

Known Unknowns: Novelty Detection in Condition Monitoring

Chris Williams

joint work with John Quinn and Neil McIntosh

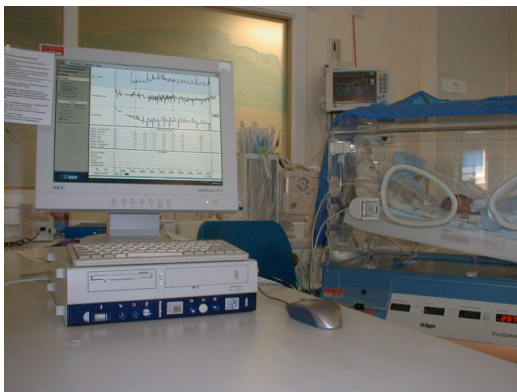
Institute for Adaptive and Neural Computation
School of Informatics, University of Edinburgh, UK

January 2008

Premature Baby Monitoring



Why model this data?

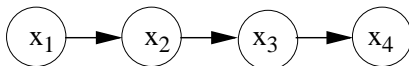


- Artifact corruption, leading to false alarms
- Our aim is to determine the baby's state of health despite these problems

Overview

- The Kalman filter
- Factors
- Factorial switching Kalman filter
- Inference
- Parameter estimation
- Results
- Modelling novel regimes

Autoregressive (AR) Processes



- Scalar AR(1) process

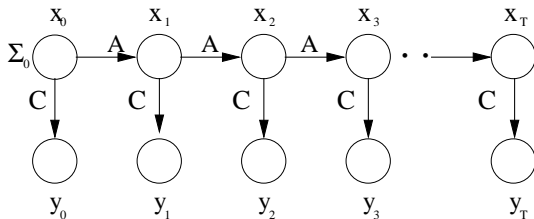
$$x_t = \alpha x_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, \gamma)$$

- Vector AR(1) process

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \Gamma)$$

- A scalar AR(p) process can be written as a vector AR(1) process one stores x_t and the previous $p - 1$ values in \mathbf{x}_t

Kalman Filter



- Dynamical model

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{w}_t$$

where $\mathbf{w}_t \sim N(\mathbf{0}, \Gamma)$ is Gaussian noise, i.e.

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) \sim N(A\mathbf{x}_{t-1}, \Gamma)$$

- A Kalman filter is a HMM with continuous state-space and observations (and Gaussian densities)

- Observation model

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t$$

where $\mathbf{v}_t \sim N(\mathbf{0}, \Sigma)$ is Gaussian noise, i.e.

$$p(\mathbf{y}_t|\mathbf{x}_t) \sim N(C\mathbf{x}_t, \Sigma)$$

- Initialization

$$p(\mathbf{x}_0) \sim N(\boldsymbol{\mu}_0, \Sigma_0)$$

Inference Problem – filtering

- As whole model is Gaussian, only need to compute means and variances

$$p(\mathbf{x}_t | \mathbf{y}_1, \dots, \mathbf{y}_t) \sim N(\boldsymbol{\mu}_t, V_t)$$

$$\boldsymbol{\mu}_t = E[\mathbf{x}_t | \mathbf{y}_1, \dots, \mathbf{y}_t]$$

$$V_t = E[(\mathbf{x}_t - \boldsymbol{\mu}_t)(\mathbf{x}_t - \boldsymbol{\mu}_t)^T | \mathbf{y}_1, \dots, \mathbf{y}_t]$$

- Recursive update split into two parts
- Time update**

$$p(\mathbf{x}_t | \mathbf{y}_1, \dots, \mathbf{y}_t) \rightarrow p(\mathbf{x}_{t+1} | \mathbf{y}_1, \dots, \mathbf{y}_t)$$

- Measurement update**

$$p(\mathbf{x}_{t+1} | \mathbf{y}_1, \dots, \mathbf{y}_t) \rightarrow p(\mathbf{x}_{t+1} | \mathbf{y}_1, \dots, \mathbf{y}_t, \mathbf{y}_{t+1})$$

- Time update

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_t$$

thus

$$\mathbb{E}[\mathbf{x}_{t+1} | \mathbf{y}_1, \dots, \mathbf{y}_t] = A\boldsymbol{\mu}_t$$

$$\text{COV}(\mathbf{x}_{t+1} | \mathbf{y}_1, \dots, \mathbf{y}_t) \stackrel{\text{def}}{=} P_t = AV_tA^T + \Gamma$$

