Functional Estimation in Systems Defined by Differential Equation using Bayesian Smoothing Methods

19th International Conference on Computational Statistics

August 24th 2010

Pharmacokinetics on theophylline

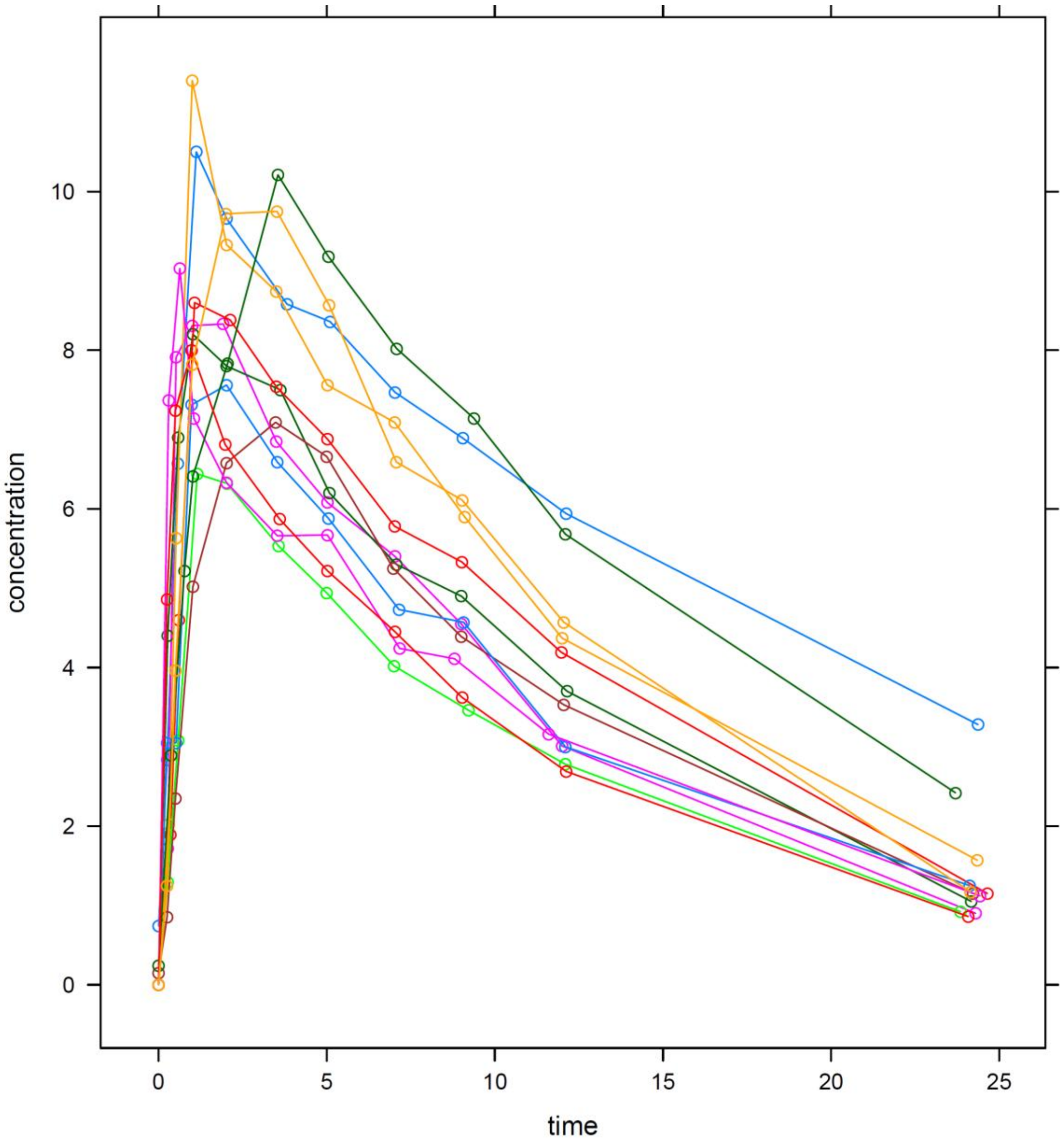


Figure 1 - Serum concentrations of the anti-asthmatic drug theophylline

Data for the **kinetics** of the anti-asthmatic drug **theophylline**.

12 subjects were given oral doses of theophylline. Serum concentrations were measured at **11 time points** over **25 hours** for each subject.

