$$
\begin{aligned}
\text{[Observation]} \quad & \mathbf{y}_t & \sim p(\mathbf{y}_t|\boldsymbol{\theta}_t, \phi), \\
\text{[State]} \quad & \boldsymbol{\theta}_t & \sim p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \phi) \\
\text{[Prior]} \quad & (\boldsymbol{\theta}_0, \phi) & \sim p(\boldsymbol{\theta}_0, \phi).
\end{aligned}
$$

The densities may be non-Gaussian and the model may be non-linear at the observation and/or system levels.

- **Example 1:** *$AR(1)$ with normal mixture structure on observational errors.*

$$
\begin{aligned}
y_t &= \mu_t + \nu_t, \quad \nu_t \sim \pi N(0, v) + (1 - \pi) N(0, \kappa^2 v), \quad \kappa > 1, \\
\mu_t &= \phi \mu_{t-1} + w_t, \quad w_t \sim N(0, w).
\end{aligned}
$$

- **Example 2:** *Univariate stochastic volatility.* P_t financial price series, and $r_t = P_t/P_{t-1} - 1$ the returns.

$$\begin{aligned} r_t &\sim N(0, \sigma_t^2), \\ \sigma_t &= \exp(\mu + x_t), \\ x_t &= \phi x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, v) \\ x_0 &\sim N(0, v/(1-\phi^2)). \end{aligned}$$

  μ : baseline log-volatility; ϕ : defines persistence in deviations in volatility from baseline; v : drives levels of activity in volatility process.

- **Example 3:** *Fat-tailed non-linear model.*

$$\begin{aligned} y_t &= \theta_t + \sqrt{\gamma_t}\nu_t, \quad \nu_t \sim N(0, v) \\ \theta_t &= \beta \frac{\theta_{t-1}}{1+\theta_{t-1}^2} + w_t, w_t \sim N(0, w), \end{aligned}$$

  with $\gamma_t \sim IG(\nu/2, \nu/2)$.

- **Example 4:** *Non-linear state-space model*

$$
\begin{aligned}
y_t &= a\theta_t^2 + \nu_t, \quad \nu_t \sim N(0, v) \\
\theta_t &= b\theta_{t-1} + c\frac{\theta_{t-1}}{1+\theta_{t-1}^2} + d\cos(\omega t) + w_t, \quad w_t \sim N(0, w)
\end{aligned}
$$

The models in examples 1 and 2 are *conditionally Gaussian dynamic linear models* (CGDLMs). The model in example 3 is a *conditionally Gaussian dynamic model* (CGDM). The model in example 4 is a nonlinear dynamic model.

Joint posterior: $p(\boldsymbol{\theta}_{1:T}, \phi | \mathcal{D}_T)$.

General MCMC algorithm:

1. Set initial values $\boldsymbol{\theta}_{1:T}^{(0)}, \phi^{(0)}$.
2. For each iteration $k = 1, 2, \ldots,$ until MCMC convergence:
   - Sample $\theta_{1:T}^{(k)}$ component by component by sampling $\boldsymbol{\theta}_t^{(k)} \sim p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{1:(t-1)}^{(k)}, \boldsymbol{\theta}_{(t+1):T}^{(k-1)}, \mathcal{D}_T)$ for each t.
   - Sample $\phi^{(k)} \sim p(\phi^{(k)} | \boldsymbol{\theta}_{1:T}^{(k)}, \mathcal{D}_T)$.

Let $\boldsymbol{\lambda}_t$ be a latent variable at time t. A *conditionally Gaussian DLM or CDLM* is a model of the form

$$\begin{aligned} y_t &= \mathbf{F}'_{\boldsymbol{\lambda}_t}\theta_t + \nu_t, \;\; \nu_t \sim N(0, v_{\boldsymbol{\lambda}_t}), \\ \theta_t &= \mathbf{G}_{\boldsymbol{\lambda}_t}\theta_{t-1} + \mathbf{w}_t, \;\; \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_{\boldsymbol{\lambda}_t}). \end{aligned}$$