- Measurement update (like posterior in Factor Analysis)

$$\boldsymbol{\mu}_{t+1} = A\boldsymbol{\mu}_t + K_{t+1}(\mathbf{y}_{t+1} - CA\boldsymbol{\mu}_t)$$

$$V_{t+1} = (I - K_{t+1}C)P_t$$

where

$$K_{t+1} = P_t C^T (C P_t C^T + \Sigma)^{-1}$$

K_{t+1} is known as the Kalman gain matrix

Simple example

$$x_{t+1} = x_t + w_t$$

$$w_t \sim N(0, 1)$$

$$y_t = z_t + v_t$$

$$v_t \sim N(0, 1)$$

$$p(z_1) \sim N(0, \sigma^2)$$

In the limit $\sigma^2 \rightarrow \infty$ we find

$$\mu_2 = \frac{5y_2 + 2y_1 + y_0}{8}$$

- Notice how later data has more weight
- Compare $y_{t+1} = y_t$ (so that w_t has zero variance); then

$$\mu_3 = \frac{y_2 + y_1 + y_0}{3}$$

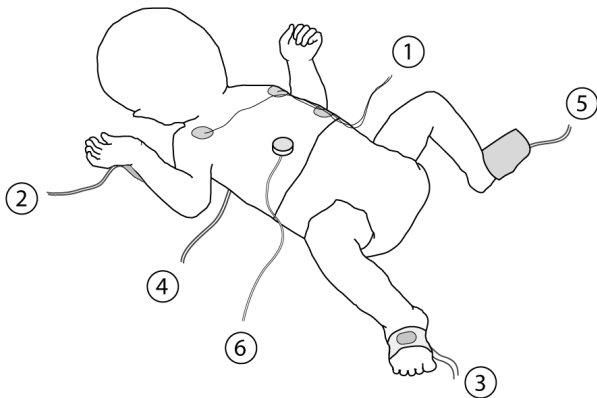
Applications

Much as a coffee filter serves to keep undesirable grounds out of your morning mug, the Kalman filter is designed to strip unwanted noise out of a stream of data.

Barry Cipra, SIAM News 26(5) 1993

- Navigational and guidance systems
- Radar tracking
- Sonar ranging
- Satellite orbit determination

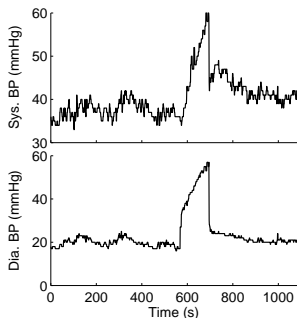
Probes



1. ECG, 2. arterial line, 3. pulse oximeter 4. core temperature, 5. peripheral temperature, 6. transcutaneous probe.

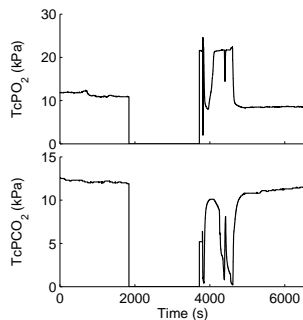
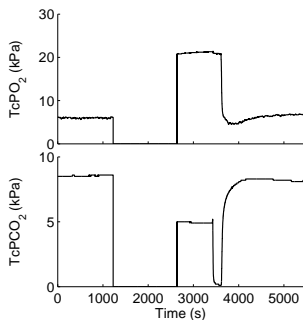
Factors affecting measurements

- The physiological **observations** are affected by different **factors**.
- Factors can be artificial or physiological.
- An arterial blood sample (artifact):



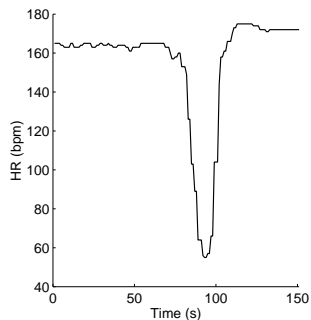
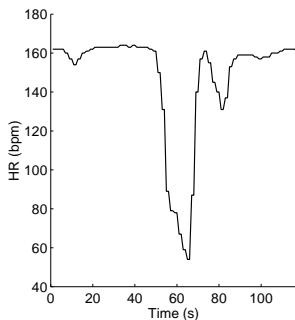
Common factor examples

