

Semiparametric Event History Models for Analyzing Human Sleep Data

Thomas Kneib & Andrea Hennerfeind

Department of Statistics
Ludwig-Maximilians-University Munich

Joint work with

Alexander Yassouridis (Max-Planck-Institute for Psychiatry Munich)

Ludwig Fahrmeir (Ludwig-Maximilians-University Munich)



1.8.2007

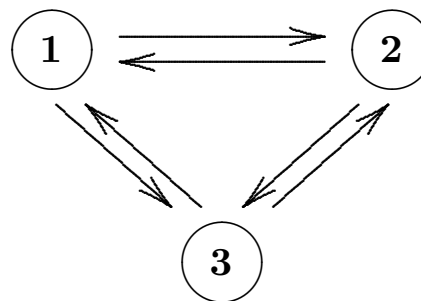


Multi-State Models

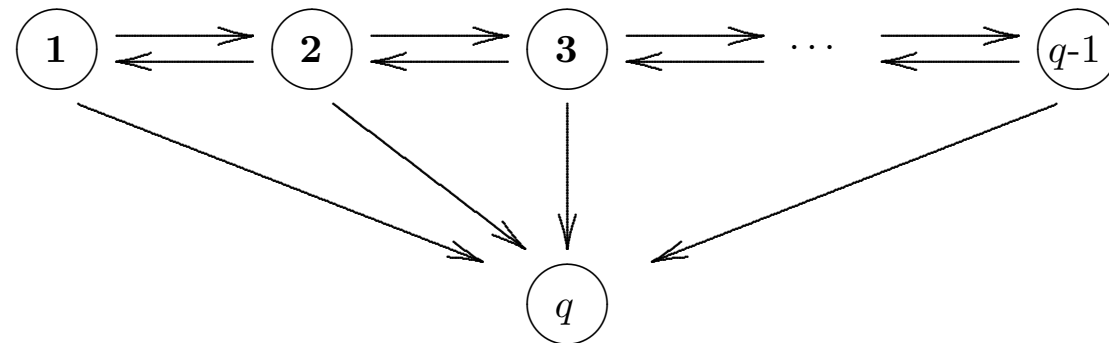
- Multi-state models form a general class for the description of the **evolution of discrete phenomena in continuous time** (i.e. event history analysis).
- We observe paths of a process

$$X = \{X(t), t \geq 0\} \quad \text{with} \quad X(t) \in \{1, \dots, q\}.$$

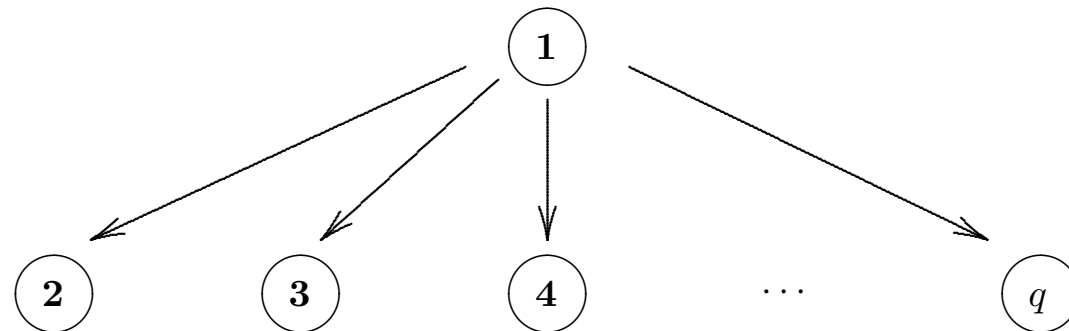
- Yields a similar data structure as for Markov processes.
- Examples:
 - Recurrent events:



– Disease progression:



– Competing risks:



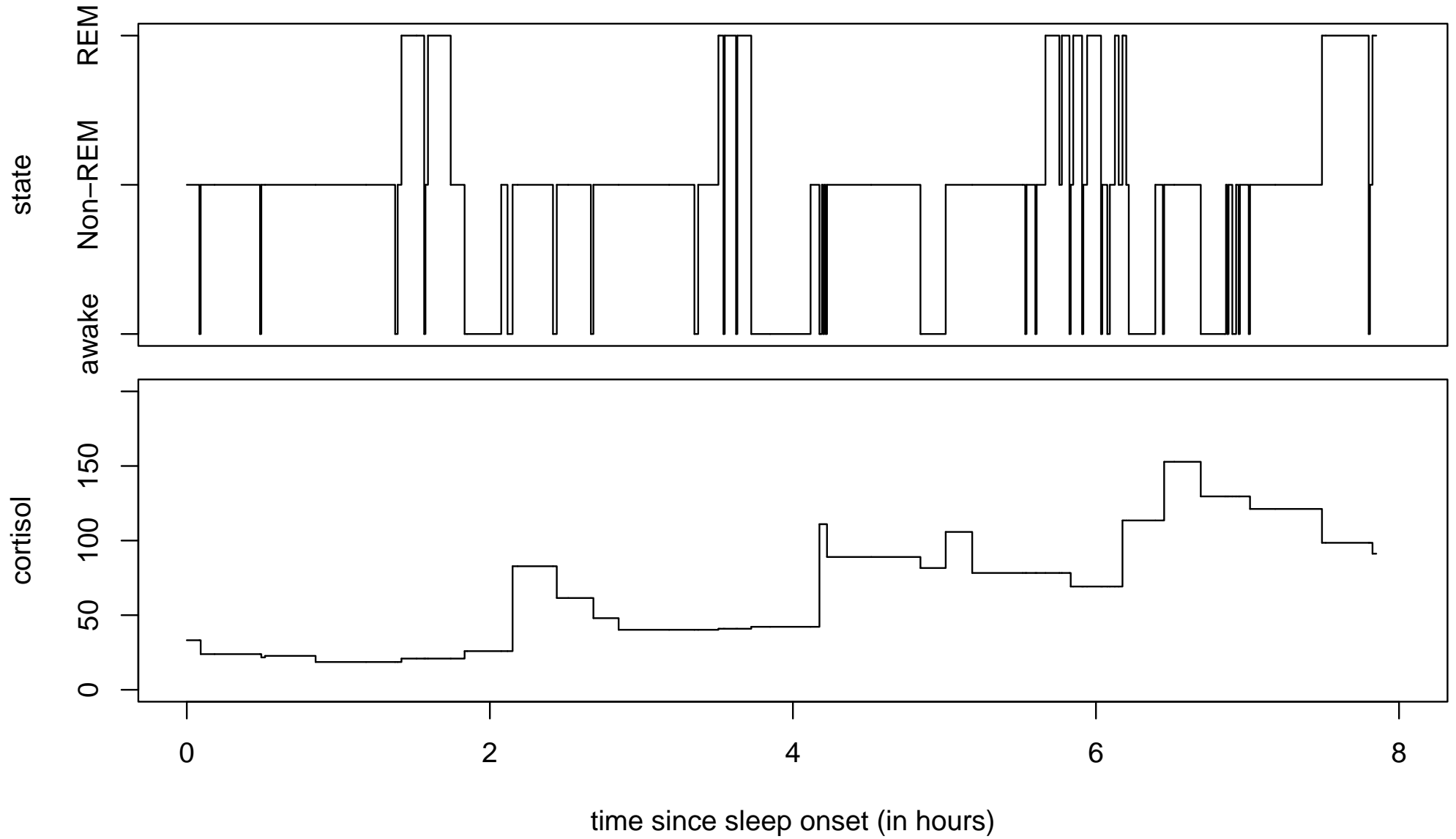
- (Homogenous) Markov processes can be compactly described in terms of the **transition intensities**