Joint posterior: $p(\boldsymbol{\theta}_{1:T}, \boldsymbol{\lambda}_{1:T} | \mathcal{D}_T)$.
MCMC inference: Begin with $\boldsymbol{\theta}_{1:T}^{(0)}$ and $\boldsymbol{\lambda}_{1:T}^{(0)}$. Obtain samples from the joint posterior via Gibbs sampling by iterating between the two conditional posteriors

$$p(\boldsymbol{\theta}_{1:T} | \boldsymbol{\lambda}_{1:T}, \mathcal{D}_T) \leftrightarrow p(\boldsymbol{\lambda}_{1:T} | \boldsymbol{\theta}_{1:T}, \mathcal{D}_T).$$

The *forward filtering backward sampling (FFBS)* algorithm (Carter & Kohn, 1994 and Frühwirth-Schnatter 1994) can be used to obtain samples from $p(\boldsymbol{\theta}_{1:T} | \boldsymbol{\lambda}_{1:T}, \mathcal{D}_T)$.

The FFBS algorithm

$$p(\boldsymbol{\theta}_{1:T}|\boldsymbol{\lambda}_{1:T}, \mathcal{D}_T) = p(\boldsymbol{\theta}_T|\boldsymbol{\lambda}_{1:T}, \mathcal{D}_T) \prod_{t=1}^{T-1} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{\lambda}_{1:T}, \mathcal{D}_T).$$

Now, $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{\lambda}_{1:T}, \mathcal{D}_T) \propto p(\boldsymbol{\theta}_t|\boldsymbol{\lambda}_{1:T}, \mathcal{D}_t) p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t, \boldsymbol{\lambda}_{1:T}, \mathcal{D}_t)$, and so for each MCMC iteration k :

1. Use the DLM filtering equations to compute $\mathbf{m}_t, \mathbf{a}_t, \mathbf{C}_t$ and $\mathbf{R}_t$ for $t = 1 : T$.
2. At time $t = T$ sample $\theta_T^{(k)} \sim N(\mathbf{m}_T, \mathbf{C}_T)$.
3. For $t = (T-1) : 1$, sample $\theta_t^{(k)} \sim N(\mathbf{l}_t, \mathbf{L}_t)$, with

$$\mathbf{l}_t = \mathbf{m}_t + \mathbf{B}_t(\theta_{t+1}^{(k)} - \mathbf{a}_{t+1}), \quad \mathbf{L}_t = \mathbf{C}_t - \mathbf{B}_t \mathbf{R}_{t+1} \mathbf{B}_t',$$

and $\mathbf{B}_t = \mathbf{C}_t \mathbf{G}_{t+1}' \mathbf{R}_{t+1}^{-1}$.

Note that the moments of these distributions depend on $\boldsymbol{\lambda}_{1:T}^{(k-1)}$.

Example 1: *$AR(1)$ with normal mixture structure on observational errors.*

$$\begin{aligned} y_t &= \mu_t + \nu_t, \quad \nu_t \sim \pi N(0, v) + (1-\pi) N(0, \kappa^2 v), \\ \mu_t &= \phi \mu_{t-1} + w_t, \quad w_t \sim N(0, w). \end{aligned}$$

Assume that π and κ are known and that $p(\phi) \propto 1$, $p(v) = IG(\alpha_{v,0}, \beta_{v,0})$ and $p(w) = IG(\alpha_{w,0}, \beta_{w,0})$. Let

$$\gamma_t = \begin{cases} 1 & \text{with probability } \pi \\ \kappa^2 & \text{with probability } 1-\pi \end{cases}$$

Given γ_t we have the DLM $\{1, \phi, v\gamma_t, w\}$.

MCMC algorithm:

- $(v|w, \phi, \mu_{0:T}, \gamma_{1:T}, \mathcal{D}_T) \sim IG(\alpha_v, \beta_v)$ with $\alpha_v = \alpha_{0,v} + T/2$, $\beta_v = \beta_{0,v} + s_v^2/2$ and

$$s_v^2 = \sum_{\gamma_t=1}(y_t - \mu_t)^2 + \sum_{\gamma_t=\kappa^2}(y_t - \mu_t)^2/\kappa^2.$$

- $(w|v, \phi, \mu_{0:T}, \gamma_{1:T}, \mathcal{D}_T) \sim IG(\alpha_w, \beta_w)$ with $\alpha_w = \alpha_{0,w} + T/2$ and $\beta_w = \beta_{w,0} + \sum_t(\mu_t - \phi\mu_{t-1})^2/2$.
- $(\mu_{0:T}|v, w, \phi, \gamma_{1:T}, \mathcal{D}_T)$ is sampled using the FFBS algorithm.

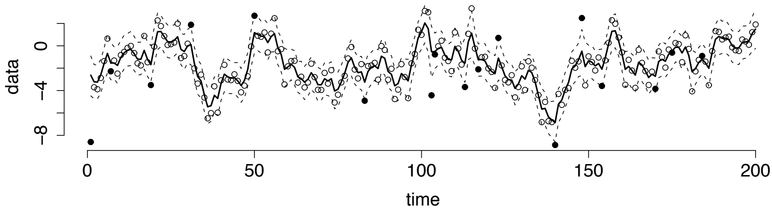
- $(\phi | v, w, \mu_{0:T}, \mathcal{D}_T) \sim N(m_\phi, C_\phi)$ with

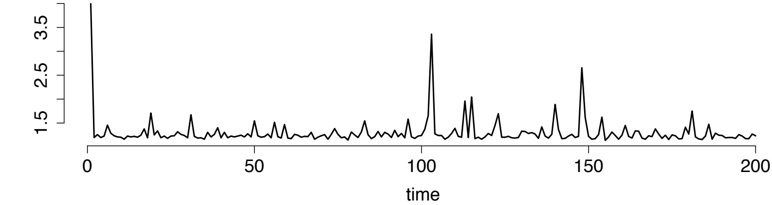
$$m_\phi = \left(\sum_{t=1}^{T} \mu_t \mu_{t-1} \right) / \sum_{t=1}^{T} \mu_{t-1}^2, \quad C_\phi = w / \sum_{t=1}^{T} \mu_{t-1}^2.$$

- $(\gamma_{1:T} | v, w, \phi, \mu_{0:T}, \mathcal{D}_T)$. At each time t, γ_t is set to 1 or κ^2 with probabilities defined as

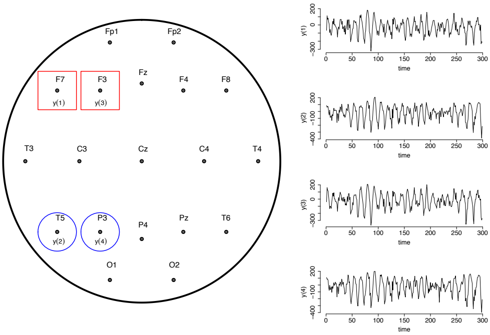
$$\frac{\Pr(\gamma_t = 1 | v, w, \mu_{0:T}, \mathcal{D}_T)}{\Pr(\gamma_t = \kappa^2 | v, w, \mu_{0:T}, \mathcal{D}_T)} = \frac{\pi}{(1-\pi)} \kappa \times \exp\{-(y_t - \mu_t)^2 (1 - \kappa^{-2}) / 2v\}.$$

Example: Simulated data, model fitted with `dlm` in `R`.





Factor model for EEG data



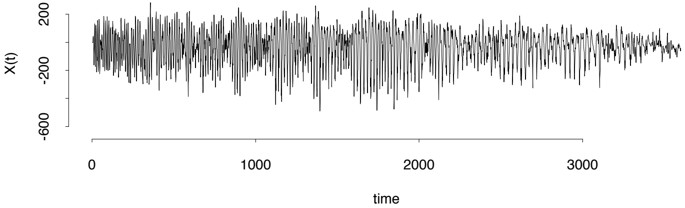
$$
\begin{aligned}
y_{t,i} &= \beta_i x_t + \nu_{t,i}, \quad \nu_{t,i} \sim N(0, v), \\
x_t &= \sum_{j=1}^{p} \phi_{t,j} x_{t-j} + \omega_t, \quad \omega_t \sim N(0, w) \\
\phi_t &= \phi_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(\mathbf{0}, \mathbf{U}_t).
\end{aligned}
$$

- $\beta_{Cz} = 1$.
- $w/v = c$ with c known.
- $\mathbf{U}_t$ specified using a discount factor δ_ϕ.

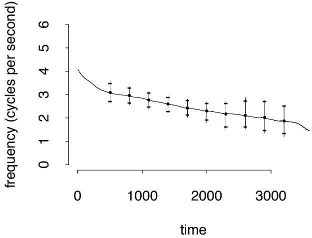
MCMC algorithm:

- Sample $x_{1:T}$ from $p(x_{1:T}|\mathbf{y}_{1:T}, \boldsymbol{\beta}_{1:m}, \phi_{1:T}, v)$ via FFBS.
- Sample $\phi_{1:T}$ from $p(\phi_{1:T}|x_{1:T}, \mathbf{U}_{1:T}, v)$ via FFBS.
- Sample $\boldsymbol{\beta}_{1:m}$ from $p(\boldsymbol{\beta}_{1:m}|\mathbf{y}_{1:T}, x_{1:T}, v)$ using a Gibbs step (conjugate priors on $\boldsymbol{\beta}_{1:m}$).
- Sample v from $p(v|\mathbf{y}_{1:T}, \boldsymbol{\beta}_{1:m}, x_{1:T})$ using a Gibbs step (conjugate prior on v).

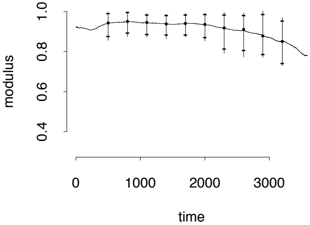
ECT data set: $p = 6$, $\delta_\phi = 0.994$, $w/v = 10$.



(a)

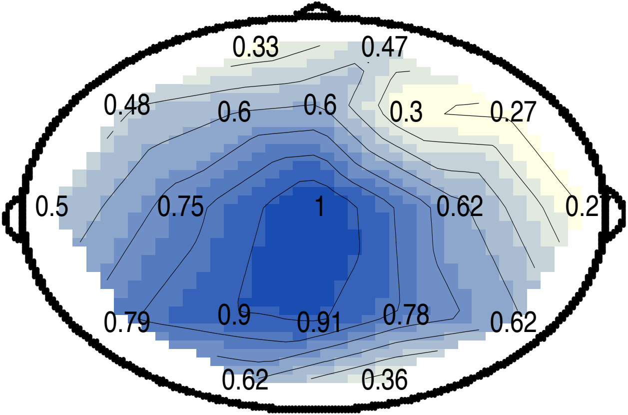


(b)



(c)

ECT data set: $p = 6$, $\delta_\phi = 0.994$, $w/v = 10$.



Dynamic lag/lead model for EEG data

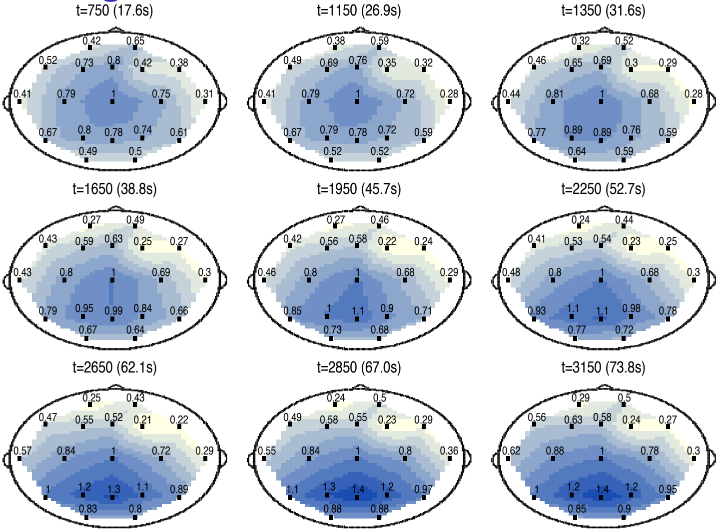
$$
\begin{aligned}
y_{i,t} &= \beta_{i,t} x_{t-l_{i,t}} + \nu_{i,t}, \quad \nu_{i,t} \sim N(0, v_i), \\
\boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{U}_t),
\end{aligned}
$$

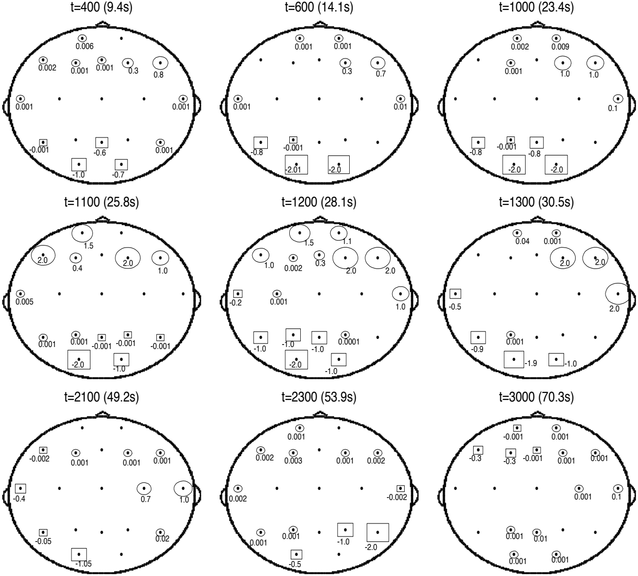
- $x_t = y_{t,Cz}$.
- $\mathbf{U}_t$ specified using a discount factor δ_β.
- $l_{i,t} \in \{-2, ..., 2\}$ with $Pr(l_{i,t} = k | l_{i,t-1} = j) = p_{jk}$ and $p_{jk} = 0.9999$ if $j = k$; $p_{jk} = 0.0001$ if $k = -1, j = -2$ or $k = 1, j = 2$; $p_{jk} = 0.0005$ if $|k - j| = 1$ and neither $k = -1, j = -2$ nor $k = 1, j = 2$; $p_{jk} = 0$ otherwise.

MCMC algorithm:

- For each $i = 1 : 19$, sample $(\boldsymbol{\beta}_{i,1:T} | \mathbf{y}_{i,1:T}, \mathbf{x}, \mathbf{l}_{i,1:T}, v_i, \mathbf{U}_t)$ using the FFBS algorithm.
- For each $i = 1 : 19$, sample $(\mathbf{l}_{i,1:T} | \mathbf{y}_{i,1:T}, \mathbf{x}, \boldsymbol{\beta}_{i,1:T}, v_i)$ using a discrete version of the FFBS algorithm (Carter and Kohn, 1994), i.e., update filtering equations for $t = 1 : T$, then sample $(l_{i,T} | \mathbf{y}_{i,1:T}, \mathbf{x}, \boldsymbol{\beta}_{i,1:T}, v_i)$ and finally, for $t = (T-1) : 1$, sample $(l_{i,t} | \mathbf{y}_{i,1:T}, \mathbf{x}, \boldsymbol{\beta}_{i,1:T}, v_i, l_{i,(t+1)})$.
- For each $i = 1 : 19$ sample $(v_i | \mathbf{y}_{i,1:T}, \mathbf{x}, \boldsymbol{\beta}_{i,1:T}, \mathbf{l}_{i,1:T})$ using Gibbs steps (conjugate priors).

Dynamic lag/lead model for EEG data




t=400 (9.4s)
0.006 0.002 0.001 0.001 0.3 0.8 0.001 0.001 -0.001 -0.6 0.001 -1.0 -0.7
t=600 (14.1s)
0.001 0.001 0.3 0.7 0.001 0.01 -0.8 -0.001 -2.01 -2.0
t=1000 (23.4s)
0.002 0.009 0.001 1.0 1.0 0.1 -0.8 -0.001 -0.8 -2.0 -2.0
t=1100 (25.8s)
1.5 2.0 0.4 2.0 1.0 0.005 0.001 0.001 -0.001 -0.001 -0.001 -2.0 -1.0
t=1200 (28.1s)
1.5 1.1 1.0 0.002 0.3 2.0 2.0 -0.2 0.001 1.0 -1.0 -1.0 -1.0 0.0001 -2.0 -1.0
t=1300 (30.5s)
0.04 0.001 2.0 2.0 -0.5 2.0 -0.9 0.001 -1.9 -1.0
t=2100 (49.2s)
-0.002 0.001 0.001 0.001 -0.4 0.7 1.0 -0.05 0.02 -1.05
t=2300 (53.9s)
0.001 0.002 0.003 0.001 0.002 0.002 -0.002 0.001 0.001 -1.0 -2.0 -0.5
t=3000 (70.3s)
-0.001 0.001 -0.3 -0.3 -0.001 0.001 0.001 0.001 0.1 0.001 0.01 0.001 0.001