- Transcutaneous probe recalibration (artifact)

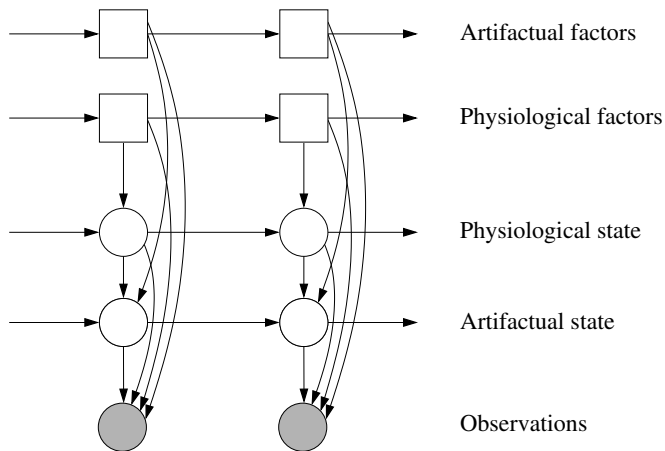


Common factor examples

- Bradycardia (physiological)



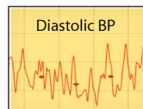
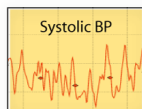
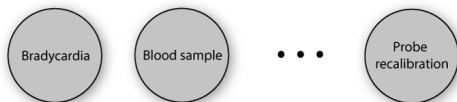
Factorial Switching Kalman Filter



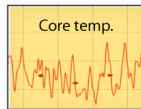
FSKF notation

- s_t is the switch variable, which indexes factor settings, e.g. 'blood sample occurring **and** first stage of TCP recalibration'.
- \mathbf{x}_t is the hidden continuous state at time t . This contains information on the true physiology of the baby, and on the levels of artifactual processes.
- $\mathbf{y}_{1:t}$ are the observations.

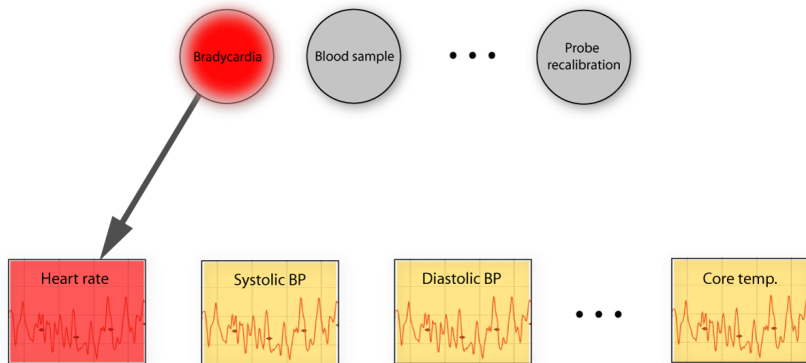
Factor interactions



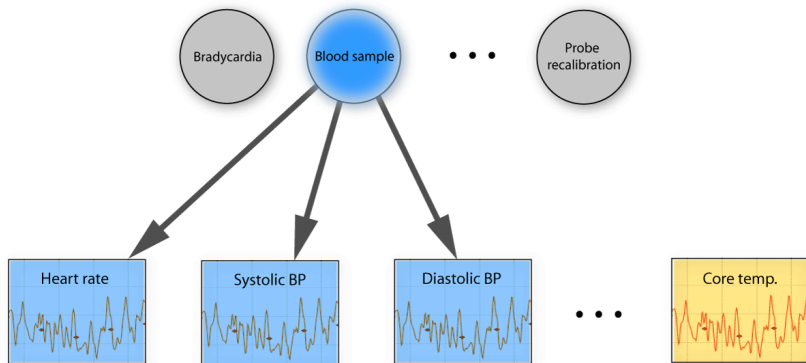
...