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P(X(t + \Delta t) = j | X(t) = i)}{\Delta t}$$

Human Sleep Data

- Human sleep can be considered an example of a recurrent event type multi-state model.
- State Space:

Awake	Phases of wakefulness
REM	Rapid eye movement phase (dream phase)
Non-REM	Non-REM phases (may be further differentiated)
- **Aims of sleep research:**
 - Describe the dynamics underlying the human sleep process.
 - Analyse associations between the sleep process and nocturnal hormonal secretion.
 - (Compare the sleep process of healthy and diseased persons.)



- **Data generation:**

- Sleep recording based on electroencephalographic (EEG) measures every 30 seconds (afterwards classified into the three sleep stages).
- Measurement of hormonal secretion based on blood samples taken every 10 minutes.
- A training night familiarizes the participants of the study with the experimental environment.

⇒ Sleep processes of 70 participants.

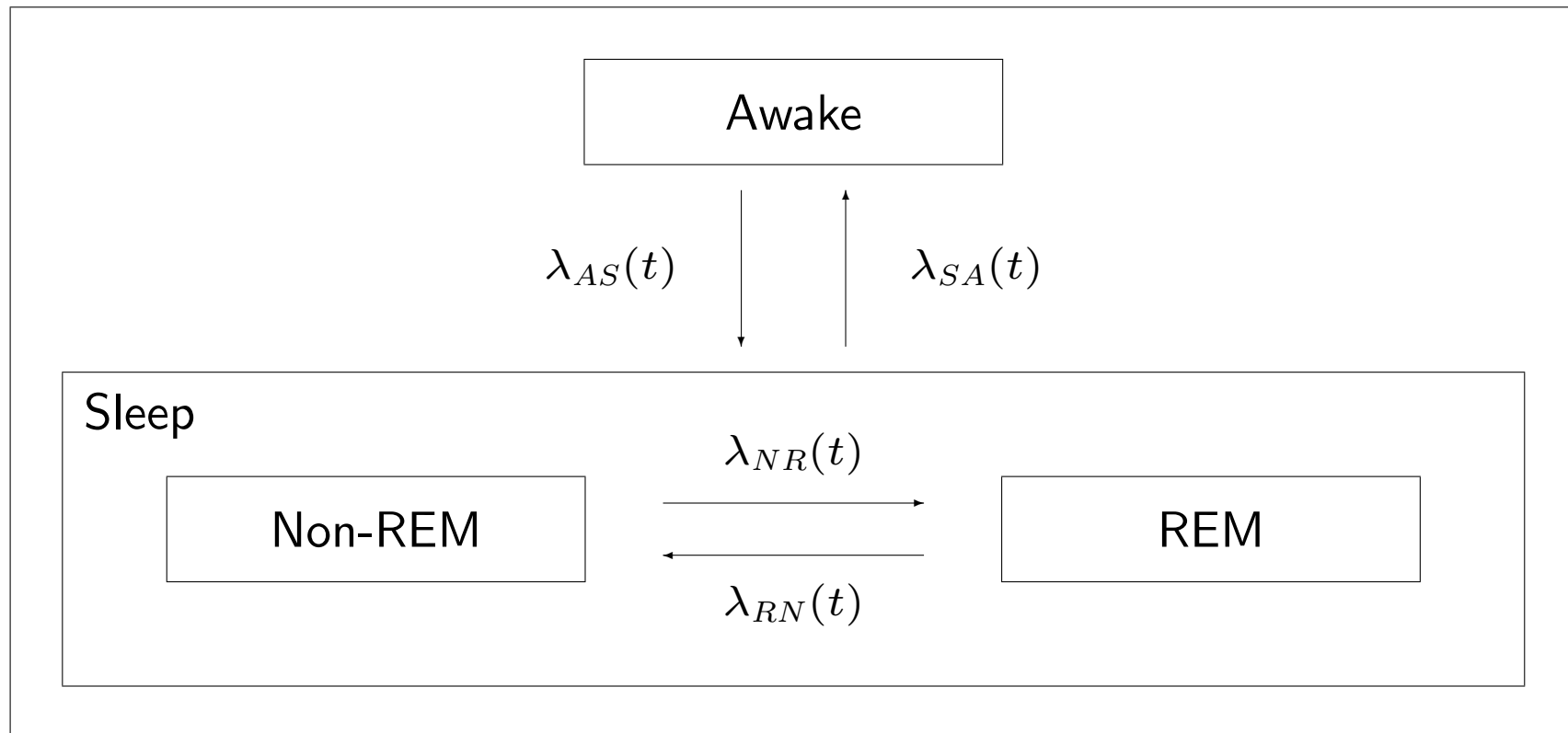
- Simple parametric approaches are not appropriate in this application due to

- **Changing dynamics** of human sleep over night.
- The **time-varying influence** of the hormonal concentration on the transition intensities.
- **Unobserved heterogeneity**.

