# Régression bayésienne sous contraintes de régularité et de forme avec splines à noeuds variables

Khader Khadraoui, Christophe Abraham

# Régression bayésienne sous contraintes de régularité et de forme avec B-splines à nœuds variables

Christophe Abraham [1] & Khader Khadraoui [2]

[1] *Montpellier SupAgro, UMR 729 MISTEA E-mail address: abraham@supagro.inra.fr*
[2] *Université de Lyon 1, Laboratoire SAF EA 2429, khader.khadraoui@univ-lyon1.fr*

**Mots-clés.** Régression bayésienne, Contraintes de forme, B-splines, nœuds variables, ...

**Abstract.** A Bayesian method for regression with free knots B-splines under shape restrictions and smoothness conditions is proposed. The shape constraints are taken into account thanks to the prior distribution. To propose statistical inference on the number and position of knots, we use a Multinomial-Dirichlet model. After specifying priors on all unknown parameters, we derive to a constant the posterior distribution. It is shown that in the case of Bayesian regression under shape constraints and with free knots, Bayesian analysis becomes complex. To handle the difficulties, we propose original simulation scheme which allows to simulate from the truncated posterior distribution with free dimension. To estimate the regression function, the posterior mode and the posterior expectation are calculated thanks to a Reversible Jump Metropolis-Hastings within Gibbs algorithm.

**Résumé.** On propose une méthode bayésienne de régression sous contraintes de régularité et de forme avec B-splines à nœuds variables. Les contraintes de forme sont prises en compte grâce à la loi a priori. Nous considérons un modèle Multinomial-Dirichlet pour l'inférence statistique sur le nombre et la position des nœuds. Une fois les lois a priori sur l'ensemble des paramètres inconnus sont spécifiées, nous dérivons à une constante près la loi a posteriori. En présence de nœuds variables, l'analyse bayésienne de la régression avec B-splines sous contraintes devient très complexe. On propose des schémas de simulation originaux permettant de générer suivant la loi a posteriori lorsque la densité tronquée des cœfficients de régression prend des dimensions variables. Pour estimer la fonction de régression, nous calculons le mode a posteriori et l'espérance a posteriori grâce à des simulations suivant la loi a posteriori par un algorithme de type Metropolis-Hastings à sauts réversibles avec l'échantillonneur de Gibbs.

## 1 Introduction

Nous considérons le modèle usuel de régression, pour un vecteur de réponse $\boldsymbol{y} = (y_1, \ldots, y_n)'$ et un vecteur explicatif $\boldsymbol{x} = (x_1, \ldots, x_n)'$,

$$\begin{cases} y_j = f(x_j) + \epsilon_j, \qquad j = 1, \ldots, n, \\ f(x) = \sum_{i=1}^{K} \beta_i B_{i,k}(x), \qquad \forall x \in [a, b], \end{cases} \tag{1}$$

où $f : [a, b] \subset \mathbb{R} \mapsto \mathbb{R}$ est une fonction inconnue et $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ est l'erreur de distribution Normale d'espérance nulle $\epsilon_j \overset{\text{iid}}{\sim} N(0, \sigma^2)$ avec $\sigma^2 < \infty$. L'entier $K < n$ désigne la dimension du vecteur des nœuds intérieurs $\mathbf{t} = (t_1, \ldots, t_\kappa)'$ plus l'ordre $k$ des fonctions B-splines $B_{i,k}$, $i = 1, \ldots, \kappa + k$.

Le problème consiste à estimer la fonction de régression $f$ sous des contraintes de régularité et de forme. Pour une estimation flexible et parcimonieuse de $f$, il s'avère important de choisir attentivement les positions des nœuds $\mathbf{t} = (t_1, \ldots, t_\kappa)'$ et leur dimension $\kappa$. En littérature, des discussions à propos du nombre et de la position des nœuds dans la modélisation par spline peuvent être trouvées dans Parker and Rice (1985), O'Sullivan (1986), Kelly and Rice (1990), Hastie (1996), Eilers and Marx (1996), Ruppert (2002), Wand (2003) et Claeskens et al. (2009). Kauermann et al. (2009) ont proposé un travail théorique en statistique bayésienne pour étudier les propriétés asymptotiques d'un estimateur bayésien construit à partir d'une spline de lissage et dont la dimension de la base B-spline augmente avec la taille de l'échantillon. Claeskens et al. (2009) ont justifié, en fonction du nombre des nœuds et la taille des données, que les propriétés théoriques des splines de régression pénalisées sont similaires à celles des splines de régression et des splines de lissage. Dans les travaux appliqués, on peut citer la méthode de sélection pas à pas des nœuds (Stone et al., 1997), l'algorithme de Leitenstorfer and Tutz (2007) avec des fonctions à bases radiales et les méthodes bayésiennes de Denison et al. (1998) et DiMatteo et al. (2001).