Pharmacokinetics on theophylline

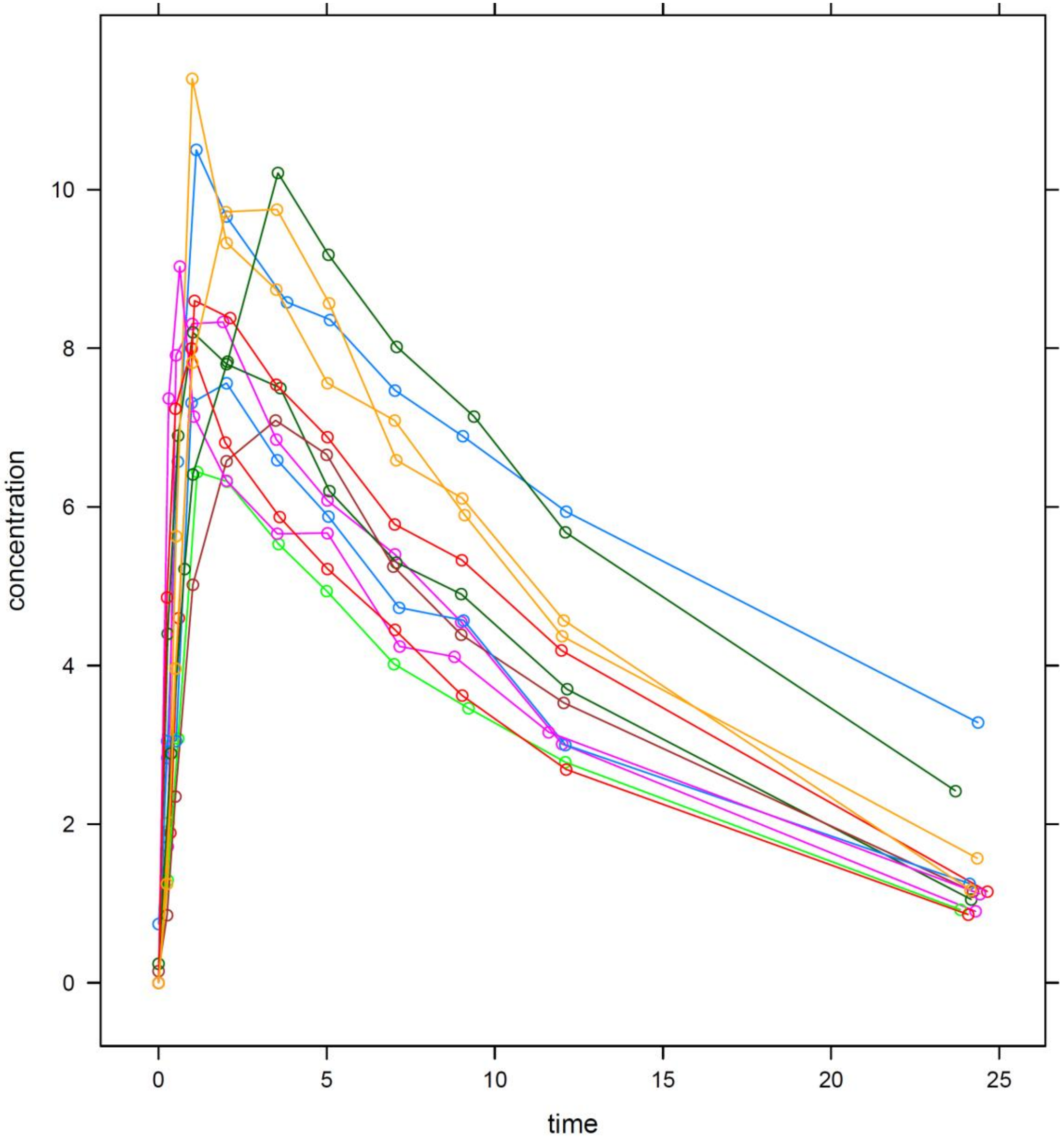


Figure 1 - Serum concentrations of the anti-asthmatic drug theophylline

Differential equations for the two compartments model:

$$\begin{cases} \frac{dQ_a(t)}{dt} & = & -k_a Q_a(t) \\ \frac{dC_e(t)}{dt} & = & \frac{k_a}{V} Q_a(t) - k_e C_e(t) \\ Q_a(0) & = & D \\ C_e(0) & = & 0 \end{cases}$$

Explicit solution of the differential equations system:

$$\begin{cases} Q_a(t) & = & D e^{-k_a t} \\ C_e(t) & = & \frac{D}{V} \frac{k_a}{k_a - k_e} (e^{-k_e t} - e^{-k_a t}) \end{cases}$$

Objectives

Introduce the concept of Bayesian ODE-penalized B-spline method in the case of linear differential equations system:

- Individual case,
- Hierarchical case.

I. **Standard Bayesian smoothing approach**

II. Hierarchical Bayesian smoothing approach

III. Illustration

IV. Conclusion & further work

Differential equation and measurement

$$\begin{cases} D\boldsymbol{x}(t) & = & f(\boldsymbol{x}(t), \boldsymbol{\theta}) \\ \boldsymbol{x}(0) & = & \boldsymbol{x_0} \end{cases}$$

With:

- $(\boldsymbol{x}(t))^{\boldsymbol{T}} = (x_1(t), \dots, x_d(t))$ the set of d **state** functions and $\boldsymbol{x_0}$ the set of **initial conditions**,
- $\boldsymbol{\theta}$ the **vector of parameters** involved in the set of differential equations.

A subset $\mathcal{J}$ of the d state functions are observed with **measurement errors** ε_j:

$$y_j = x_j(t) + \varepsilon_j$$

Basis function expansion

$$\tilde{x}_j(t) \quad = \quad \left(\boldsymbol{B_j}(t)\right)^T \boldsymbol{c_j}$$

With:

- $\boldsymbol{B_j}(t)$ the vector of B-spline basis functions at time t,
- $\boldsymbol{c_j}$ the vector of spline coefficients.