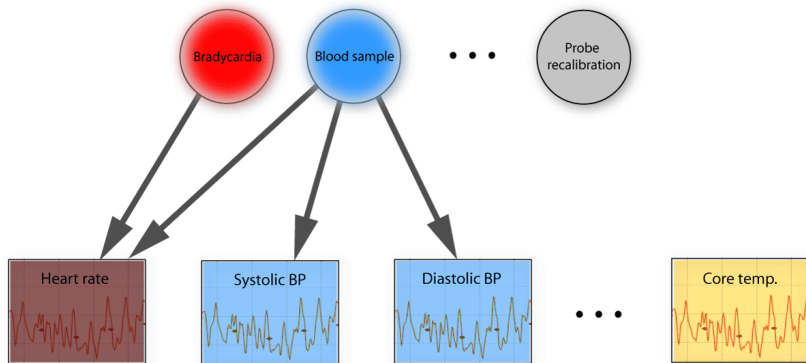
Factor interactions



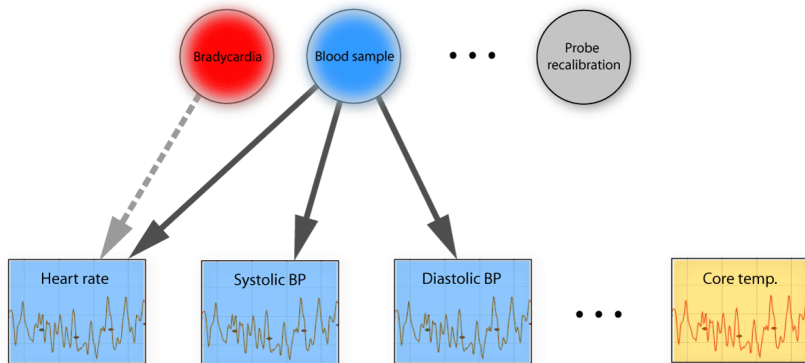
Factor interactions



Factor interactions



Factor interactions



Related work

- Switching linear dynamical models have been studied by many authors, e.g. Alspach and Sorenson (1972), Ghahramani and Hinton (1996).
- Applications include fault detection in mobile robots (de Freitas et al., 2004), speech recognition (Droppo and Acero, 2004), industrial monitoring (Morales-Menedez et al., 2002).
- A two-factor FSKF was used for speech recognition by Ma and Deng (2004). Factorised SKF also used for musical transcription (Cemgil et al., 2006).
- There has been previous work on condition monitoring in the ICU, though we are unaware of any studies that use a FSKF.

Kalman filtering

- Continuous hidden state affects some observations:

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{A}\mathbf{x}_{t-1}, \mathbf{Q})$$

$$\mathbf{y}_t \sim \mathcal{N}(\mathbf{C}\mathbf{x}_t, \mathbf{R})$$

- Kalman filter equations can be used to work compute $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$
- Done iteratively by *predicting* and *updating*

Switching dynamics

- The switch variable s_t selects the dynamics for a particular combination of factor settings:

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{A}^{(s_t)}\mathbf{x}_{t-1}, \mathbf{Q}^{(s_t)})$$

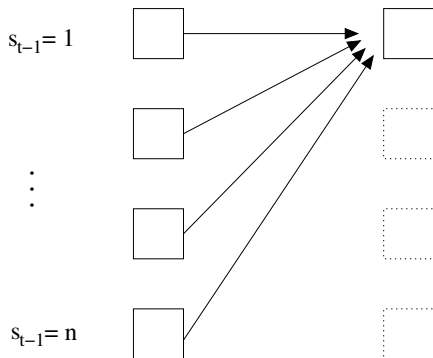
$$\mathbf{y}_t \sim \mathcal{N}(\mathbf{C}^{(s_t)}\mathbf{x}_t, \mathbf{R}^{(s_t)})$$

- For each setting of s_t , the Kalman filter equations give a predictive distribution for \mathbf{x}_t .

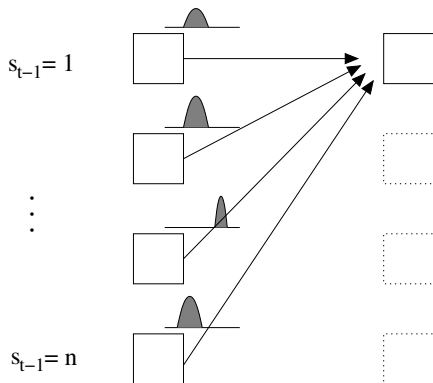
Inference

- For this application, we are interested in filtering, inferring $p(s_t, \mathbf{x}_t | \mathbf{y}_{1:t})$.
- Exact inference is intractable.
- Using two inference methods:
 - Gaussian Sum (Alspach and Sorenson, 1972), analytical approximation
 - Rao-Blackwellised particle filtering.

