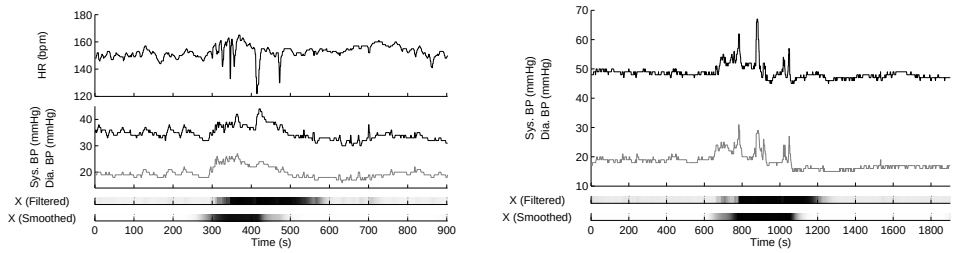


Figure 12.17: Panel (a) shows filtered inference, (b) shows smoothed inference. The gray lines show diastolic blood pressure.

to the ground truth interval. This again highlights a difficulty with filtered inference, since at time just over 1000 it is difficult to tell that this is the beginning of a significant change in dynamics without the benefit of hindsight. In panel (b) a blood sample is correctly picked up by the blood sample factor, while a later period of physiological disturbance on the same measurement channels is correctly picked up by the X-factor. Panel (c) shows another example of the bradycardia factor operating with the X-factor, where this time the onset of the first bradycardia is before the onset of the X-factor. The X-factor picks up a desaturation, a common pattern which is already familiar from panel (a). In panel (d), an interaction between the X-factor and the 'Incubator open' factor can be seen. From time 270 to 1000 the incubator has been opened, and all variation including the spike in HR at $t = 420$ are attributed to handling of the baby. Once the incubator appears to have been closed, further physiological disturbance is no longer explained as an effect of handling and is picked up by the X-factor.

While the previous examples all use filtering inference, Fig. 13.17 shows the effect of using smoothing to infer periods of novelty. Two periods of novel dynamics are shows, with inference results for filtering (upper) and smoothing (lower). Filtered inferences of the X-factor tend to 'trail off', because without information about future observations it is difficult to tell when an unusual period of dynamics has ended. Smoothed inferences correct this, so that the inferred period of normality correspond more closely to periods of clinically significant change.

12.7 Summary

This chapter has presented a general framework for inferring hidden factors from monitoring data, and has shown its successful application to the significant real-world task of monitoring the condition of a premature infant receiving intensive care. We have shown how knowledge engineering and learning can be successfully combined in this framework. Our formulation of an additional factor (the "X-factor") allows the model to handle novel dynamics. Experimental demonstration has shown that these methods are effective when applied to genuine monitoring data.

There are a number of directions in which this work could be continued. The set of known factors presented here is limited, and more could usefully be added to the model given training data. Also, experiments with the X-factor have shown that there are a significant number of non-normal regimes in the data which have not yet been formally analysed. Future work might therefore look at learning what different

regimes are claimed by the X-factor. This could be cast as an unsupervised or semi-supervised learning problem within the model. Another possible avenue would be to look at the incorporation of nonlinear dynamics within the switching state space framework for physiological monitoring, using generalised linear models for state transitions or observations.

Acknowledgments

We thank Neil McIntosh for supplying annotation of the data and providing expert medical input on the experimental design. We also thank Birgit Wefers for supplying additional annotation of the data, and Jim Hunter for modifying the Time Series Workbench software for use in this research. Author JQ was funded by the premature baby charity BLISS. The work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778.

Bibliography

[1] N. de Freitas, R. Dearden, F. Hutter, R. Morales-Menedez, J. Mutch, and D. Poole. Diagnosis by a waiter and a Mars explorer. *Proceedings of the IEEE*, 92(3), 2004.

[2] Z. Ghahramani and G. E. Hinton. Parameter Estimation for Linear Dynamical Systems. Technical report, Department of Computer Science, University of Toronto, 1996.

[3] Z. Ghahramani and G. E. Hinton. Variational learning for switching state-space models. *Neural Computation*, 12(4):963–996, 1998.

[4] Z. Ghahramani and M. Jordan. Factorial Hidden Markov Models. *Machine Learning*, 29:245–273, 1997.

[5] J. R. W. Hunter. TSNet A Distributed Architecture for Time Series Analysis. In N. Peek and C. Combi, editors, *Intelligent Data Analysis in bioMedicine and Pharmacology (IDAMAP 2006)*, pages 85–92, 2006.

[6] U. Lerner and R. Parr. Inference in Hybrid Networks: Theoretical Limits and Practical Algorithms. In *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence*, pages 310–318, 2001.

[7] K. Murphy and S. Russell. Rao-Blackwellised particle filtering for dynamic Bayesian networks. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo in Practice*. Springer-Verlag, 2001.

[8] J. A. Quinn. *Bayesian Condition Monitoring in Neonatal Intensive Care*. PhD thesis, University of Edinburgh, 2007.

[9] J. A. Quinn. Neonatal condition monitoring demonstration code. `http://www.cit.mak.ac.ug/staff/jquinn/software.html`, 2007.

[10] J. A. Quinn, C. K. I. Williams, and N. McIntosh. Factorial switching linear dynamical systems applied to condition monitoring in neonatal intensive care. To appear in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.

[11] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer-Verlag, 2000.

[12] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, 1999.

[13] P. C. Woodland. Hidden Markov models using vector linear prediction and discriminative output functions. In *Proceedings of 1992 IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 509–512. IEEE, 1992.