Dans ce travail, nous proposons d'étudier une problématique plus complexe que ce qui a été étudié en littérature jusqu'à présent. Il s'agit d'un problème de régression bayésienne sous contraintes de forme qui s'ajoute au problème des nœuds variables. On traitera essentiellement l'aspect pratique du problème où différentes difficultés sont à résoudre. La première difficulté consiste à spécifier judicieusement des priors pour tous les paramètres inconnus du modèle. La seconde difficulté est de proposer des schémas de simulations suivant la loi a posteriori. En effet, nous proposons de simuler suivant la loi a posteriori par un algorithme de type Metropolis-Hastings à sauts réversibles couplé à l'échantillonneur de Gibbs. Le calcul de la probabilité d'acceptation/rejet de cet algorithme exige le calcul d'un rapport de deux priors sur les cœfficients de la régression. Pour introduire les contraintes de forme, la densité du prior des cœfficients est tronquée à un ensemble quelconque et ainsi connue uniquement à une constante près. Pour simuler suivant la loi a posteriori, on a besoin de calculer le rapport de prior sur les cœfficients ce qui revient à calculer le rapport de deux densités tronquées de dimensions différentes et connues uniquement à deux constantes multiplicatives près.

Cet article est organisé comme suit: dans une première partie, nous présentons l'inférence bayésienne sous contraintes en donnant le modèle complet et le posterior. Dans une deuxième partie, nous abordons les schémas des simulations suivant la loi a posteriori et nous expliquons comment on peut varier la configuration des nœuds en nombre et en position grâce à un modèle Multinomial-Dirichlet. En particulier, on montre qu'il est possible de générer suivant la loi a posteriori à partir d'un algorithme de type Metropolis-Hastings à sauts réversibles avec l'échantillonneur de Gibbs et que l'estimateur bayésien peut être calculé à partir de ses simulations suivant la loi a posteriori.

## 2 Le modèle complet et le posterior

Nous rappelons que le modèle (1) s'écrit matriciellement $\boldsymbol{y} = B\beta + \epsilon$, où on note par $B$ une matrice $n \times (\kappa + k)$ qui désigne la base de B-spline, $K = (\kappa + k)$ avec $0 \leq \kappa \leq n$, $\beta = (\beta_1, \ldots, \beta_K)'$ et $\epsilon \sim \mathcal{N}_n(0, \sigma^2 \mathbb{I}_n)$. On désigne par $(\beta, \sigma^2, \mathbf{t}, \kappa)$ les paramètres inconnus et à estimer dans le modèle avec $\kappa$ nœuds. Soit les nœuds $a = t_0 < t_1 < \cdots < t_\kappa < t_{\kappa+1} = b$. Pour donner les lois a priori sur $(\beta, \sigma^2, \mathbf{t}, \kappa)$, nous considérons les contraintes de forme spécifiées dans Abraham and Khadraoui (2012). Soit $S$ l'ensemble des vecteurs $\beta = (\beta_i)_{i=1}^K$ tels que $f$ respecte la contrainte de forme. Par exemple, si on impose à $f$ d'être croissante, on a :

$$S = \{f \mid \beta_i \in \mathbb{R}, \beta_1 \leq \beta_2 \leq \cdots \leq \beta_K\},$$

pour une régression unimodale concave, on a :

$$S = \{f \mid \beta_i \in \mathbb{R}, \beta_1 < \beta_2, \beta_{K-1} > \beta_K, \beta_i - 2\beta_{i-1} + \beta_{i-2} \leq 0\},$$

et pour une régression unimodale, on a :

$$S = \{f \mid \beta_i \in \mathbb{R}, \beta_1 = \beta_2 < \beta_3 \leq \cdots \leq \beta_l \geq \beta_{l+1} \geq \cdots \geq \beta_{K-1} > \beta_K, \text{pour } l = 4, \ldots, K-1\}.$$