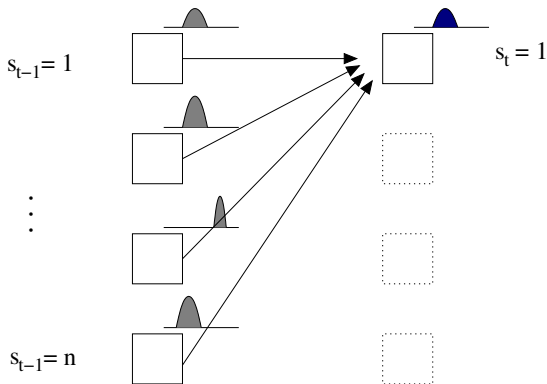
Gaussian Sum approximation



Gaussian Sum approximation



Gaussian Sum approximation

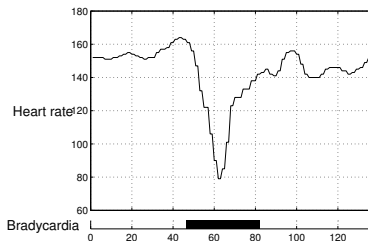


Parameter estimation

- We need to estimate a dynamical model for each continuous state variable for each setting of the factors
- We use AR/ARMA/ARIMA modelling, e.g. an AR(p) process

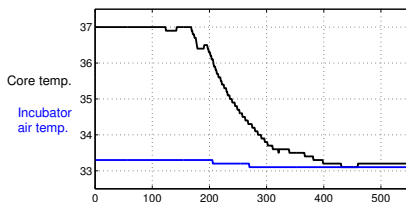
$$x_i(t) = \sum_{j=1}^p \alpha_{ij} x_i(t-j) + \epsilon_t$$

- Fortunately, annotated training data is available



- The hidden continuous state in this application is interpretable, and domain knowledge can be used to help parameterize the dynamical models for each factor.

Parameter estimation example



- For example, we know that the falling temperature measurements caused by a probe disconnection will follow an exponential decay
- Therefore we can model these dynamics as an AR(1) process, and set parameters by solving the Yule-Walker equations.

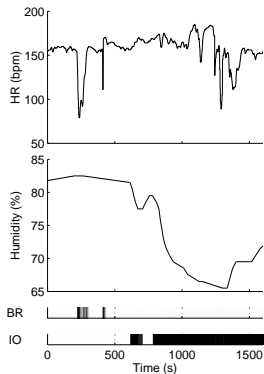
Learning stable physiological dynamics

- Each observation channel has different dynamics when the baby is 'stable' (self regulating) and no artifactual factors are active
- By analysing examples of stable data, dynamical models can be found for each channel with the Box-Jenkins approach and EM.
- For example, a hidden ARIMA(2,1,0) model is a good fit to baseline heart rate data.

Known factor classification demo

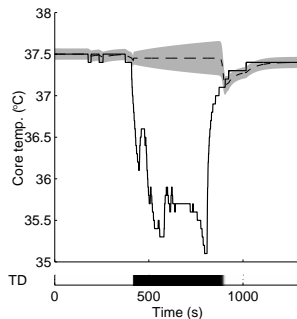
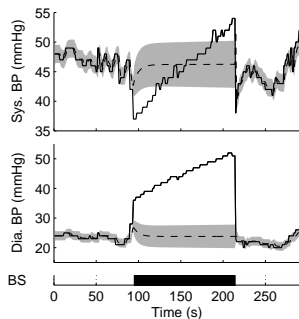
Inference results

- Inference of bradycardia and incubator open factors. Note that heart rate variation while incubator is open is attributed to handling of the baby (BR factor suppressed)