⇒ **Model transition intensities nonparametrically.**

Specification of Transition Intensities

- To reduce complexity, we consider a simplified transition space:



- Model specification:

$$\begin{aligned}\lambda_{AS,i}(t) &= \exp \left[\gamma_0^{(AS)}(t) + b_i^{(AS)} \right] \\ \lambda_{SA,i}(t) &= \exp \left[\gamma_0^{(SA)}(t) + b_i^{(SA)} \right] \\ \lambda_{NR,i}(t) &= \exp \left[\gamma_0^{(NR)}(t) + c_i(t)\gamma_1^{(NR)}(t) + b_i^{(NR)} \right] \\ \lambda_{RN,i}(t) &= \exp \left[\gamma_0^{(RN)}(t) + c_i(t)\gamma_1^{(RN)}(t) + b_i^{(RN)} \right]\end{aligned}$$

where

$$c_i(t) = \begin{cases} 1 & \text{cortisol} > 60 \text{ n mol/l at time } t \\ 0 & \text{cortisol} \leq 60 \text{ n mol/l at time } t, \end{cases}$$

$$b_i^{(j)} \sim N(0, \tau_j^2) = \text{transition- and individual-specific frailty terms.}$$

- Penalized splines for baseline and time-varying effects:
 - Approximate $\gamma(t)$ by a weighted sum of **B-spline basis** functions

$$\gamma(t) = \sum_j \xi_j B_j(t).$$

- Employ a large number of basis functions to enable flexibility.
- **Penalize k -th order differences** between parameters of adjacent basis functions to ensure smoothness:

$$\text{Pen}(\xi|\tau^2) = \frac{1}{2\tau^2} \sum_j (\Delta^k \xi_j)^2.$$

- Bayesian interpretation: Assume a k -th order **random walk prior** for ξ_j , e.g.

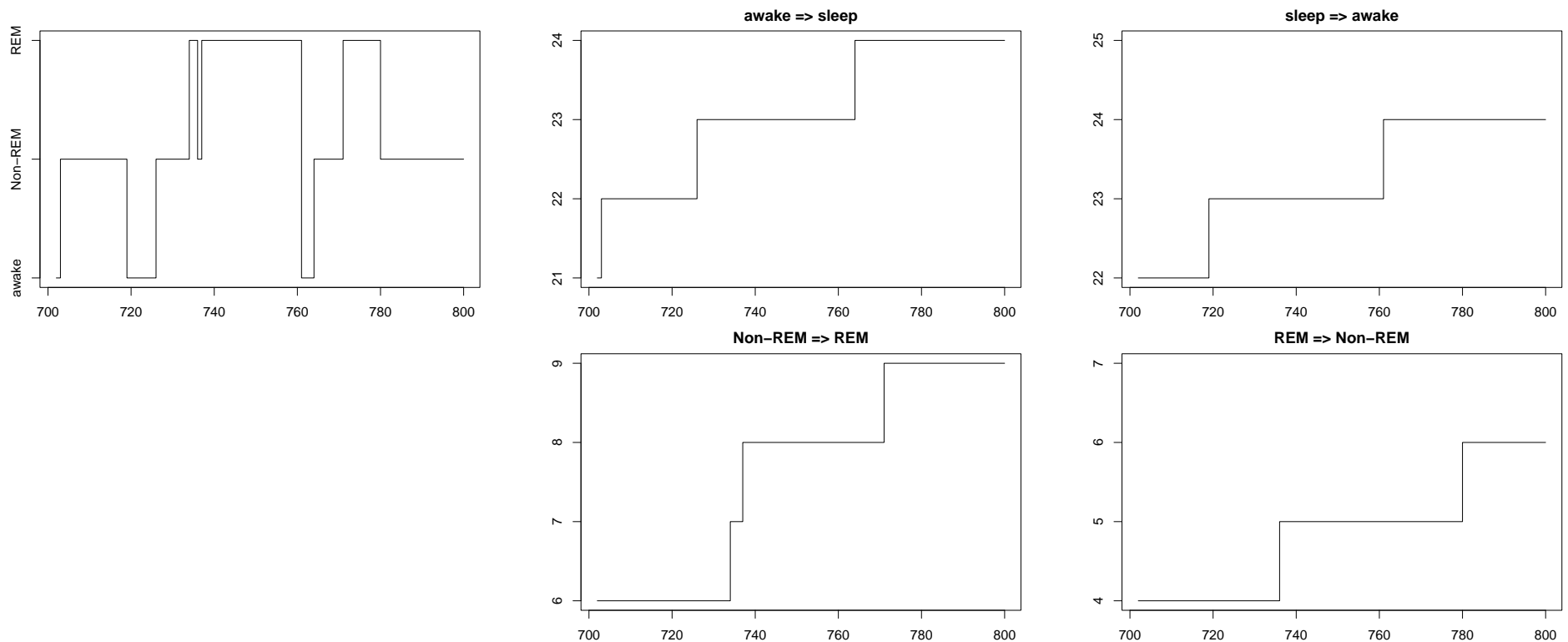
$$\xi_j = 2\xi_{j-1} - \xi_{j-2} + u_j, \quad u_j \sim N(0, \tau^2) \quad (\text{RW2}).$$

- This yields the **prior distribution**:

$$p(\xi|\tau^2) \propto \exp\left(-\frac{1}{2\tau^2} \xi' K \xi\right).$$

Counting Process Representation

- A multi-state model with k different types of transitions can be equivalently expressed in terms of k counting processes $N_h(t)$, $h = 1, \dots, k$ counting these transitions.



- From the counting process representation we can derive the likelihood contributions.
- The counting process representation also provides a possibility for model validation based on **martingale residuals**.
- Every counting process is a submartingale and can therefore be (Doob-Meyer-) decomposed as

$$\begin{aligned} N_{hi}(t) &= A_{hi}(t) + M_{hi}(t) \\ &= \int_0^t \lambda_{hi}(u) Y_{hi}(u) du + M_{hi}(t), \end{aligned}$$

where $M_{hi}(t)$ is a martingale and $A_{hi}(t)$ is the (predictable) compensator process of $N_{hi}(t)$.

- The martingales $M_{hi}(t)$ can be interpreted as **continuous-time residuals**.
- Plots of $M_{hi}(t)$ against t can be used to compare models, evaluate the model fit, etc.