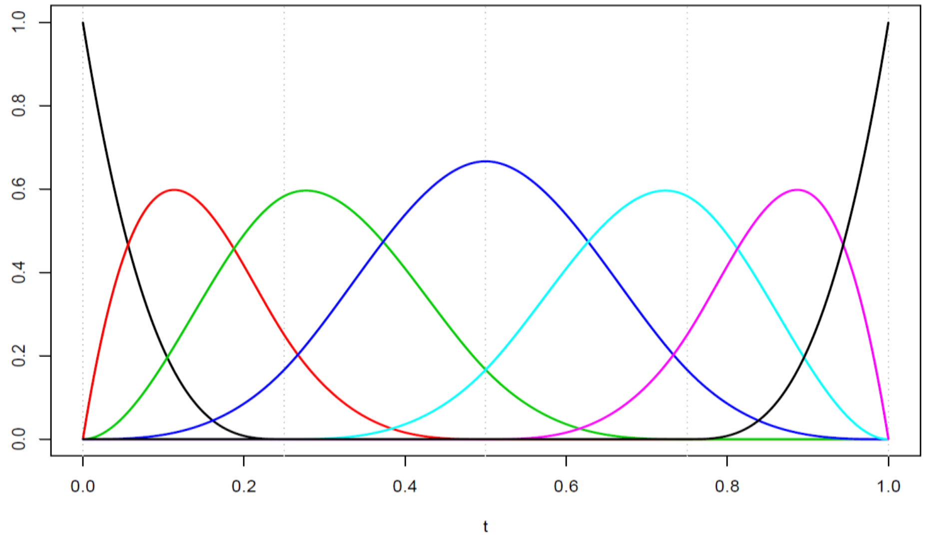


Figure 2 – B-spline basis function of order 4 with knots at 0.25, 0.5 and 0.75

Penalty

The penalty for the j-th equation asses the proximity of the approximation $\tilde{x}_j(t)$ from the solution $x_j(t)$

$$\begin{aligned} PEN_j &= \gamma_j \int \left(D\tilde{x}_j(t) - f_j(\tilde{\boldsymbol{x}}(t), \boldsymbol{\theta}) \right)^2 dt \\ PEN &= \sum_{j=1}^{d} PEN_j \\ &= \boldsymbol{c}^T \boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}) \boldsymbol{c} \end{aligned}$$

Where $\boldsymbol{\gamma}^T = (\gamma_1, \dots, \gamma_d)$ is the ODE-adhesion parameters vector and $\boldsymbol{c}^T = \left(\boldsymbol{c}_1^T, \dots, \boldsymbol{c}_d^T\right)$

Fitting criterion

$$J(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{\tau} | \boldsymbol{\gamma}, \boldsymbol{y}) = \sum_{j \in \mathcal{J}} \left\{ \frac{n_j}{2} \log(\tau_j) - \frac{\tau_j}{2} \sum_{k=1}^{n_j} \left(y_{jk} - \tilde{x}_j(t_{jk}) \right)^2 \right\} - \frac{1}{2} PEN$$

J is a trade-off between **goodness-of-fit** and **solving the system of differential equations**.

Bayesian model

$$\begin{cases} y_{jk}|\boldsymbol{c_j},\tau_j & \sim & \mathcal{N}\left(\left(\boldsymbol{B_j}(t_{jk})\right)^T \boldsymbol{c_j};\tau_j^{-1}\right) & j \in \mathcal{J}, k=1,\dots,n_j \\ \pi(\boldsymbol{c}|\boldsymbol{\theta},\boldsymbol{\gamma}) & \propto & \exp\left(-\frac{1}{2}PEN - \frac{1}{2}\{\boldsymbol{c}^T\boldsymbol{\Sigma_c}^{-1}\boldsymbol{c} - 2\boldsymbol{c}^T\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c}\}\right) & \\ \gamma_j & \sim & \mathcal{Ga}\left(a_{\gamma_j};b_{\gamma_j}\right) & j=1,\dots,d \\ \tau_j & \sim & \mathcal{Ga}\left(a_{\tau_j};b_{\tau_j}\right) & j \in \mathcal{J} \\ \boldsymbol{\theta} & \sim & \pi(\boldsymbol{\theta}) & \end{cases}$$

The second term in $\pi(\boldsymbol{c}|\boldsymbol{\theta},\boldsymbol{\gamma})$ expresses **uncertainty** w.r.t. **initial conditions** of the state function.

Constant of normalization for prior distribution of spline coefficients $\boldsymbol{c}$:

$$(\det(\boldsymbol{M_1}))^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{v_1}^T\boldsymbol{M_1}^{-1}\boldsymbol{v_1}\right)$$

Where:

- $\boldsymbol{M_1} = \boldsymbol{M_1}(\boldsymbol{\theta},\boldsymbol{\gamma}) = \boldsymbol{R}(\boldsymbol{\theta},\boldsymbol{\gamma}) + \boldsymbol{\Sigma_c}^{-1}$
- $\boldsymbol{v_1} = \boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c}$

Conditional posterior distributions for $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$:

Marginalization of the joint posterior distribution w.r.t. the spline coefficients to avoid correlation between ODE parameters chains and spline coefficients chains.

Metropolis-Hastings steps for ODE-adhesion parameters $\gamma_j, j = 1, \dots, d$, the precision parameters $\tau_j, j \in \mathcal{J}$ and for differential equation parameter $\boldsymbol{\theta}$ using **adaptive proposals** to reduce the rejection rate.

If necessary, use of **rotation and translation** to avoid correlation between components in $\boldsymbol{\theta}$.

After convergence of MCMC-chains for $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$:

If needed, sample directly from the conditional posterior distribution of the spline coefficients $\boldsymbol{c}$ using a **multivariate Gaussian distribution**.