Inference results

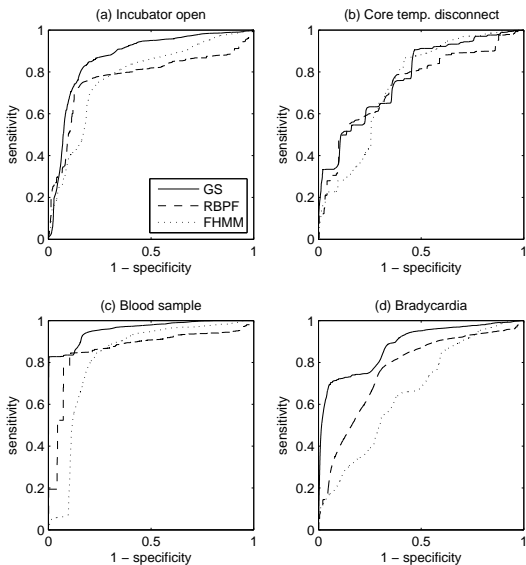
- Can examine variance of estimates of true physiology \hat{x}_t , e.g. for blood sample (left) and temperature probe disconnection (right):



Quantitative Evaluation

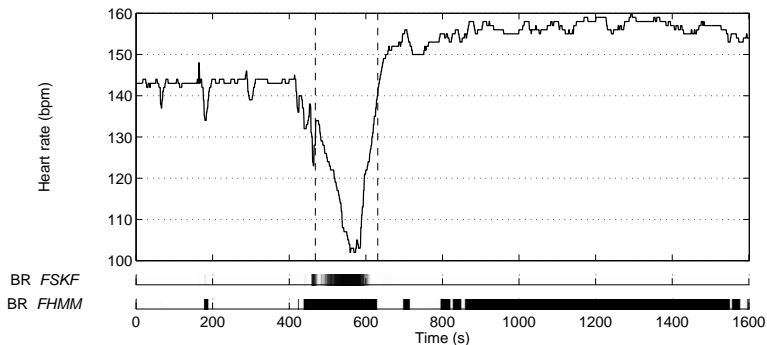
- 3-fold cross validation on 360 hours of monitoring data from 15 babies.
- FHMM has the same factor structure as the FSKF, with no hidden continuous state.

Inference type		Incu. open	Core temp.	Blood sample	Bradycardia
GS	AUC	0.87	0.77	0.96	0.88
	EER	0.17	0.34	0.14	0.25
RBPF	AUC	0.77	0.74	0.86	0.77
	EER	0.23	0.32	0.15	0.28
FHMM	AUC	0.78	0.74	0.82	0.66
	EER	0.25	0.32	0.20	0.37



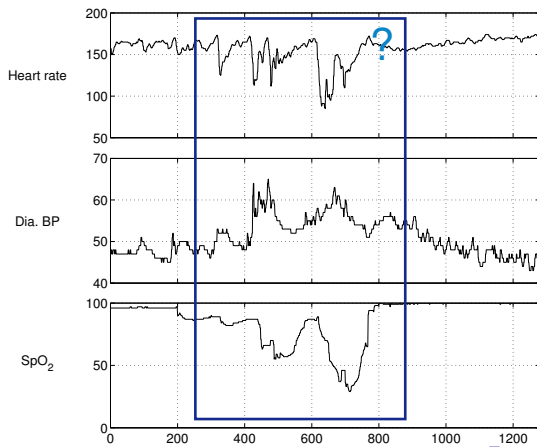
Comparison with FHMM model

- FSKF can handle drift in baseline levels:



Novel dynamics

- There are many other factors influencing the data: drugs, sepsis, neurological problems...



Known Unknowns

- Add a factor to represent abnormal dynamics

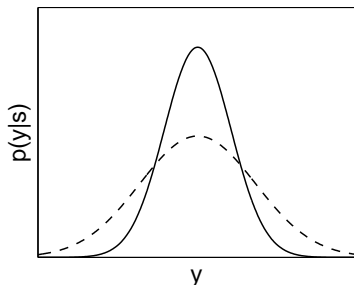
Known Unknowns

- Add a factor to represent abnormal dynamics



X-factor for static 1-D data

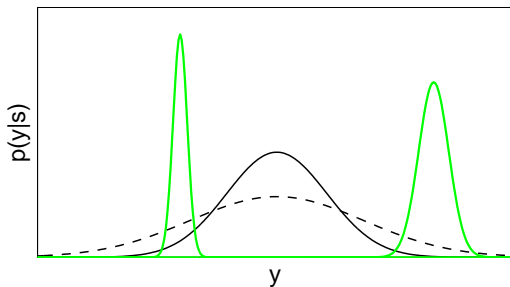
- For static data, we can use a model \mathcal{M}_* representing 'abnormal' data points.