Bayesian Inference

- In principle, a multi-state model consists of several duration time models
⇒ Adopt methodology developed for nonparametric hazard regression.
- Fully Bayesian inference based on Markov Chain Monte Carlo simulation techniques (Hennerfeind, Brezger & Fahrmeir, 2006):
 - Assign inverse gamma priors to the variance and smoothing parameters.
 - Metropolis-Hastings update for the regression coefficients (based on IWLS-proposals).
 - Gibbs sampler for the variances (inverse gamma with updated parameters).
 - Efficient algorithms make use of the sparse matrix structure of the matrices involved.

- **Mixed model** based empirical Bayes inference (Kneib & Fahrmeir, 2007):
 - Consider the variances and smoothing parameters as **unknown constants** to be estimated by mixed model methodology.
 - Problem: The P-spline priors are **partially improper**.
 - **Mixed model representation**: Decompose the vector of regression coefficients as

$$\xi = X\beta + Zb,$$

where

$$p(\beta) \propto \text{const} \quad \text{and} \quad b \sim N(0, \tau^2 I).$$

$\Rightarrow \beta$ is a **fixed effect** and b is an **i.i.d. random effect**.

- **Penalized likelihood** estimation of the regression coefficients in the mixed model (posterior modes).
- **Marginal likelihood** estimation of the variance and smoothing parameters (Laplace approximation).

Software

- Implemented in BayesX.
- Free software package for Bayesian inference in geadditive and related models.

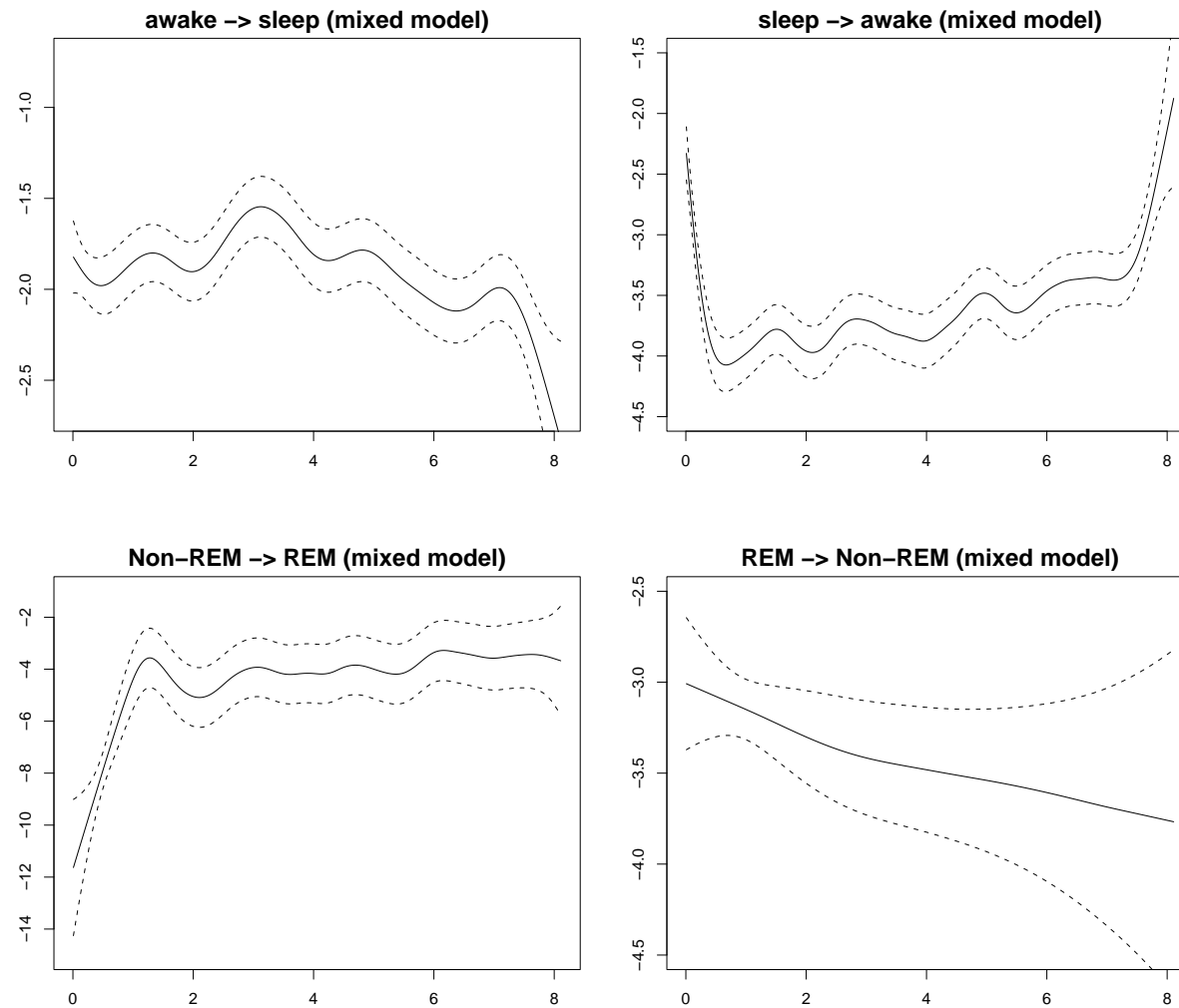


- Available from

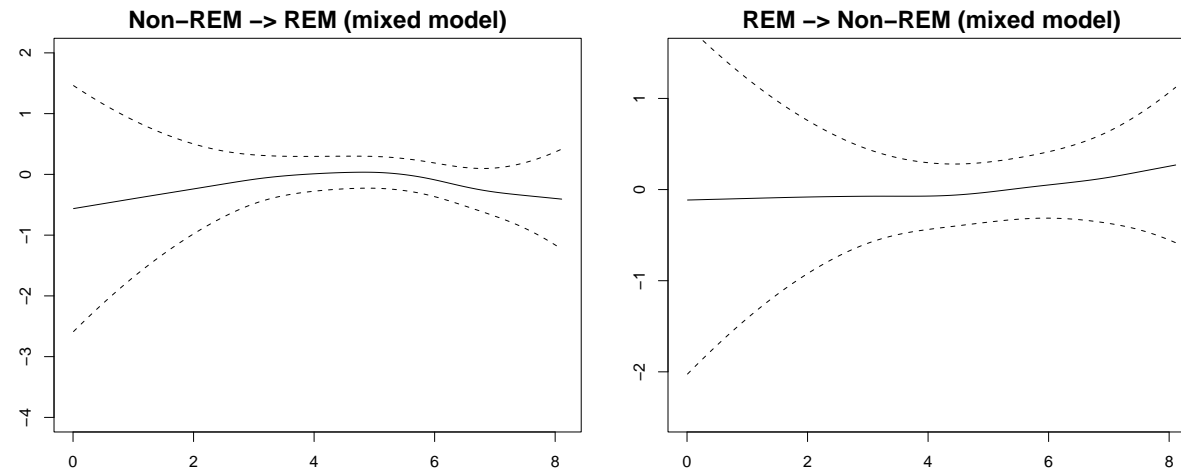
<http://www.stat.uni-muenchen.de/~bayesx>

Human Sleep Data II

- Baseline effects:

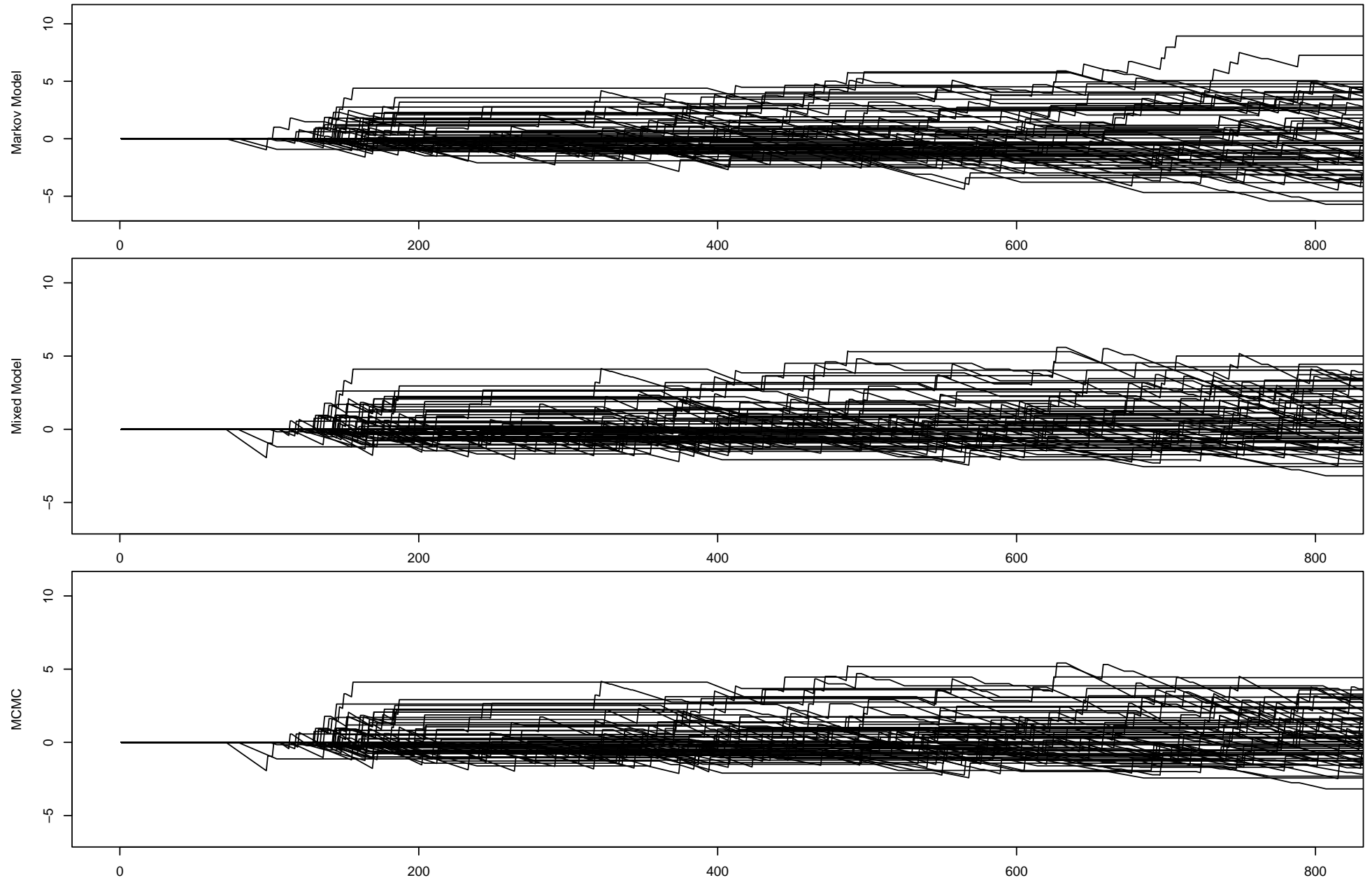


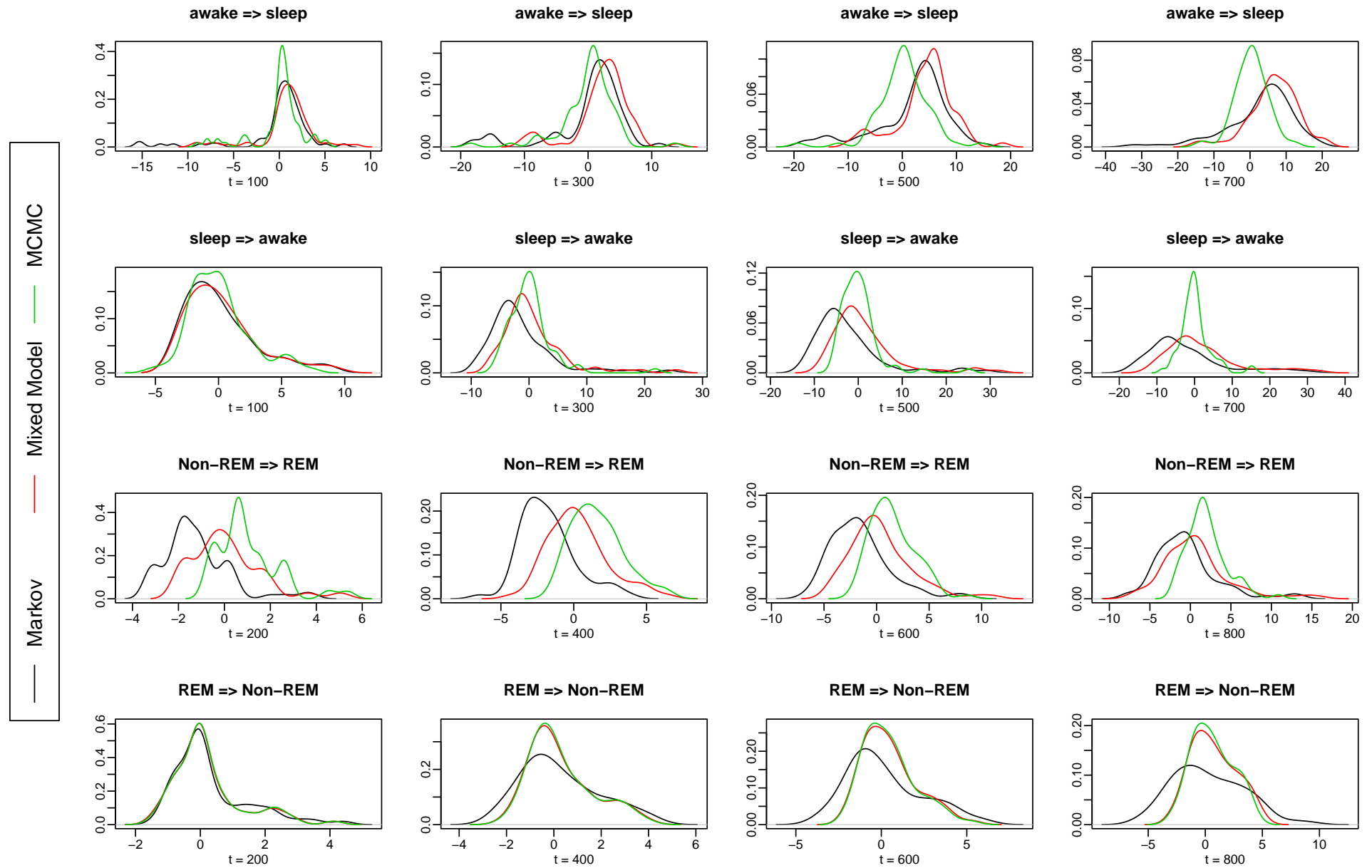
- Time-varying effects for a high level of cortisol:



- The fully Bayesian approach detects individual-specific variation for all transitions.
- The empirical Bayes approach only detects individual-specific variation for the transition between REM and Non-REM.

Martingale residuals REM => Non-REM





Summary and Outlook

- Computationally feasible semiparametric multi-state models for the analysis of event-history data.
- Fully Bayesian and empirical Bayes inference.
- Model validation based on martingale residuals.
- Directly extendable to more complicated models including
 - Nonparametric effects of continuous covariates.
 - Spatial effects.
 - Interaction surfaces and varying coefficients.
- Future work:
 - Application to larger data sets and different types of multi-state models.
 - Consider coarsened observations, i.e. interval censored multi-state data.

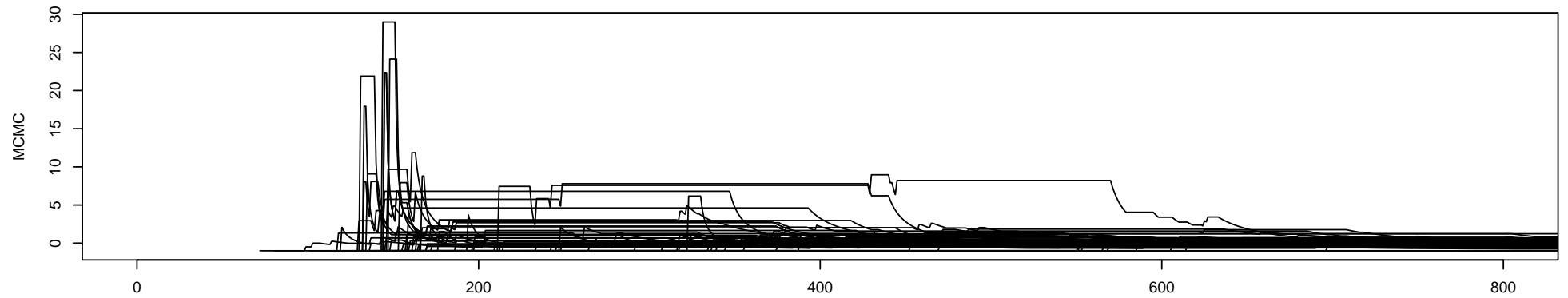
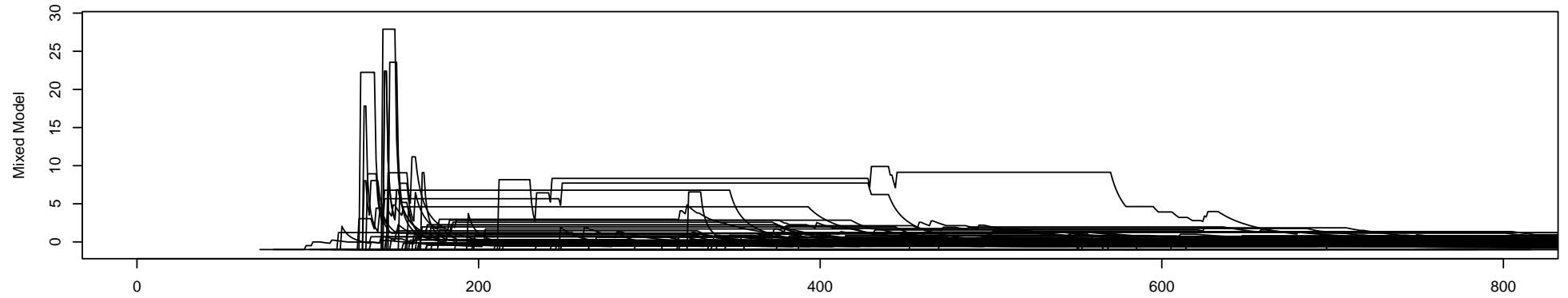
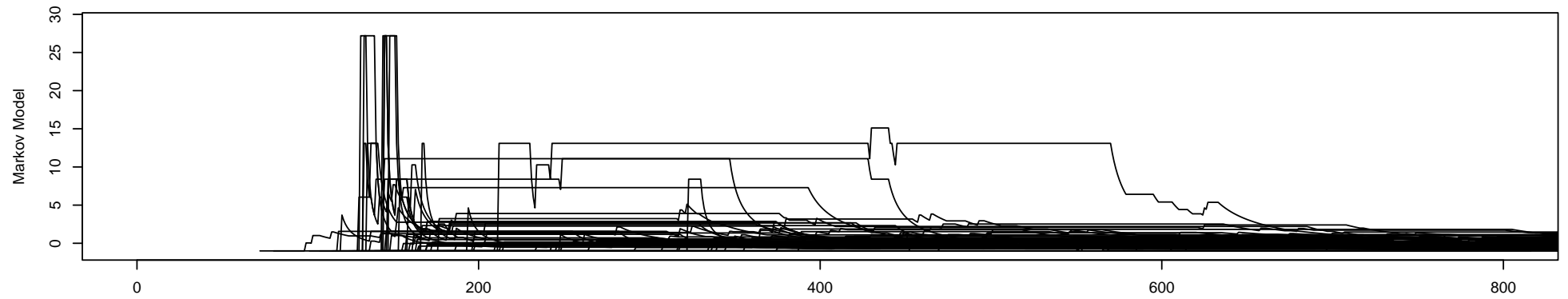
References

