Variational filtering and DEM

EPSRC Symposium Workshop on

Computational Neuroscience

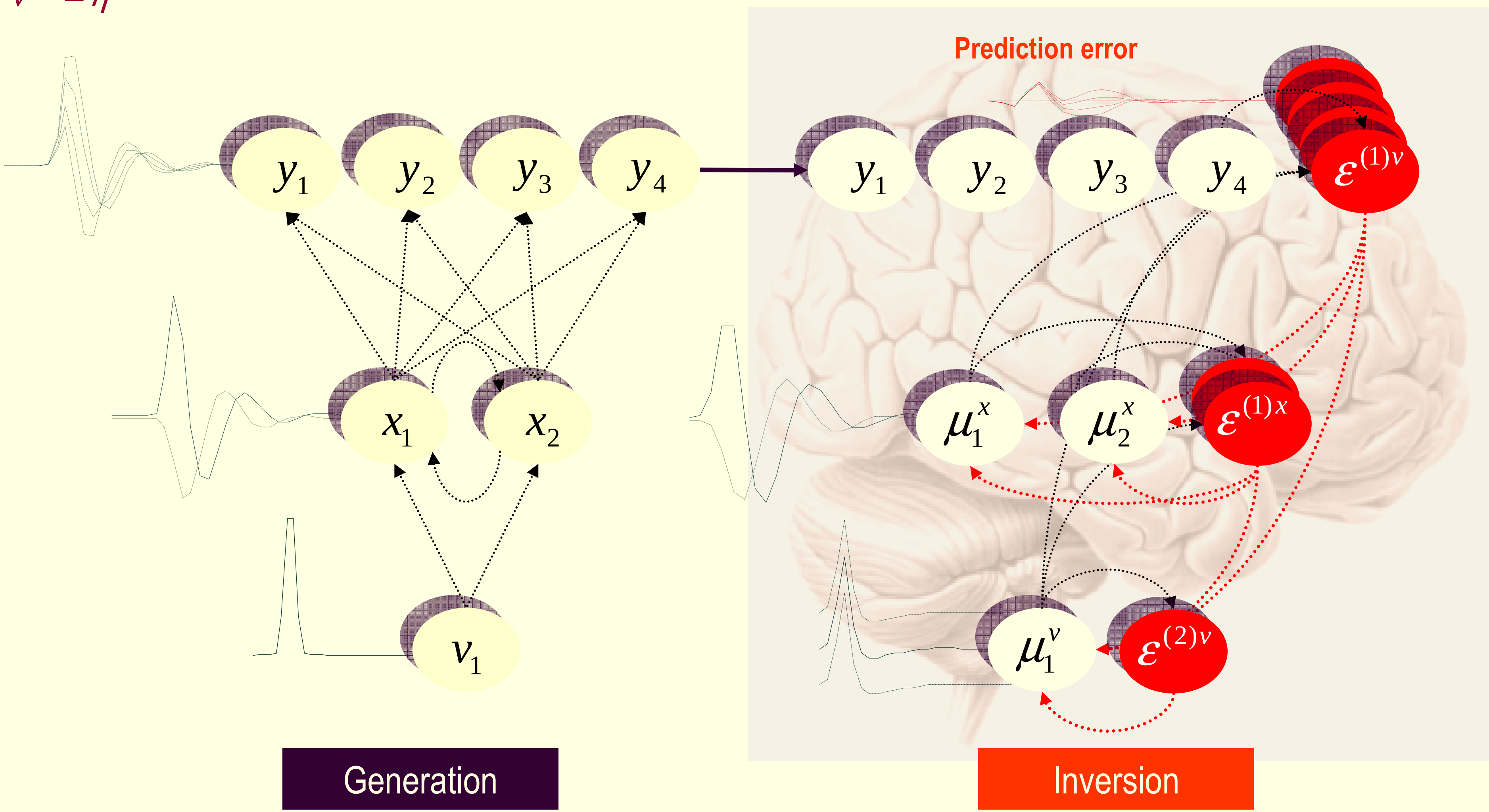
Monday 8 – Thursday 11, December 2008

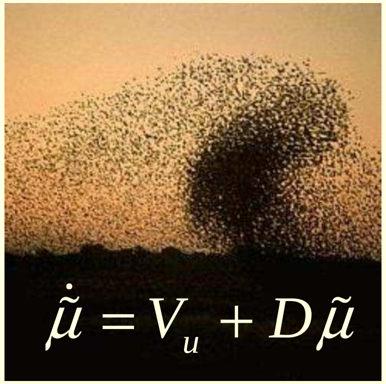
Abstract

This presentation reviews variational treatments of dynamic models that furnish time-dependent conditional densities on the path or trajectory of a system's states and the time-independent densities of its parameters. These obtain by maximizing a variational action with respect to conditional densities. The action or path-integral of free-energy represents a lower-bound on the model's log-evidence or marginal likelihood required for model selection and averaging. This approach rests on formulating the optimization in generalized coordinates of motion. The resulting scheme can be used for online Bayesian inversion of nonlinear hierarchical dynamic causal models and is shown to outperform existing approaches, such as Kalman and particle filtering. Furthermore, it provides for multiple inference on a models states, parameters and hyperparameters using exactly the same principles. Free-form (Variational filtering) and fixed form (Dynamic Expectation Maximization) variants of the scheme will be demonstrated using simulated (bird-song) and real data (from hemodynamic systems studied in neuroimaging).

$$y = g(x^{(1)}, v^{(1)}, \theta^{(1)}) + z^{(1)}$$
$$\dot{x}^{(1)} = f(x^{(1)}, v^{(1)}, \theta^{(1)})$$
$$v^{(1)} = \eta$$

Bayesian filtering and the brain





Overview

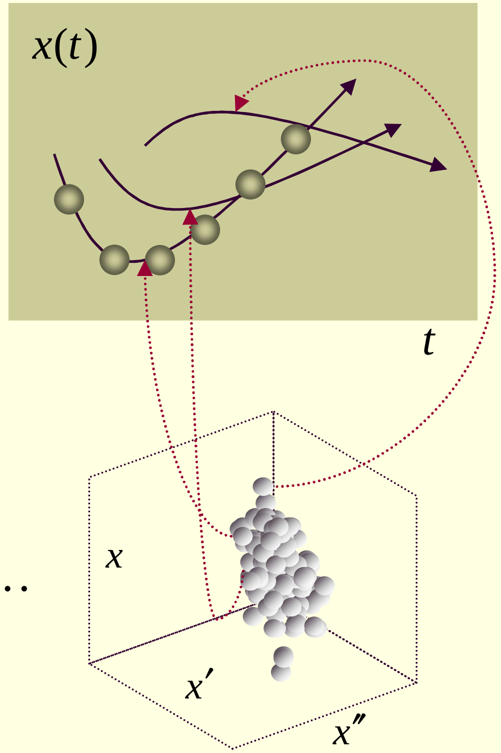
Variational learning and free-energy

Hierarchical dynamic models
- Generalised coordinates (dynamical priors)
- Hierarchal forms (structural priors)

Model inversion
- Variational filtering (free-form)
- Laplace approximation and DEM (fixed-form)

Comparative evaluations

Hemodynamics and Bird songs



$$\dot{x} := \frac{dx}{dt} \quad x' := x^{[1]} \quad f_x := \frac{\partial f}{\partial x} = \nabla_x f \quad \tilde{x} := x, x', x'', x''' \ldots$$

Variational learning, free-energy and action

Aim: *To optimise the path-integral (Action) of a free-energy bound on model evidence with respect to a recognition density* q

$$\partial_t \bar{F} = F(y, q(\vartheta))$$

Free-energy:
$$F = \ln p(y \mid m) - D(q(\vartheta) \,\|\, p(\vartheta \mid y, m))$$
$$= G - H$$

Expected energy:
$$G = \langle \ln p(y, \vartheta) \rangle_q = \langle U(\vartheta) \rangle_q$$

Entropy:
$$H = \langle \ln q(\vartheta) \rangle_q$$

When optimised, the recognition density approximates the true conditional density and Action becomes a bound approximation to the integrated log-evidence; these can then be used for inference on parameters and models respectively

The mean-field approximation

$$q(\vartheta)=\prod_i q(\vartheta^i)=q(u(t))q(\theta)q(\lambda)$$

Lemma : *The free energy is maximised with respect to* $q(\vartheta^i)$ *when* $\delta_{q(\vartheta^i)}\bar{F}=0 \Rightarrow$

Recognition density

$$q(u(t)) \propto \exp(V(t)) \qquad V(t)=\langle U(t)\rangle_{q(\theta,\lambda)}$$

$$q(\theta) \propto \exp(\bar{V}^\theta) \qquad \bar{V}^\theta=\int \langle U(t)\rangle_{q(u,\lambda)}dt+U^\theta$$

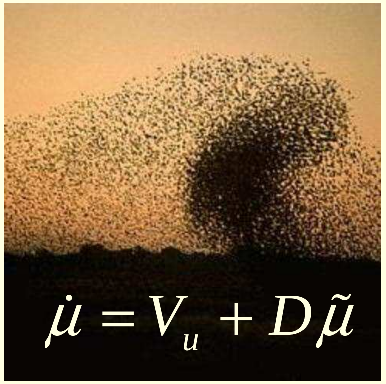
$$q(\lambda) \propto \exp(\bar{V}^\lambda) \qquad \bar{V}^\lambda=\int \langle U(t)\rangle_{q(u,\theta)}dt+U^\lambda$$

Variational energy and actions

Where $U^\theta=\ln p(\theta)$ and $U^\lambda=\ln p(\lambda)$ are the prior energies

and the instantaneous energy is specified by a generative model

$$U(t)=\ln p(\tilde{y},u\,|\,\theta,\lambda)$$



Overview

Variational learning and free-energy

Hierarchical dynamic models
- Generalised coordinates (dynamical priors)
- Hierarchal forms (structural priors)

Model inversion
- Variational filtering (free-form)
- Laplace approximation and DEM (fixed-form)