- The high-variance model wins when the data is not well explained by the original model

X-factor with known factors

- The X-factor can be applied to the static data in conjunction with known factors (green):



X-factor for dynamic data

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{A}\mathbf{x}_{t-1}, \mathbf{Q})$$

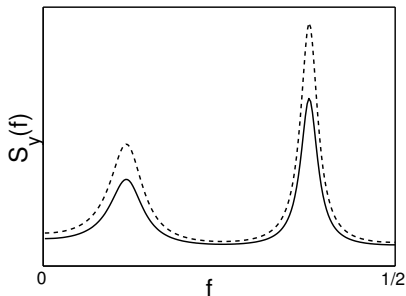
$$\mathbf{y}_t \sim \mathcal{N}(\mathbf{C}\mathbf{x}_t, \mathbf{R})$$

- Can construct an ‘abnormal’ dynamic regime analogously:

Normal dynamics: $\{\mathbf{A}, \mathbf{Q}, \mathbf{C}, \mathbf{R}\}$

X-factor dynamics: $\{\mathbf{A}, \xi\mathbf{Q}, \mathbf{C}, \mathbf{R}\}, \quad \xi > 1.$

Spectral view of the X-factor

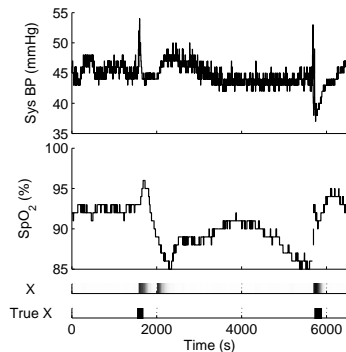
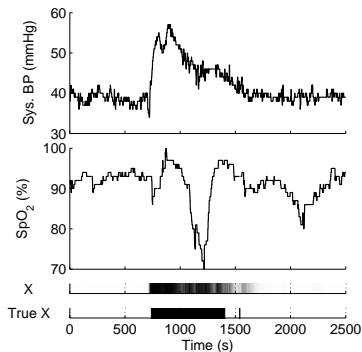


- Plot shows the spectrum of a hidden AR(5) process, and accompanying X-factor
- More power at every frequency
- Dynamical analogue of the static 1-D case

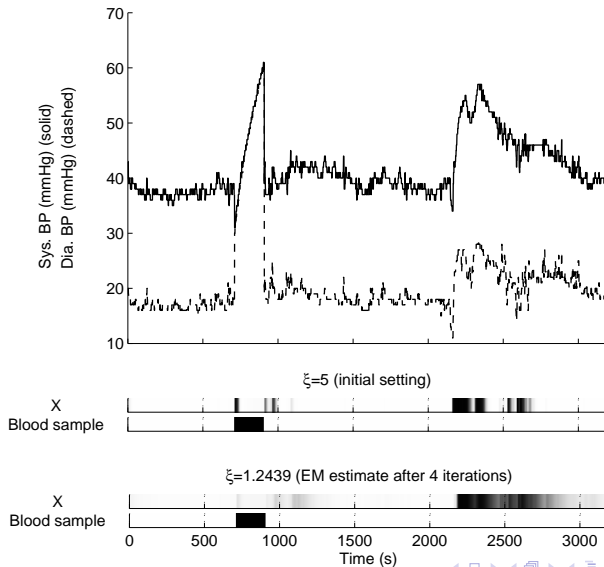
X-factor demo

More inference results

- Classification of periods of clinically significant cardiovascular disturbance:



EM for novel regimes



Conclusions

- FSKF successfully applied to complex physiological monitoring data
- FSKF can be applied more generally to condition monitoring problems
- Interpretable structure
- Knowledge engineering used to parameterize dynamic models
- Allows monitoring of known and novel dynamics (supervised and unsupervised learning)

References

- Factorial Switching Kalman Filters for Condition Monitoring in Neonatal Intensive Care.
Christopher K. I. Williams, John Quinn, Neil McIntosh. In Advances in Neural Information Processing Systems 18, MIT Press (2006)
- Known Unknowns: Novelty Detection in Condition Monitoring.
John A. Quinn, Christopher K. I. Williams. Proc 3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2007)
- Both available from
http://www.dai.ed.ac.uk/homes/ckiw/online_pubs.html

Models for stable physiology

- A fitted model can be verified by comparing real physiological data against a sample from that model, e.g. for heart rate:

