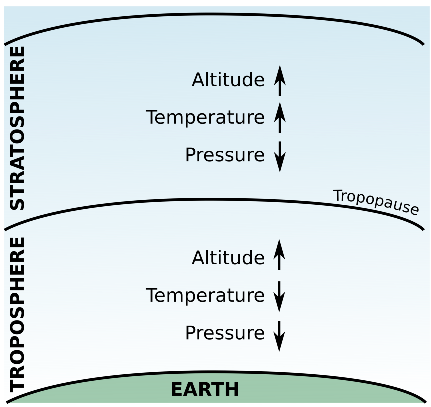
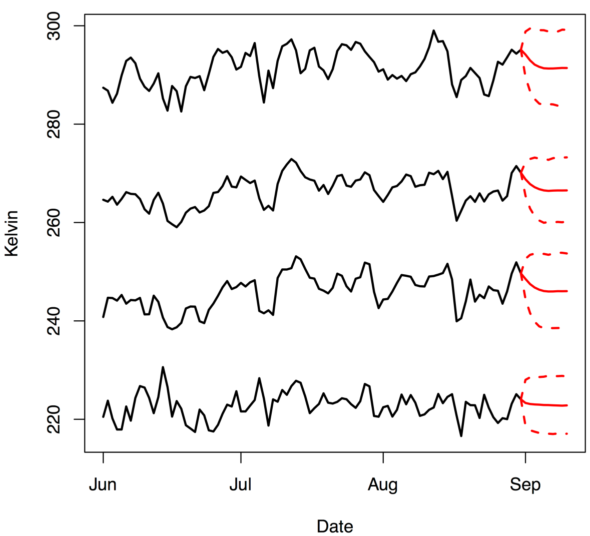
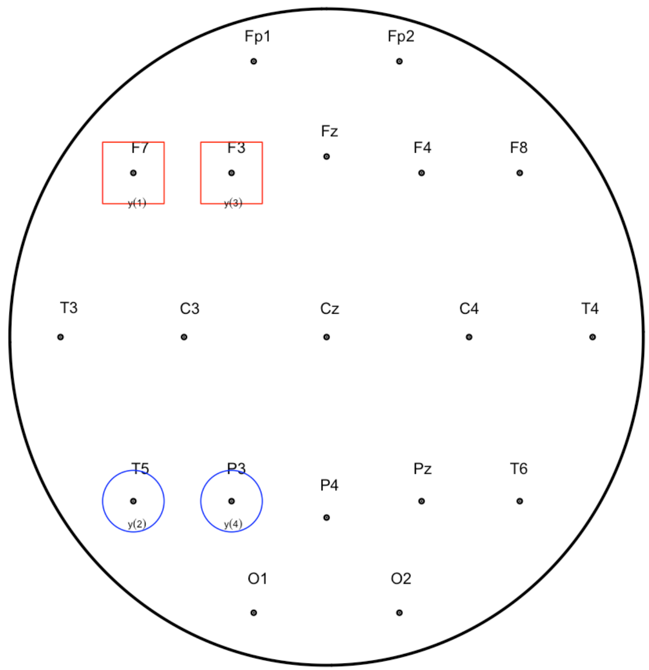
Bayesian Multivariate Time Series Analysis for Environmental Science and Neuroscience

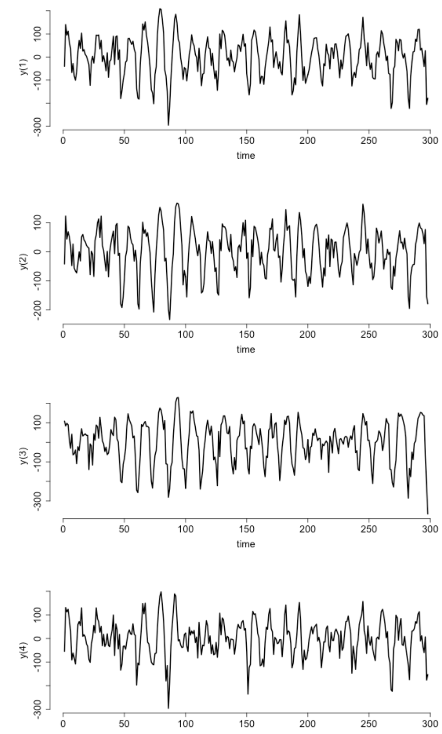
Marco Ferreira
Virginia Tech

Raquel Prado
University of California Santa Cruz









ISBA World Meeting, Montreal, 2022

Course Structure

Part I

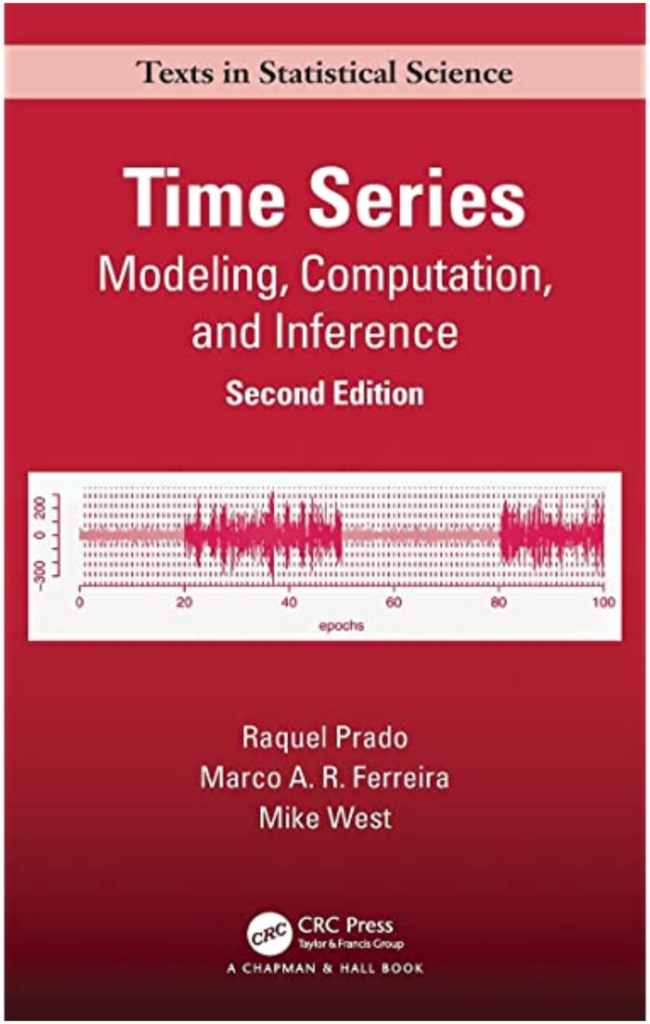
- **Introduction to Dynamic Linear Models**
  - Model building: trend, seasonal and regression models; superposition
  - Inference: filtering, smoothing and forecasting;
  - Case studies
- **CGDLM and Models for Multiple Time Series**
  - MCMC for CGDLMs
  - Hierarchical DLMs
  - Case study: multi-trial EEG analysis
- **Models for Multivariate Time Series**
  - Multivariate DLMs
  - VAR, TV-VAR, TV-VPARCOR; sparsity priors
  - Case studies: multi-location EEG; multi-location wind component data

Part II

- **Static Factor Models**
  - Latent factor models with one factor
  - MCMC computations
  - Latent k-factor models
  - Choice of the number of factors
  - Case study: Troposphere temperature

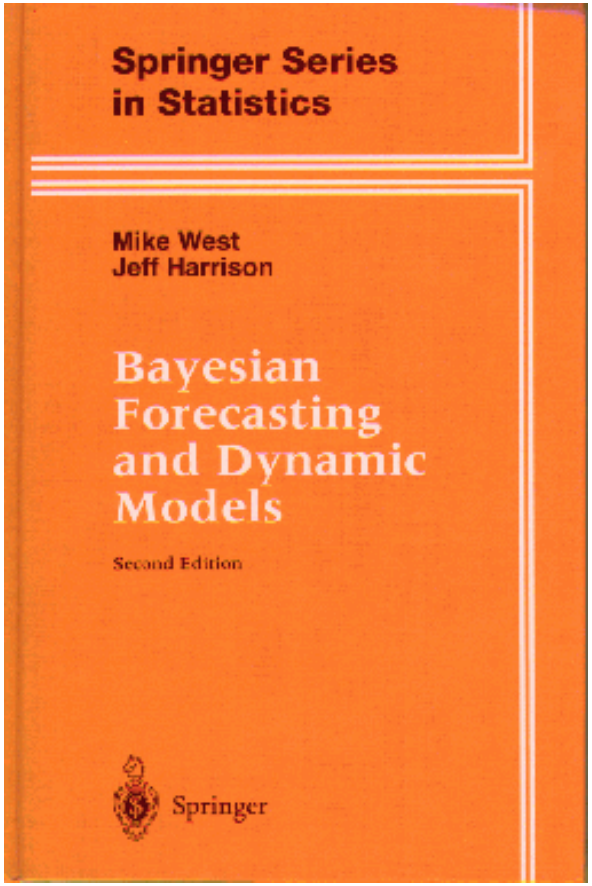
- **Multiple Dynamic Factor Models**
  - Dynamic latent factor models
    - Example: Time series troposphere temperature
  - Dynamic spatio-temporal factor model
    - Example: Spatio-temporal troposphere temperature

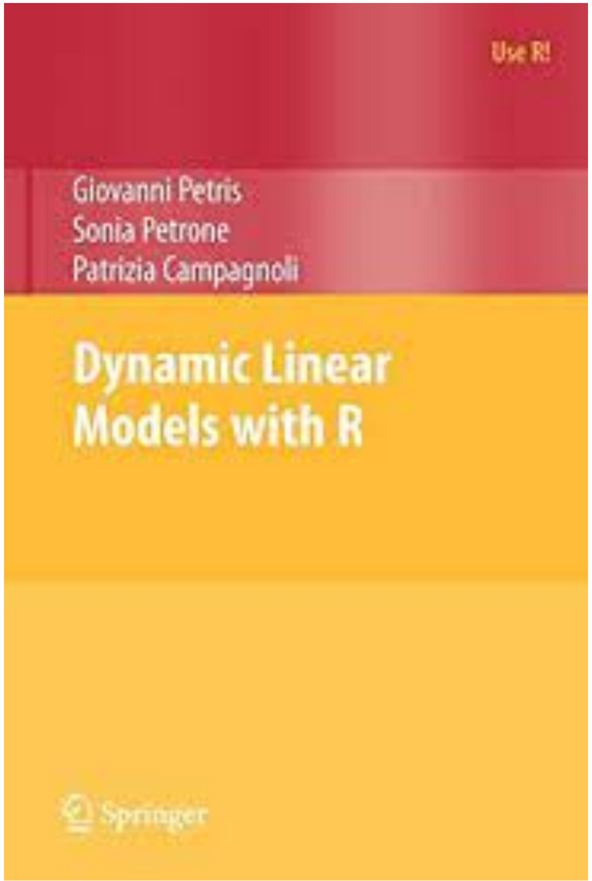
References



Promo code: ASA18 for a 30% discount

References


Springer Series in Statistics
Mike West
Jeff Harrison
Bayesian Forecasting and Dynamic Models
Second Edition
Springer


Use R!
Giovanni Petris
Sonia Petrone
Patrizia Campagnoli
Dynamic Linear Models with R
Springer

Introduction - Dynamic linear models (DLMs)

$$y_t = \beta x_t + \epsilon_t, \quad t = 1,2,\ldots$$

- β : Prior beliefs expressed through $p(\beta)$
- **Sequential learning:** $p(\beta \mid y_{1:t}, x_{1:t})$
- **Forecasting:** $p(y_{t+h} \mid y_{1:t}, x_{1:(t+h)})$

As time passes, we may want to characterize β as evolving to adapt to changes over time, e.g.,

$$y_t = \beta_t x_t + \epsilon_t$$
$$\beta_t = \beta_{t-1} + \nu_t$$

Filtering

$$p(\beta_t \mid y_{1:t}, x_{1:t}),$$

Smoothing

$$p(\beta_t \mid y_{1:T}, x_{1:T}), T \geq t,$$

Forecasting

$$p(y_{t+h} \mid y_{1:t}, x_{1:(t+h)})$$

Introduction - Dynamic linear models (DLMs)

General Dynamic Linear Model

$$
\begin{aligned}
&\text{[Observation]} \quad & y_t &= \mathbf{F}_t' \boldsymbol{\theta}_t + \nu_t, \\
&\text{[System]} \quad & \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \\
&\text{[Prior]} \quad & \boldsymbol{\theta}_0 &\sim p(\boldsymbol{\theta}_0 \mid \mathscr{D}_0)
\end{aligned}
$$

- $\boldsymbol{\theta}_t : p \times 1$ state vector
- $\mathbf{F}_t : p \times 1$ vector of constants
- $\mathbf{G}_t : p \times p$ evolution matrix
- ν_t : observation noise (scalar); $\mathbf{w}_t : p \times 1$ evolution noise

Normal DLM (NDLM): observational & evolution noises independent, mutually independent

$$\nu_t \sim N(0, v_t), \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t), \quad (\boldsymbol{\theta}_0 \mid \mathscr{D}_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$$

Introduction - Dynamic linear models (DLMs)

- *Short hand notation:*

$$\{\overset{p\times 1}{\mathbf{F}_t},\ \overset{p\times p}{\mathbf{G}_t},\ v_t,\ \overset{p\times p}{\mathbf{W}_t}\}$$

$$y_t$$

Univariate DLM

$$\{\overset{p\times n}{\mathbf{F}_t},\ \overset{p\times p}{\mathbf{G}_t},\ \overset{n\times n}{\mathbf{V}_t},\ \overset{p\times p}{\mathbf{W}_t}\}$$

$$\mathbf{y}_t \quad n\times 1$$

Multivariate DLM

- *h-steps ahead forecast function h ≥ 1:*

$$f_t(h) = E(y_{t+h}|\mathcal{D}_t) = \mathbf{F}'_{t+h}\mathbf{G}_{t+h}\ldots\mathbf{G}_{t+1}E(\boldsymbol{\theta}_t|\mathcal{D}_t)$$

If $\mathbf{G}_t = \mathbf{G} \Rightarrow f_t(h) = \mathbf{F}'_{t+h}\mathbf{G}^h E(\boldsymbol{\theta}_t|\mathcal{D}_t)$

DLMs: Polynomial Trend Models

- **First order polynomial model**

$$\begin{aligned} y_t &= \theta_t + \nu_t, \\ \theta_t &= \theta_{t-1} + w_t. \end{aligned} \qquad \{1, 1, v_t, \mathrm{w}_t\}$$

Forecast function: $f_t(h) = a_{t,0}$

- **Second order polynomial model**

$$\begin{aligned} y_t &= \theta_{t,1} + \nu_t, \\ \theta_{t,1} &= \theta_{t-1,1} + \theta_{t-1,2} + w_{t,1}, \\ \theta_{t,2} &= \theta_{t-1,2} + w_{t,2}. \end{aligned}$$

Forecast function: $f_t(h) = a_{t,0} + a_{t,1} h$

DLMs: Polynomial Trend Models

Note that in this case we have $\{\mathbf{E}_2, \mathbf{J}_2(1), v_t, \mathbf{W}_t\}$ with $\mathbf{E}_2 = (1,0)'$ and

$$\mathbf{J}_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

- **p-th order polynomial model** (canonical representation):

$\{\mathbf{E}_p, \mathbf{J}_p(1), v_t, \mathbf{W}_t\}$ with

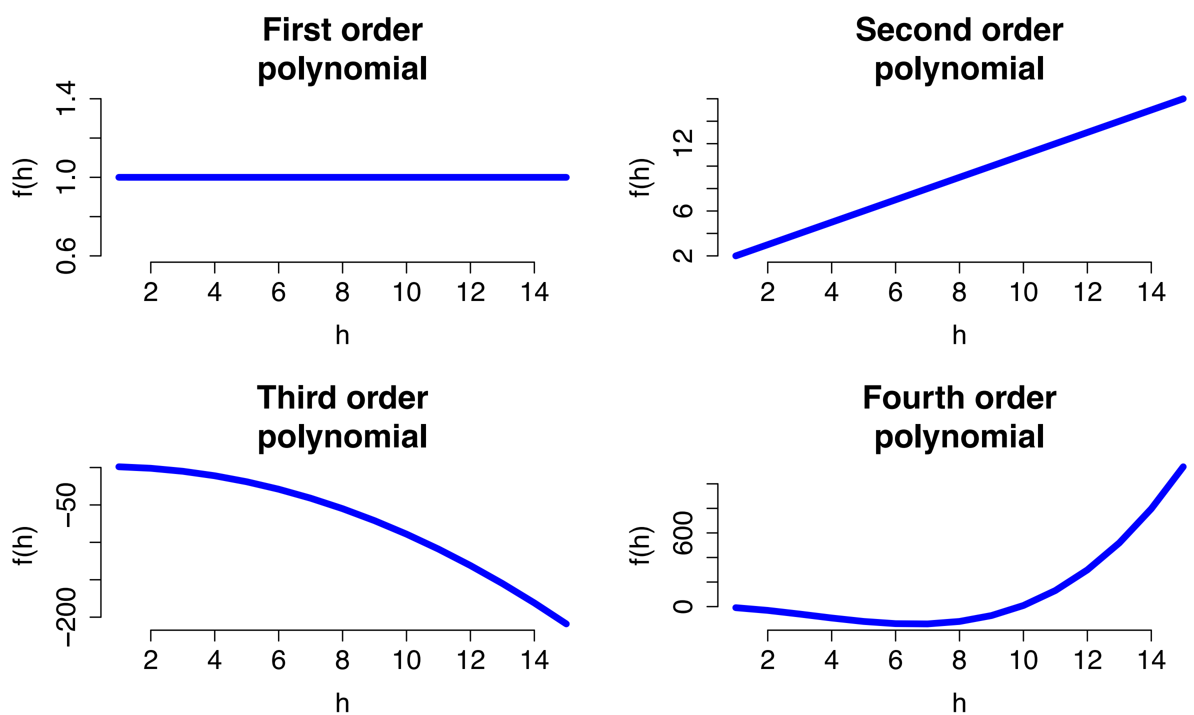
$$\mathbf{E}_p = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{J}_p(1) = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

<u>*Forecast function:*</u> $f_t(h) = a_{t,0} + a_{t,1}h + \cdots + a_{t,p-1}h^{p-1}.$

DLMs: Polynomial Trend Models

- Alternative representation for the **p-th order polynomial DLM:** $\{\mathbf{E}_p, \mathbf{L}_p, \cdot, \cdot\}$ with

$$\mathbf{L}_p = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$



DLMs : Similar Models

$$\mathcal{M}_1 \qquad\qquad \mathcal{M}_2$$

$$\{\mathbf{E}_p, \mathbf{J}_p(1), v_t, \mathbf{W}_t\} \qquad \{\mathbf{E}_p, \mathbf{L}_p, \cdot, \cdot\}$$

Similar Models

$$\mathcal{M}_1 \sim \mathcal{M}_2$$

$\mathbf{J}_p(1)$ and $\mathbf{L}_p$ are similar matrices

Forecast functions have the same form

Seasonal Factor Models

$$\{\mathbf{E}_p, \mathbf{P}, v_t, \mathbf{W}_t\}$$

- p: period
- $\boldsymbol{\theta}_t$: $p \times 1$ state vector of seasonal levels; first component is the level at the current time.
- The evolution matrix is

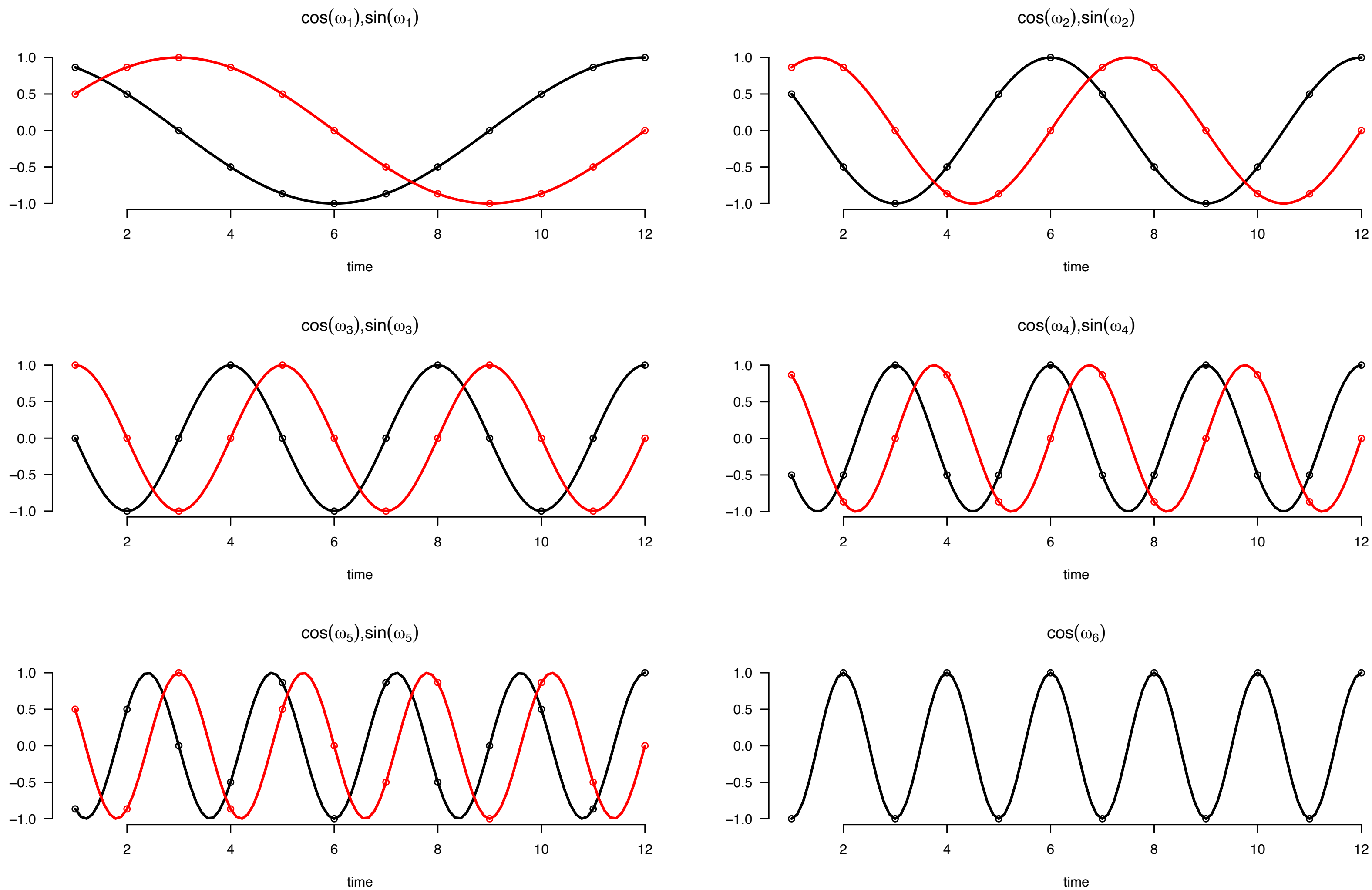
$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Forecast function: $f_t(h) = \mathbf{E}_p'\mathbf{P}^h\mathbf{m}_t = m_{t,j}, \quad j = p|h$

Seasonal effects: $\sum_{i=1}^{p} \theta_{t,i} = 0$

DLMs: Seasonal Models

E.g., p=12: the 12 seasonal factors can be represented in terms of 6 harmonic components related to the frequencies 2π/12, 2π/6, 2π/4, 2π/3, 5π/6, and π (Nyquist).



DLMs: Seasonal Models

- **Component Fourier Representation:** For a frequency $\omega \in (0,\pi)$ a corresponding harmonic component DLM is $\{\mathbf{E}_2, \mathbf{J}_2(1,\omega), \cdot\,, \cdot\,\}$, with

$$\mathbf{J}_2(1,\omega) = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}.$$

For $\omega = \pi$ we have $\{1, -1, \cdot\,, \cdot\,\}$.

Forecast function:

$$f_t(h) = a_t \cos(\omega h) + b_t \sin(\omega h) = A_t \cos(\omega h + \gamma_t),$$

with A_t the amplitude and γ_t the phase for $\omega \in (0,\pi)$ and

$$f_t(h) = (-1)^h A_t$$

for $\omega = \pi$.

Complete Fourier Representation

- <u>$p = 2m - 1$ odd:</u>

$$\mathbf{F} = (\mathbf{E}_2', \ldots, \mathbf{E}_2')'$$
$$\mathbf{G} = \text{blockdiag}(\mathbf{G}_1, \ldots, \mathbf{G}_{m-1}), \quad \mathbf{G}_j = \mathbf{J}_2(1, \omega_j),$$
$$\omega_j = 2\pi j/p, \quad j = 1 : (m-1)$$

- <u>$p = 2m$ even:</u>

$$\mathbf{F} = (\mathbf{E}_2', \ldots, \mathbf{E}_2', 1)'$$
$$\mathbf{G} = \text{blockdiag}(\mathbf{G}_1, \ldots, \mathbf{G}_{m-1}, -1)$$

<u>*Forecast function:*</u>

$$f_t(h) = \sum_{j=1}^{m-1} A_{t,j} \cos(\omega_j h + \gamma_{t,j}) + (-1)^h A_{t,m},$$

$A_{t,m} = 0$ if p odd.

DLMs: Regression Models

- **General regression DLM:** $\{\mathbf{F}_t, \mathbf{G}_t, v_t, \mathbf{W}_t\}$

$$\mathbf{F}_t = (x_{t,1}, \ldots, x_{t,k})', \quad \mathbf{G}_t = \mathbf{I}_k$$

Forecast function: $f_t(h) = a_{t,1}x_{t+h,1} + \ldots + a_{t,k}x_{t+h,k}$

- **Time-varying autoregression**

$$y_t = \sum_{i=1}^{p} \phi_{t,i} y_{t-i} + \nu_t,$$

$$\phi_{t,i} = \phi_{t-1,i} + w_{t,i}, \quad i = 1:p.$$

Forecast function:

$$f_t(h) = a_{t,1}y_{t+h-1} + \ldots + a_{t,p}y_{t+h-p}.$$

DLMs: Superposition

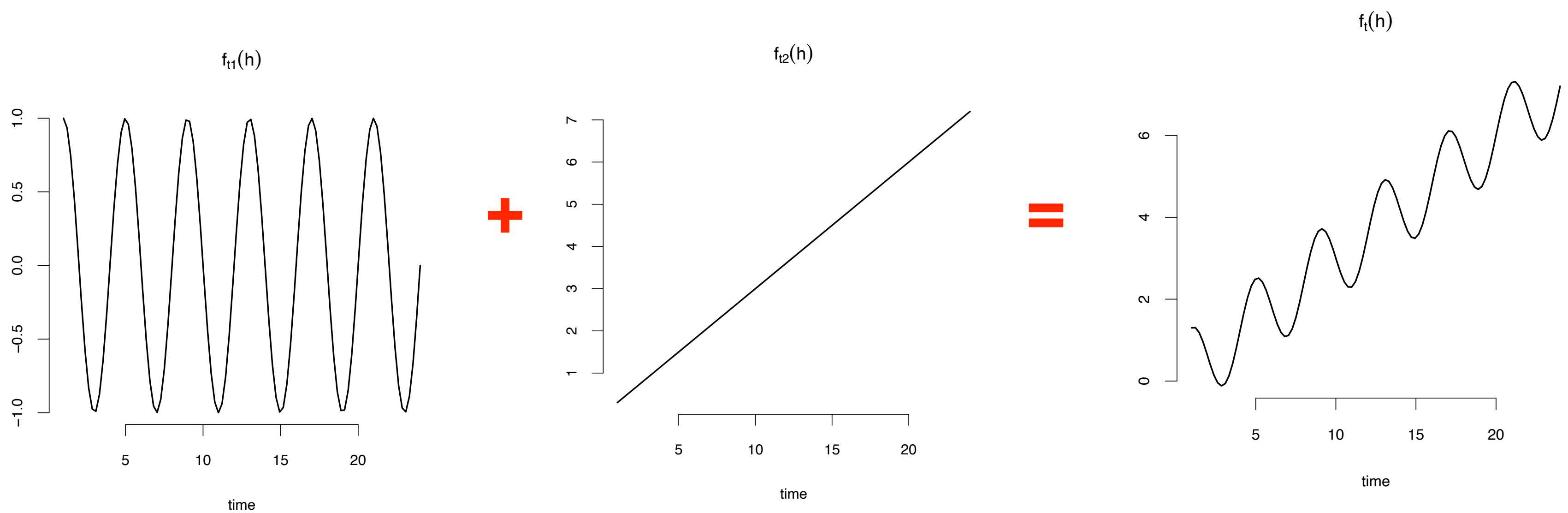
$\{\mathbf{F}_{i,t}, \mathbf{G}_{i,t}, v_{i,t}, \mathbf{W}_{i,t}\}$ for $i = 1 : m \Rightarrow y_t = \sum_{i=1}^{m} y_{i,t}$

has a DLM representation $\{\mathbf{F}_t, \mathbf{G}_t, v_t, \mathbf{W}_t\}$ given by

$$\mathbf{F}_t = (\mathbf{F}'_{1,t}, \ldots, \mathbf{F}'_{m,t})', \qquad \mathbf{G}_t = \text{blockdiag}(\mathbf{G}_{1,t}, \ldots, \mathbf{G}_{m,t})$$

$$v_t = \sum_{i=1}^{m} v_{i,t} \qquad \mathbf{W}_t = \text{blockdiag}(\mathbf{W}_{1,t}, \ldots, \mathbf{W}_{m,t})$$

Forecast function: $f_t(h) = \sum_{i=1}^{m} f_{i,t}(h)$



Example: Linear Trend + Seasonal Component $p = 4$

- Linear trend component: $\{\mathbf{F}_1, \mathbf{G}_1, \cdot, \cdot\}$

$$\mathbf{F}_1 = (1,0)' \qquad \mathbf{G}_1 = \mathbf{J}_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

- Full seasonal component: $\{\mathbf{F}_2, \mathbf{G}_2, \cdot, \cdot\}$
  $\omega_1 = 2\pi/4 = \pi/2,\ \ \omega_2 = \pi,$
  $\mathbf{F}_2 = (1,0,1)'$

$$\mathbf{G}_2 = \begin{pmatrix} \cos(\pi/2) & \sin(\pi/2) & 0 \\ -\sin(\pi/2) & \cos(\pi/2) & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

DLMs: Superposition

The resulting DLM is a 5-dimensional model $\{\mathbf{F}, \mathbf{G}, \cdot, \cdot\}$ with

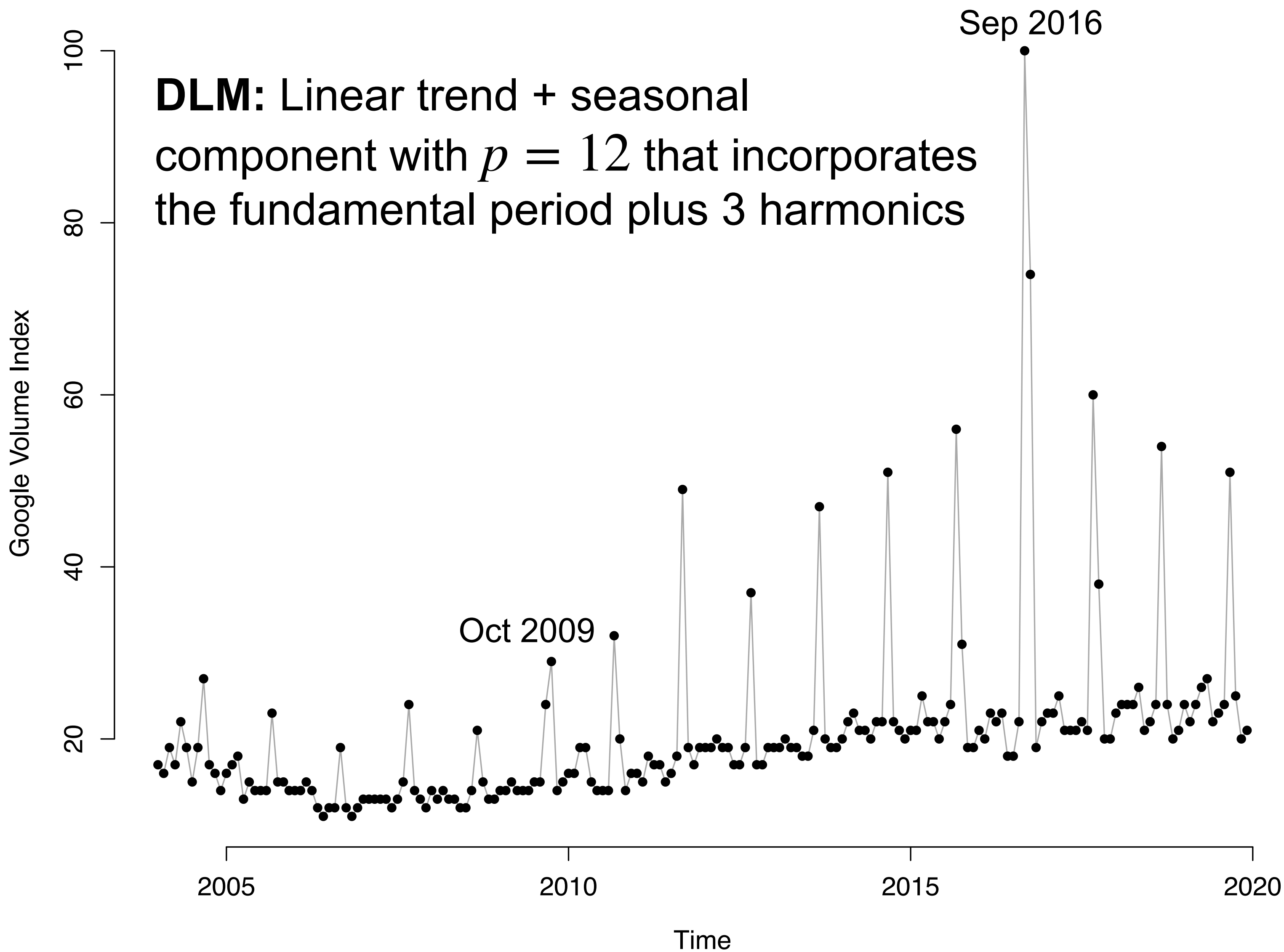
$$\mathbf{F} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Forecast function:

$$f_t(h) = (a_{t,1} + a_{t,2}h) + a_{t,3}\cos(\pi h/2) + a_{t,4}\sin(\pi h/2) + a_{t,5}(-1)^h$$

DLMs: Trend + Seasonal

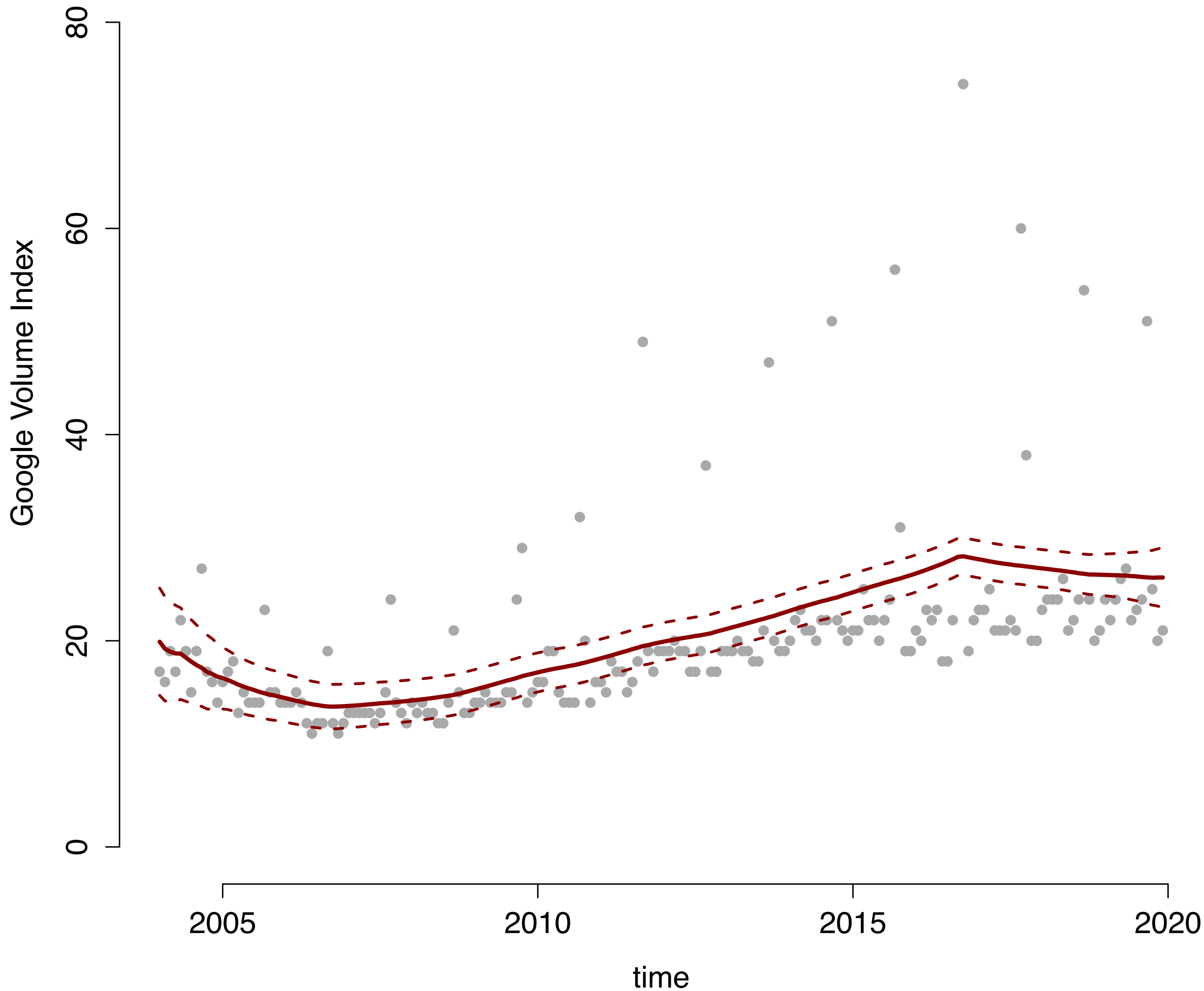
Google trends: “time series”



DLM: Linear trend + seasonal component with $p = 12$ that incorporates the fundamental period plus 3 harmonics

DLMs: Trend + Seasonal

Smoothing distribution: linear trend component



DLMs: Sequential Learning and Forecasting

Learning/Filtering

Case 1: observational variance <u>and</u> system covariance known

$$
\begin{aligned}
y_t &= \mathbf{F}_t' \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, \underline{v_t}), \\
\boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(0, \underline{\mathbf{W}_t})
\end{aligned}
$$

$$
\begin{array}{ccccccccccccc}
\boldsymbol{\theta}_0 & \rightarrow & \boldsymbol{\theta}_1 & \rightarrow & \boldsymbol{\theta}_2 & \rightarrow & \cdots & \rightarrow & \boldsymbol{\theta}_{t-1} & \rightarrow & \boldsymbol{\theta}_t & \rightarrow & \boldsymbol{\theta}_{t+1} & \rightarrow \\
& & \uparrow & & \uparrow & & & & \uparrow & & \uparrow & & \uparrow & \\
& & y_1 & & y_2 & & & & y_{t-1} & & y_t & & y_{t+1} &
\end{array}
$$

Begin with $(\boldsymbol{\theta}_0 | \mathcal{D}_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$ and sequentially obtain the filtering and forecasting distributions via Bayes' rule:

$$
p(\boldsymbol{\theta}_t | \mathcal{D}_t) \propto p(y_t | \boldsymbol{\theta}_t) \times p(\boldsymbol{\theta}_t | \mathcal{D}_{t-1}),
$$

$$
p(\boldsymbol{\theta}_t | \mathcal{D}_{t-1}) = \int p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) p(\boldsymbol{\theta}_{t-1} | \mathcal{D}_{t-1}) d\boldsymbol{\theta}_{t-1}.
$$

DLMs: Sequential Learning and Forecasting

- Evolution equation + $(\boldsymbol{\theta}_{t-1}|\mathcal{D}_{t-1})$ leads to

$$(\boldsymbol{\theta}_t|\mathcal{D}_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t), \quad \mathbf{a}_t = \mathbf{G}_t\mathbf{m}_{t-1}, \quad \mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t$$

- Observation equation + $(\boldsymbol{\theta}_t|\mathcal{D}_{t-1})$ leads to

$$(y_t|\mathcal{D}_{t-1}) \sim N(y_t|f_t, q_t), \quad f_t = \mathbf{F}_t'\mathbf{a}_t, \quad q_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + v_t$$

- After observing y_t the forecast error $e_t = y_t - f_t$ is obtained. Using Bayes' rule we have:

$$(\boldsymbol{\theta}_t|\mathcal{D}_t) \sim N(\mathbf{m}_t, \mathbf{C}_t)$$

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t, \quad \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t\mathbf{A}_t' q_t, \quad \mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t/q_t$$

DLMs: Sequential Learning and Forecasting

Case 2: observational variance constant but unknown and system covariance known (scaled by the observational variance):

$$\begin{aligned} y_t &= \mathbf{F}_t'\boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, v), \\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, v\mathbf{W}_t^*), \end{aligned}$$

with prior:

$$\begin{aligned} (\boldsymbol{\theta}_0 | \mathcal{D}_0) &\sim N(\mathbf{m}_0, v\mathbf{C}_0^*) \\ (\phi | \mathcal{D}_0) &\sim G(n_0/2, n_0 S_0/2), \quad \phi = v^{-1} \end{aligned}$$

leading to the following distributions…

DLMs: Sequential Learning and Forecasting

- $(\boldsymbol{\theta}_t | \mathcal{D}_{t-1}) \sim T_{n_{t-1}}(\mathbf{a}_t, \mathbf{R}_t)$

- $(y_t | \mathcal{D}_{t-1}) \sim T_{n_{t-1}}(f_t, q_t)$ with $q_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + S_{t-1}$

- $(\phi | \mathcal{D}_t) \sim G(n_t/2, n_t S_t/2)$

$$n_t = n_{t-1} + 1, \quad S_t = S_{t-1} + \frac{S_{t-1}}{n_t} \left(\frac{e_t^2}{q_t} - 1 \right)$$

- $(\boldsymbol{\theta}_t | \mathcal{D}_t) \sim T_{n_t}(\mathbf{m}_t, \mathbf{C}_t)$

$$\mathbf{C}_t = \frac{S_t}{S_{t-1}} \left(\mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' q_t \right)$$

DLMs: Sequential Learning and Forecasting

Forecasting

- $v_t, \mathbf{W}_t$ **known:**

$$(\boldsymbol{\theta}_{t+h}|\mathcal{D}_t) \sim N(\mathbf{a}_t(h), \mathbf{R}_t(h)), \quad (y_{t+h}|\mathcal{D}_t) \sim N(f_t(h), q_t(h)),$$

$$\begin{aligned} \mathbf{a}_t(h) &= \mathbf{G}_{t+h}\mathbf{a}_t(h-1), \ \ \mathbf{a}_t(0) = \mathbf{m}_t, \\ \mathbf{R}_t(h) &= \mathbf{G}_{t+h}\mathbf{R}_t(h-1)\mathbf{G}'_{t+h} + \mathbf{W}_{t+h}, \ \ \mathbf{R}_t(0) = \mathbf{C}_t \end{aligned}$$

$$f_t(h) = \mathbf{F}'_{t+h}\mathbf{a}_t(h), \ \ q_t(h) = \mathbf{F}'_{t+h}\mathbf{R}_t(h)\mathbf{F}_{t+h} + v_{t+h}.$$

- $v_t = v$ **unknown and** $\mathbf{W}_t$ **known:**

$$\begin{aligned} (\boldsymbol{\theta}_{t+h}|\mathcal{D}_t) &\sim T_{n_t}(\mathbf{a}_t(h), \mathbf{R}_t(h)), \\ (y_{t+h}|\mathcal{D}_t) &\sim T_{n_t}(f_t(h), q_t(h)), \end{aligned}$$

with $q_t(h) = \mathbf{F}'_{t+h}\mathbf{R}_t(h)\mathbf{F}_{t+h} + S_t$

DLMs: Retrospective Analysis/Smoothing

- **Smoothing:** observational variances and system covariances known. Let t < T and $\mathbf{a}_T(0) = \mathbf{m}_T, \quad \mathbf{R}_T(0) = \mathbf{C}_T.$
  Then, for $t = (T-1), (T-2), \ldots,$

$$(\boldsymbol{\theta}_t | \mathcal{D}_T) \sim N(\mathbf{a}_T(t-T), \mathbf{R}_T(t-T))$$

$$\begin{aligned} \mathbf{a}_T(t-T) &= \mathbf{m}_t - \mathbf{B}_t[\mathbf{a}_{t+1} - \mathbf{a}_T(t-T+1)] \\ \mathbf{R}_T(t-T) &= \mathbf{C}_t - \mathbf{B}_t[\mathbf{R}_{t+1} - \mathbf{R}_T(t-T+1)]\mathbf{B}_t' \end{aligned}$$

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

- **Smoothing:** observational variance unknown but fixed and system covariances known

$$(\boldsymbol{\theta}_t | \mathcal{D}_T) \sim T_{n_T}(\mathbf{a}_T(t-T), \mathbf{R}_T(t-T) S_T / S_t)$$

DLMs: Filtering & Smoothing

$$y_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t)$$
$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t)$$

$\{F_t, G_t, v_t, W_t\}$

$T = 3$

$$\begin{array}{ccccccccc}
 & & y_1 & & y_2 & & y_3 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\theta_0 & & \theta_1 & & \theta_2 & & \theta_3 & & \theta_4 \\
\underline{\mathbf{m}_0, \mathbf{C}_0} & \rightarrow & \mathbf{a}_1, \mathbf{R}_1 & \rightarrow & \underline{\mathbf{a}_2, \mathbf{R}_2} & \rightarrow & \underline{\mathbf{a}_3, \mathbf{R}_3} & \rightarrow & \mathbf{a}_4, \mathbf{R}_4 \\
 & & f_1, q_1 & & f_2, q_2 & & f_3, q_3 & & f_4, q_4 \\
 & & \underline{\mathbf{m}_1, \mathbf{C}_1} & & \underline{\mathbf{m}_2, \mathbf{C}_2} & & \underline{\mathbf{m}_3, \mathbf{C}_3} & & \\
\mathbf{a}_3(-3), \mathbf{R}_3(-3) & \leftarrow & \underline{\mathbf{a}_3(-2), \mathbf{R}_3(-2)} & & \underline{\mathbf{a}_3(-1), \mathbf{R}_3(-1)} & \leftarrow & \underline{\mathbf{a}_3(0), \mathbf{R}_3(0)} & &
\end{array}$$

Smoothing: $(\boldsymbol{\theta}_t \mid \mathcal{D}_T) \sim N(\mathbf{a}_T(t-T), \mathbf{R}_T(t-T)), \qquad t \leq T$

$t = (T-1), \ldots, 0:$ $\qquad \mathbf{a}_T(0) = \mathbf{m}_T, \ \mathbf{R}_T(0) = \mathbf{C}_T$

$$\mathbf{a}_T(t-T) = \mathbf{m}_t + \mathbf{B}_t[\mathbf{a}_T(t-T+1) - \mathbf{a}_{t+1}]$$

$$\mathbf{R}_T(t-T) = \mathbf{C}_t + \mathbf{B}_t[\mathbf{R}_T(t-T+1) - \mathbf{R}_{t+1}]\mathbf{B}_t', \qquad \mathbf{B}_t = \mathbf{C}_t\mathbf{G}_{t+1}'\mathbf{R}_{t+1}^{-1}$$

Discounting at the system level

Prior variance when $\mathbf{W}_t = \mathbf{0}$

$$\mathbf{R}_t = V(\boldsymbol{\theta}_t | \mathcal{D}_{t-1}) = \mathbf{P}_t + \mathbf{W}_t,$$

where $\mathbf{P}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t'$.

Assume

$$\mathbf{R}_t = \frac{\mathbf{P}_t}{\delta}, \quad \delta \in (0, 1].$$

$$\mathbf{W}_t = \frac{(1-\delta)}{\delta} \mathbf{P}_t.$$

Note that $\mathbf{W}_t$ is identified for all t given δ and $\mathbf{C}_0$.

Choosing δ

- $\delta \in (0.7,1]$ typically relevant in practical settings

- $\delta = 1$ corresponds to a static model; smaller values of δ lead to larger variation over time

- δ can be chosen to maximize

$$\log(\delta) \equiv \log[p(y_{1:T}|\mathcal{D}_0, \delta)] = \sum_{t=1}^{T} \log[p(y_t|\mathcal{D}_{t-1}, \delta)],$$

or to minimize the MSE/MAD

- Smoothing distributions are available and depend on δ

Component Discounting

If m DLMs are superposed, $\{\mathbf{F}_{i,t}, \mathbf{G}_{i,t}, v_{i,t}, \mathbf{W}_{i,t}\}, \;\; i = 1:m$, a component discount DLM can be considered defining

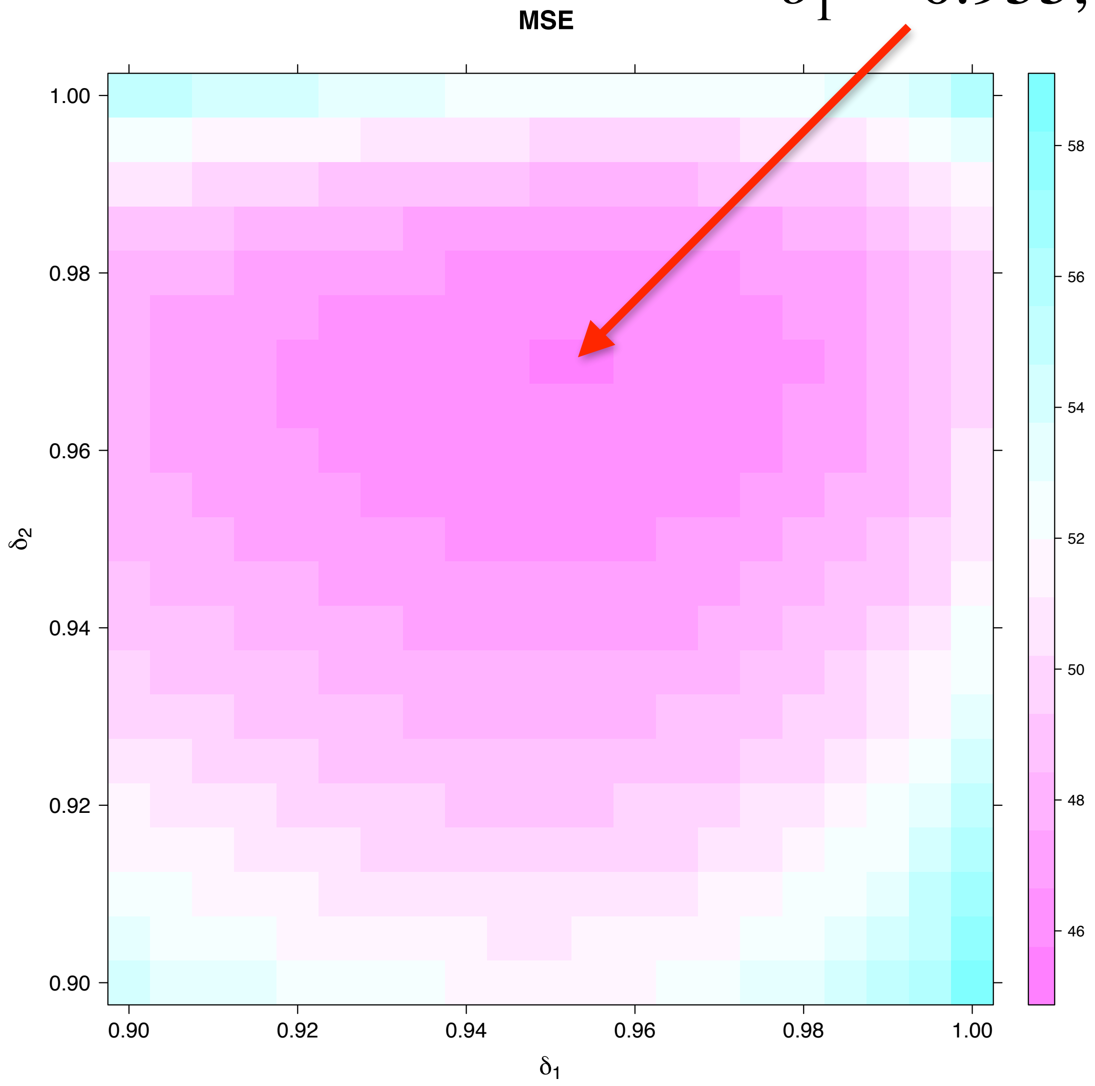
$$\mathbf{W}_{i,t} = \frac{(1-\delta_i)}{\delta_i}\mathbf{P}_{i,t},$$

with $\delta_i \in (0,1]$.

This allows greater flexibility, e.g., trend components may require more/less variation over time than seasonal components.

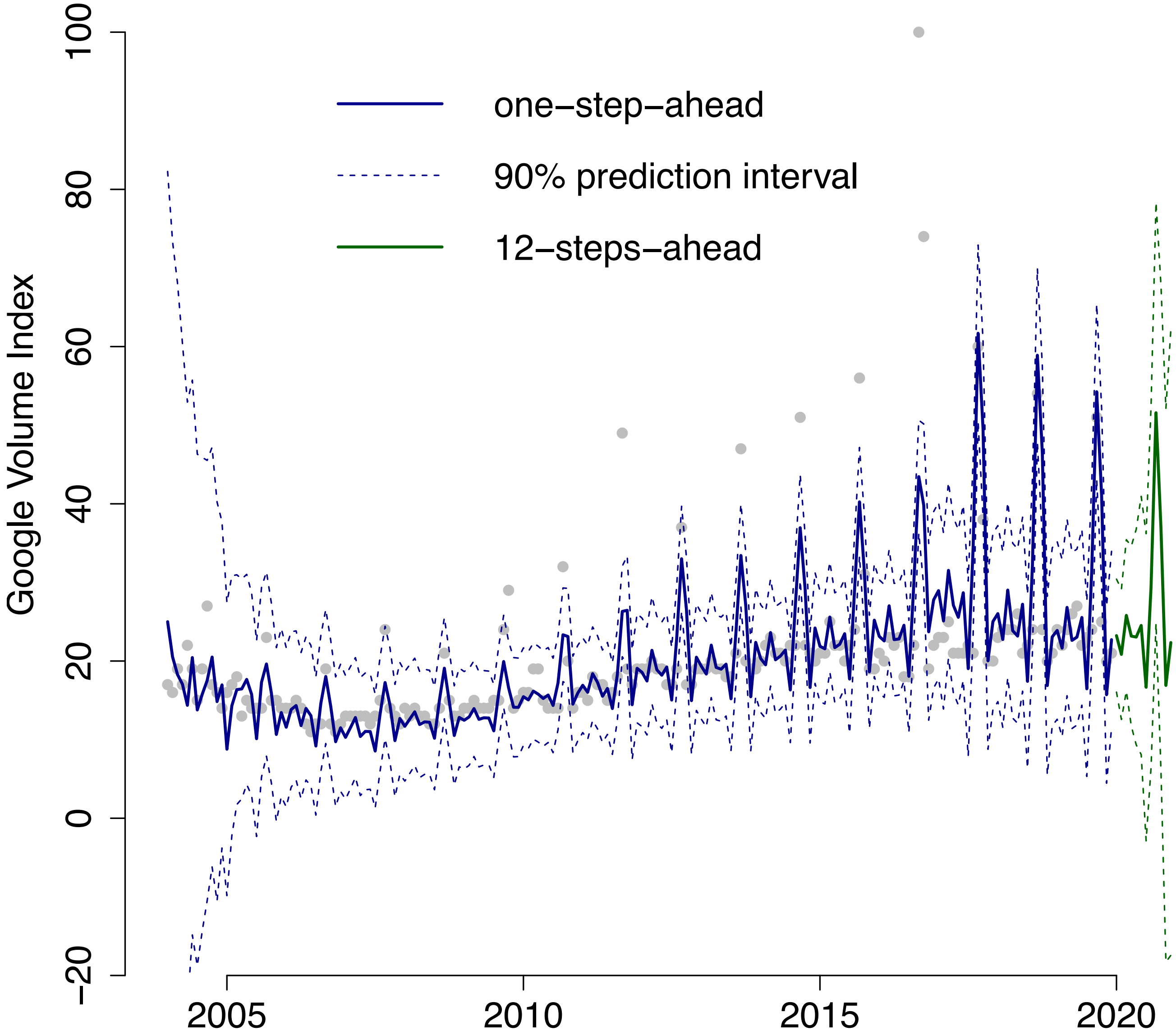
Google trends: “time series”

$$\hat{\delta}_1 = 0.955, \ \ \hat{\delta}_2 = 0.97$$



Component discounting:

- $\delta_1 \in [0.9, 1]$ for the linear trend
- $\delta_2 \in [0.9, 1]$ for the seasonal component



one−step−ahead
90% prediction interval
12−steps−ahead
Google Volume Index
100
80
60
40
20
0
−20
2005
2010
2015
2020

DLMs: Discounting at the observational level

$$y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t)$$
$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t)\,.$$

Assume that $(v_{t-1} \mid \mathscr{D}_{t-1}) \sim \mathrm{IG}(n_{t-1}/2, d_{t-1}/2)$, and that the evolution equation for v_t is given by

$$v_t = \frac{\delta_v v_{t-1}}{\eta_t},$$

$$(\eta_t \mid \mathscr{D}_{t-1}) \sim \mathrm{Beta}(\delta_v n_{t-1}/2, (1-\delta_v) n_{t-1}/2)$$

with $\delta_v \in (0,1]$ a discount factor

DLMs: Discounting at the observational level

$$(v_t \mid \mathscr{D}_t) \sim \mathrm{IG}(n_t/2, d_t/2),$$

$$n_t = \delta_v n_{t-1} + 1, \quad d_t = \delta_v d_{t-1} + S_{t-1} e_t^2 / q_t, \quad S_{t-1} = d_{t-1}/n_{t-1}$$

- The **filtering distributions** are computed as before
- The **smoothing distributions** are approximately inverse-Gamma,

$$(v_t \mid \mathscr{D}_T) \approx \mathrm{IG}(n_T(t-T)/2,\, d_T(t-T)/2),$$

$$n_T(t-T) = (1-\delta_v) n_t + \delta_v n_T(t-T+1), \quad d_T(t-T) = n_T(t-T) S_T(t-T),$$

$[S_T(t-T)]^{-1} = (1-\delta_v) S_t^{-1} + \delta_v [S_T(t-T+1)]^{-1}$, initialized at $n_T(0) = n_T, \quad S_T(0) = S_T$.

DLMs: Decompositions

$$y_t = x_t + \nu_t, \quad x_t = \mathbf{F}'\boldsymbol{\theta}_t, \quad \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t$$

Assume that $\mathbf{G}_t$ is a $p \times p$ matrix with p different eigenvalues $\alpha_{t,1}, \ldots, \alpha_{t,p}$ such that:

- C **pairs of complex eigenvalues:**

$$r_{t,j} \exp(\pm i\omega_{t,j}), \quad j = 1 : C$$

- $R = p - 2C$ **real eigenvalues:**

$$r_{t,j}, \quad j = (C+1) : (C+R)$$

$$\mathbf{G}_t = \mathbf{B}_t\mathbf{A}_t\mathbf{B}_t^{-1}$$

matrix of eigenvectors

diagonal matrix of eigenvalues

DLMs: AR models

AR(p)

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \epsilon_t, \quad \epsilon_t \sim N(0, v)$$

AR characteristic polynomial:

$$\Phi(u) = 1 - \phi_1 u - \ldots - \phi_p u^p = \prod_{i=1}^{p} (1 - \alpha_i u),$$

$\alpha_1, \ldots, \alpha_p$ are the reciprocal roots of this polynomial.

$|\alpha_i| < 1$ for all i $\longrightarrow$ the process is <u>stationary</u>

State-space (DLM) representations

(1) $\{\mathbf{F}_t, \mathbf{I}_p, v, \mathbf{0}\}, \quad \mathbf{F}_t' = (y_{t-1}, \ldots, y_{t-p})$

State-space (DLM) representations

(2) $\{\mathbf{E}_p, \mathbf{G}(\phi), 0, \mathbf{W}\}, \quad \mathbf{E}_p = (1, 0, \ldots, 0)'$

$$\boldsymbol{\theta}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})', \quad \mathbf{w}_t = (\epsilon_t, 0, \ldots, 0)',$$

$$\mathbf{G}(\phi) = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

and $\mathbf{W} = \mathrm{diag}(v, 0, \ldots, 0)$

This representation allows us to explore the underlying latent structure/decomposition of y_t

AR(p) decomposition:

$$\mathbf{G} = \mathbf{B}\mathbf{A}\mathbf{B}^{-1}$$

with $\mathbf{A} = \mathrm{diag}(\alpha_1, \ldots, \alpha_p)$ which leads to:

$$y_t = \sum_{j=1}^{C} x_{t,j} + \sum_{j=1}^{R} z_{t,j},$$

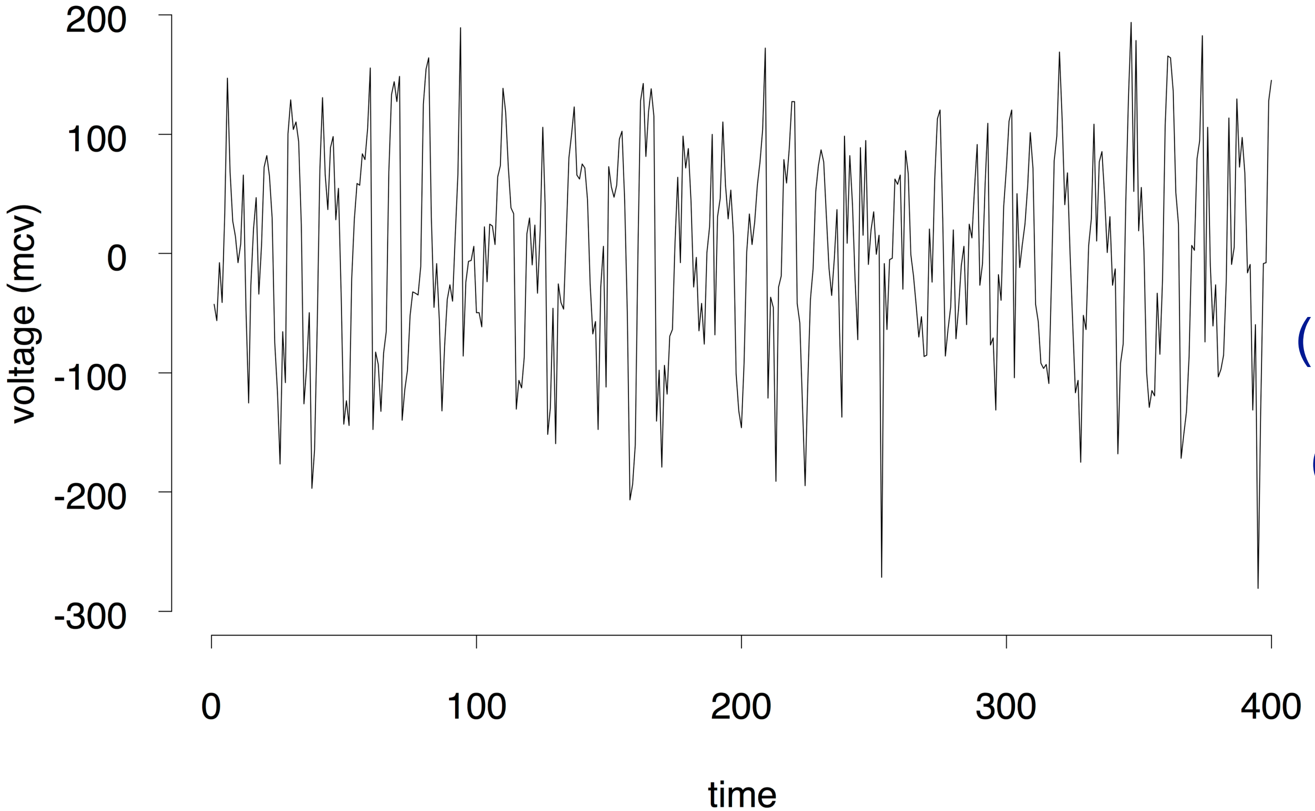
- $x_{t,j} \sim \mathrm{ARMA}(2,1)$ with constant moduli r_j and frequency ω_j for $j = 1,\ldots,C$
- $z_{t,j} \sim \mathrm{AR}(1)$ with constant moduli r_j, $j = C+1,\ldots,C+R$

AR models: EEG data

Bayesian inference in AR models

$$y_t = \sum_{i=1}^{8} \phi_i y_{t-i} + \epsilon_t, \quad \epsilon_t \sim N(0, v)$$

$$p(\boldsymbol{\phi}, v) \propto 1/v$$



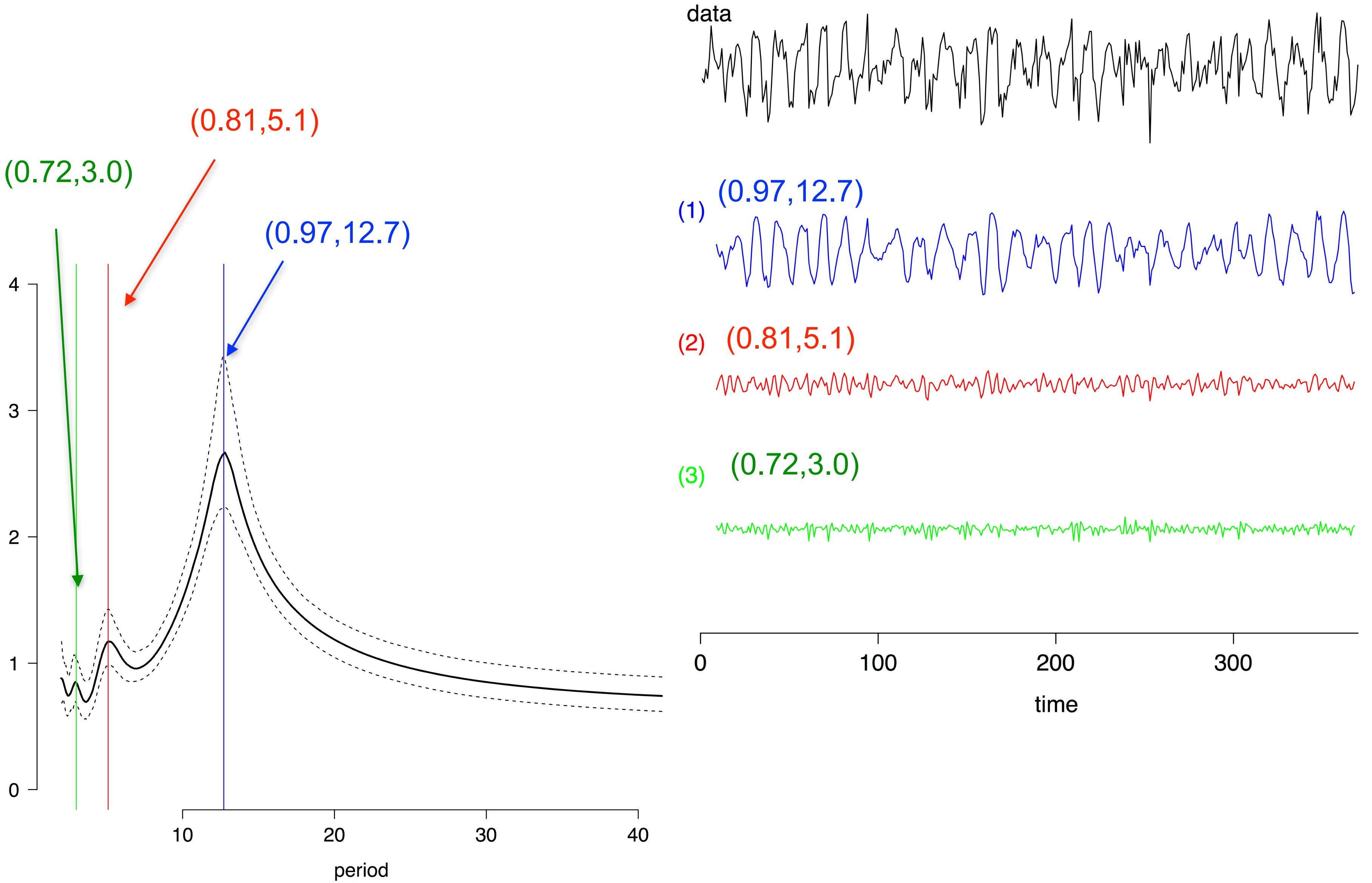
reciprocal roots, moduli and periods:

(0.97,12.7); (0.81,5.1)

(0.72,3.0); (0.63,2.2)

$E(\boldsymbol{\phi} \mid y_{9:400})$: (0.27,0.07,-0.13,-0.15,-0.11,-0.15,-0.23,-0.14)'

AR models: EEG data



DLMs: TVAR models

$$y_t = \phi_{t,1} y_{t-1} + \ldots + \phi_{t,p} y_{t-p} + \nu_t, \quad \nu_t \sim N(0,v)$$

$$\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, v\mathbf{W}_t^*),$$

State-space (DLM) representations

(1) $\mathbf{F}_t' = (y_{t-1}, \ldots, y_{t-p}), \quad \boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p})', \quad \mathbf{G}_t = \mathbf{I},$

$\mathbf{W}_t^*$ can be specified using a discount factor $\delta_W \in (0,1]$. A more general model with time-varying observational variance v_t can be considered, with the evolution of v_t specified using another discount factor $\delta_v \in (0,1]$.

(2) Conditioning on $\boldsymbol{\phi}_t = (\phi_{1,t}, \ldots, \phi_{p,t})'$ and v, we can also obtain the following representation of a TVAR process:

$$\{\mathbf{E}_p, \mathbf{G}_t, 0, \mathbf{W}\},$$

with $\mathbf{E}_p = (1,0,\ldots,0)$, $\boldsymbol{\theta}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, $\mathbf{w}_t = (\nu_t,0,\ldots,0)'$, $\mathbf{W} = \text{diag}(v,0,...,0)$, and

$$\mathbf{G}_t = \begin{pmatrix} \phi_{1,t} & \phi_{2,t} & \phi_{3,t} & \cdots & \phi_{p-1,t} & \phi_{p,t} \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

Decomposition: In the TVAR case we have

$$\mathbf{G}_t = \mathbf{B}_t \mathbf{A}_t \mathbf{B}_t^{-1},$$

with $\mathbf{A}_t = \mathrm{diag}(\alpha_{t,1}, \ldots, \alpha_{t,p})$, and $\alpha_{t,j}$ corresponding to the reciprocal roots of the AR characteristic polynomial at time t. This leads to the following decomposition of y_t :

$$y_t = \sum_{j=1}^{C} x_{t,j} + \sum_{j=1}^{R} z_{t,j},$$

DLMs: TVAR models

- $x_{t,j} \approx$ TVAR(2) with time-varying modulus $r_{t,j}$ and frequency $\omega_{t,j}$, $j = 1,\ldots,C$

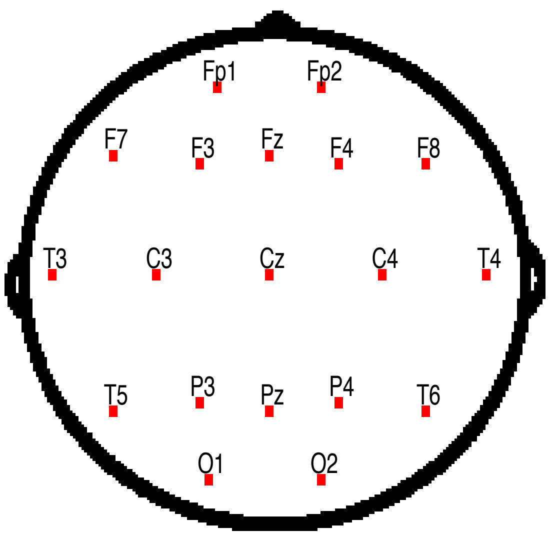
- $z_{t,j} \approx$ TVAR(1) with time-varying modulus $r_{t,j+C}$, $j = 1,\ldots,R$

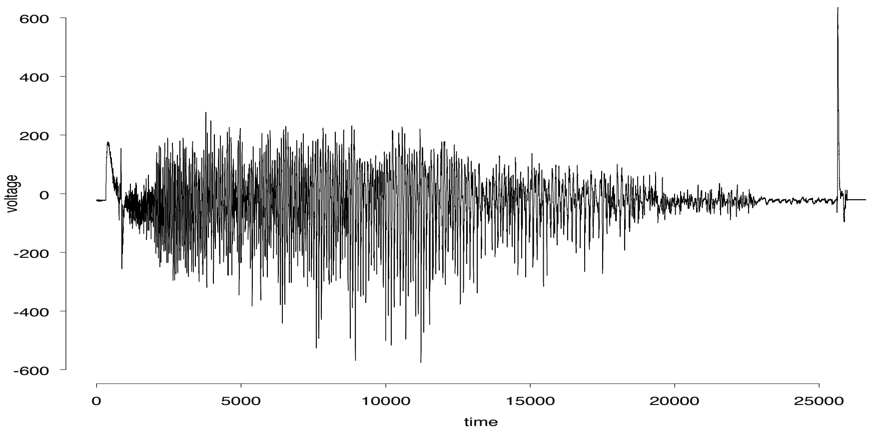
- The underlying structure of these processes above is <u>approximate</u> but the calculation of the components and the decomposition is <u>exact</u>

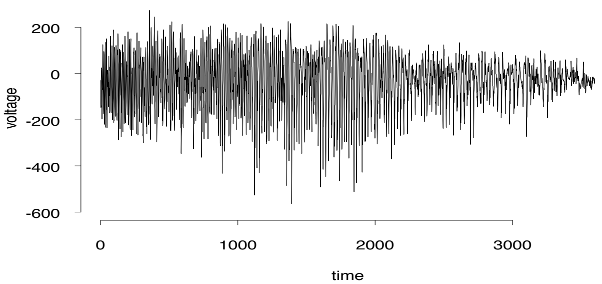
- Note that R and C do not need to be constant over time

Inferring latent structure in EEG data

- **Data:** 3,600 EEG observations recorded on a patient who received electroconvulsive therapy (ECT). The time series corresponds to a subsampled central portion of the entire data at channel *Cz*. 18 additional channels are available.







Inferring latent structure in EEG data

$$
\begin{aligned}
y_t &= \sum_{j=1}^{p} \phi_{j,t} y_{t-j} + \nu_t, \quad \nu_t \sim N[0, v_t(\delta_v)], \\
\boldsymbol{\phi}_t &= \boldsymbol{\phi}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N[\mathbf{0}, v_t(\delta_v)\mathbf{W}_t^*(\delta_\phi)],
\end{aligned}
$$

- δ_v, δ_ϕ discount factors in $(0,1]$.
- Optimal values of δ_v, δ_ϕ and p obtained by maximizing:

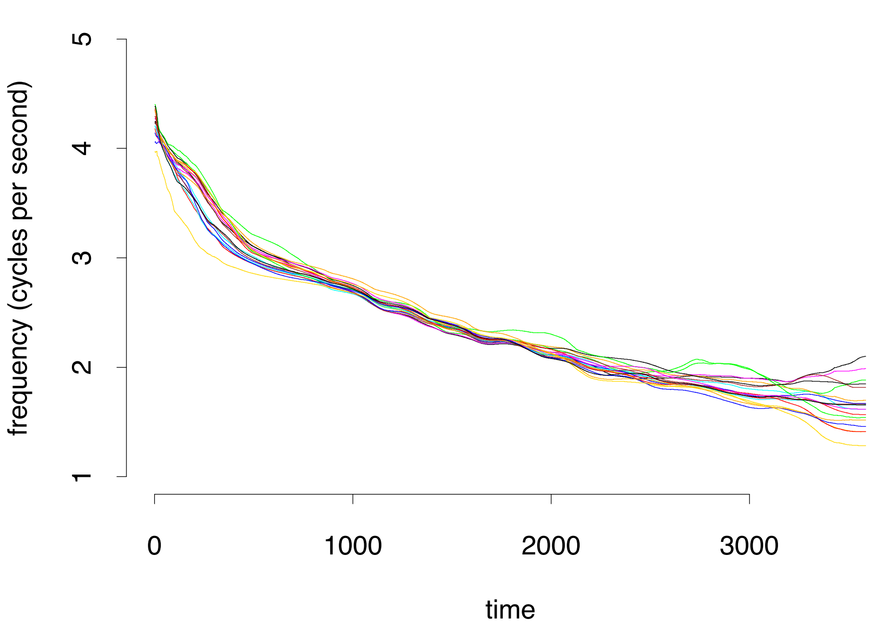
$$
\log(\delta_\phi, \delta_v, p) = \log[(y_{(p^*+1):T} | \mathcal{D}_{p^*})] = \sum_{t=p^*+1}^{T} \log[p(y_t | \mathcal{D}_{t-1})]
$$

over $[0.9,1] \times [0.9,1] \times [4,20]$ and $p^* = 20$.

Optimal values: $\hat{\delta}_\phi = 0.994, \ \hat{\delta}_v = 0.95, \ p = 12.$

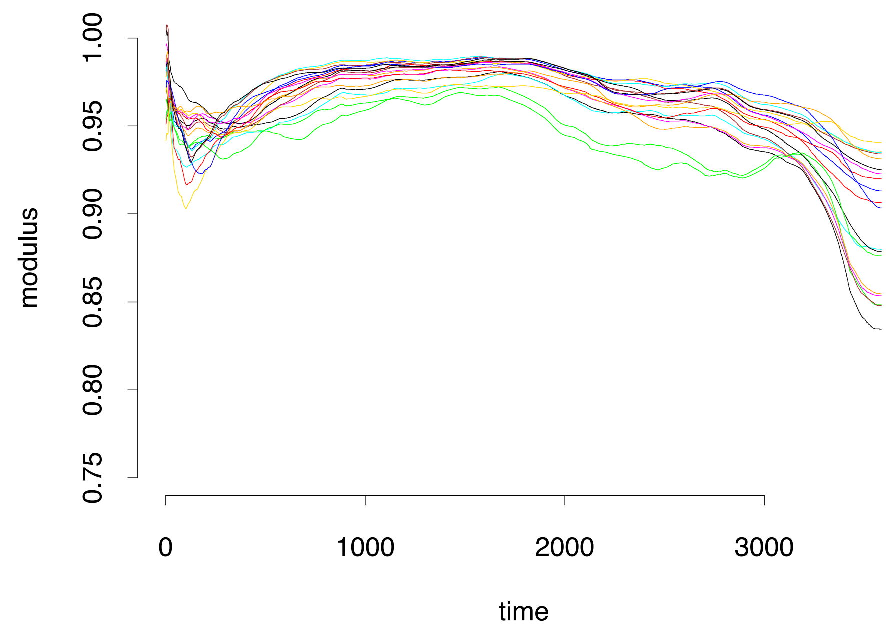
Univariate TVAR(12) analyses

EEG data: Trajectories of dominant frequencies



(a)

EEG data: Moduli of dominant frequencies



(b)

Models for multiple time series

- The 19 EEG series share a common underlying structure, however, individual time series analysis does not allow us to explore relationships across multiple time series.

- We explore a **dynamic factor model** for **multiple time series**

- In this case we do not have methods that allow us to obtain filtering/smoothing in closed form.

- This model is a **conditionally Gaussian dynamic linear models**

Conditionally Gaussian DLMs

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

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

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

Forward Filtering Backward Sampling 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 $\boldsymbol{\theta}_T^{(k)} \sim N(\mathbf{m}_T, \mathbf{C}_T)$.
3. For $t = (T-1) : 1$, sample $\boldsymbol{\theta}_t^{(k)} \sim N(\mathbf{l}_t, \mathbf{L}_t)$, with

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

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

Factor model for EEG data

CGDLM

CGDLM

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

- $\beta_{Cz} = 1$.
- $w/v = c$ with c known.
- $\mathbf{U}_t$ specified using a discount factor δ_ϕ.
- Conjugate priors for β, v and ϕ.

Factor model for EEG data: MCMC

- **Sampling** $x_t's$:

$$y_{i,t} = \beta_i x_t + \nu_{t,i}, \quad \nu_{t,i} \sim N(0,v),$$

$$x_t = \sum_{j=1}^{p} \phi_{t,j} x_{t-j} + \omega_t, \quad \omega_t \sim N(0,w),$$

CGDLM $\rightarrow$ FFBS

- Sampling $\boldsymbol{\phi}_t's$:

$$x_t = \sum_{j=1}^{p} \phi_{t,j} x_{t-j} + \omega_t, \quad \omega_t \sim N(0,w),$$

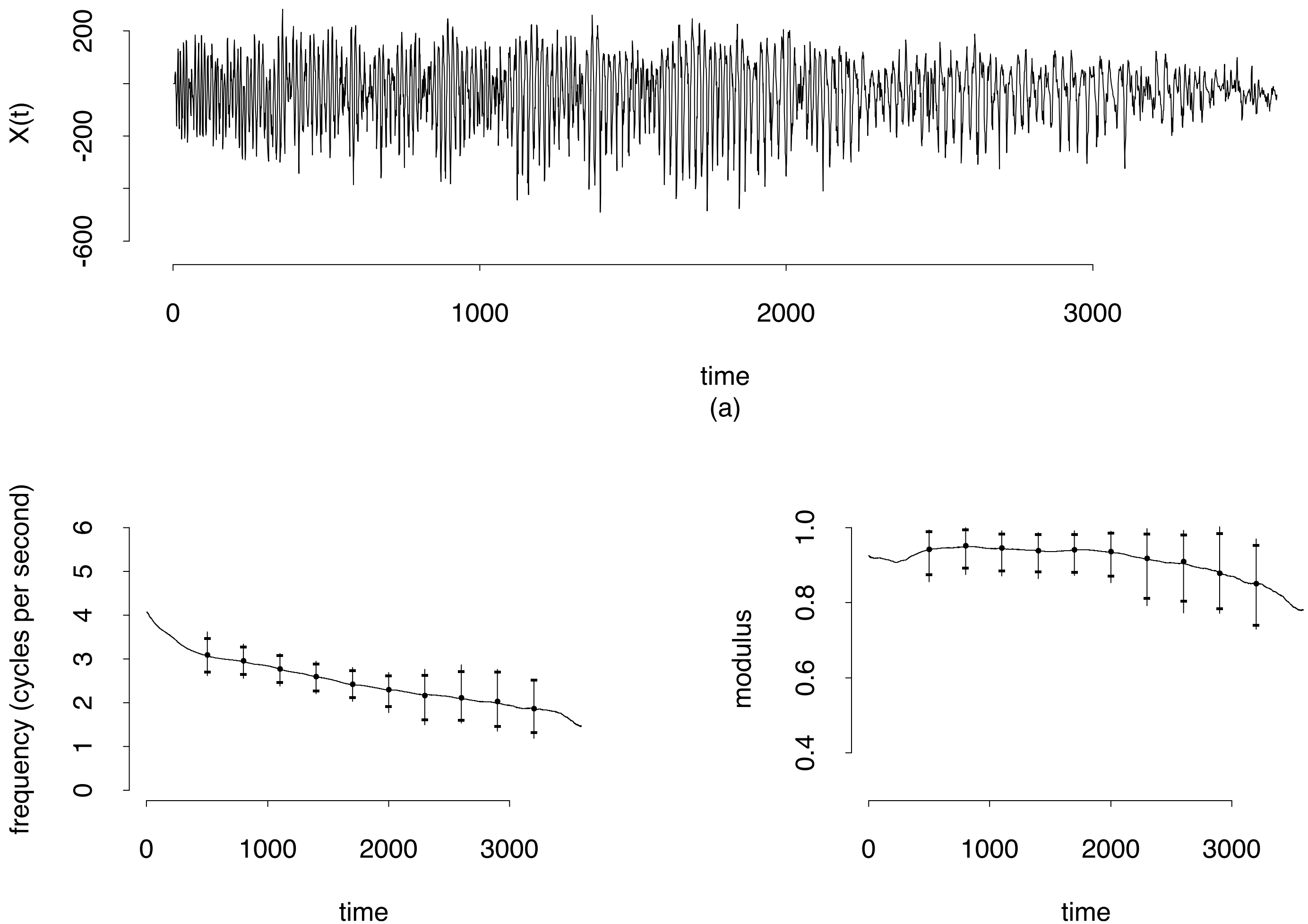
$$\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{U}_t)$$

CGDLM $\rightarrow$ FFBS

- Sample $\beta_{1:m}$ from $p(\beta_{1:m} \,|\, \mathbf{y}_{1:t}, x_{1:T}, v)$ (Gibbs step)
- Sample v from $p(v \,|\, \mathbf{y}_{1:t}, \beta_{1:m}, x_{1:T})$ (Gibbs step)

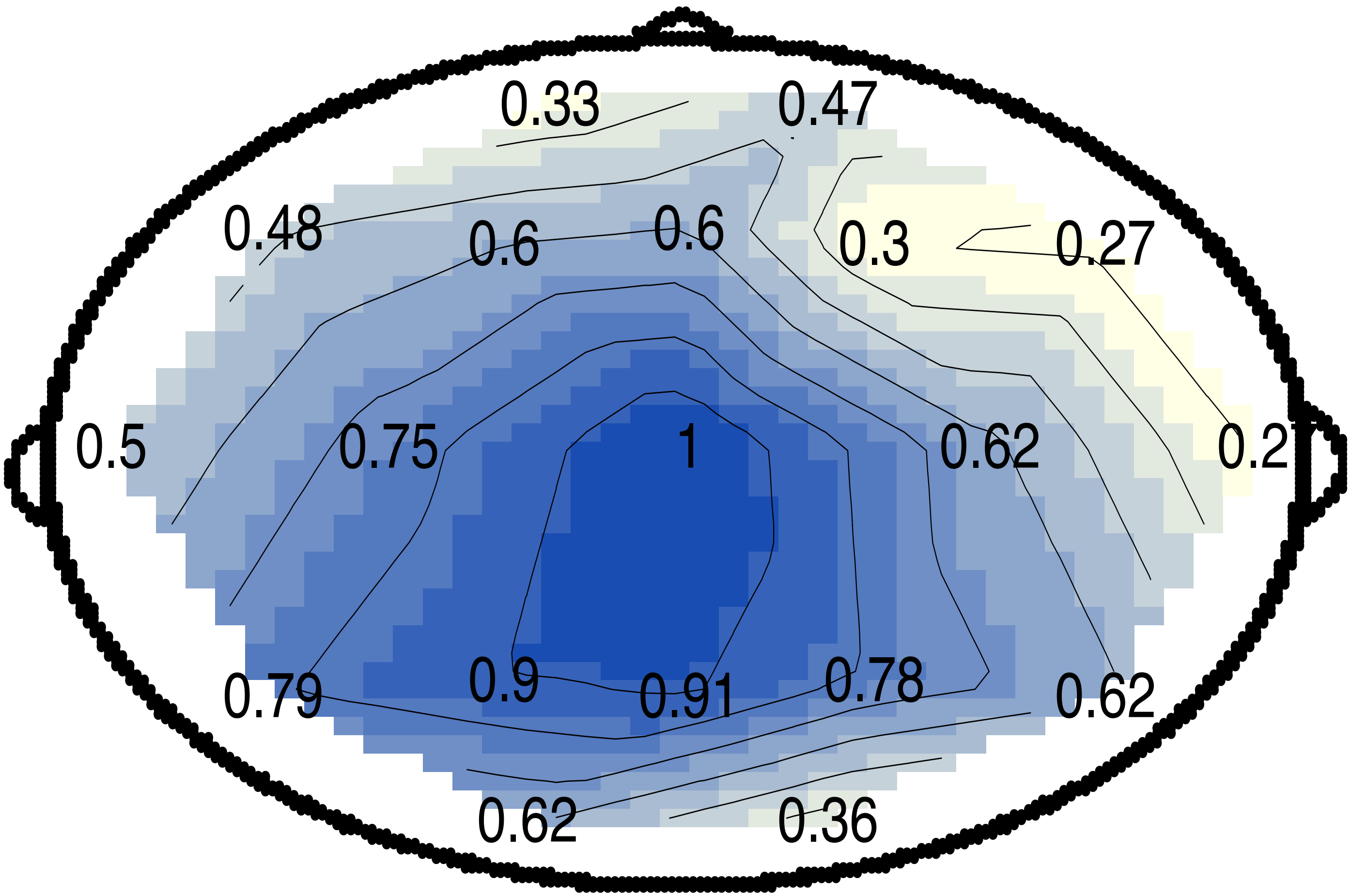
Factor model for EEG data

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



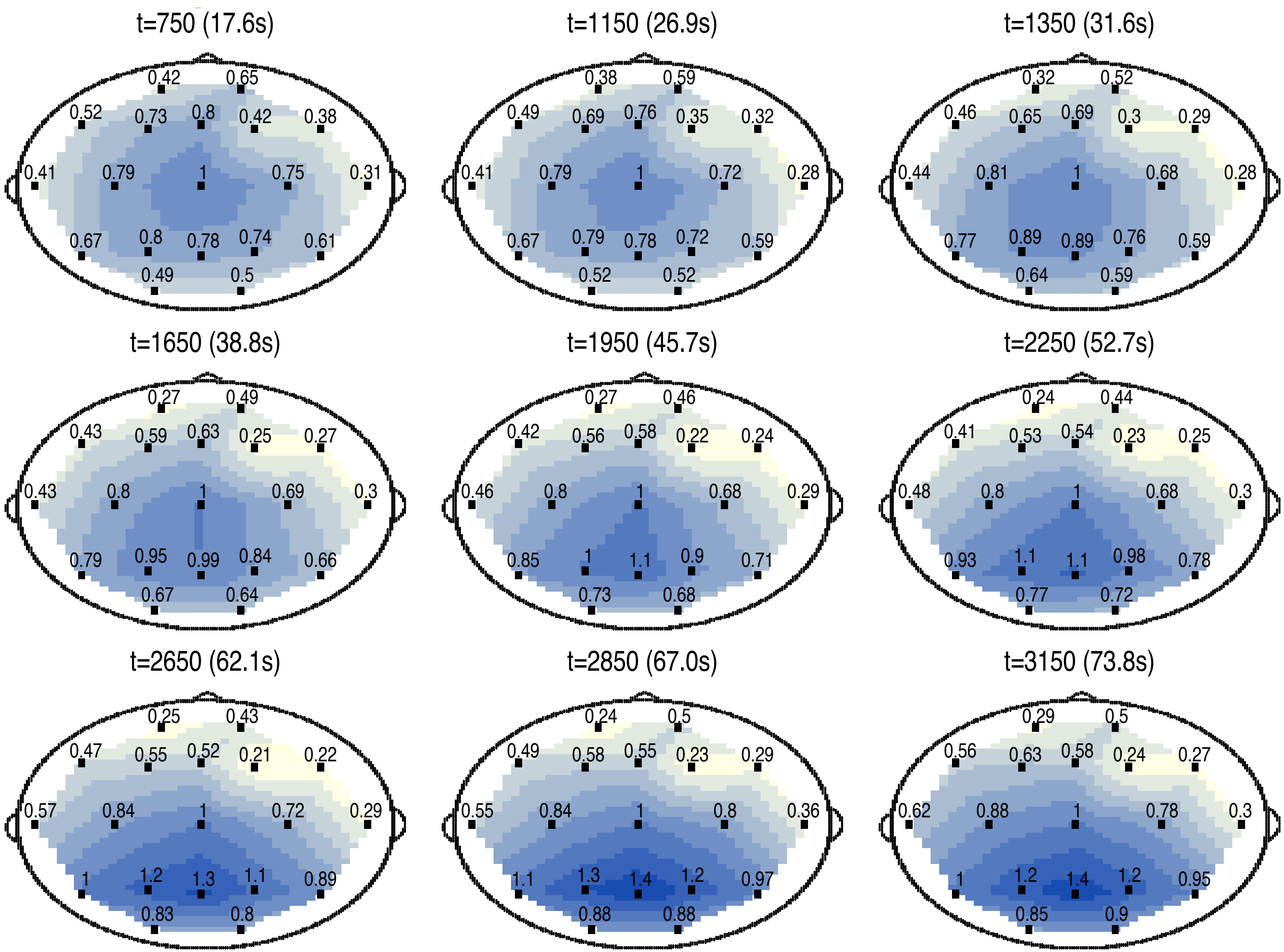
Factor model for EEG data

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



Factor model for EEG data: Extensions

Time-varying $\boldsymbol{\beta}_{i,t}$



A hierarchical DLM (Gamerman & Migon, 1993) can be specified as follows:

- **Observation equation:**

$$\mathbf{y}_t = \mathbf{F}_{1t}\boldsymbol{\theta}_{1t} + \mathbf{v}_{1t}, \quad \mathbf{v}_{1t} \sim N(\mathbf{0}, \mathbf{V}_{1t})$$

with $\dim(\mathbf{y}_t) = n \times 1$, $\dim(\boldsymbol{\theta}_{1t}) = r_1 \times 1$.

- **Structural equations:**

$$\boldsymbol{\theta}_{1t} = \mathbf{F}_{2t}\boldsymbol{\theta}_{2t} + \mathbf{v}_{2t}, \quad \mathbf{v}_{2t} \sim N(\mathbf{0}, \mathbf{V}_{2t}),$$
$$\boldsymbol{\theta}_{2t} = \mathbf{F}_{3t}\boldsymbol{\theta}_{3t} + \mathbf{v}_{3t}, \quad \mathbf{v}_{3t} \sim N(\mathbf{0}, \mathbf{V}_{3t}),$$

with $\dim(\boldsymbol{\theta}_{2t}) = r_2 \times 1$, $\dim(\boldsymbol{\theta}_{3t}) = r_3 \times 1$. Some models have only the first structural equation, or $\mathbf{F}_{3t} = \mathbf{I}$ and $\mathbf{V}_{3t} = \mathbf{0}$.

- Evolution/system equation:

$$\boldsymbol{\theta}_{3t} = \mathbf{G}_t \boldsymbol{\theta}_{3,t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t) .$$

Example: **Random effects DLM**

$$y_{t,i} = \beta_{t,i} + v_{t,i,1}, \quad v_{t,i,1} \sim N(0,\sigma^2), \quad i = 1:n$$

$$\beta_{t,i} = \mu_t + v_{t,i,2}, \quad v_{t,i,2} \sim N(0,\tau^2),$$

$$\mu_t = \mu_{t-1} + w_t, \quad w_t \sim N(0,W_t) .$$

Hierarchical DLMs: Inference

(1) Close form filtering and smoothing equations when (a) $\mathbf{V}_{1t}, \mathbf{V}_{2t}, \mathbf{V}_{3t}$ are known and $\mathbf{W}_t$ is known or specified via discount factors or (b) $\mathbf{V}_{1t} = \sigma^2\mathbf{I}$ and $\mathbf{V}_{it}, \mathbf{W}_t$ and the prior covariance matrix $\mathbf{C}_{K,0}$ are scaled by σ^2. We can use discount factors again.

(2) Alternatively, inverse-Wishart formulations for the covariance matrices can be used and inference can be obtained via MCMC or approximate inference.

Hierarchical DLMs: Inference: Case 1(a) K stages

$$\mathbf{y}_t \mid \boldsymbol{\theta}_{1t} \sim N(\mathbf{F}_{1t}\boldsymbol{\theta}_{1t}, \mathbf{V}_{1t}),$$

$$\boldsymbol{\theta}_{i,t} \mid \boldsymbol{\theta}_{i+1,t}, \mathscr{D}_{t-1} \sim N(\mathbf{F}_{i+1,t}\boldsymbol{\theta}_{i+1,t}, \mathbf{V}_{i+1,t}), \quad i = 1, \ldots, K-1$$

$$\boldsymbol{\theta}_{K,t} \mid \boldsymbol{\theta}_{K,t-1}, \mathscr{D}_{t-1} \sim N(\mathbf{G}_t\boldsymbol{\theta}_{K,t-1}, \mathbf{W}_t),$$

Prior: $\boldsymbol{\theta}_{K,0} \mid \mathscr{D}_0 \sim N(\mathbf{m}_{K,0}, \mathbf{C}_{K,0})\,.$

Relationship between consecutive stages i **and** j, $j > (i+1)$:

$$\boldsymbol{\theta}_{it} \mid \boldsymbol{\theta}_{jt}, \mathscr{D}_{t-1} \sim N(\mathbf{E}_{i,j,t}\boldsymbol{\theta}_{j,t}, \mathbf{V}^*_{i,j,t}),$$

$$\mathbf{E}_{i,j,t} = \mathbf{F}_{i+1,t} \times \ldots \times \mathbf{F}_{jt}, \quad 0 \le i < j \le K,$$

$$\mathbf{E}_{i,i,t} = \mathbf{I}_{r_i} \quad 1 \le i < K, \quad \mathbf{V}^*_{i,j,t} = \sum_{l=(i+1)}^{j} \mathbf{E}_{i,l-1,t}\mathbf{V}_{lt}\mathbf{E}'_{i,l-1,t}\,.$$

Hierarchical DLMs: Inference: Case 1(a) K stages

- **Prior at time t**

$$\boldsymbol{\theta}_{i,t} \mid \mathcal{D}_{t-1} \sim N(\mathbf{a}_{i,t}, \mathbf{R}_{i,t}), \quad k = 1, \ldots, K,$$

$$\mathbf{a}_{i,t} = \mathbf{F}_{i+1,t}\mathbf{a}_{i+1,t}, \quad \mathbf{R}_{i,t} = \mathbf{F}_{i+1,t}\mathbf{R}_{i+1,t}\mathbf{F}'_{i+1,t} + \mathbf{V}_{i+1,t}, \quad i = 1, \ldots, K-1,$$

$$\mathbf{a}_{K,t} = \mathbf{G}_t\mathbf{m}_{K,t-1}, \quad \mathbf{R}_{K,t} = \mathbf{G}_t\mathbf{C}_{k,t-1}\mathbf{G}'_t + \mathbf{W}_t,$$

- **Predictive distributions (one-step ahead)**

$$\mathbf{y}_t \mid \mathcal{D}_{t-1} \sim N(\mathbf{f}_t, \mathbf{Q}_t),$$

$$\mathbf{f}_t = \mathbf{F}_{1,t}\mathbf{a}_{1,t}, \quad \mathbf{Q}_t = \mathbf{F}_{1,t}\mathbf{R}_{1,t}\mathbf{F}'_{1,t} + \mathbf{V}_{1,t},$$

- **Posterior at time t**

$$\boldsymbol{\theta}_{i,t} \mid \mathcal{D}_t \sim N(\mathbf{m}_{i,t}, \mathbf{C}_{i,t}), \quad k = 1, \ldots, K,$$

$$\mathbf{m}_{i,t} = \mathbf{a}_{i,t} + \mathbf{S}_{i,t}\mathbf{Q}_t^{-1}(\mathbf{y}_t - \mathbf{f}_t), \quad \mathbf{S}_{i,t} = \mathbf{R}_{i,t}\mathbf{E}'_{0,i,t},$$

$$\mathbf{C}_{i,t} = \mathbf{R}_{i,t} - \mathbf{S}_{i,t}\mathbf{Q}_t^{-1}\mathbf{S}'_{i,t}\,.$$

Hierarchical DLMs: Inference: Case 1(a) smoothing

$$\boldsymbol{\theta}_{i,t} \,|\, \mathscr{D}_T \sim N(\mathbf{m}_{i,T}(t-T), \mathbf{R}_{i,T}(t-T)), \quad t = 1,\ldots,T, \;\; k = 1,\ldots,K,$$

$$\mathbf{m}_{i,T}(t-T) = \mathbf{m}_{i,t} + \mathbf{A}'_{i,t}\mathbf{R}^{-1}_{i,t+1}(\mathbf{m}_{i,t}(t-T+1) - \mathbf{a}_{i,t+1})$$

$$\mathbf{C}_{i,T}(t-T) = \mathbf{C}_{i,t} - \mathbf{A}'_{i,t}\mathbf{R}^{-1}_{i,t+1}[\mathbf{R}_{i,t} - \mathbf{C}_{i,T}(t-T+1)]\mathbf{R}^{-1}_{i,t+1}\mathbf{A}_{i,t},$$

$$\mathbf{A}_{i,t} = \mathbf{G}_t\mathbf{C}_{i,t}, \quad \mathbf{m}_{i,T}(0) = \mathbf{m}_{i,T}, \quad \mathbf{C}_{i,T}(0) = \mathbf{C}_{i,T}\,.$$

Hierarchical DLMs: Inference: Case 1(b)

Assuming $\mathbf{V}_{1,t} = \sigma^2\mathbf{I},$ and
$\mathbf{V}_{i,t} = \sigma^2\mathbf{V}^*_{i,t},\ i > 1,\ \ \mathbf{W}_t = \sigma^2\mathbf{W}^*_t,\ \mathbf{C}_{K,0} = \sigma^2\mathbf{C}^*_{K,0},$ with
$\mathbf{V}^*_{i,t},\ \mathbf{W}^*_t,\ \mathbf{C}^*_{K,0}$ known and

$$\sigma^2 \,|\, \mathcal{D}_{t-1} \sim \mathrm{IG}(n_{t-1}/2, d_{t-1}/2),$$

we obtain closed form filtering and smoothing with normal distributions conditional on σ^2 and

$$\sigma^2 \,|\, \mathcal{D}_t \sim \mathrm{IG}(n_t/2, d_t/2),$$

with

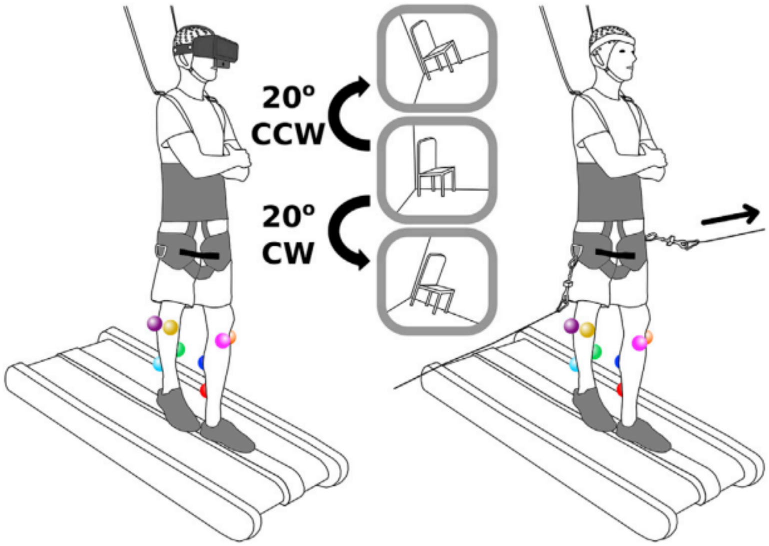
$$n_t = n_{t-1} + n, \quad d_t = d_{t-1} + (\mathbf{y}_t - \mathbf{f}_t)'\mathbf{Q}_t^{-1}(\mathbf{y}_t - \mathbf{f}_t)\,.$$

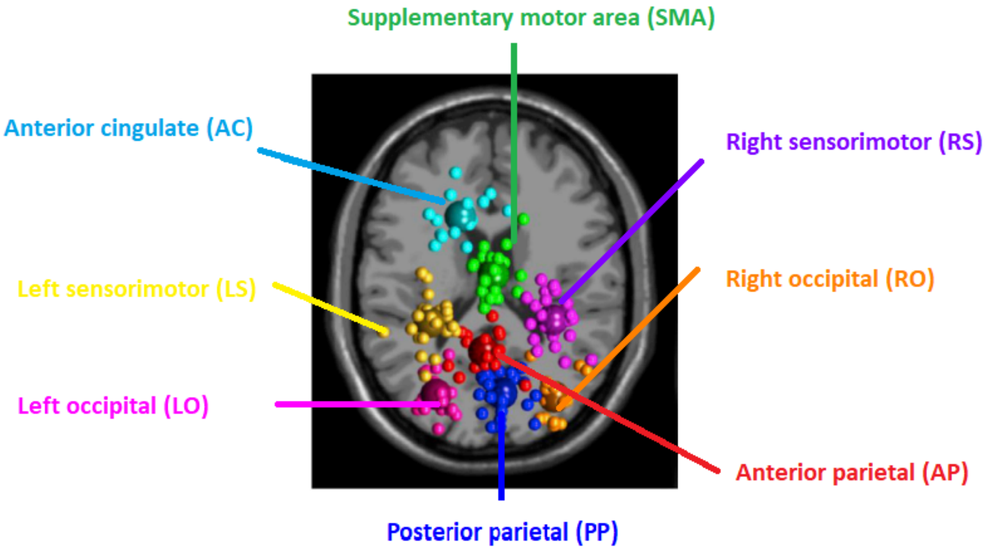
Discount factors can be used to specify $\mathbf{V}^*_{i,t}$ and $\mathbf{W}^*_t$.

Case study: Multi-trial EEG

EEG data from Peterson and Ferris (2019)

- <u>Experimental conditions:</u> walk/stand, pull/rotate; perturbations can be physical (1s) or visual (0.5s)
- 136 EEG channels summarized in 8 cortical clusters; sampling rate 128 Hz
- We analyze multi-trial data from a single subject

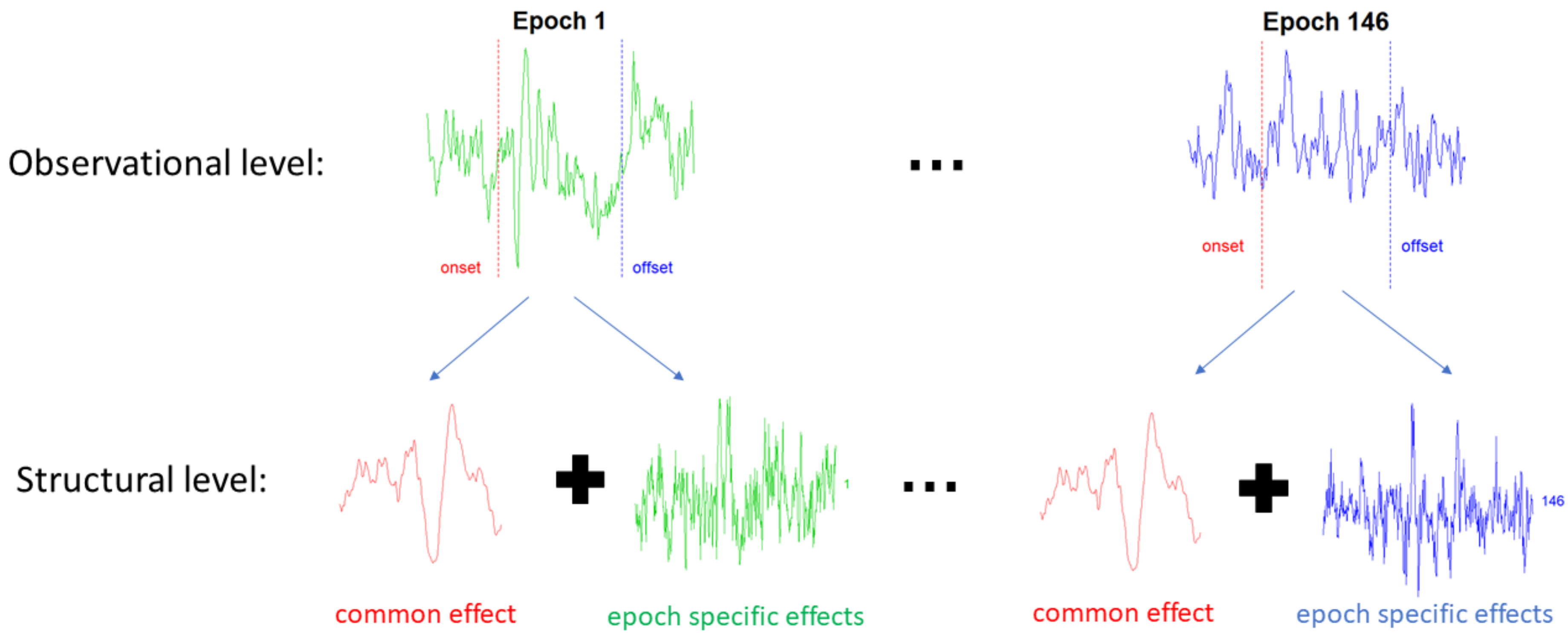




Case study: Multi-trial EEG

- Data split into epochs with going from -1s to +2s from perturbation onset
- 146 epochs

Subject 25: supplementary motor area, stand pull



Goal: Infer the latent common time-frequency characteristics of the multi-trial data for each cluster

Case study: Multi-trial EEG

We assume a TVAR model structure for each trial i :

$$y_{i,t} = \sum_{j=1}^{P} a_{i,j,t}^{(P)} y_{i,t-j} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma^2),$$

- A DLM representation of a TVAR model requires state-space vectors of dimension P; therefore, filtering and smoothing require operations (including inversions) with matrices of dimension $P \times P$
- We represent this model in the partial autocorrelation (PARCOR) domain. We extend Yang, Holan and Wikle (2016) to consider a hierarchical structure that allow us to jointly analyze the multi-trial data

TV-PARCOR representation of a TVAR

TVAR model in the PARCOR domain that describes the variation in the PARCOR coefficients through a DLM

Standard TVAR model of order P:

Observation equation: $$y_t = \sum_{j=1}^{P} \phi_{j,t}^{(P)} y_{t-j} + \nu_t, \quad \nu_t \sim N(0, v_t)$$

Evolution equation: $$\boldsymbol{\phi}_t^{(P)} = \boldsymbol{\phi}_{t-1}^{(P)} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(0, \mathbf{W}_t)$$

This representation requires operations with $P \times P$ matrices for filtering and smoothing.

TV-PARCOR representation of a TVAR

Forward prediction error

$$f_t^{(P)} = y_t - \sum_{j=1}^{P} \phi_{j,t}^{(P)} y_{t-j}$$

Backward prediction error

$$b_t^{(P)} = y_t - \sum_{j=1}^{P} \alpha_{j,t}^{(P)} y_{t+j}$$

For $m = 1, \ldots, P$, the m-th stage of the lattice filter is given by:

Forward DLM

$$f_t^{(m-1)} = \beta_{f,t}^{(m)} b_{t-m}^{(m-1)} + f_t^{(m)}, \quad f_t^{(m)} \sim N(0, \sigma_{f,m}^2)$$

$$\beta_{f,t}^{(m)} = \beta_{f,t-1}^{(m)} + w_{f,t}^{(m)}, \quad w_{f,t}^{(m)} \sim N(0, W_{f,t,m})$$

Backward DLM

$$b_t^{(m-1)} = \beta_{b,t}^{(m)} f_{t+m}^{(m-1)} + b_t^{(m)}, \quad b_t^{(m)} \sim N(0, \sigma_{b,m}^2)$$

$$\beta_{b,t}^{(m)} = \beta_{b,t-1}^{(m)} + w_{b,t}^{(m)}, \quad w_{b,t}^{(m)} \sim N(0, W_{b,t,m})$$

TV-PARCOR representation of a TVAR

- The TV-PARCOR model fully avoids matrix computations. $2P$ DLMs with univariate state-space parameters are fitted.
- Closed form filtering and smoothing available if conjugate priors are used and variances are specified via discount factors.
- Estimates of the TVAR coefficients are obtained using the Durbin-Levinson recursions:

For $m = 1, \ldots, P$,

$$\phi_{m,t}^{(m)} = \beta_{f,t}^{(m)}, \quad \alpha_{m,t}^{(m)} = \beta_{b,t}^{(m)},$$

$$\phi_{k,t}^{(m)} = \phi_{k,t}^{(m-1)} - \phi_{m,t}^{(m)} \alpha_{m-k,t}^{(m-1)},$$

$$\alpha_{k,t}^{(m)} = \alpha_{k,t}^{(m-1)} - \alpha_{m,t}^{(m)} \phi_{m-k,t}^{(m-1)},$$

- We can also obtain estimates of any functions of these coefficients, such as the spectral density of the TVAR process.

TV-PARCOR representation of a TVAR

Algorithm of Yang, Holan & Wikle (2016):

- Step 1. Set values of the discount factors for the observational and system variances for all $m = 1 : P$ and set $f_t^{(0)} = b_t^{(0)} = y_t$.

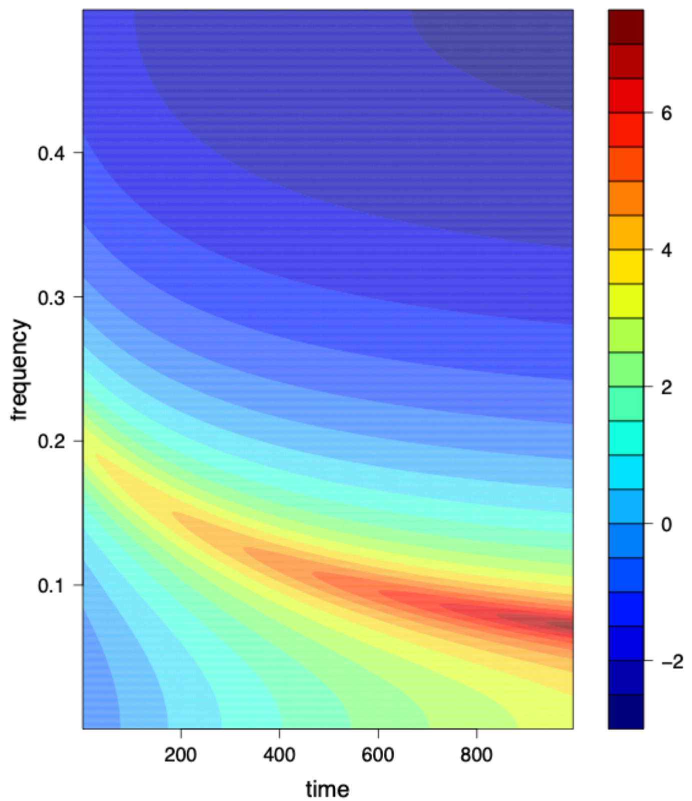
- Step 2. (a) Obtain estimates of $\beta_{f,t}^{(1)}, f_t^{(1)}$ and $\sigma_{f,1}^2$ using the forward DLM model at stage $m = 1$ (b) Obtain estimates of $\beta_{b,t}^{(1)}, b_t^{(1)}$ and $\sigma_{b,1}^2$ using the backward DLM model at stage $m = 1$.

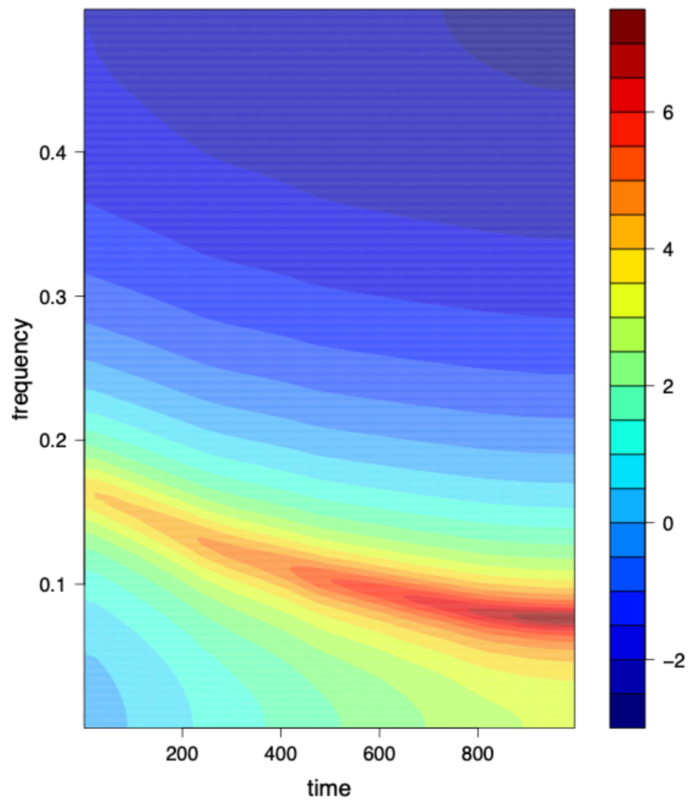
- Step 3. Repeat step 2(a) and step 2(b) for $m = 2, \ldots, P$.

- Step 4. Use the Durbin-Levinson recursions to obtain estimates of the TVAR coefficients.

TV-PARCOR representation of a TVAR

- A time series of 1000 data points was simulated from a TVAR(2) with a linearly increasing modulus from 0.9 to 0.97 and a linearly decreasing period from 5 to 14, or, equivalently, a linearly decreasing frequency from 0.2 to 0.07.





TV-PARCOR hierarchical model

Forward prediction error

$$f_{i,t}^{(P)} = y_{i,t} - \sum_{j=1}^{P} \phi_{i,j,t}^{(P)} y_{i,t-j}$$

Backward prediction error

$$b_{i,t}^{(P)} = y_{i,t} - \sum_{j=1}^{P} \alpha_{i,j,t}^{(P)} y_{i,t+j}$$

<u>Observational equations:</u> Lattice filter for $m = 1, \ldots, P$

$$f_{i,t}^{(m-1)} = \beta_{i,f,m,t}^{(m)} b_{i,t-m}^{(m-1)} + f_{i,t}^{(m)}, \quad f_{i,t}^{(m)} \sim N(0, \sigma_{f,m}^2)$$

$$b_{i,t}^{(m-1)} = \beta_{i,b,m,t}^{(m)} f_{i,t+m}^{(m-1)} + b_{i,t}^{(m)}, \quad b_{i,t}^{(m)} \sim N(0, \sigma_{b,m}^2)$$

2 DLMs with univariate state vectors for each m $\Rightarrow$

No matrix operations

Case study: Multi-trial EEG

Structural equations:

$$\beta^{(m)}_{i,f,m,t} = \beta^{(m)}_{f,m,t} + \nu_{i,f,m,t}, \quad \nu_{i,f,m,t} \sim N(0, v_{i,f,m,t})$$

$$\beta^{(m)}_{i,b,m,t} = \beta^{(m)}_{b,m,t} + \nu_{i,b,m,t}, \quad \nu_{i,b,m,t} \sim N(0, v_{i,b,m,t})$$

System/evolution equations:

$$\beta^{(m)}_{f,m,t} = \beta^{(m)}_{f,m,t-1} + w_{\beta,f,m,t} \sim N(0, W_{\beta,f,m,t})$$

$$\beta^{(m)}_{b,m,t} = \beta^{(m)}_{b,m,t-1} + w_{b,m,t} \sim N(0, W_{\beta,b,m,t})$$

Hierarchical forward model

$$\boldsymbol{\theta}^{(m)}_{1,f,m,t} = \mathbf{F}_2 \boldsymbol{\theta}^{(m)}_{2,f,m,t} + \boldsymbol{\nu}_{1,f,m,t}, \quad \boldsymbol{\nu}_{1,f,m,t} \sim N(\mathbf{0}, \sigma^2_{f,m} \mathbf{V}^*_{f,m,t}),$$

$$\boldsymbol{\theta}^{(m)}_{1,f,m,t} = \mathbf{F}_2 \boldsymbol{\theta}^{(m)}_{2,f,m,t} + \boldsymbol{\nu}_{1,f,m,t}, \quad \boldsymbol{\nu}_{1,f,m,t} \sim N(\mathbf{0}, \sigma^2_{f,m} \mathbf{V}^*_{f,m,t}),$$

$$\boldsymbol{\theta}^{(m)}_{2,f,m,t} = \boldsymbol{\theta}^{(m)}_{2,f,m,t-1} + \mathbf{w}_{2,f,m,t}, \quad \mathbf{w}_{2,f,m,t} \sim N(\mathbf{0}, \sigma^2_{f,m} \mathbf{W}^*_{f,m,t}) \, .$$

- $\mathbf{V}^*_{f,m,t}$ and $\mathbf{W}^*_{f,m,t}$ are **specified via discount factors**
- $\boldsymbol{\theta}_{2,f,m,0} \sim N(\mathbf{0}, \sigma^2_{f,m} \mathbf{C}^*_{f,m,0}),$
- $\sigma^2_{f,m} \sim IG(n_{f,m,0}/2, h_{f,m,0}/2) \, .$

We have a similar **hierarchical backward model**

Case study: Multi-trial EEG

- Model order selection is achieved by computing the deviance information criterion (DIC)

- Spectral density estimates for each series i can be obtained via

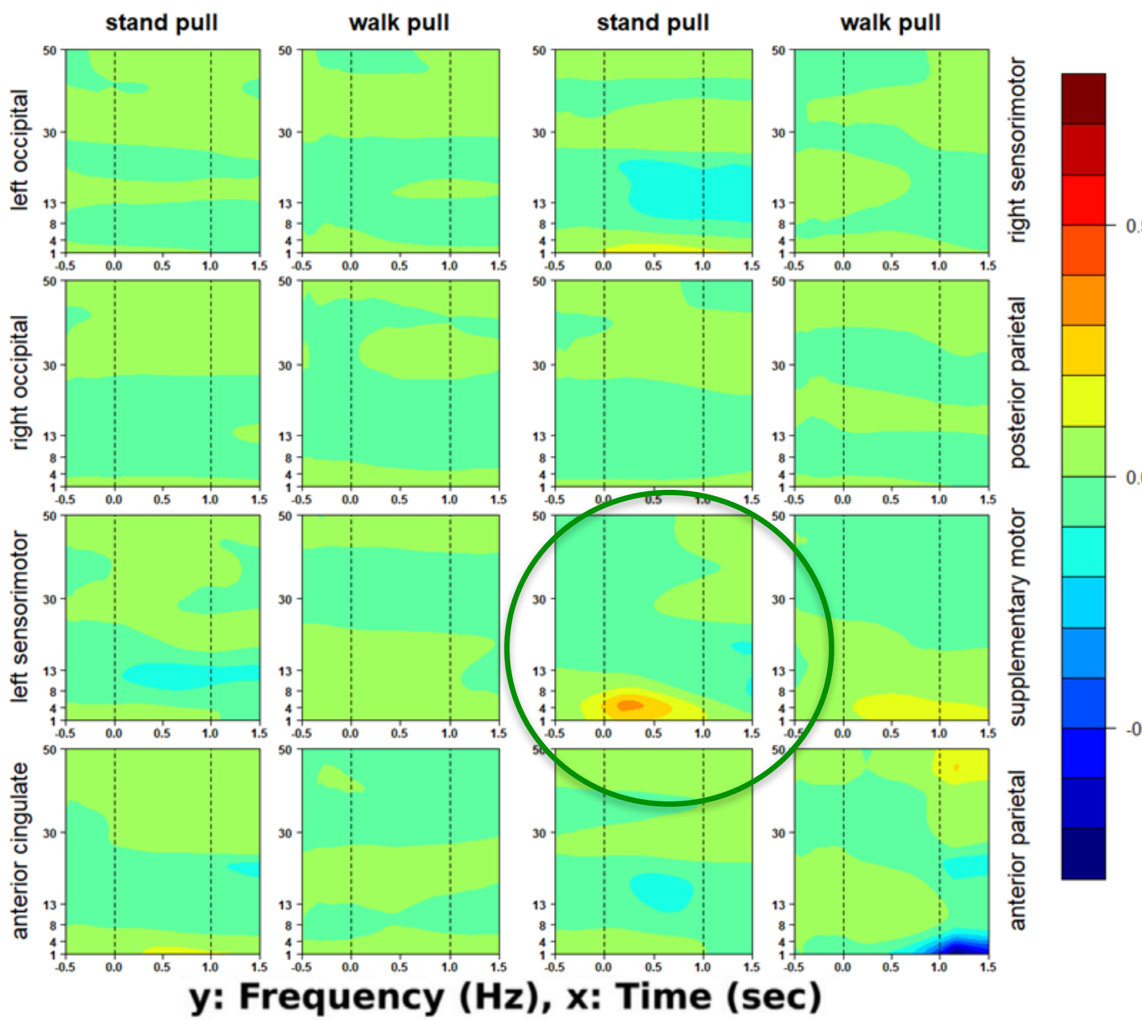
$$g_i(t,\omega) = \frac{\sigma^2_{f,P}}{|\,1 - \sum_{j=1}^{P} \phi^{(P)}_{i,j,t} \exp\{-2\pi i j\omega\}\,|^2}$$

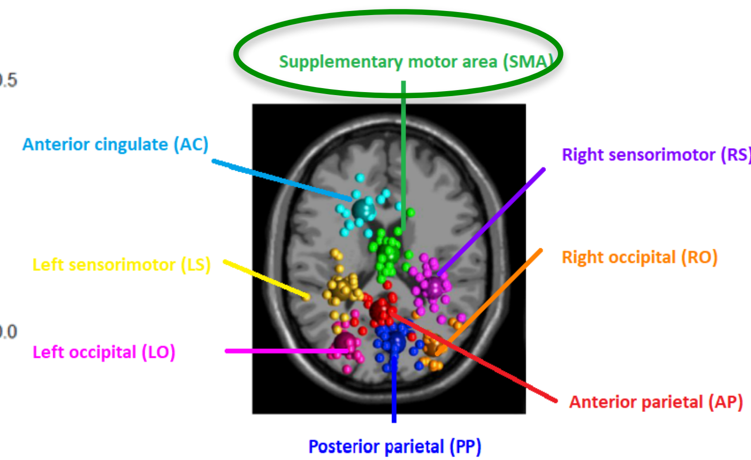
  using estimates of the TVAR coefficients obtained from applying the Durbin-Levinson recursion to the $\beta_{i,f,m,t}$.

- Estimates of the baseline TVAR coefficients $\bar{\phi}^{(P)}_{1,t},\ldots,\bar{\phi}^{(P)}_{P,t}$, and their corresponding spectral density, $\bar{g}(t,\omega)$, can be obtained by applying the Durbin-Levinson recursion to the $\beta_{m,f,t}$.

Case study: Multi-trial EEG

Subject 25: Stand/walk under pull perturbations

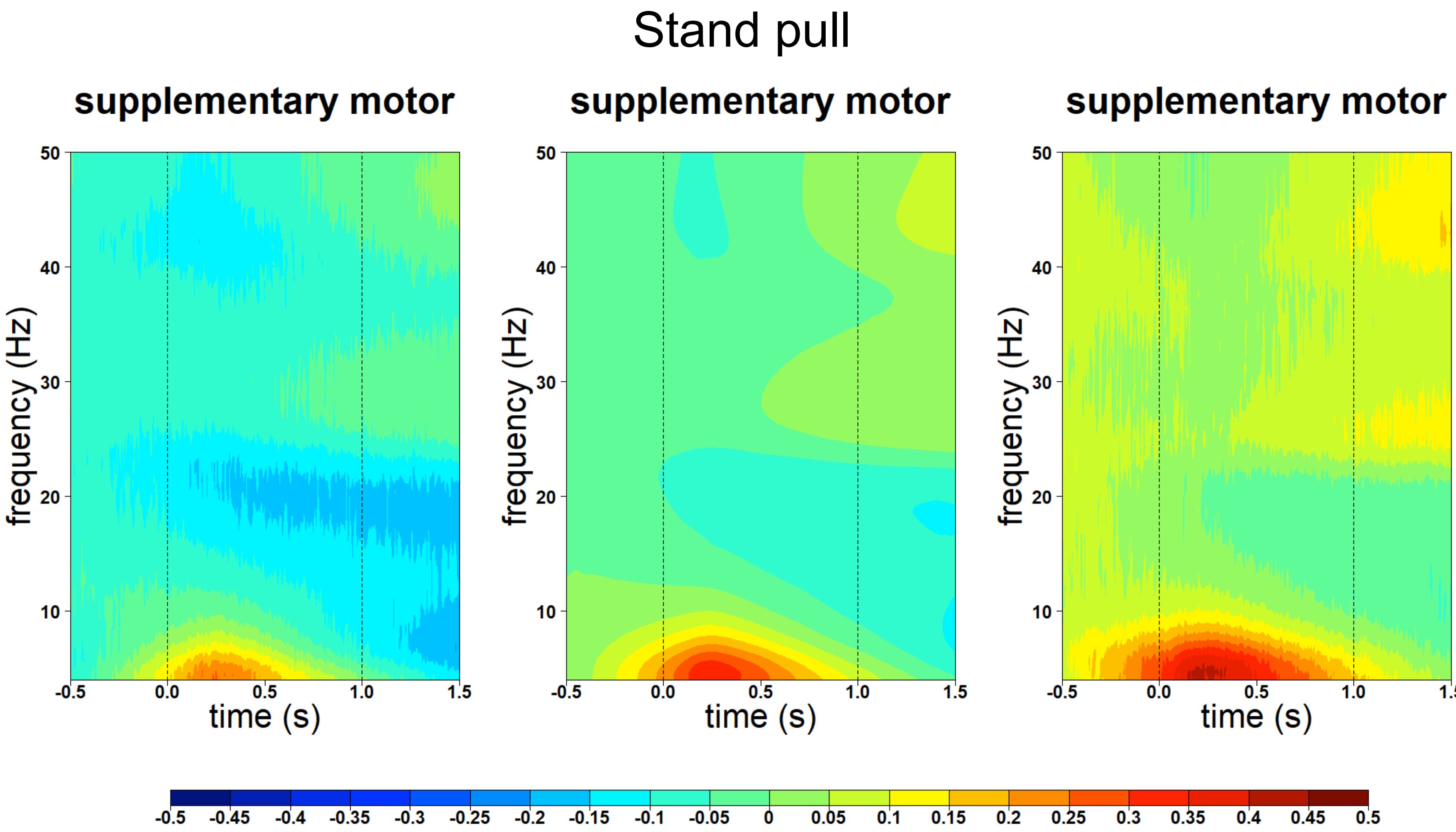




Delta band: 1-4 Hz
Theta band: 4-8 Hz

Case study: Multi-trial EEG

Uncertainty quantification



Multivariate DLMs

$$\mathbf{y}_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{V}_t),$$
$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t),$$

- $\mathbf{y}_t$ is a k-dimensional vector and $\boldsymbol{\theta}_t$ is a p-dimensional vector
- Close form filtering and smoothing equations are available when $\mathbf{V}_t$ and $\mathbf{W}_t$ are known (or $\mathbf{W}_t$ is specified via discount factors), assuming a conjugate prior structure of the form $(\boldsymbol{\theta}_0 \mid \mathscr{D}_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$.
- If $\mathbf{W}_t$ is assumed known or specified using discount factors, it is possible to obtain approximate filtering and smoothing for the case in which $\mathbf{V}_t = \mathbf{V}$.

MNDLMs: Approximate filtering and smoothing

- Let $\mathbf{S}_0$ be the prior expectation of $\mathbf{V}$ with n_0 degrees of freedom
- Assume $(\boldsymbol{\theta}_{t-1} \mid \mathscr{D}_{t-1}) \approx N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$. Then $(\boldsymbol{\theta}_t \mid \mathscr{D}_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t), \quad (\mathbf{y}_t \mid \mathscr{D}_{t-1}) \approx N(\mathbf{f}_t, \mathbf{Q}_t)$, with

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}, \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t,$$

$$\mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_t, \quad \mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{S}_{t-1}$$

- Triantafyllopoulos (2017): $\mathbf{S}_t$ will approximate $\mathbf{V}$ with

$$\mathbf{S}_t = \frac{1}{(n_0 + t)} \left(n_0 \mathbf{S}_0 + \sum_{i=1}^{t} \mathbf{S}_{i-1}^{1/2} \mathbf{Q}_i^{-1/2} \mathbf{e}_i \mathbf{e}_i' \mathbf{Q}_i^{-1/2} \mathbf{S}_{i-1}^{1/2} \right)$$

and $\mathbf{e}_t = \mathbf{y}_t - \mathbf{a}_t$.

- Then, $(\boldsymbol{\theta}_t \mid \mathscr{D}_t) \approx N(\mathbf{m}_t, \mathbf{C}_t)$ with $\mathbf{m}_t$ and $\mathbf{C}_t$ functions of the quantities above

VAR

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V}),$$

- $\mathbf{y}_t$ is a k-dimensional vector, $\mathbf{A}_i$ are $k \times k$ matrices of VAR coefficients and $\mathbf{V}$ is a $k \times k$ covariance matrix
- The process is stable and stationary if the characteristic polynomial

$$\mathbf{A}(u) = \det(\mathbf{I}_k - \mathbf{A}_1 u - \cdots - \mathbf{A}_p u^p),$$

has no roots within or on the complex unit circle.

TV-VAR

$$\mathbf{y}_t = \mathbf{A}_{1,t}\mathbf{y}_{t-1} + \ldots + \mathbf{A}_{p,t}\mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V}),$$

- The process is not stationary but we can talk about instantaneous stability and stationary if the characteristic polynomial at time t, $\mathbf{A}_t(u)$ has all the roots outside the unit complex circle.

State-space/DLM representations

Inference: We can vectorize the matrices of VAR/TV-VAR coefficients and represent the VAR/TV-VAR as a multivariate DLM and proceed with Bayesian inference.

Example: 3-dimensional TV-VAR(2)

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = \begin{pmatrix} a_{1,11,t} & a_{1,12,t} & a_{1,13,t} \\ a_{1,21,t} & a_{1,22,t} & a_{1,23,t} \\ a_{1,31,t} & a_{1,32,t} & a_{1,33,t} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{pmatrix} + \begin{pmatrix} a_{2,11,t} & a_{2,12,t} & a_{2,13,t} \\ a_{2,21,t} & a_{2,22,t} & a_{2,23,t} \\ a_{2,31,t} & a_{2,32,t} & a_{2,33,t} \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \\ y_{3,t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}$$

$$\mathbf{y}_t = \mathbf{A}_{1,t} \mathbf{y}_{t-1} + \mathbf{A}_{2,t} \mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t$$

DLM with and 18-dimensional state vector:

$$\boldsymbol{\theta}_t = (a_{1,11,t}, \ldots, \ldots, a_{2,33,t})'$$

VAR, TV-VAR models

- $\mathbf{F}_t'$ is given by:

$$\mathbf{F}_t' = \begin{pmatrix} y_{1,t-1} & y_{2,t-1} & y_{3,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & y_{1,t-2} & y_{2,t-2} & y_{3,t-2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{1,t-1} & y_{2,t-1} & y_{3,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & y_{1,t-2} & y_{2,t-2} & y_{3,t-2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{1,t-1} & y_{2,t-1} & y_{3,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & y_{1,t-2} & y_{2,t-2} & y_{3,t-2} \end{pmatrix}$$

- $\mathbf{G} = \mathbf{I}_{18}$

Then,

$$\mathbf{y}_t = \mathbf{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V})$$
$$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t)$$

Note that this representation requires operations with matrices of dimension $(k^2p) \times (k^2p)$

State-space/DLM representations

Decompositions: An alternative representation allows for decompositions of the components of $\mathbf{y}_t$ in terms of latent processes with simpler structures.

In this case we define $\boldsymbol{\theta}_t' = (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p+1}')$, leading to the following representation:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{F}'\boldsymbol{\theta}_t, \\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \end{aligned}$$

with $\mathbf{F}' = [\mathbf{I}_k \ \ \mathbf{0}_k \ \ \cdots \ \ \mathbf{0}_k], \mathbf{w}_t' = (\boldsymbol{\epsilon}_t', \mathbf{0}', \ldots, \mathbf{0}')$ and $\mathbf{G}_t$ given by…

$$\mathbf{G}_t = \begin{pmatrix} \mathbf{A}_{1,t} & \mathbf{A}_{2,t} & \cdots & \mathbf{A}_{p-1,t} & \mathbf{A}_{p,t} \\ \mathbf{I}_k & \mathbf{0}_k & \cdots & \mathbf{0}_k & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{I}_k & \cdots & \mathbf{0}_k & \mathbf{0}_k \\ \vdots & & \ddots & & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{0}_k & \cdots & \mathbf{I}_k & \mathbf{0}_k \end{pmatrix}$$

Note that the eigenvalues of $\mathbf{G}_t$ satisfy the equation

$$\mathbf{A}_t(u) = \det(\mathbf{I}_k - \mathbf{A}_{1,t}u - \ldots - \mathbf{A}_{p,t}u^p),$$

Therefore, the eigenvalues of $\mathbf{G}_t$ correspond to the reciprocal roots of the TV-VAR characteristic polynomial.

In the VAR case the matrix $\mathbf{G}_t = \mathbf{G}$ for all t.

Multivariate TV-VPARCOR

We can consider a TV-VPARCOR representation of a TV-VAR process of order P. Assume we have a K-dimensional time series $\mathbf{y}_t$. The forward MDLMs are given by:

$$\mathbf{f}_t^{(m-1)} = \mathbf{B}_{t-m}^{(m-1)} \boldsymbol{\beta}_{f,m,t}^{(m)} + \mathbf{f}_t^{(m)}, \quad \mathbf{f}_t^{(m)} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{f,m,t})$$
$$\boldsymbol{\beta}_{f,m,t}^{(m)} = \boldsymbol{\beta}_{f,m,t-1}^{(m)} + \boldsymbol{\omega}_{f,m,t}, \quad \boldsymbol{\omega}_{f,m,t} \sim N(\mathbf{0}, \mathbf{W}_{f,m,t})$$

- $\mathbf{f}_t^{(m-1)}$ is a K-dimensional vector
- $\mathbf{B}_{t-m}^{(m-1)} = (\mathbf{b}_{t-m}^{(m-1)}) \otimes \mathbf{I}_{K \times K}$ is a $K \times K^2$ matrix
- $\boldsymbol{\beta}_{f,t,m}^{(m)}$ is a K^2-dimensional vector
- $m = 1 : P$

We have a similar set of backward DLMs

Multivariate TV-VPARCOR

For a fixed model order P and K time series:

- **TV-VPARCOR:** requires fitting multivariate $2P$ MDLMs with state-space vectors of dimension K^2
- **TV-VAR:** requires fitting a multivariate DLM with state-space vectors of dimension $K^2 P$
- To find the optimal order & tuning parameters we need to consider all possible models up to $P_{\max}$

Example: $K = 10, \;\; P_{\max} = 10$

- Maximum dimension of state-space vector:

TV-VPARCOR: 100 **TV-VAR:** 1000

- Operations with matrices of dimension:

TV-VPARCOR: 100 x 100 **TV-VAR:** 1000 x 1000

Multivariate TV-VPARCOR

We optimally choose the discount factor values and model order. Then, using estimates/samples of the model parameters we can compute estimates/samples of:

- The spectral density matrix: $\mathbf{g}(t, \omega)$; the spectral density of the ith component: $g_{ii}(\omega, t)$

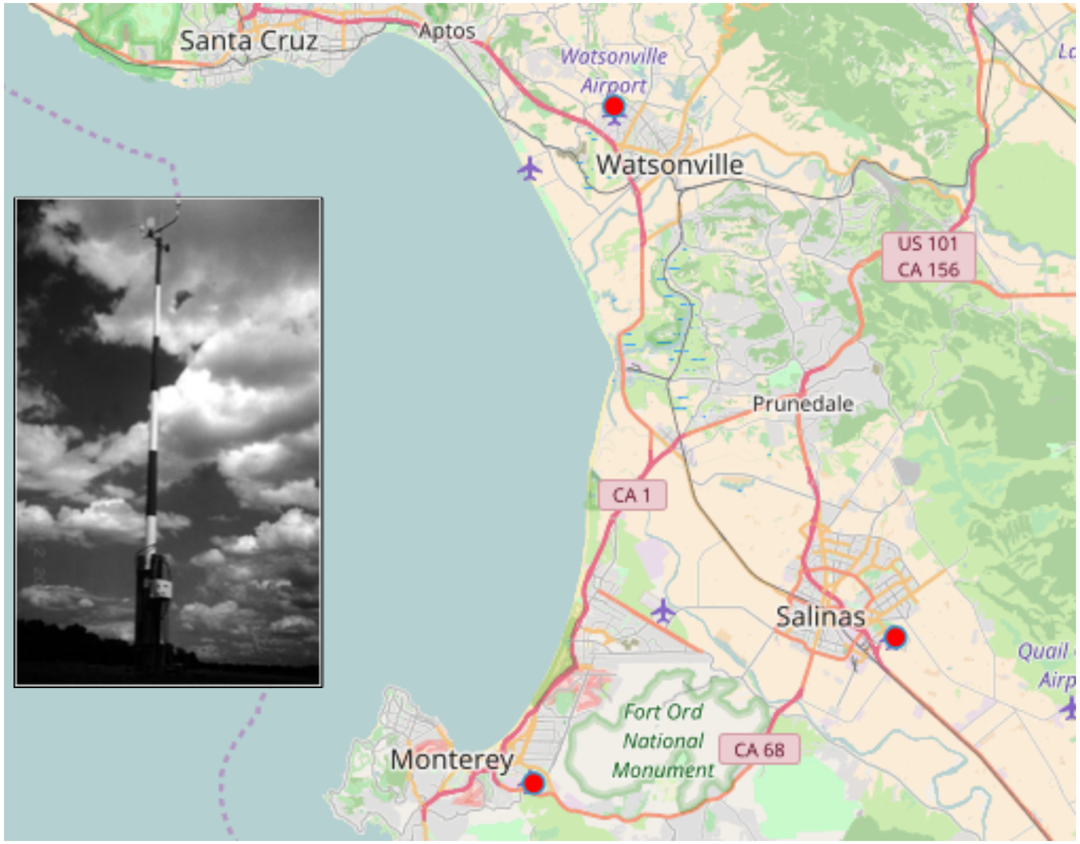
- The squared coherence:

- $$\rho_{ij}^2(t, \omega) = \frac{|\, g_{ij}(t, \omega)\, |^2}{g_{ii}(t, \omega) \times g_{jj}(t, \omega)}$$

- Frequency/period collapsed spectral measurements

Multi-location wind component data

- IEM (Iowa Environmental Mesonet)-ASOS (automated surface observing system) data; <u>CA:</u> Monterey, Salinas and Watsonville

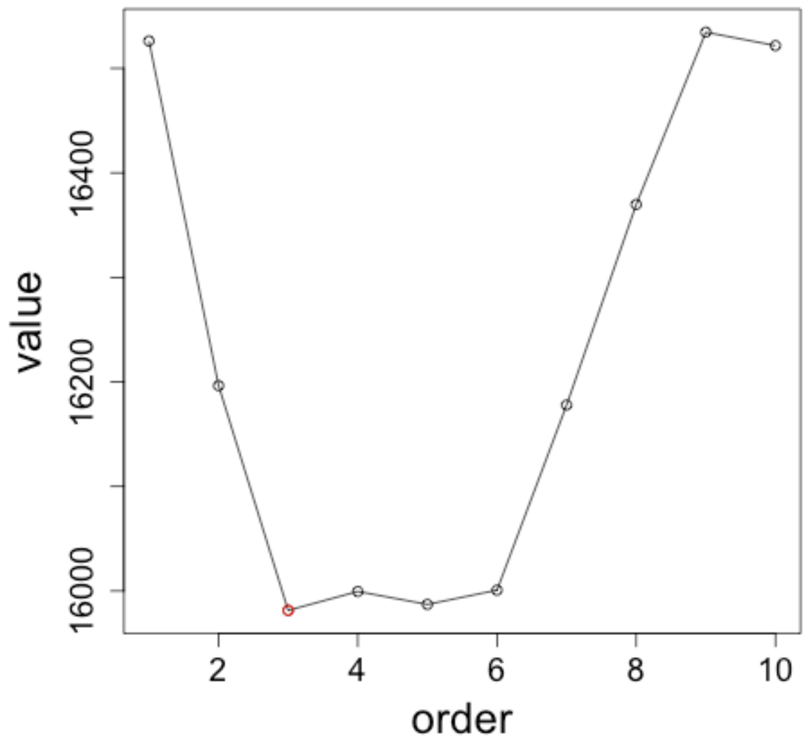


- June 1st to August 31 2010

- East-West (X) and North-South (Y) wind components derived from median wind speed and direction measurements (rotating cup anemometer) taken every 4 hours. Wind direction is measured in angles relative to the true north.

Multi-location wind component data

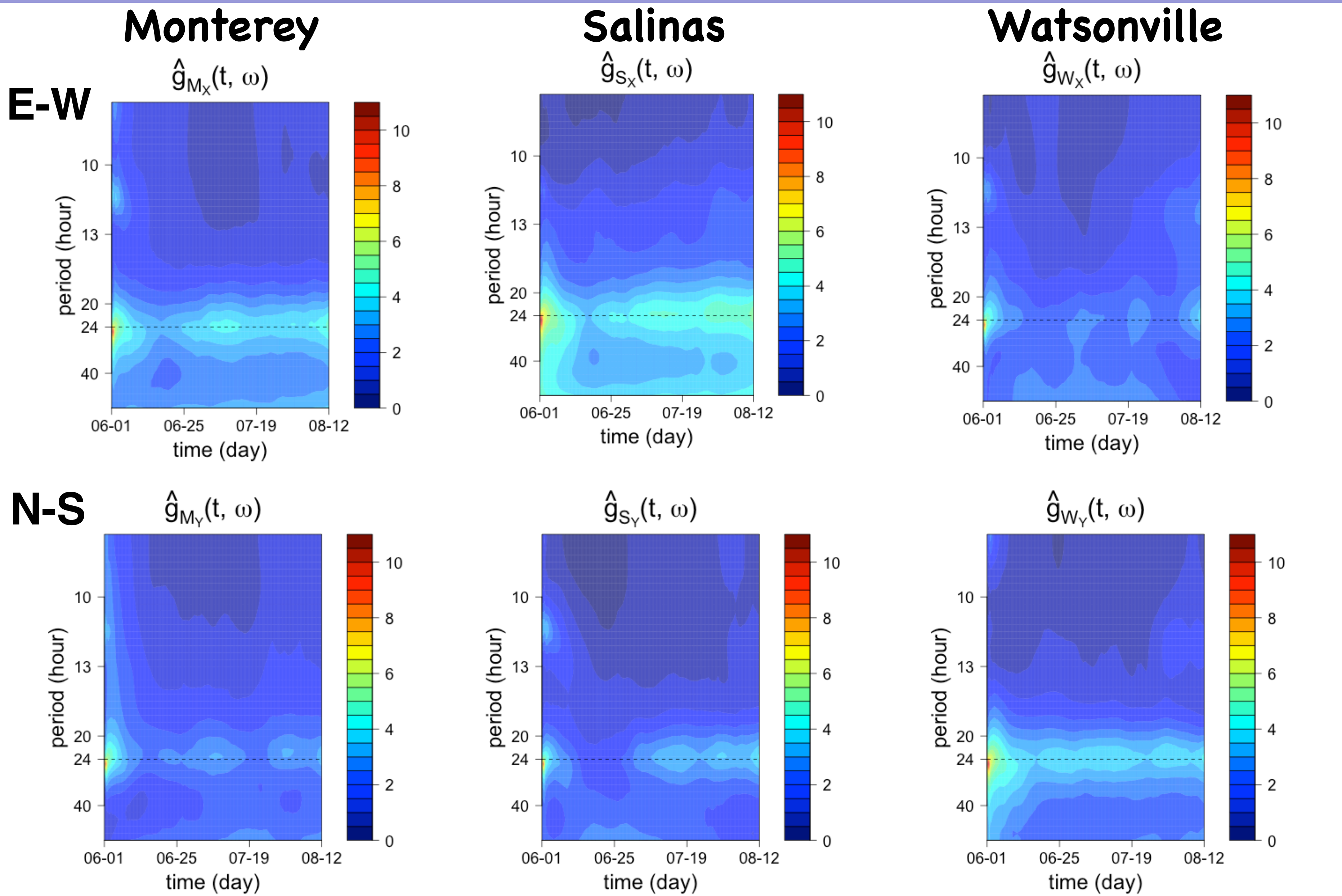
- $P_{max} = 10$; Optimal order: $P = 3$



- Optimal discount factors: 0.97, 0.98 and 0.99.

- Similar results were obtained with models of orders 4-6

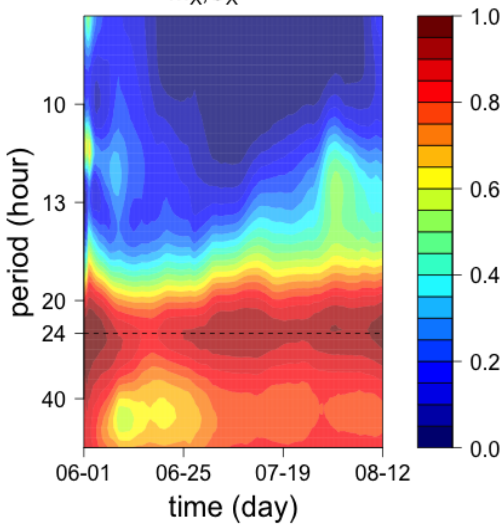
Multi-location wind component data



Multi-location wind component data

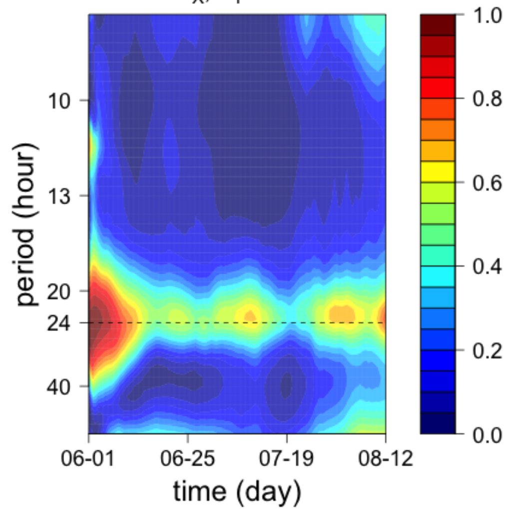
SNS E-W

$\hat{\rho}_{M_X,S_X}(t, \omega)$



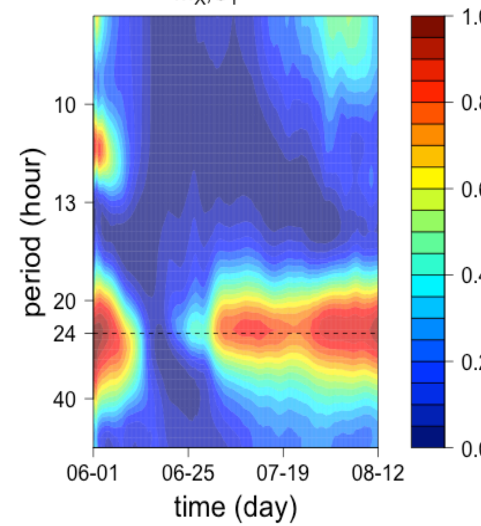
MRY N-S

$\hat{\rho}_{M_X,M_Y}(t, \omega)$



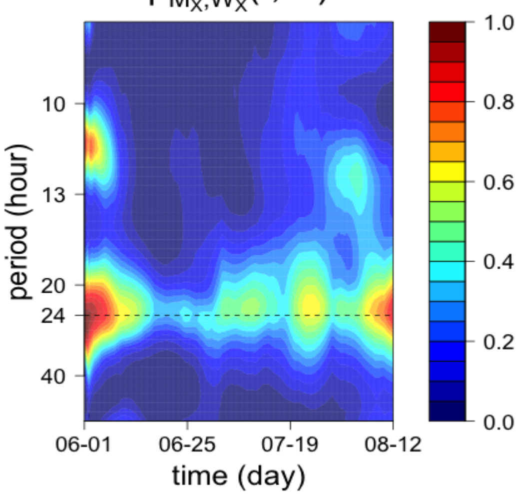
SNS N-S

$\hat{\rho}_{M_X,S_Y}(t, \omega)$

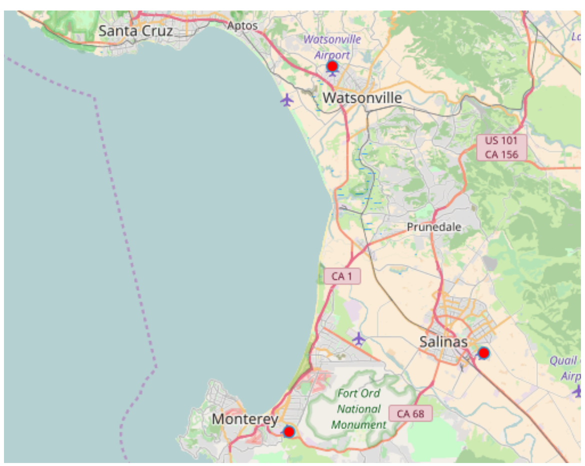


WVI E-W

$\hat{\rho}_{M_X,W_X}(t, \omega)$

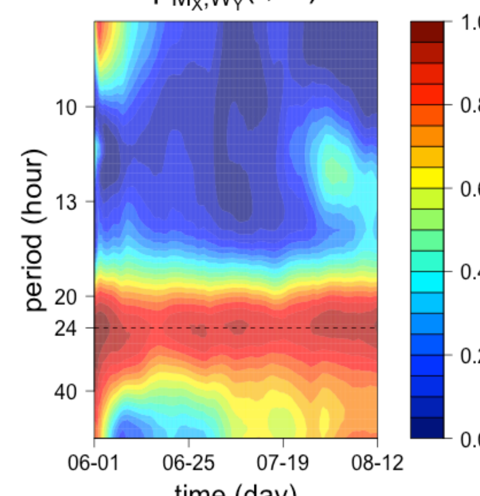


MRY E-W

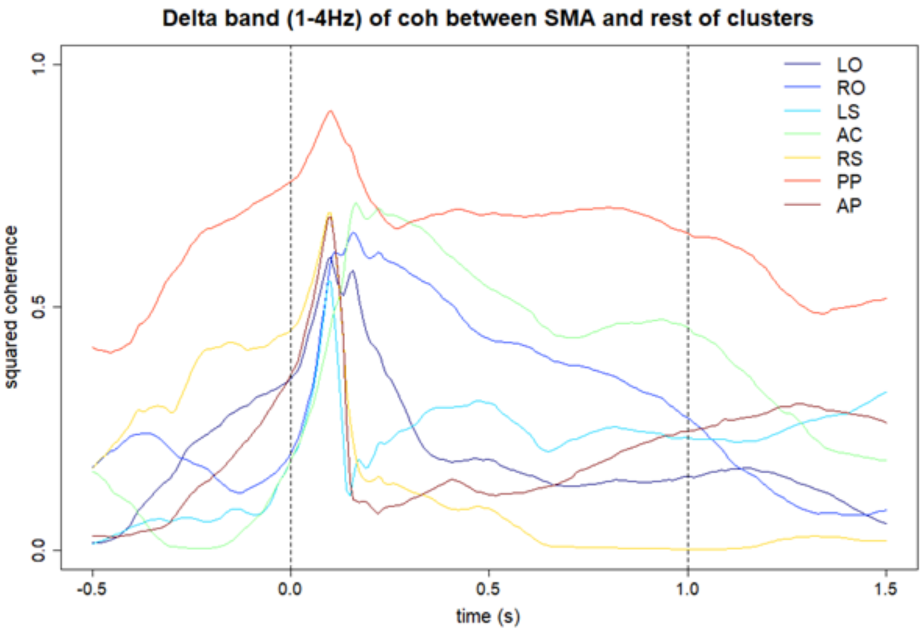


WVI N-S

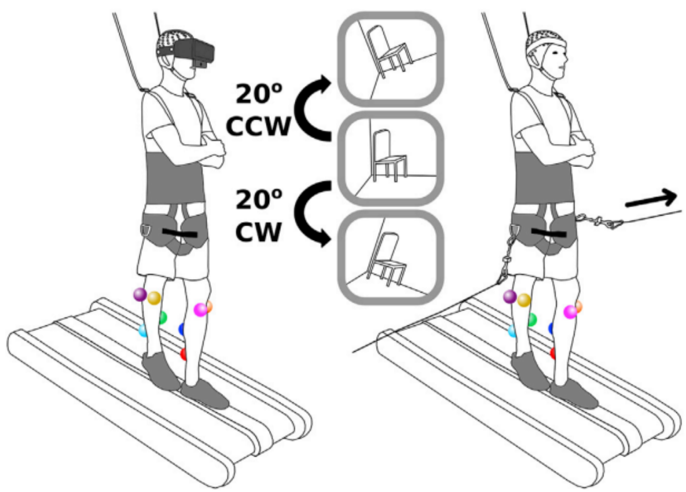
$\hat{\rho}_{M_X,W_Y}(t, \omega)$

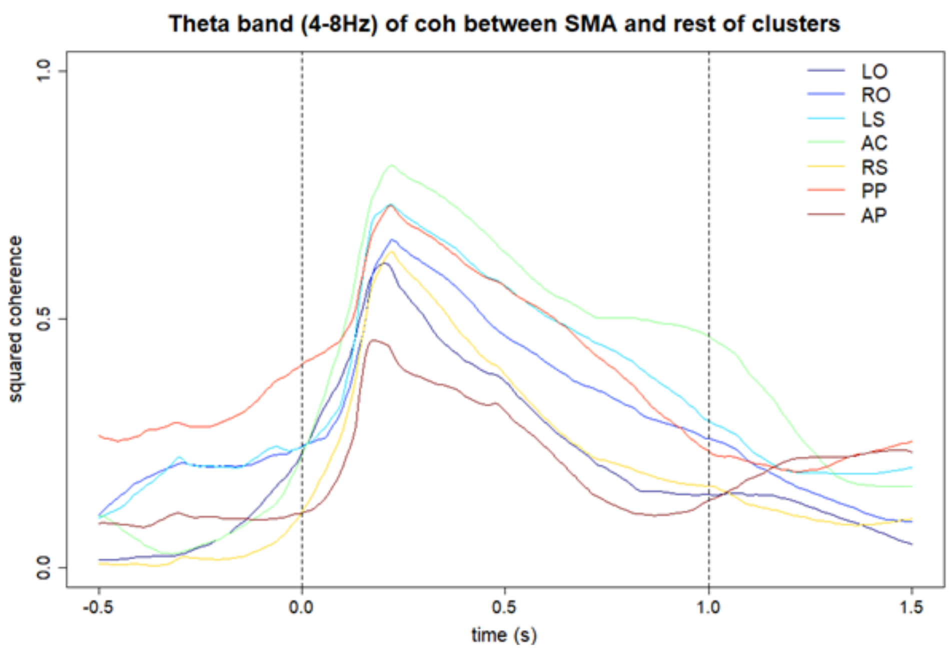


Multi-location EEG data: Peterson & Ferris (2019)



LO: Left occipital
RO: Right sensorimotor
LS: Left sensorimotor
AC: Anterior cingulate
RS: Right sensorimotor
PP: Posterior parietal
AP: Anterior parietal





Bitto and Frühwirth-Schnatter (2019)

$$
\begin{aligned}
y_t &= \mathbf{F}_t'\boldsymbol{\theta}_t + \nu_t \quad \nu_t \sim N(0, v_t), \\
\boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}),
\end{aligned}
$$

with $\mathbf{F}_t' = (x_{t,1}, \ldots, x_{t,d})$ and $\boldsymbol{\theta}_t = (\theta_{t,1}, \ldots, \theta_{t,d})'$. Also, $\boldsymbol{\theta}_0 \sim N(\boldsymbol{\theta}, \mathbf{P}_0\mathbf{W})$, with $\mathbf{P}_0$ and $\mathbf{W}$ diagonal matrices, then:

$$
\theta_{t,j} = \theta_{t-1,j} + \omega_{t,j} \quad \omega_{t,j} \sim N(0, w_j),
$$

$$
\theta_{0,j} \sim N(\theta_j, w_j P_{0,jj}) \, .
$$

A shrinkage prior on w_j allows to pull $\theta_{1,j}, \ldots, \theta_{T,j}$ towards θ_j if the effect of $x_{t,j}$ is not changing over time.

Shrinkage priors DLMs

Non-centered parameterization

$$y_t = \mathbf{F}_t'\boldsymbol{\theta} + \mathbf{F}_t'\text{Diag}(\sqrt{w_1}, \ldots, \sqrt{w_d})\tilde{\boldsymbol{\theta}}_t + \nu_t \qquad \nu_t \sim N(0, v_t),$$

$$\tilde{\theta}_{t,j} = \tilde{\theta}_{t-1,j} + \tilde{\omega}_{t,j}, \qquad \tilde{\omega}_{t,j} \sim N(0,1)$$

- <u>Shrinkage towards static parameters</u>:

$$\sqrt{w}_j \sim N(0, \xi_j^2), \quad \xi_j^2 \sim \text{Gamma}(a^\xi, a^\xi \kappa^2/2), \quad \kappa^2 \sim \text{Gamma}(d_1, d_2)$$

$a^\xi = 1$ corresponds to Bayesian Lasso prior. Alternatively:

$a^\xi \sim \text{Exp}(b^\xi)$

- κ^2 controls the global level of shrinkage (larger κ^2 stronger shrinkage)
- a^ξ controls the local adaptation to the global shrinkage

- <u>Shrinkage towards zero:</u>

$$\theta_j \sim N(0, \tau_j^2), \quad \tau_j^2 \sim \text{Gamma}(a^\tau, a^\tau \lambda^2/2),$$

$$\lambda^2 \sim \text{Gamma}(e_1, e_2), \quad a^\tau \sim \text{Exp}(b^\tau)$$

Shrinkage priors MDLMs

$$\mathbf{y}_t \quad = \quad \boldsymbol{\Theta}_t \mathbf{F}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t),$$

Here $\mathbf{y}_t$ is an K-dimensional time series, $\boldsymbol{\Theta}_t$ is a $K \times d$ matrix of time-varying parameters (regression coefficients).

$$\theta_{t,i,j} = \theta_{t-1,i,j} + \omega_{t,i,j}, \quad \omega_{t,i,j} \sim N(0, w_{i,j}),$$

Non-centered parameterization:

$$\theta_{t,i,j} = \theta_{i,j} + \sqrt{w}_{i,j} \tilde{\theta}_{t,i,j}$$

$$\tilde{\theta}_{t,i,j} = \tilde{\theta}_{t-1,i,j} + \tilde{\omega}_{t,i,j}, \quad \tilde{\omega}_{t,i,j} \sim N(0,1)$$

Shrinkage priors MDLMs

- Shrinkage towards static parameters:

$$\sqrt{w}_{i,j} = \mathrm{N}\left(0, \xi_{i,j}^2\right), \quad \xi_{i,j}^2 \sim \mathrm{Gamma}\left(a_i^{\xi}, a_i^{\xi}\kappa_i^2/2\right),$$

$$\kappa_i^2 \sim \mathrm{Gamma}(d_1, d_2), \quad a_i^{\xi} \sim \mathrm{Exp}(b^{\xi})$$

- Shrinkage towards zero:

$$\theta_{i,j} \sim N(0, \tau_{i,j}^2), \quad \tau_{i,j}^2 \sim \mathrm{Gamma}(a_i^{\tau}, a_i^{\tau}\lambda_i^2/2),$$

$$\lambda_i^2 \sim \mathrm{Gamma}(e_1, e_2), \quad a_i^{\tau} \sim \mathrm{Exp}(b^{\tau})$$

Shrinkage priors in MDLM

TV-VPARCOR Forward model:

$$\mathbf{f}_t^{(m-1)} = \mathbf{B}_{t-m}^{(m-1)} \boldsymbol{\beta}_{f,m,t}^{(m)} + \mathbf{f}_t^{(m)}, \quad \mathbf{f}_t^{(m)} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{f,m,t})$$

$$\boldsymbol{\beta}_{f,m,t}^{(m)} = \boldsymbol{\beta}_{f,m,t-1}^{(m)} + \boldsymbol{\omega}_{f,m,t}, \quad \boldsymbol{\omega}_{f,m,t} \sim N(\mathbf{0}, \mathbf{W}_{f,m,t})$$

- Rewrite the model using non-centered parameterization and use the shrinkage priors on $\boldsymbol{\beta}_{f,m,t}^{(m)}$. This allows us to reduce time-varying parameters to static ones.
- Do the same for the backward model.

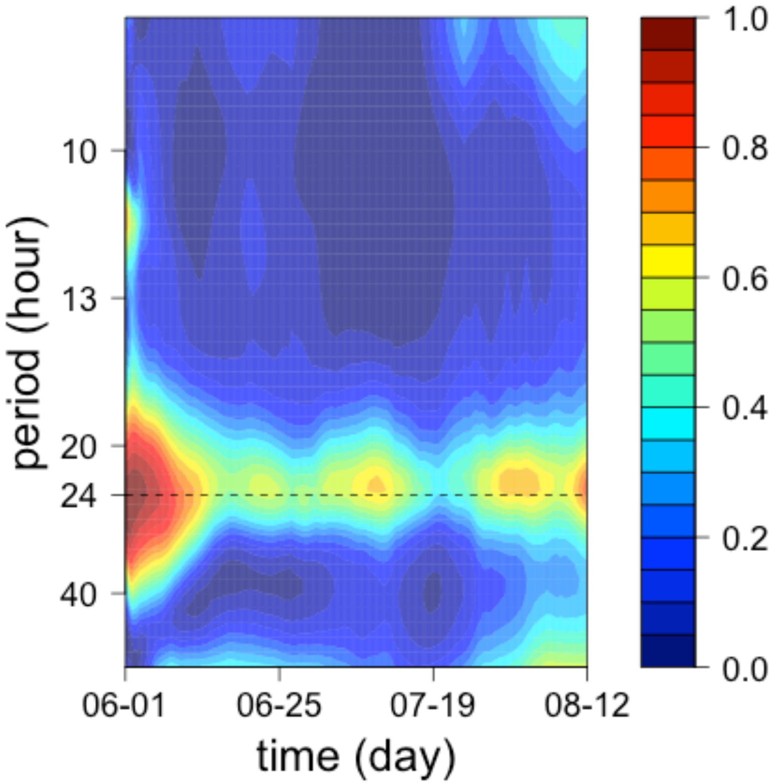
Approximate inference using an importance sampling variational Bayes algorithm:

- Each iteration goes over all the m stages
- It uses the mean field approximation
- It implements Bayesian lasso and double-gamma priors
- It uses the approximations of Triantafyllopoulos (2007) to estimate the observational variances in the forward and backward models

Wind Component Data

Coherence between MRY NS and EW components

No shrinkage



Shrinkage