Differential equation and measurement for the subject $\boldsymbol{i} = 1, \dots, \boldsymbol{I}$

$$\begin{cases} D\boldsymbol{x_i}(t) & = & f(\boldsymbol{x_i}(t), \boldsymbol{\theta_i}) \\ \boldsymbol{x_i}(0) & = & \boldsymbol{x_{i,0}} \end{cases}$$

For each subject i, the same subset $\mathcal{J}$ of state functions are observed with **measurement errors** ε_{ij}:

$$y_{ij} = x_{ij}(t) + \varepsilon_{ij}$$

Basis function expansion for the $\boldsymbol{j}$ state function of the subject $\boldsymbol{i}$

$$\tilde{x}_{ij}(t) = \left(\boldsymbol{B_{ij}}(t)\right)^T \boldsymbol{c_{ij}}$$

Penalty

The individual penalty term and the overall penalty term are similar from the standard approach:

$$PEN_i = \boldsymbol{c}_i^T \boldsymbol{R}_i(\boldsymbol{\theta}_i, \boldsymbol{\gamma}) \boldsymbol{c}_i$$

$$PEN = \sum_{i=1}^{I} PEN_i = \boldsymbol{c}^T \boldsymbol{R}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_I, \boldsymbol{\gamma}) \boldsymbol{c}$$

Bayesian model

$$\begin{cases} y_{ijk}|\boldsymbol{c_{ij}},\tau_j & \sim & \mathcal{N}\left(\left(\boldsymbol{B_{ij}}(t_{ijk})\right)^T \boldsymbol{c_{ij}}; \tau_j^{-1}\right) & i=1,\dots,I \quad j\in\mathcal{J} \quad k=1,\dots n_{ij} \\ \pi(\boldsymbol{c}|\boldsymbol{\theta_1},\dots,\boldsymbol{\theta_I},\boldsymbol{\gamma}) & \propto & \exp\left(-\frac{1}{2}PEN-\frac{1}{2}\{\boldsymbol{c}^T\boldsymbol{\Sigma_c}^{-1}\boldsymbol{c}-2\boldsymbol{c}^T\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c}\}\right) & \\ \boldsymbol{\theta_i}|\boldsymbol{\theta},\boldsymbol{P_\theta} & \sim & \mathcal{N}(\boldsymbol{\theta};\boldsymbol{P_\theta}^{-1}) & i=1,\dots,I \\ \gamma_j & \sim & \mathcal{G}a\left(a_{\gamma_j};b_{\gamma_j}\right) & j=1,\dots,d \\ \tau_j & \sim & \mathcal{G}a\left(a_{\tau_j};b_{\tau_j}\right) & j\in\mathcal{J} \\ \boldsymbol{P_\theta} & \sim & \mathcal{W}_q(\boldsymbol{V}^{-1};r) & \\ \boldsymbol{\theta} & \sim & \mathcal{N}(\mu;\boldsymbol{\Lambda}^{-1}) & \end{cases}$$

Constant of normalization:

$$(\det(\boldsymbol{M_1}))^{\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{v_1}^T\boldsymbol{M_1}^{-1}\boldsymbol{v_1}\right)$$

Where $\boldsymbol{M_1}=\boldsymbol{R}(\boldsymbol{\theta_1},\dots,\boldsymbol{\theta_I},\boldsymbol{\gamma})+\boldsymbol{\Sigma_c}^{-1}$ and $\boldsymbol{v_1}=\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c}$

Conditional posterior distributions for $\boldsymbol{\gamma}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_I, \boldsymbol{\tau}, \boldsymbol{\theta}$ and $\boldsymbol{P}_{\boldsymbol{\theta}}$

Marginalization of the joint posterior distribution w.r.t. the spline coefficients to avoid correlation between individual ODE parameters chains and individual spline coefficients chains.

Metropolis-Hastings steps for ODE-adhesion parameters $\gamma_j, j = 1, \dots, d$, the precision parameters $\tau_j, j \in \mathcal{J}$ and for each differential equation parameter $\boldsymbol{\theta}_i$ using **adaptive proposals** to reduce the rejection rate.

If necessary, use of **rotations and translations** to avoid correlation between components in each individual parameter $\boldsymbol{\theta}_i$

Gaussian and Wishart distribution for the conditional posterior distribution of the **mean population parameter** $\boldsymbol{\theta}$ and **precision parameter** $\boldsymbol{P}_{\boldsymbol{\theta}}$ of random effects

After convergence of the MCMC-chains for $\boldsymbol{\gamma}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_I, \boldsymbol{\tau}, \boldsymbol{\theta}$ and $\boldsymbol{P}_{\boldsymbol{\theta}}$

If needed, sample directly from the conditional posterior distribution of the spline coefficients $\boldsymbol{c}$ using a multivariate Gaussian distribution.

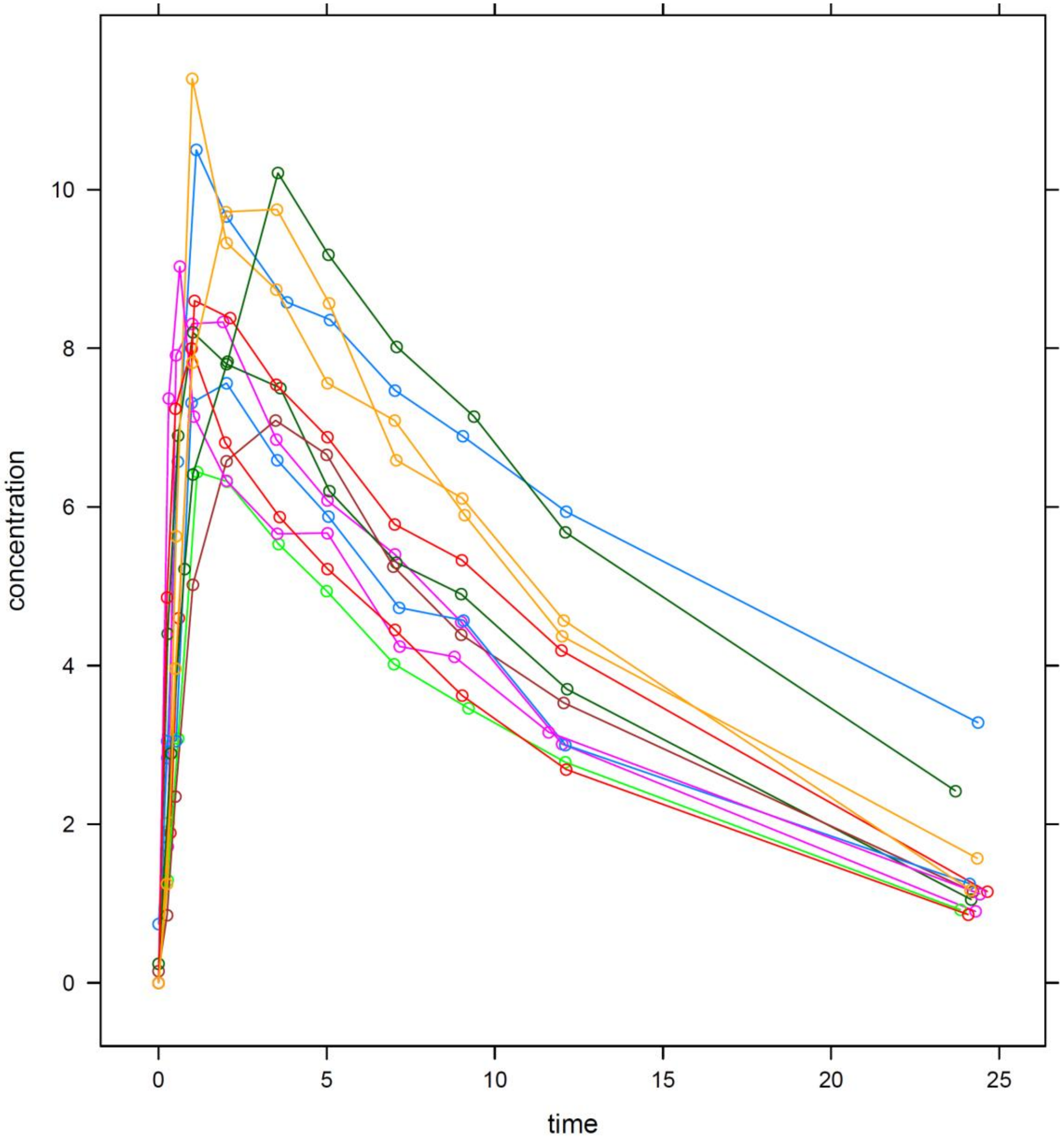


Figure 3 - Serum concentrations of the anti-asthmatic drug theophylline

Differential equation:

$$\begin{cases} \frac{dQ_a(t)}{dt} & = & -k_a Q_a(t) \\ \frac{dC_e(t)}{dt} & = & \frac{k_a}{V} Q_a(t) - k_e C_e(t) \\ Q_a(0) & = & D \\ C_e(0) & = & 0 \end{cases}$$

Data distribution, parameterization and random effects:

Additive Gaussian error measurements.
Log-parameterization for the PK parameters.
Gaussian random effects on the log-PK parameters.

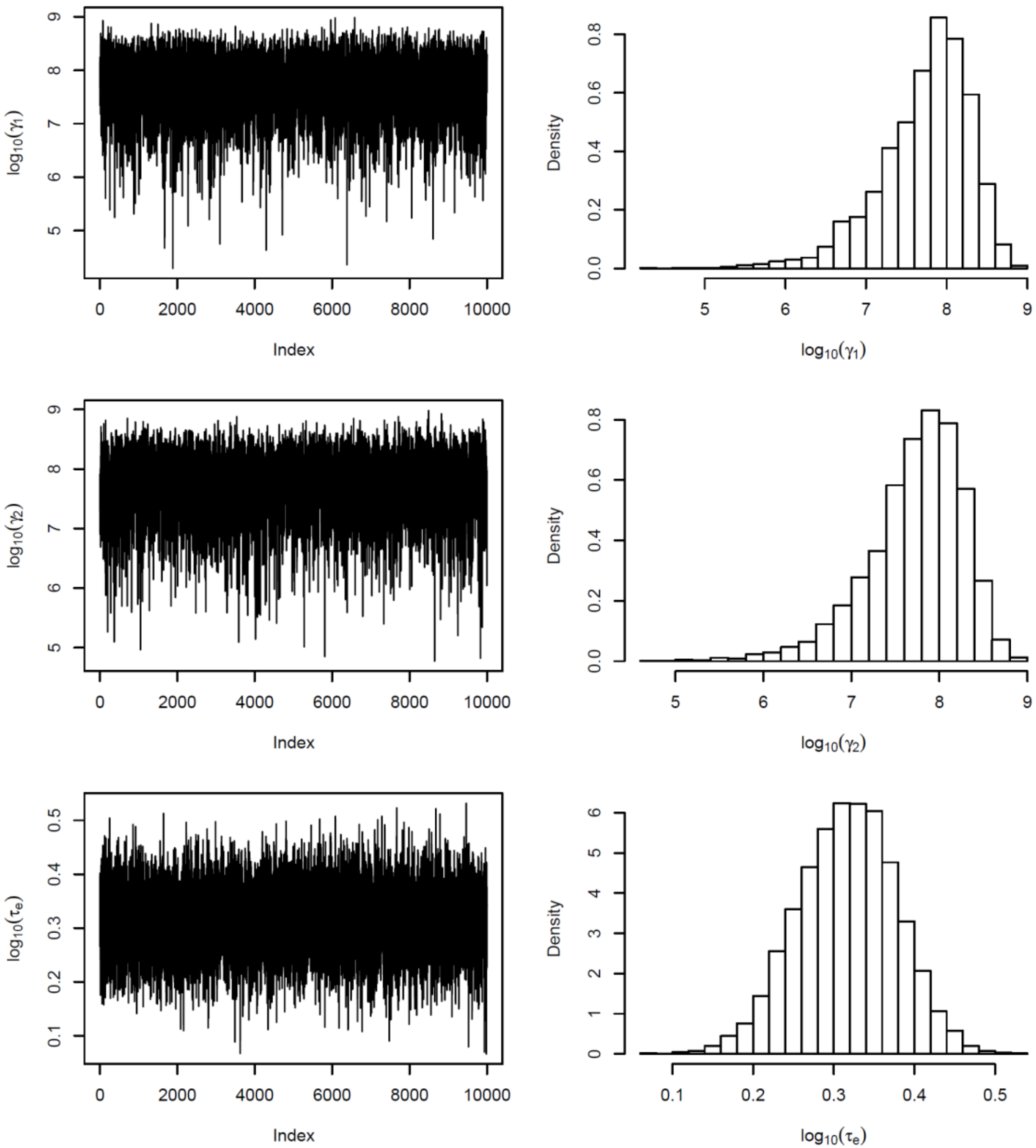


Figure 4 - Traces and histograms for $\log_{10}(\gamma_1)$, $\log_{10}(\gamma_2)$ and $\log_{10}(\tau_e)$

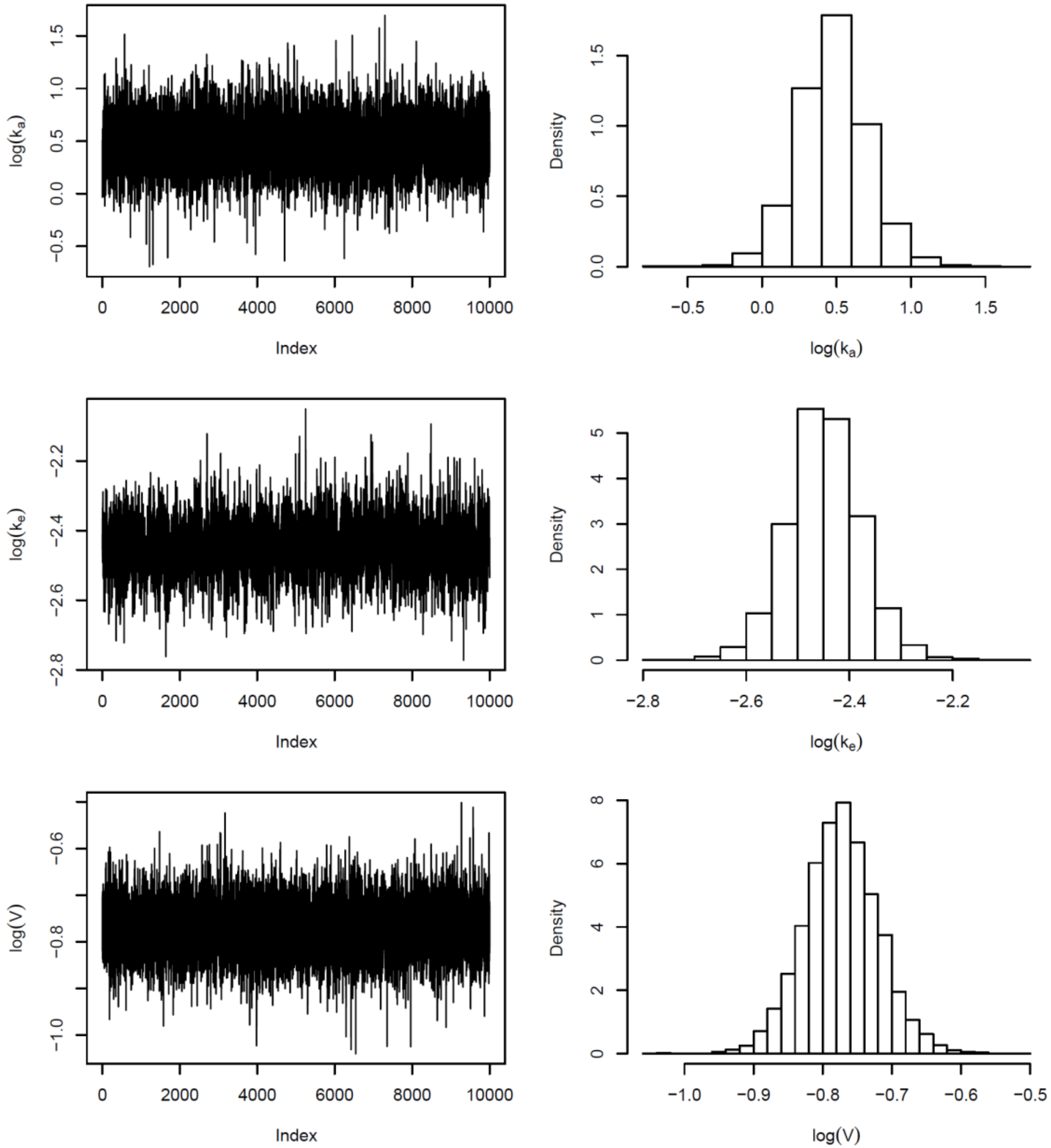


Figure 5 - Traces and histograms for $\log(k_a)$, $\log(k_e)$ and $\log(V)$.

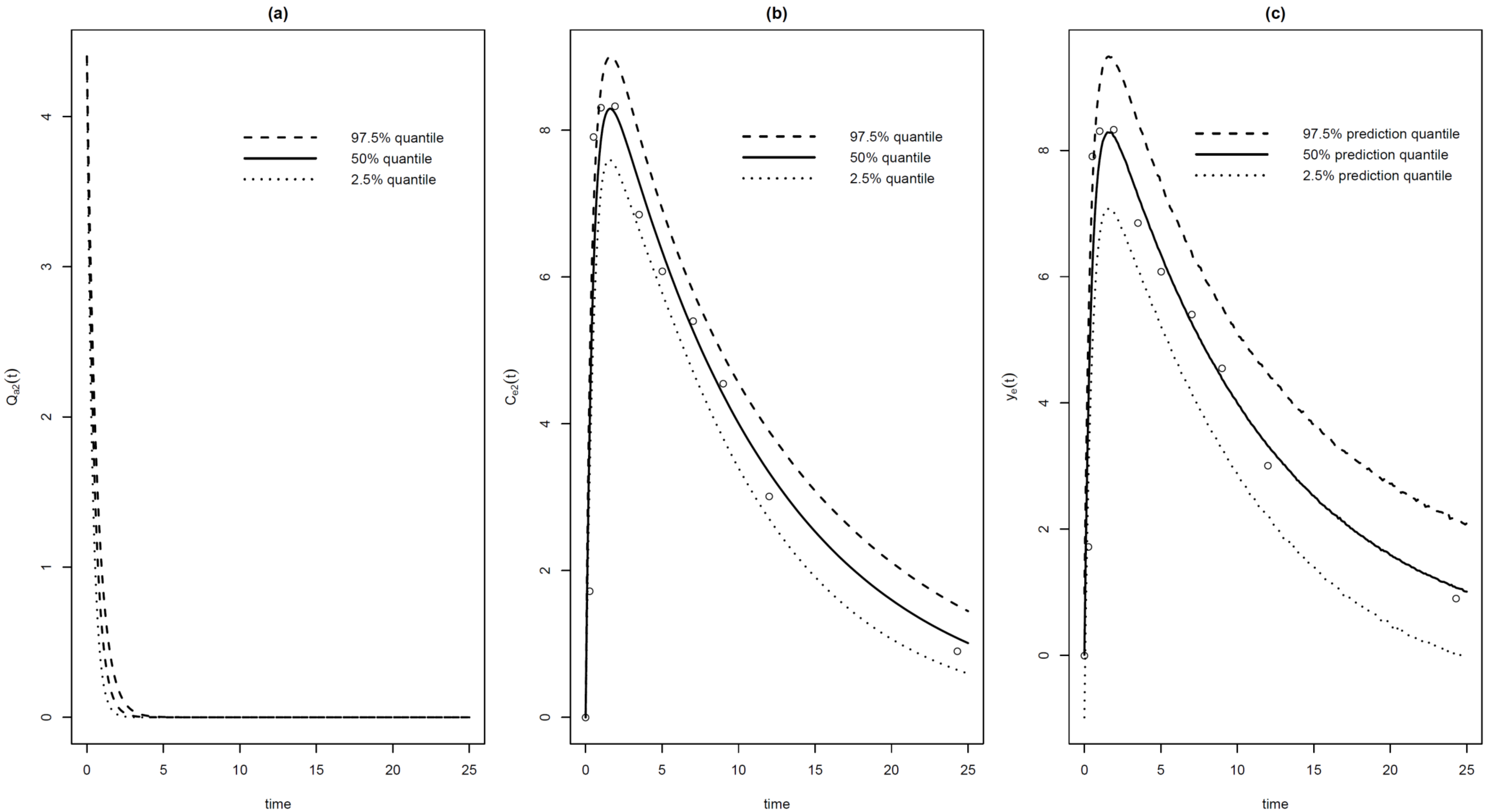


Figure 6 – (a): Credibility interval for the posterior mean of $Q_{a,2}(t)$. (b): Credibility interval for the posterior mean of $C_{e,2}(t)$. (c): Credibility interval for the posterior predictive distribution of $C_{e,2}(t)$ (subject #2).

I. Standard Bayesian smoothing approach

II. Hierarchical Bayesian smoothing approach

III. Illustration

IV. Conclusion & further work

Conclusion

- **Powerful tool** that overcomes solving the DE using a numerical method,
- **Convenient implementation** of the Bayesian generalized profiling estimation for DE,
- Simple method to include **prior information** about DE parameters,
- Possibility to express **uncertainty** with respect to initial conditions,
- Hierarchical part only a **simple generalization** of the standard approach

Further work

- Consider other **data distributions**,
- Generalize this method to **nonlinear differential equations**,
- **Optimal design** for the data collection,
- Differential equation model with **lagged effects** e.g. $Dx(t) = f(x(t-\delta_1), u(t-\delta_2), \theta)$.

[1] Berry S.M., Carroll R.J. and Ruppert D., Bayesian smoothing and regression splines for measurement error problems, Journal of the American Statistical Association, 97:160-169 (2002)

[2] Poyton A.A, Varziri M.S., McAuley K.B., McLellan P.J. and Ramsay J.O., Parameter estimation in continuous-time dynamic models using principal differential analysis, Computers and Chemical Engineering, 30:698-708 (2006)

[3] Ramsay J.O., Hooker G., Campbell D. and Cao J., Parameter estimation for differential equations: a generalized smoothing approach, Journal of the Royal Statistical Society, Series B, 69:741-796 (2007)

[4] Campbell D., Bayesian collocation tempering and generalized profiling for estimation of parameters from differential equation models, PhD Thesis (2007)

[5] Cai B., Meyer R and Perron F., Metropolis-Hastings algorithms with adaptive proposals, Statistics and Computing, 18:421-433 (2008)

[6] Lambert P., Archimedean copula estimation using Bayesian splines smoothing techniques, Computational Statistics & Data Analysis, **51**:6307-6320 (2007)

Appendix A: B-splines definition & properties

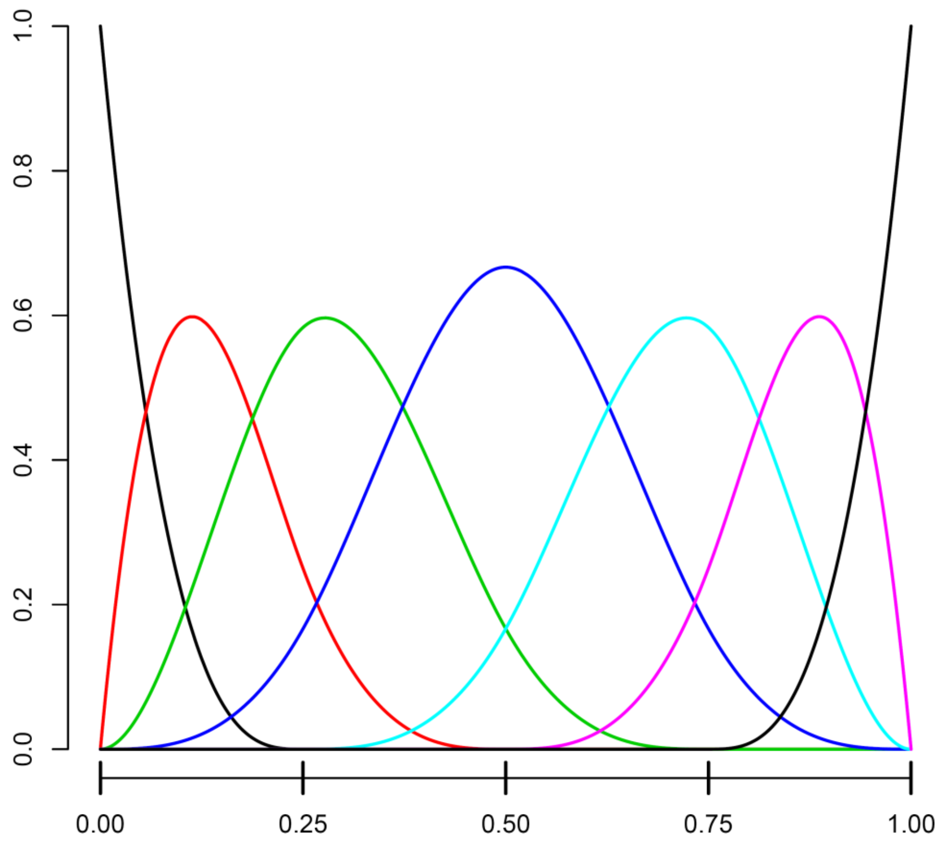
Definition

B-spline basis function defined using:

- order p,
- m inner knots at $\tau_1 \leq \cdots \leq \tau_m$,
- p-multiple knots τ_0 and τ_{m+1},
- **Recursive definition** for each function $\boldsymbol{B_k}(t,p)$.

Properties

- $B_k(t,p)$ is a **piecewise polynomial** of degree $p-1$,
- Derivatives up to order $p-2$ are continuous,
- **Sum** of all non-zero basis function is **1**,
- Number of basis function is $\boldsymbol{K = m + p}$.



Appendix B: Bayesian Smoothing method: Credibility intervals

	2.5% quantile	Mean	Median	97.5% quantile
$\log_{10}(\gamma_1)$	6.3824	7.7267	7.8266	8.5639
$\log_{10}(\gamma_2)$	6.3958	7.7263	7.8091	8.5516
$\log_{10}(\tau_e)$	0.1934	0.3162	0.3181	0.4325
lk_a	0.0172	0.4761	0.4744	0.9516
τ_{lk_a}	0.6393	2.1304	1.9383	4.7129
lk_e	-2.5893	-2.4482	-2.4495	-2.3050
τ_{lk_e}	9.8007	94.2905	47.8619	505.1716
lV	-0.8793	-0.7732	-0.7742	-0.6645
τ_{lV}	13.9389	45.9627	40.6656	109.9077

Table 1 – Posterior mean, median and credibility intervals for