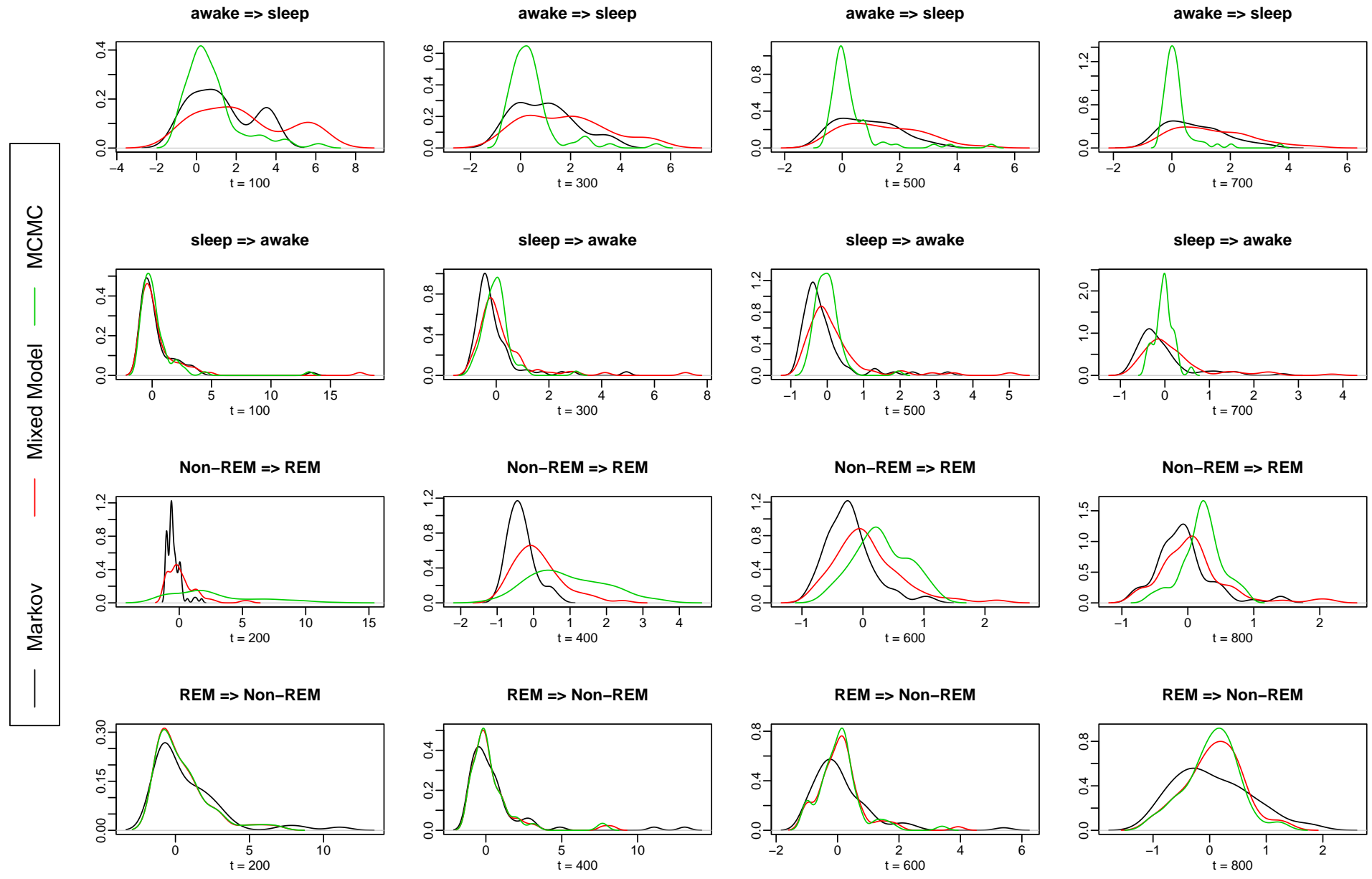
- BREZGER, KNEIB & LANG (2005): BayesX: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*, **14** (11).
- HENNERFEIND, BREZGER, AND FAHRMEIR (2006): Geoadditive survival models. *Journal of the American Statistical Association*, **101**, 1065-1075.
- KNEIB & FAHRMEIR (2007): A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, **34**, 207-228.
- KNEIB & HENNERFEIND (2007): Bayesian semiparametric multi-state models. *Statistical Modelling*, to appear.
- A place called home:

<http://www.stat.uni-muenchen.de/~kneib>

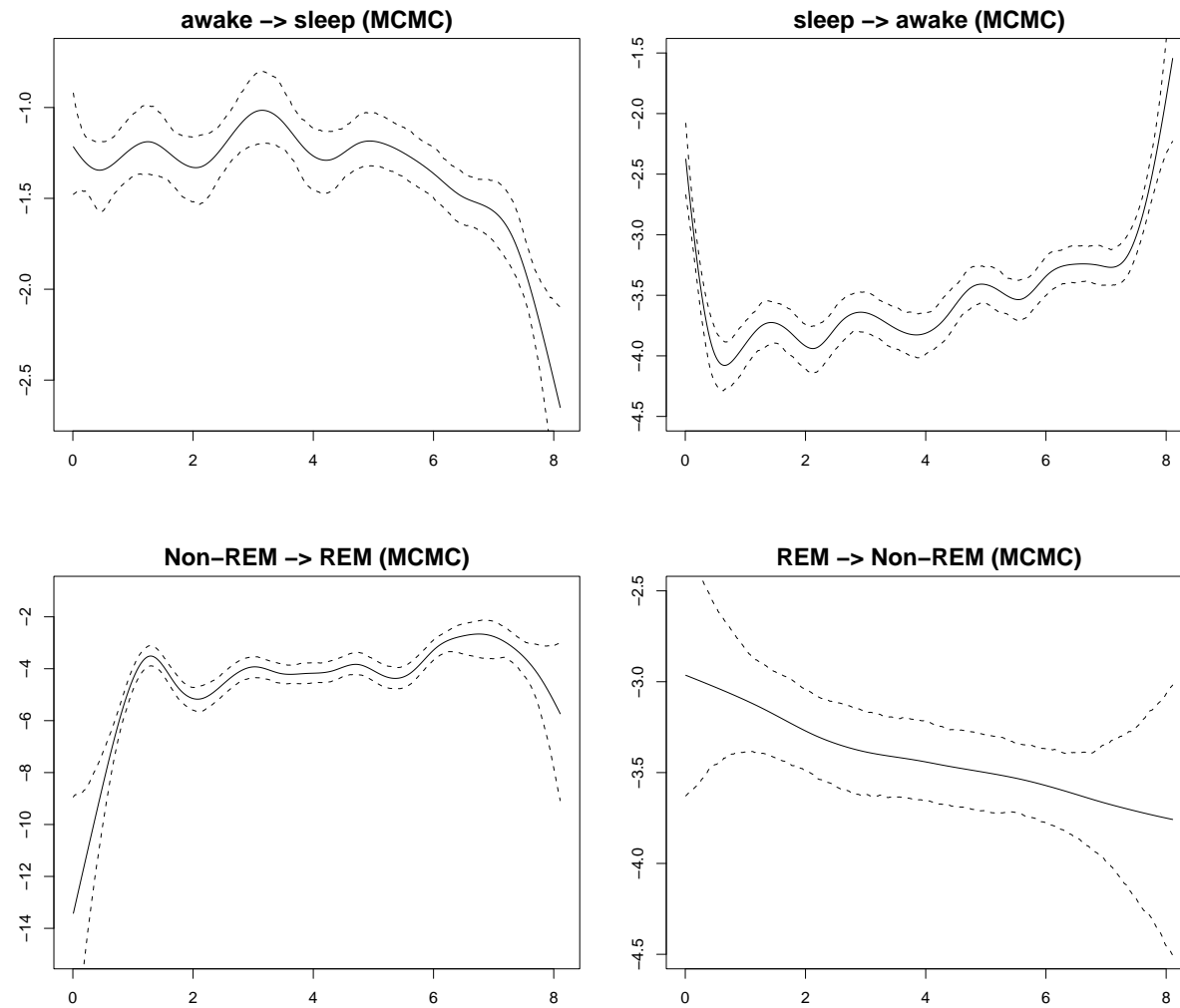
$$\begin{aligned}
l_i &= \sum_{h=1}^k \left[\int_0^{T_i} \log(\lambda_{hi}(t)) dN_{hi}(t) - \int_0^{T_i} \lambda_{hi}(t) Y_{hi}(t) dt \right] \\
&= \sum_{j=1}^{n_i} \sum_{h=1}^k \left[\delta_{hi}(t_{ij}) \log(\lambda_{hi}(t_{ij})) - Y_{hi}(t_{ij}) \int_{t_{i,j-1}}^{t_{ij}} \lambda_{hi}(t) dt \right].
\end{aligned}$$

- k number of possible transitions.
 $N_{hi}(t)$ counting process for type h event and individual i .
 $Y_{hi}(t)$ at risk indicator for type h event and individual i .
 t_{ij} event times of individual i .
 n_i number of events for individual i .
 $\delta_{hi}(t_{ij})$ transition indicator for type h transition.

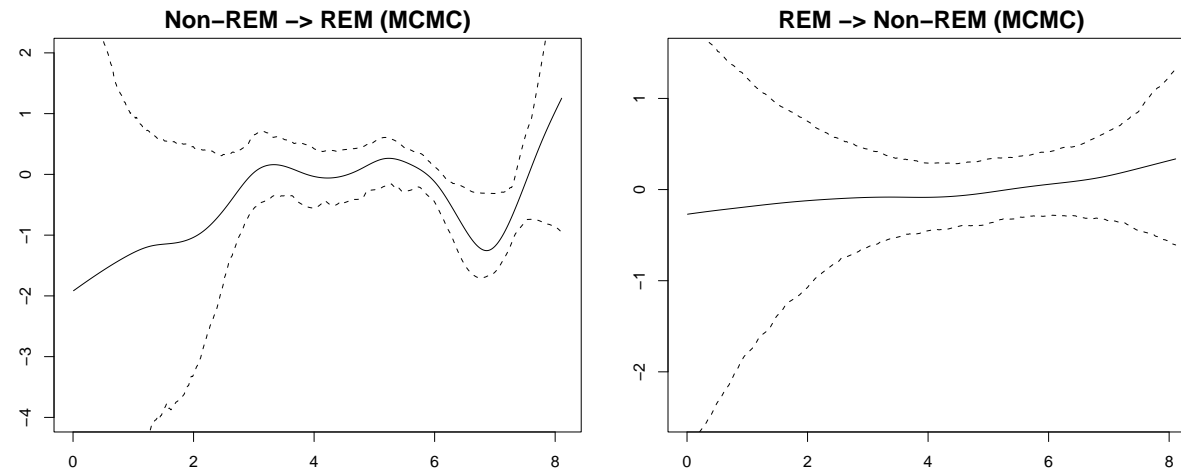




- Baseline effects:



- Time-varying effects for a high level of cortisol:



- Extensions for **signal regression**:
 - Sleep stages are obtained from the aggregated EEG signals.
 - Penalised splines can also be used in the **analysis of the raw data**.
 - Example: Analyse differences in signal intensities between healthy and depressed study participants.
 - **Logit signal regression**:

$$P(Y = 1) = h(Z\gamma)$$

where Z is the matrix of signal intensities and γ the corresponding regression coefficient.

- Assume that the effect of the signals varies smoothly over time
 - ⇒ Model $\gamma = \gamma(t)$ as a P-spline.
- First promising results have been achieved with this approach.