Comparative evaluations

Hemodynamics and Bird songs

Generalised coordinates and dynamic models

$$y = g(x,v) + z$$
$$\dot{x} = f(x,v) + w$$

$$\begin{aligned} y &= g(x,v) + z \\ y' &= g_x x' + g_v v' + z' \\ y'' &= g_x x'' + g_v v'' + z'' \\ &\vdots \\ \tilde{y} &= \tilde{g} + \tilde{z} \end{aligned}$$

$$\begin{aligned} x' &= f(x,v) + w \\ x'' &= f_x x' + f_v v' + w' \\ x''' &= f_x x'' + f_v v'' + w'' \\ &\vdots \\ D\tilde{x} &= \tilde{f} + \tilde{w} \end{aligned}$$

$$D = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} \otimes I$$

$$\begin{aligned} g &= g(x,v) \\ g' &= g_x x' + g_v v' \\ g'' &= g_x x'' + g_v v'' \\ &\vdots \end{aligned}$$

$$\begin{aligned} f &= f(x,v) \\ f' &= f_x x' + f_v v' \\ f'' &= f_x x'' + f_v v'' \\ &\vdots \end{aligned}$$

Likelihood

$$p(\tilde{y} \mid \tilde{x}, \tilde{v}) = N(\tilde{g}, \Sigma^z)$$

(Dynamical) prior

$$p(D\tilde{x} \mid \tilde{v}) = N(\tilde{f}, \Sigma^w)$$

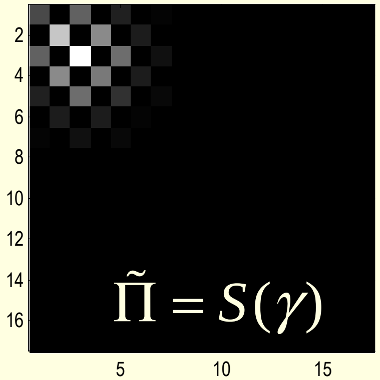
Energies and generalised precisions

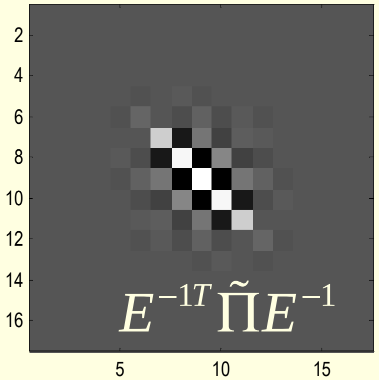
Instantaneous energy
a simple function of prediction error

$$U(t) := \ln p(\tilde{y}, u) = \ln p(\tilde{y} \mid \tilde{x}, \tilde{v})\, p(\tilde{x} \mid \tilde{v})\, p(\tilde{v})$$
$$= \tfrac{1}{2} \ln \left|\tilde{\Pi}\right| - \tfrac{1}{2} \tilde{\varepsilon}^T \tilde{\Pi} \tilde{\varepsilon}$$

$$\tilde{\Pi} = \begin{bmatrix} \tilde{\Pi}^z & \\ & \tilde{\Pi}^w \end{bmatrix} \quad \tilde{\varepsilon} = \begin{bmatrix} \tilde{\varepsilon}^v = \tilde{y} - \tilde{g} \\ \tilde{\varepsilon}^x = D\tilde{x} - \tilde{f} \end{bmatrix}$$

Precision matrices in generalised coordinates and time





$$\vec{y} = \tilde{E}\tilde{y}(t)$$

$$\tilde{\Pi} = S(\gamma) \otimes \Pi(\lambda)$$

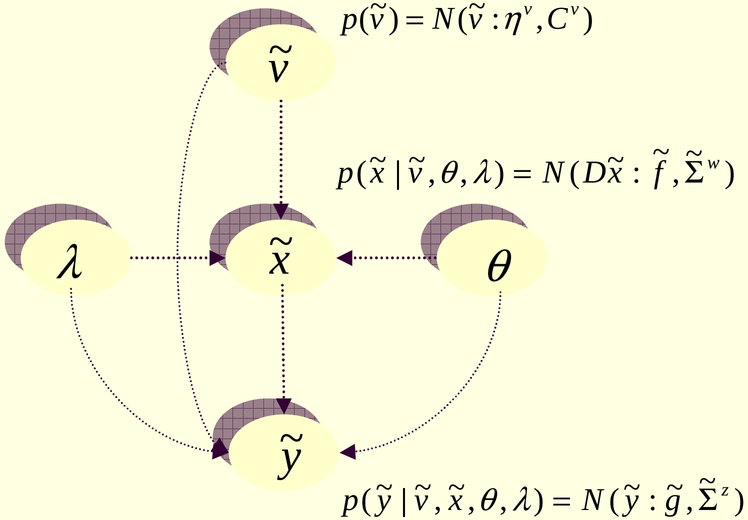
$$S(\gamma)^{-1} = \begin{bmatrix} 1 & 0 & \ddot{\rho}(0) & \cdots \\ 0 & -\ddot{\rho}(0) & 0 & \\ \ddot{\rho}(0) & 0 & \ddot{\ddot{\rho}}(0) & \\ \vdots & & & \ddots \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\frac{1}{2}\gamma & \cdots \\ 0 & \frac{1}{2}\gamma & 0 & \\ -\frac{1}{2}\gamma & 0 & \frac{3}{4}\gamma^2 & \\ \vdots & & & \ddots \end{bmatrix}$$

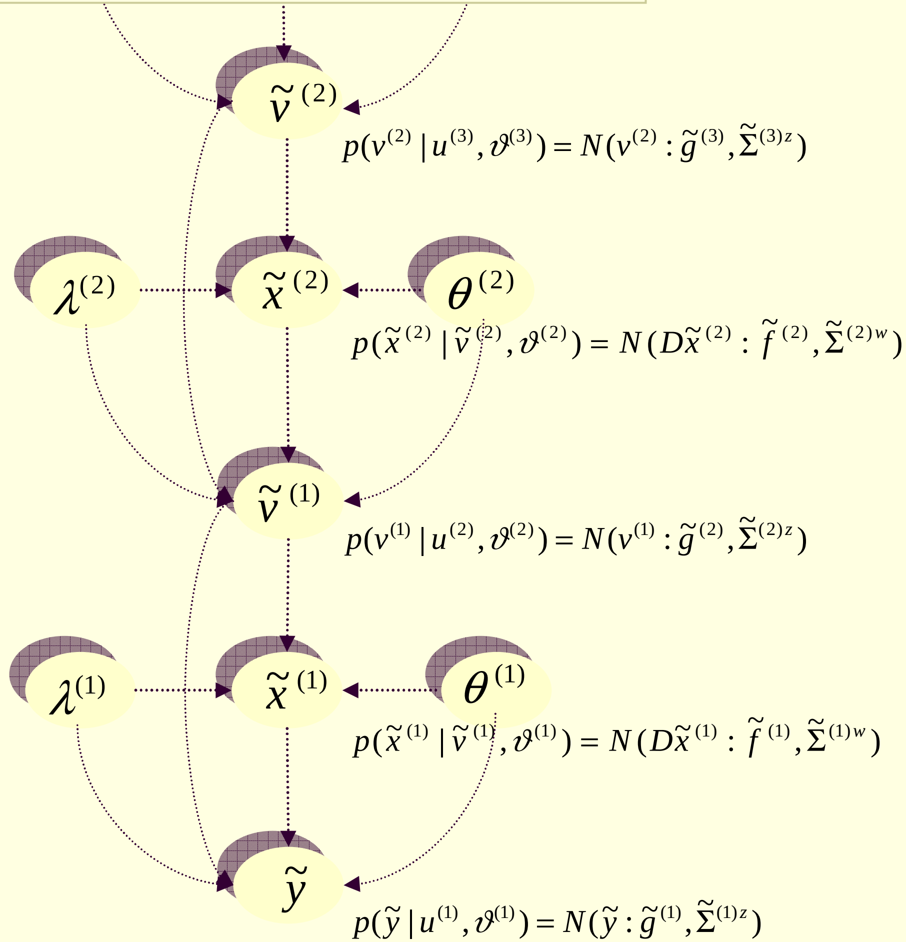
General and Gaussian forms

Hierarchical forms

$$\begin{aligned} y &= g(x,v) + z \\ \dot{x} &= f(x,v) + w \end{aligned} \quad \Rightarrow \quad \begin{aligned} \tilde{y} &= \tilde{g} + \tilde{z} \\ D\tilde{x} &= \tilde{f} + \tilde{w} \end{aligned}$$

$$\begin{aligned} v^{(i-1)} &= g(x^{(i)}, v^{(i)}) + z^{(i)} \\ \dot{x}^{(i)} &= f(x^{(i)}, v^{(i)}) + w^{(i)} \end{aligned} \quad \Rightarrow \quad \begin{aligned} \tilde{v}^{(i-1)} &= \tilde{g}^{(i)} + \tilde{z}^{(i)} \\ D\tilde{x}^{(i)} &= \tilde{f}^{(i)} + \tilde{w}^{(i)} \end{aligned}$$





$$\vartheta = \{u, \theta, \lambda\}$$

$$u = \{\tilde{x}, \tilde{v}\}$$

Hierarchal forms and empirical priors

$$
\begin{aligned}
y &= g(x^{(1)}, v^{(1)}) + z^{(1)} \\
\dot{x}^{(1)} &= f(x^{(1)}, v^{(1)}) + w^{(1)} \\
&\vdots \\
v^{(i-1)} &= g(x^{(i)}, v^{(i)}) + z^{(i)} \\
\dot{x}^{(i)} &= f(x^{(i)}, v^{(i)}) + w^{(i)} \\
&\vdots \\
v^{(m)} &= \eta + z^{(m+1)}
\end{aligned}
$$

$$
\begin{aligned}
U(t) &= \ln p(\tilde{y}, \tilde{x}, \tilde{v} \mid \theta, \lambda) \\
&= \tfrac{1}{2} \ln \left| \tilde{\Pi} \right| - \tfrac{1}{2} \tilde{\varepsilon}^T \tilde{\Pi} \tilde{\varepsilon}
\end{aligned}
$$

The instantaneous energy function of prediction error

$$
\varepsilon^v = \begin{bmatrix} y \\ v^{(1)} \\ \vdots \\ v^{(m)} \end{bmatrix} - \begin{bmatrix} g^{(1)} \\ \vdots \\ g^{(m)} \\ \eta \end{bmatrix}
$$

$$
p(\tilde{y}, \tilde{x}, \tilde{v}) = p(\tilde{y} \mid \tilde{x}, \tilde{v})\, p(\tilde{x}, \tilde{v})
$$

$$
p(\tilde{x}, \tilde{v}) = p(\tilde{v}^{(m)}) \prod_{i=1}^{m-1} p(\tilde{x}^{(i)} \mid \tilde{v}^{(i)})\, p(\tilde{v}^{(i)} \mid \tilde{x}^{(i+1)}, \tilde{v}^{(i+1)})
$$

$$
p\left(\tilde{x}^{(i)} \mid \tilde{v}^{(i)}\right) = N(D\tilde{x}^{(i)} : \tilde{f}^{(i)}, \tilde{\Sigma}^{(i)w})
$$

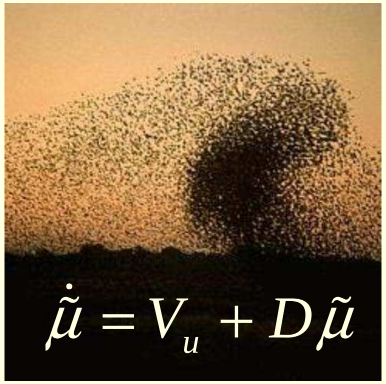
Dynamical priors (empirical)

$$
p\left(\tilde{v}^{(i)} \mid \tilde{x}^{(i+1)}, \tilde{v}^{(i+1)}\right) = N(\tilde{v}^{(i)} : \tilde{g}^{(i+1)}, \tilde{\Sigma}^{(i)z})
$$

Structural priors (empirical)

$$
p\left(\tilde{v}^{(m)}\right) = N(\tilde{v}^{(m)} : \tilde{\eta}, \tilde{\Sigma}^{v})
$$

Priors (full)



Overview

Variational learning and free-energy

Hierarchical dynamic models
- Generalised coordinates (dynamical priors)
- Hierarchal forms (structural priors)

Model inversion
- Variational filtering (free-form)
- Laplace approximation and DEM (fixed-form)

Comparative evaluations

Hemodynamics and Bird songs

Variational filtering (free-form)

Aim: To approximate the recognition density given the variational (instantaneous) energy

Lemma: $q(u(t)) \propto \exp(V(t))$ *is the stationary solution, in a moving frame of reference, for an ensemble of particles, whose equations of motion and ensemble dynamics are*

$$\begin{aligned}\dot{u} &= V(t)_v + \mu' + \Gamma \\ \dot{u}' &= V(t)_{v'} + \mu'' + \Gamma \\ \dot{u}'' &= \ldots\end{aligned} \quad \Rightarrow \quad \boxed{\dot{u} = V(t)_u + D\tilde{\mu} + \Gamma}$$

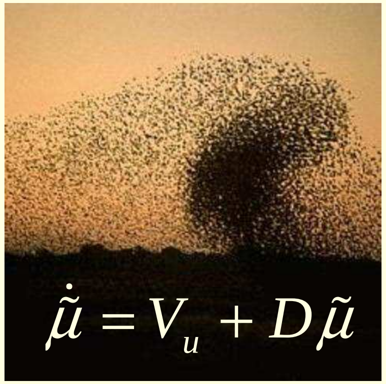
$$\dot{q} = \nabla_u \cdot [\nabla_u q - q\nabla_u V(t)] - \nabla_u \cdot qD\tilde{\mu}$$

Proof: Substituting the recognition density $q \propto \exp(V(t))$ gives

$$\dot{q} = -\nabla_u \cdot qD\tilde{\mu}$$

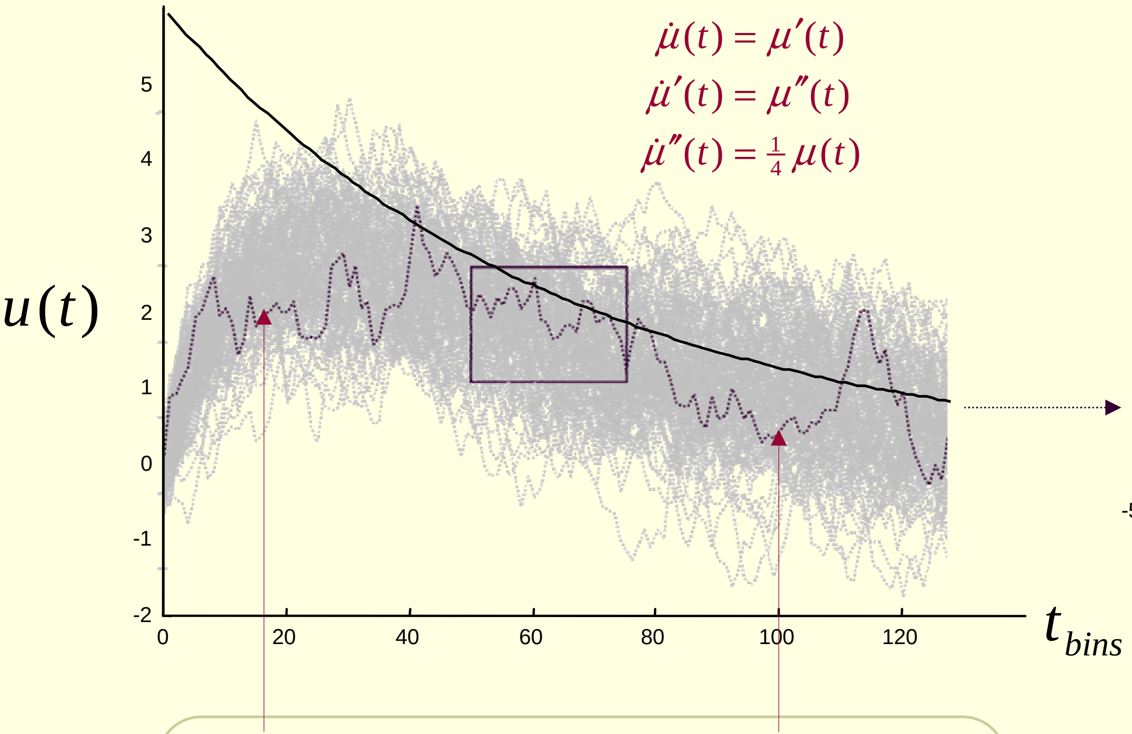
This describes a stationary density under a moving frame of reference, with velocity $D\tilde{\mu}$ as seen using the co-ordinate transform

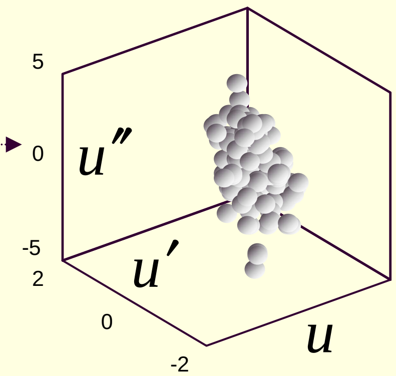
$$\begin{aligned}\dot{\upsilon} &= D\tilde{\mu} \Rightarrow \\ \dot{q}(\upsilon) &= \dot{q}(u) + \nabla_u \cdot q\dot{\upsilon} = 0\end{aligned}$$



A toy example

$$V(\tilde{u},t) = 4(u-\mu)^2 + (u'-\mu')^2 + \tfrac{1}{4}(u''-\mu'')^2$$
$$\dot{\mu}(t) = \mu'(t)$$
$$\dot{\mu}'(t) = \mu''(t)$$
$$\dot{\mu}''(t) = \tfrac{1}{4}\mu(t)$$





$$\begin{aligned} \dot{u} &= V(t)_u + \mu' + \Gamma(t) & \dot{\mu} &= \mu' \\ \dot{u}' &= V(t)_{u'} + \mu'' + \Gamma(t) \quad \Rightarrow & \dot{\mu}' &= \mu'' = \ddot{\mu} \\ \dot{u}'' &= \ldots & \dot{\mu}'' &= \ldots \end{aligned}$$

Optimizing free-energy under the Laplace approximation

Mean-field approximation: $q(\vartheta) = \prod_i q(\vartheta^i) = q(u(t))q(\theta)q(\lambda)$

Laplace approximation: $q(\vartheta^i) = N(\mu^i, \Sigma^i)$

The Laplace approximation enables us the specify the sufficient statistics of the recognition density very simply

Conditional modes	Conditional precisions
$\tilde{\mu}(t) = \arg\max_u V(t)$	$\Pi^u = -U(t)_{uu}$
$\mu^\theta = \arg\max_\theta \bar{V}^\theta$	$\Pi^\theta = -\int U(t)_{\theta\theta} dt - U^\theta_{\theta\theta}$
$\mu^\lambda = \arg\max_\lambda \bar{V}^\lambda$	$\Pi^\lambda = -\int U(t)_{\lambda\lambda} dt - U^\lambda_{\lambda\lambda}$

Under these approximations, all we need to do is optimise the conditional modes

Fixed form schemes … a gradient ascent in moving coordinates

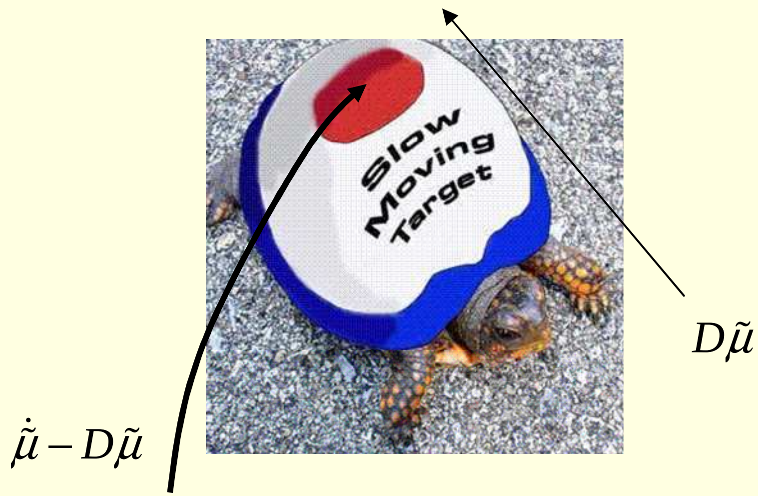
Taking the expectation of the ensemble dynamics, we get:

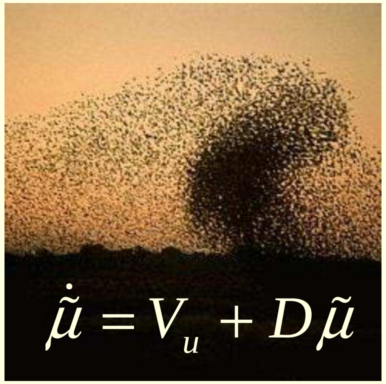
$$\dot{u} = V(t)_u + D\tilde{\mu} + \Gamma \Rightarrow$$
$$\dot{\tilde{\mu}} = V(t)_u + D\tilde{\mu} \Leftrightarrow \dot{\tilde{\mu}} - D\tilde{\mu} = V(t)_u$$

Here, $\dot{\tilde{\mu}} - D\tilde{\mu}$ can be regarded as a gradient ascent in a frame of reference that moves along the trajectory encoded in generalised coordinates. The stationary solution, in this moving frame of reference, maximises variational action.

$$\dot{\tilde{\mu}} - D\tilde{\mu} = 0 \Rightarrow$$
$$\partial_u V(t) = 0 \Leftrightarrow \delta_u \bar{V}^u = 0$$

by the Fundamental lemma; c.f., Hamilton's principle of stationary action.





$$\dot{\tilde{\mu}} = V_u + D\tilde{\mu}$$

Dynamic expectation maximization

D-Step
inference
$q(u(t))$

$$\Im = V(t)_{uu} + D$$
$$\Delta\tilde{\mu} = (\exp(\Im) - I)\Im^{-1}(V(t)_u + D\tilde{\mu})$$
$$\Pi^u = -U(t)_{uu}$$

local linearisation (Ozaki 1992)

E-Step
learning
$q(\theta)$

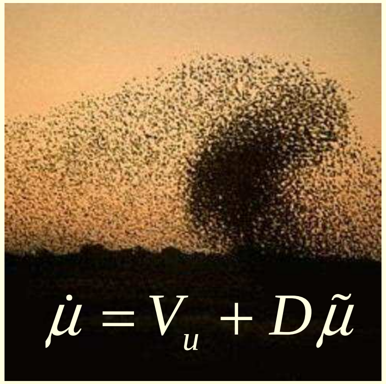
$$\Delta\mu^\theta = -\bar{V}_{\theta\theta}^{\theta\,-1}\bar{V}_\theta^\theta$$
$$\Pi^\theta = -\bar{U}_{\theta\theta}^\theta$$

M-Step
uncertainty
$q(\lambda)$

$$\Delta\mu^\lambda = -\bar{V}_{\lambda\lambda}^{\lambda\,-1}\bar{V}_\lambda^\lambda$$
$$\Pi^\lambda = -\bar{U}_{\lambda\lambda}^\lambda$$

$$q(\vartheta) = q(u(t))q(\theta)q(\lambda)$$

A dynamic recognition system that minimises prediction error



Overview

Variational learning and free-energy

Hierarchical dynamic models
- Generalised coordinates (dynamical priors)
- Hierarchal forms (structural priors)

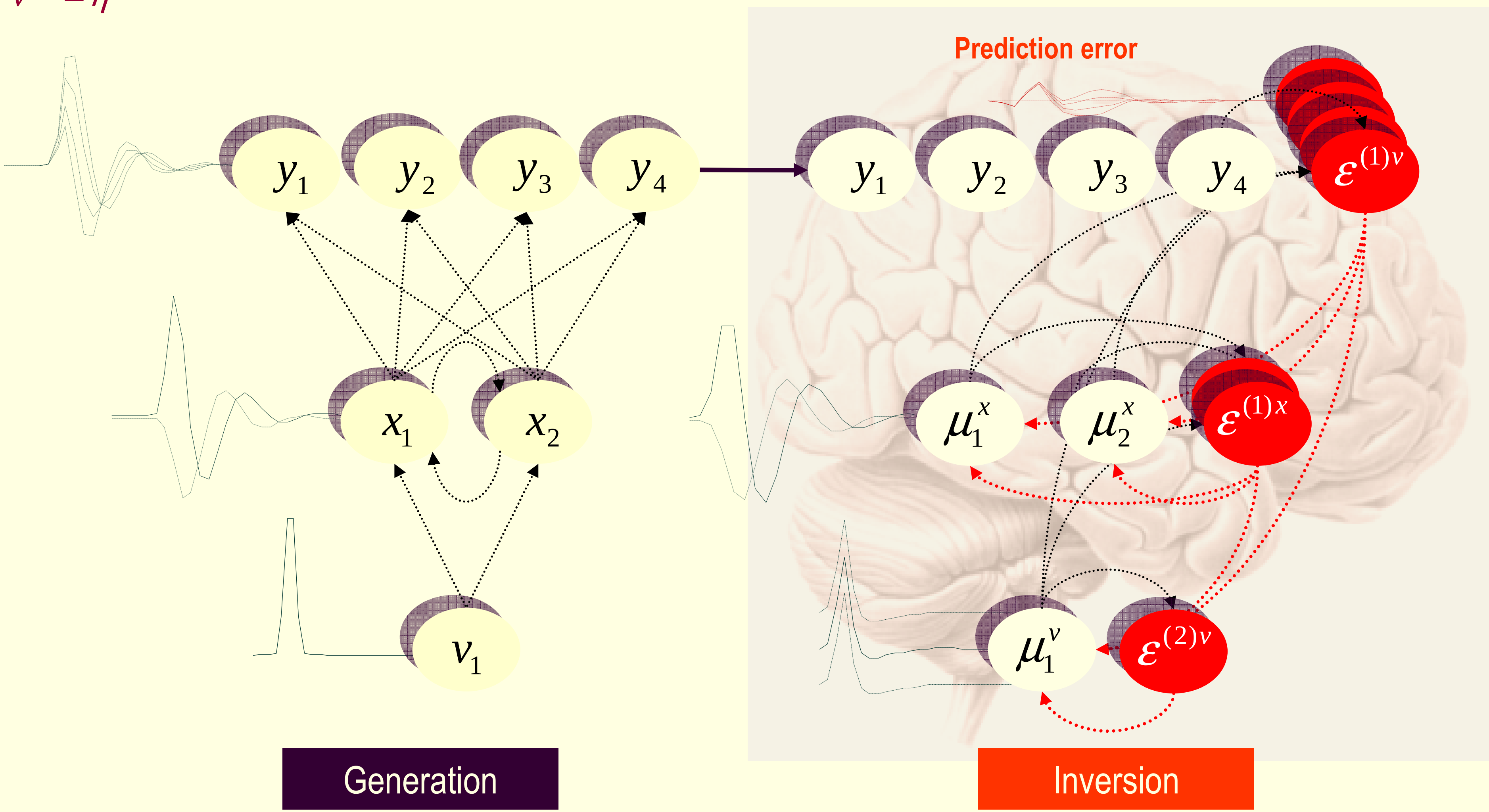
Model inversion
- Variational filtering (free-form)
- Laplace approximation and DEM (fixed-form)

Comparative evaluations

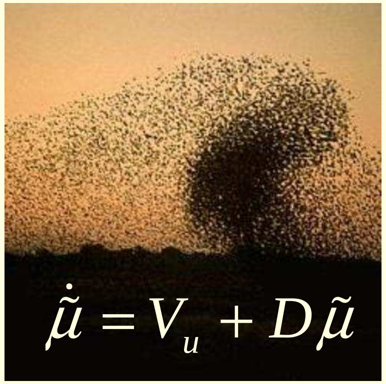
Hemodynamics and Bird songs

$$y = g(x^{(1)}, v^{(1)}, \theta^{(1)}) + z^{(1)}$$
$$\dot{x}^{(1)} = f(x^{(1)}, v^{(1)}, \theta^{(1)})$$
$$v^{(1)} = \eta$$

A linear convolution model

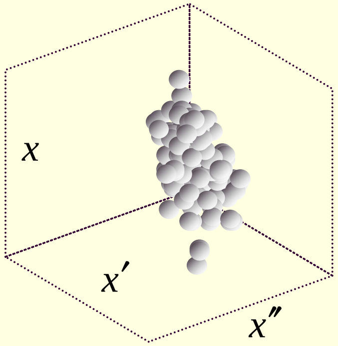


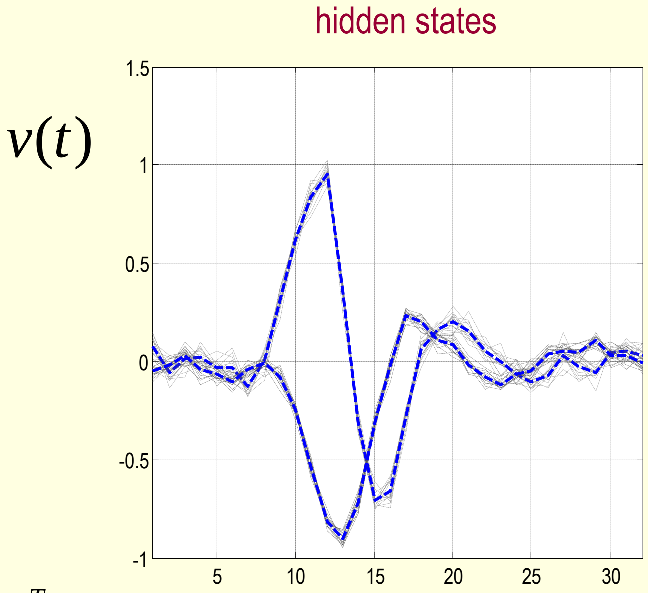
Variational filtering on states and causes

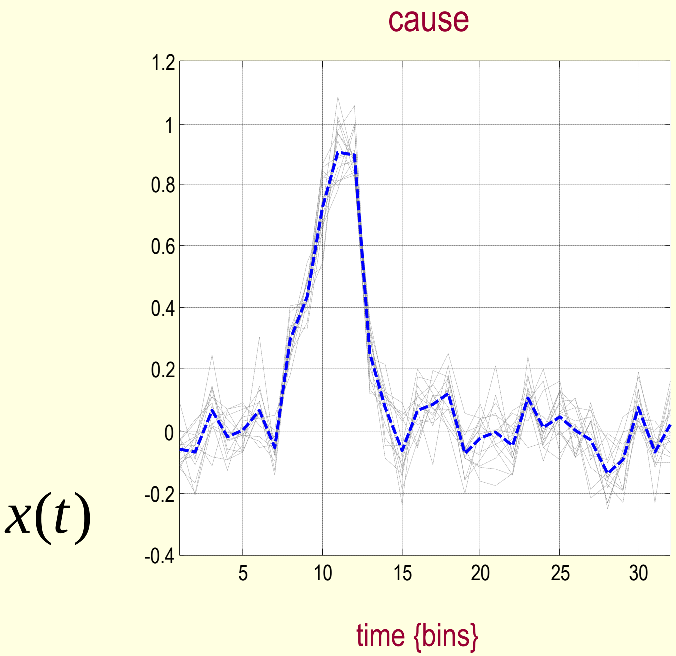


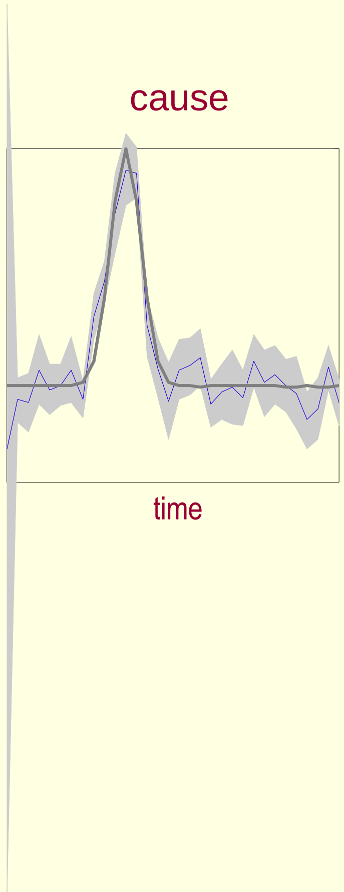
$$\dot{u}(t) = V(t)_u + D\tilde{\mu}(t) + \Gamma(t)$$

$$u(t) = [v, v', v'', \ldots x, x', x'', \ldots]^T$$



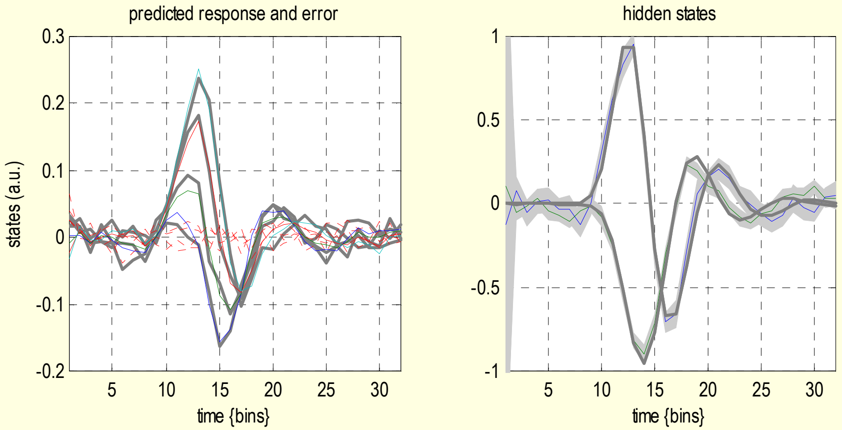


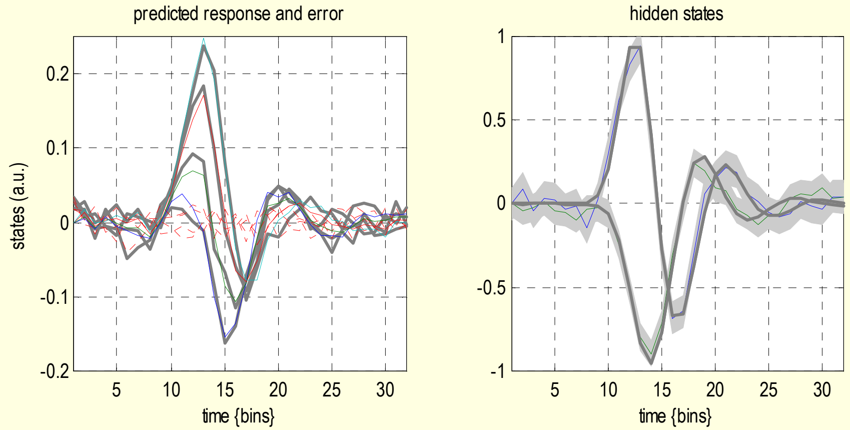




$$\dot{u}(t) = V(t)_u + D\tilde{\mu}(t) + \Gamma(t)$$

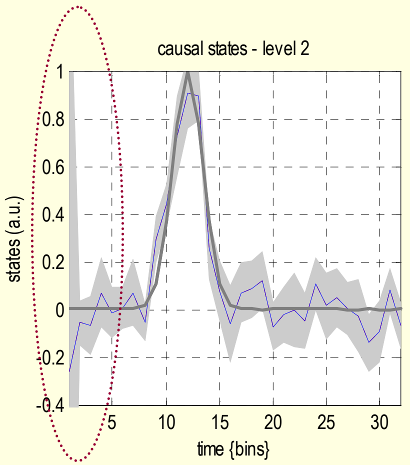
$$\dot{\tilde{\mu}}(t) = V(t)_u + D\tilde{\mu}(t)$$

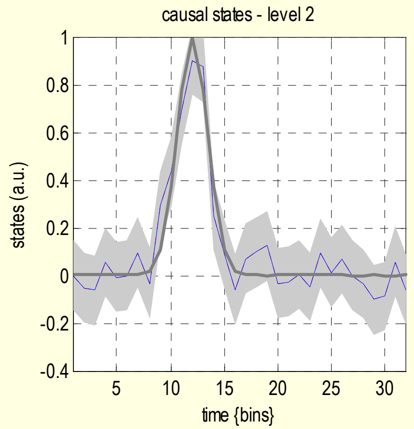


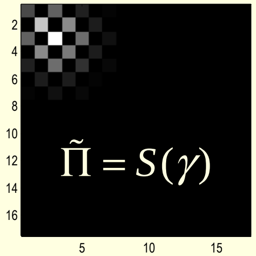


Linear deconvolution with variational filtering (SDE) – free-form

Linear deconvolution with Dynamic expectation maximisation (ODE) – fixed-form

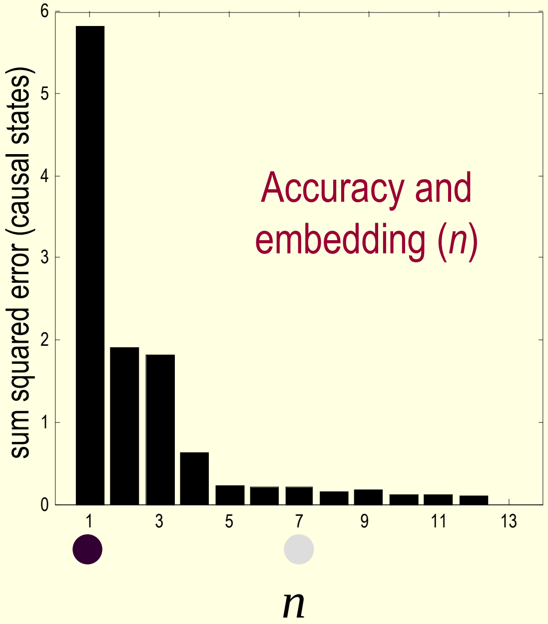


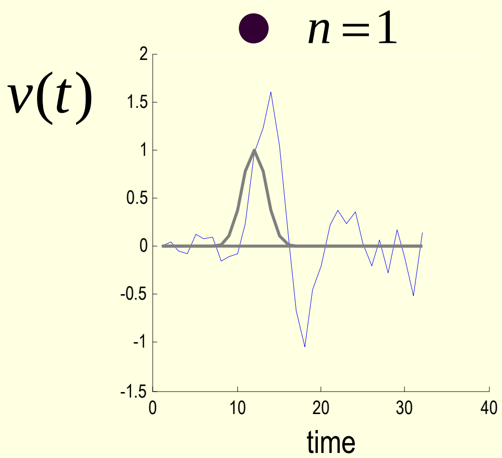


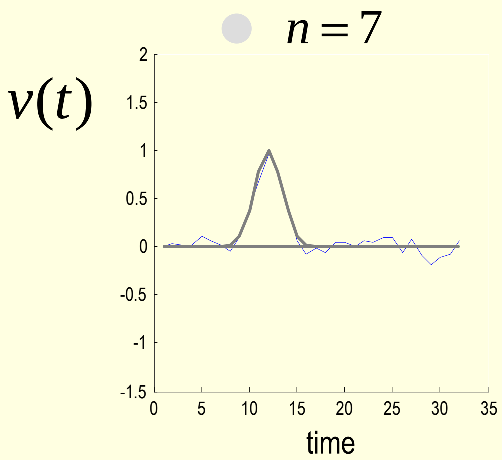


The order of generalised motion

Precision in generalised coordinates

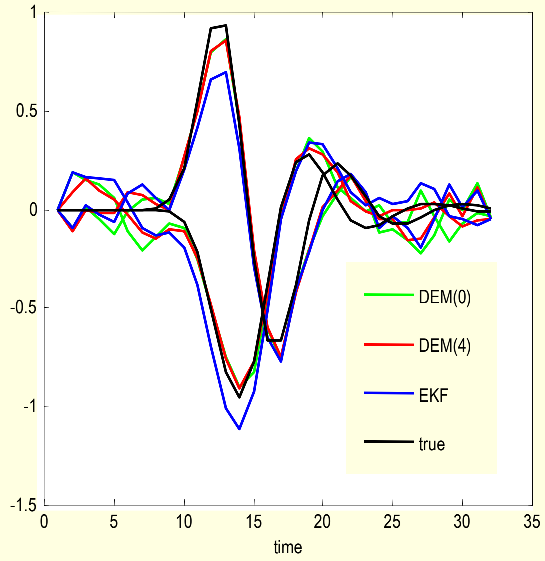




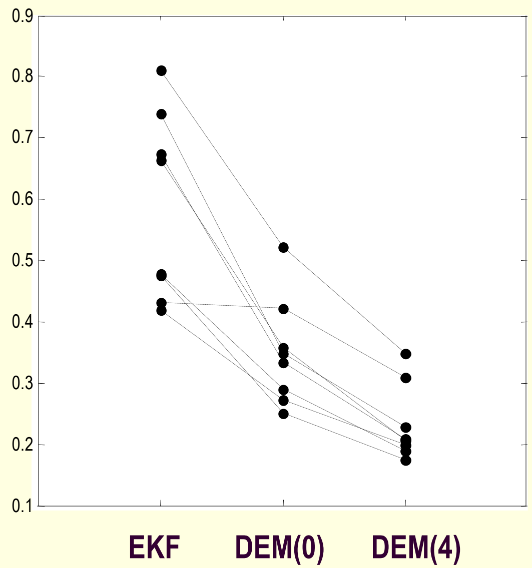


DEM and extended Kalman filtering

hidden states



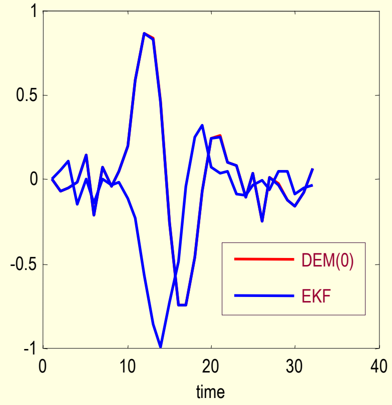
sum of squared error (hidden states)



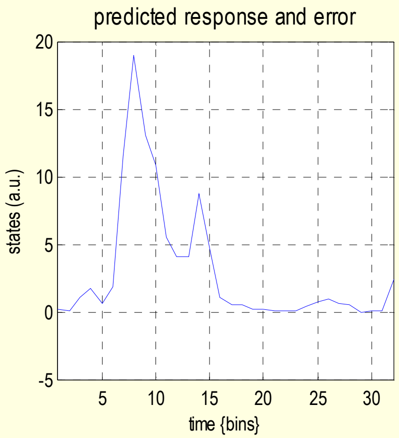
With convergence when

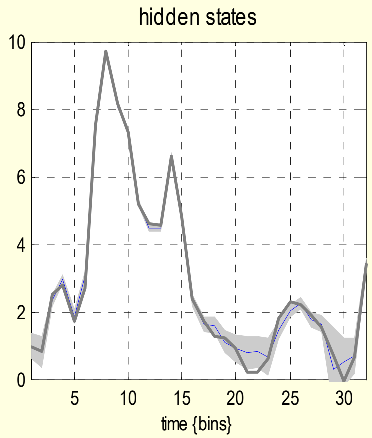
$$\gamma \rightarrow \infty \equiv n = 1$$

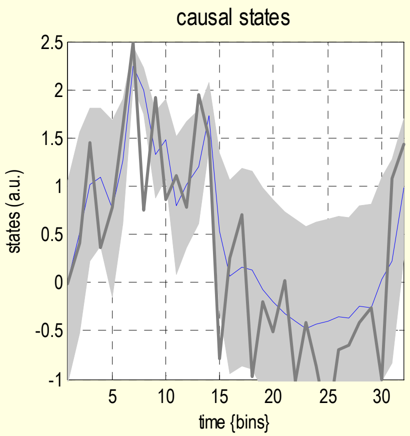
hidden states



A nonlinear convolution model



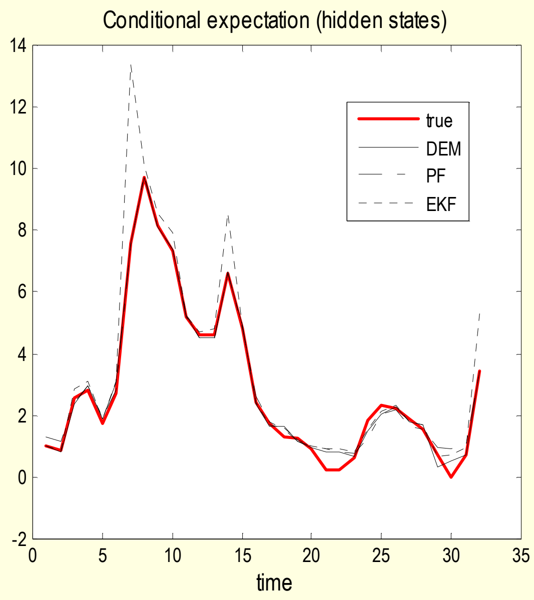


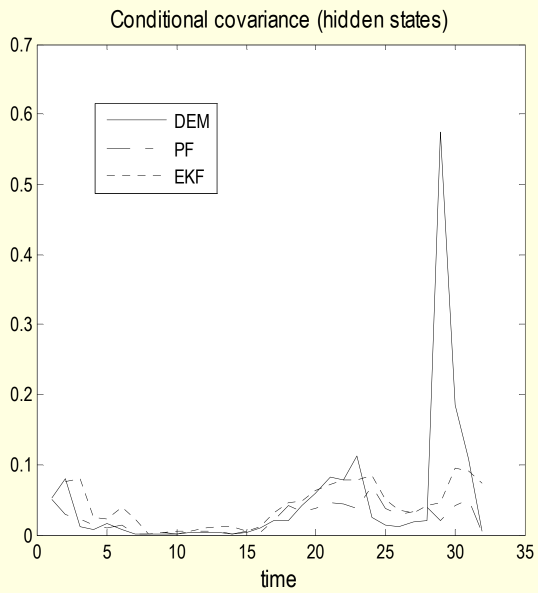


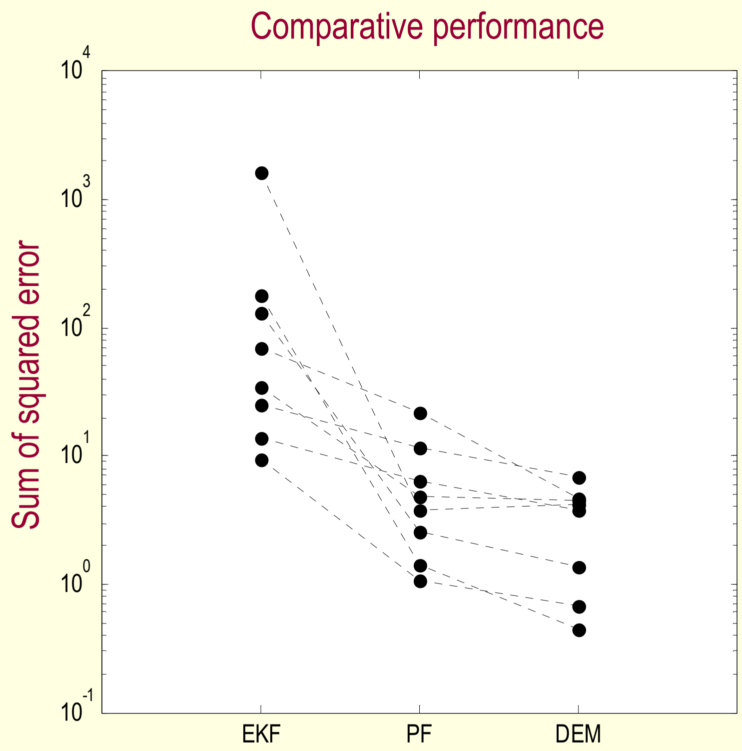
level	$g(x,v)$	$f(x,v)$	Π^z	Π^w	η^v
$m=1$	$\frac{1}{5}x^2$	$e^v - x\ln 2$	e^4	e^{16}	
$m=2$			2		$\frac{1}{2}+\sin(\frac{1}{16}\pi t)$

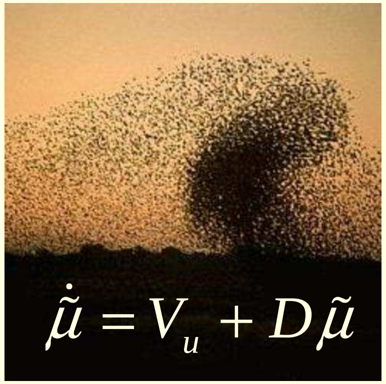
This system has a slow sinusoidal input or cause that excites increases in a single hidden state. The response is a quadratic function of the hidden states (c.f., Arulampalam *et al* 2002).

DEM and particle filtering

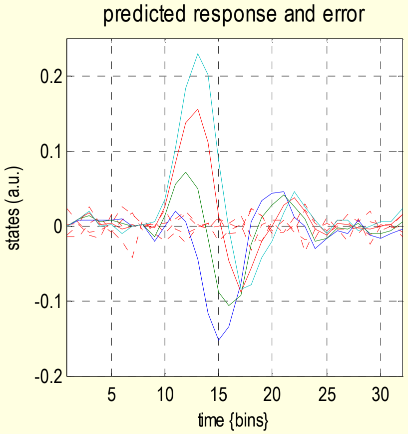


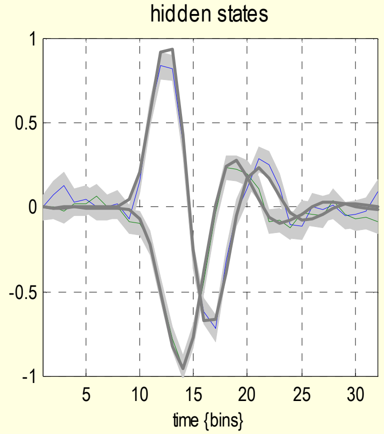






g



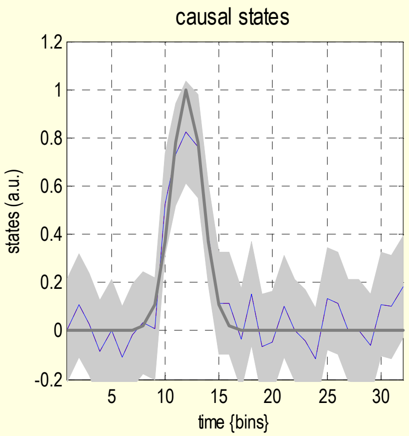


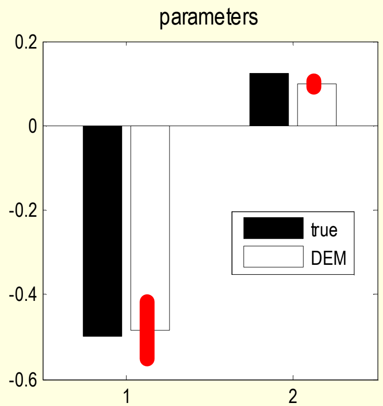
μ^x

Inference on states

Triple estimation (DEM)

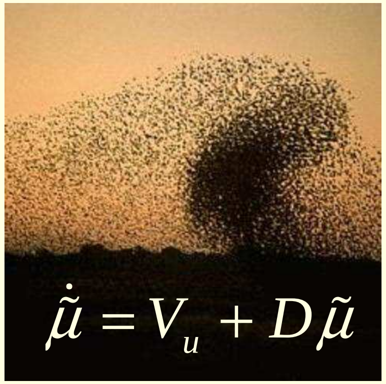
μ^v





μ^θ

Learning parameters



Overview

Variational learning and free-energy

Hierarchical dynamic models
- Generalised coordinates (dynamical priors)
- Hierarchal forms (structural priors)

Model inversion
- Variational filtering (free-form)
- Laplace approximation and DEM (fixed-form)

Comparative evaluations

Hemodynamics and Bird songs

An fMRI study of attention

Stimuli 250 radially moving dots at 4.7 degrees/s

Pre-Scanning

5 x 30s trials with 5 speed changes (reducing to 1%)

Task: detect change in radial velocity

Scanning (no speed changes)

4 x 100 scan sessions; each comprising 10 scans of 4 different conditions

F A F N F A F N S ....

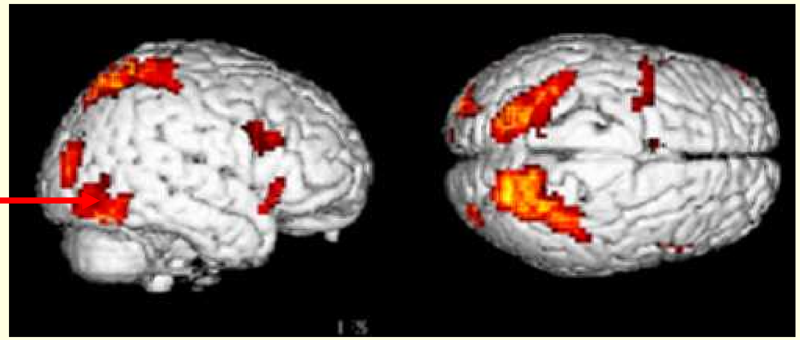
A – dots, motion and attention (detect changes)

N – dots and motion

S – dots

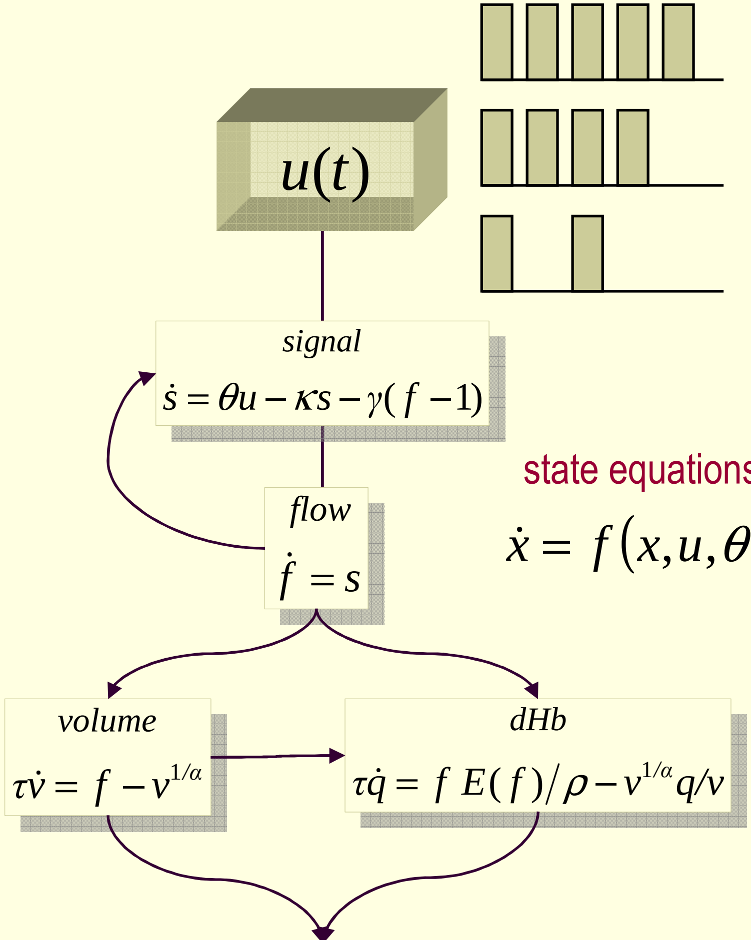
F – fixation

V5 (motion sensitive area)



Buchel *et al* 1999

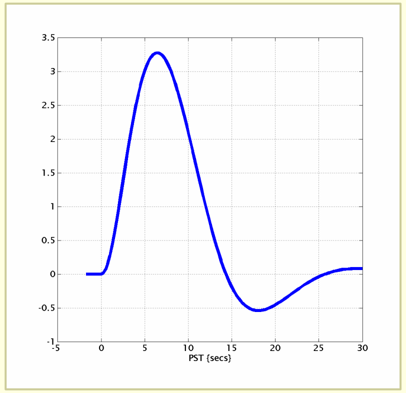
A hemodynamic model



state equations

$$\dot{x} = f(x, u, \theta)$$

convolution kernel

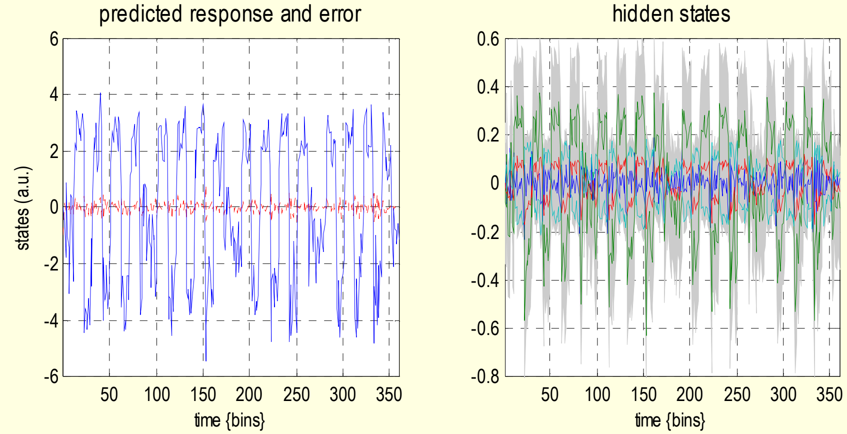


Output: a mixture of intra- and extravascular signal

$$y(t) = g(v, q) = V_0(k_1(1-q) + k_2(1 - q/v) + k_3(1-v))$$

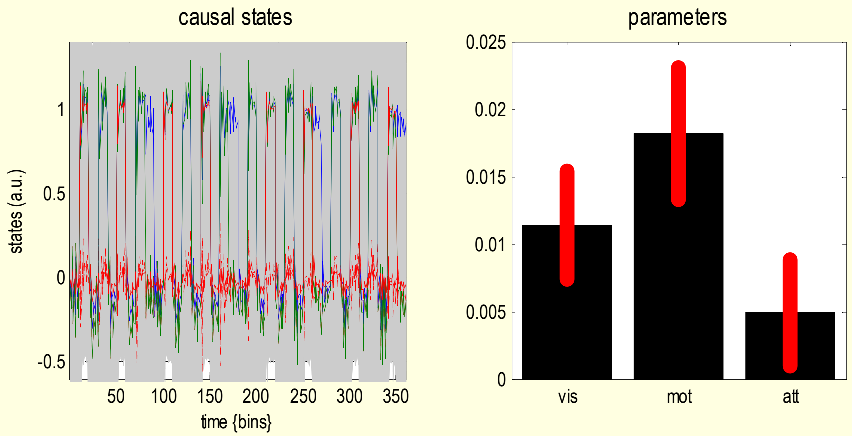
output equation

$$y = g(x, \theta)$$



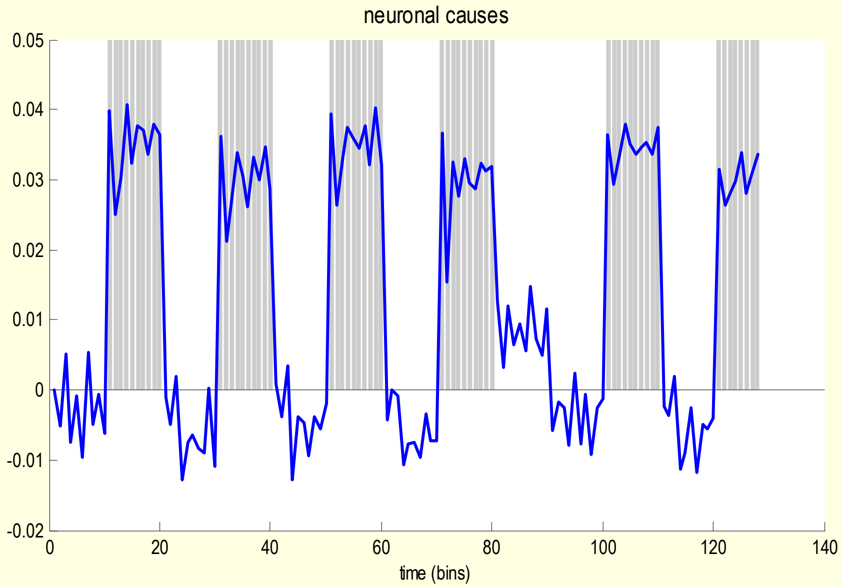
Inference on states

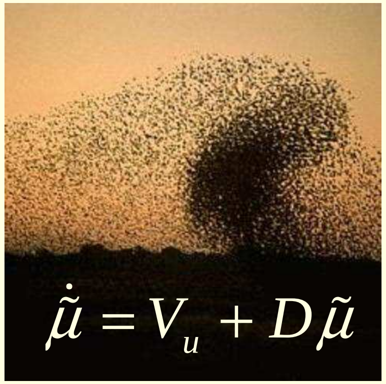
Hemodynamic deconvolution (V5)



Learning parameters

… and a closer look at the states





Overview

Variational learning and free-energy

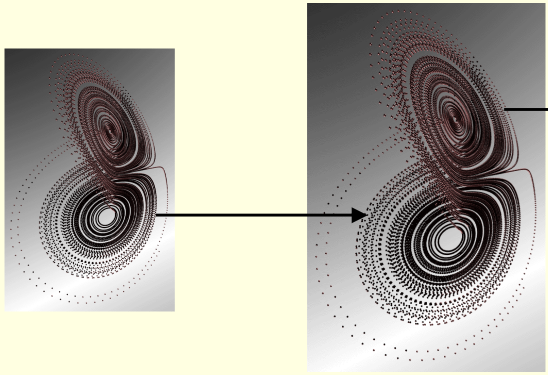
Hierarchical dynamic models
- Generalised coordinates (dynamical priors)
- Hierarchal forms (structural priors)

Model inversion
- Variational filtering (free-form)
- Laplace approximation and DEM (fixed-form)

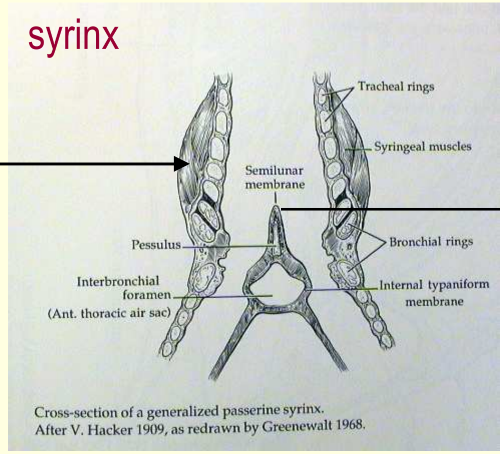
Comparative evaluations

Hemodynamics and Bird songs

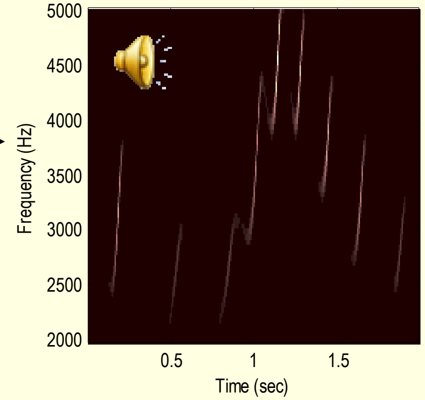
Synthetic song-birds

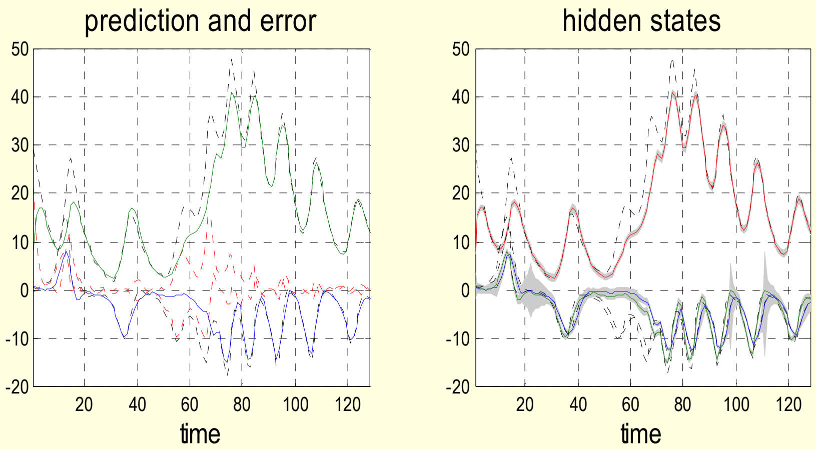


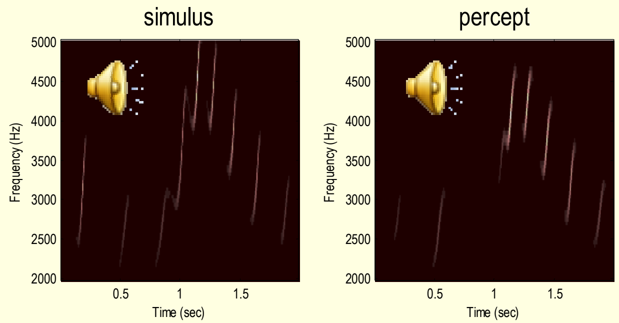
hierarchy of Lorenz attractors



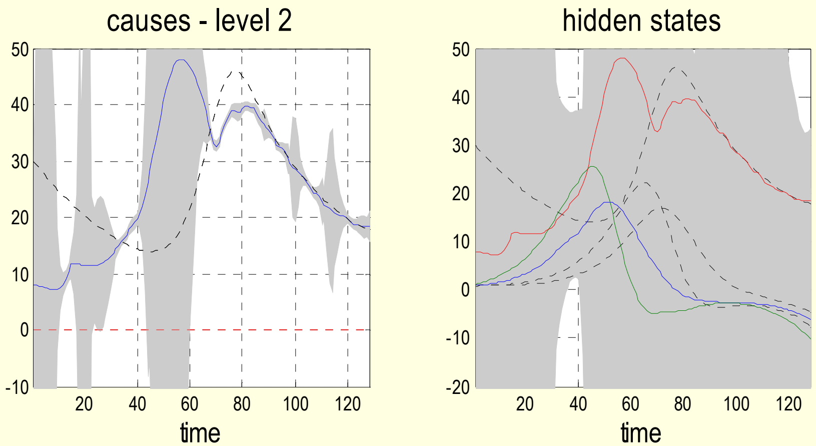
simulus



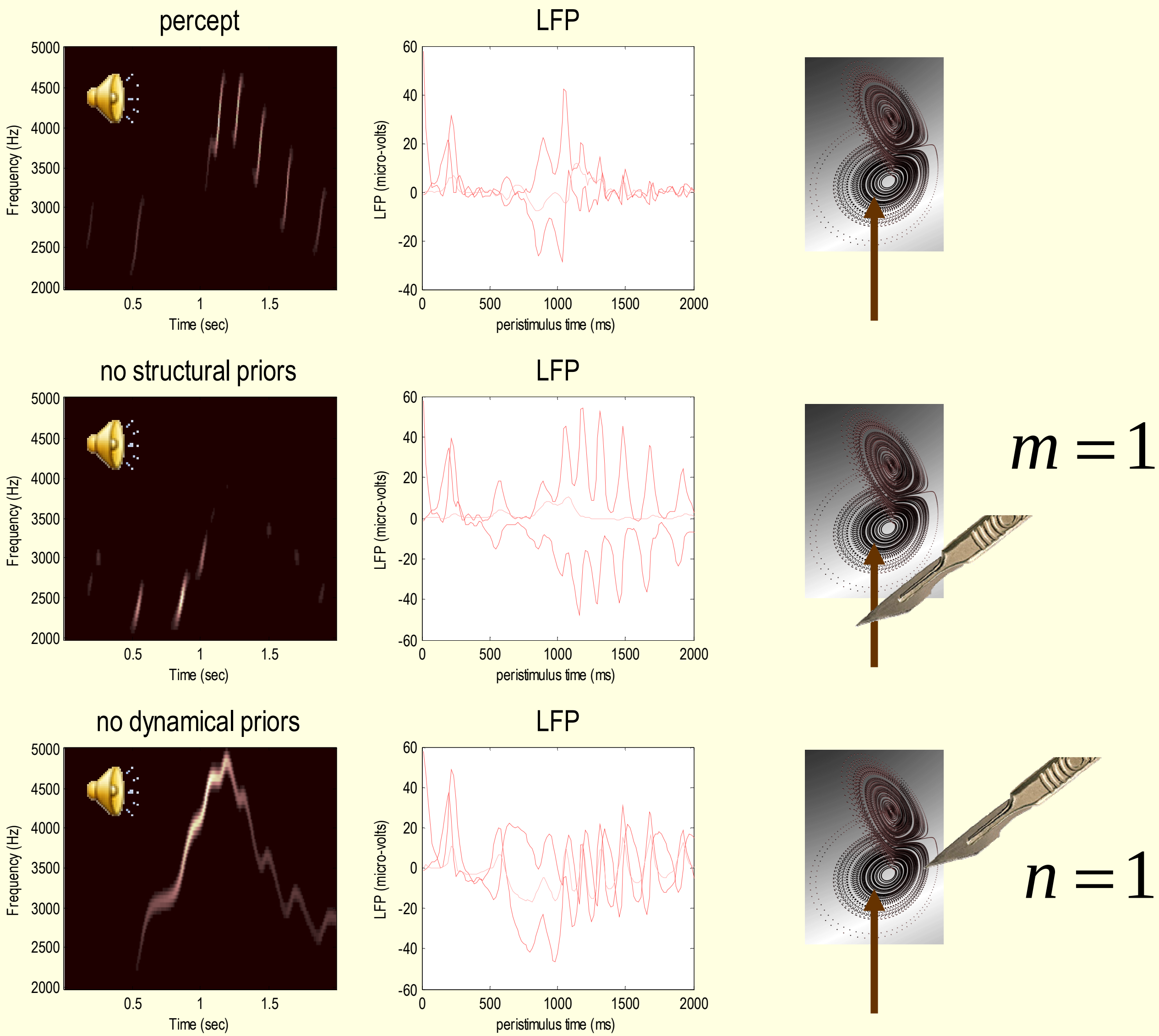


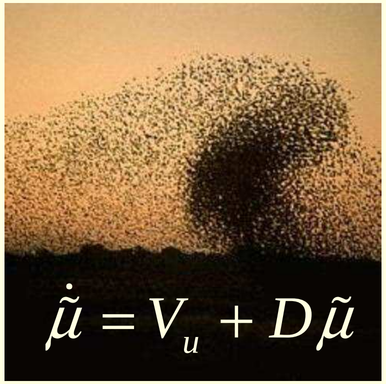


Song recognition with DEM



… and broken birds





Summary

Variational learning and free-energy

Hierarchical dynamic models

- Generalised coordinates (dynamical priors)
- Hierarchal forms (structural priors)

Model inversion

- Variational filtering (free-form)
- Laplace approximation and DEM (fixed-form)

Comparative evaluations

Hemodynamics and Bird songs