Dans la contrainte de l'unimodalité, on restreint $l$ à l'ensemble $\{4, \ldots, K-1\}$ pour contrôler la pente de la courbe au début où $\beta_1 = \beta_2 < \beta_3$ et à la fin où $\beta_{K-1} > \beta_K$. Les distributions a priori sur $(\beta, \sigma^2, \mathbf{t}, \kappa)$ seront spécifiées comme suit. On suppose que $\beta$ et $\sigma^2$ sont a priori dépendants et que l'espérance a priori des cœfficients est $\mathbf{m} = (0, \ldots, 0)$:

$$\begin{cases} \beta|\sigma^2 \sim \mathcal{N}_K^S(\mathbf{m}, \sigma^2\mathbf{V}) \text{ de densité proportionnelle à } \frac{(2\pi\sigma^2)^{K/2}}{|\mathbf{V}|^{1/2}} \exp\{-\frac{\beta'\mathbf{V}^{-1}\beta}{2\sigma^2}\}\mathbf{1}_{\{\beta\in S\}}; \\ \sigma^2 \sim IG(\xi, \varsigma) \text{ de densité égale à } \frac{\varsigma^\xi}{\Gamma(\xi)}(\sigma^2)^{\xi+1}\exp\{-\frac{\varsigma}{\sigma^2}\}; \\ \mathbf{u}|\kappa \sim \mathcal{D}(1, \ldots, 1) \text{ de densité égale à } (\kappa-1)!; \\ \kappa \sim \mathcal{P}(\lambda) \text{ de densité égale à } \exp(\lambda)\frac{\lambda^\kappa}{\kappa!}, \end{cases} \tag{2}$$

où $\mathbf{V}$ est une matrice $K \times K$ qui sera définie, dans la proposition 3.1, en fonction des contraintes de forme imposées sur les cœfficients. Le paramètre $\mathbf{u} = (u_0, \ldots, u_\kappa)$ est relié à la position des nœuds par la relation $u_i = (t_{i+1} - t_i)/(b - a)$ où $i = 0, \ldots, \kappa$. La loi a priori $p(\beta|\sigma^2)$, où $p$ est une notation générique d'une densité de probabilité quelconque, sur les cœfficients de régression assure que la contrainte de forme est garantie par la fonction indicatrice $\mathbf{1}_{\{\beta\in S\}}$ alors que la régularité est contrôlée par l'ordre $k$ de la spline. Par l'inférence bayésienne, nous obtenons, à une constante près, la distribution a posteriori

$$\begin{aligned} p(\beta, \sigma^2, \mathbf{u}, \kappa|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\beta, \sigma^2, \mathbf{u}, \kappa)p(\beta|\sigma^2)p(\sigma^2)p(\mathbf{u}|\kappa)p(\kappa) \\ &\propto (\sigma^2)^{-(\xi^*+(K/2)+1)} \exp\Big( -\frac{(\beta - \mathbf{m}^*)'(\mathbf{V}^*)^{-1}(\beta - \mathbf{m}^*) + 2\varsigma^*}{2\sigma^2} \\ &\qquad + \lambda + \kappa\log(\lambda) - \log(\kappa!)\Big)\mathbf{1}_{\{\beta\in S\}}, \end{aligned} \tag{3}$$

où $\mathbf{m}^*, \mathbf{V}^*, \xi^*$ et $\varsigma^*$ sont obtenus par le modèle conjugué standard (normal inverse gamma) de régression.

# 3 La régression bayésienne sous contraintes

Pour estimer la fonction de régression, nous calculons le mode a posteriori et l'espérance a posteriori grâce à des simulations suivant la loi a posteriori par un algorithme de type Metropolis-Hastings à sauts réversibles. Pour pouvoir simuler suivant la distribution tronquée (3), nous devons calculer, entre autres, un rapport de prior pour les cœfficients de régression qui vérifient bien entendu la contrainte de forme ($\beta \in S$). Ce calcul revient à calculer le rapport de deux distributions normales tronquées à $S$ connues uniquement à des constantes près et éventuellement de dimension différentes. Soit $\beta^c$ un vecteur candidat proposé dans une itération au niveau des simulations. Pour simplifier la présentation du problème, nous supposons qu'en partant d'un vecteur $\beta$, nous insérons $d$ cœfficients pour obtenir $\beta^c$.

**Proposition 3.1.** *Sous le prior (2), on considère que $\beta^c$ est obtenu par une simple insertion de $d$ coefficients par rapport à $\beta$.*
*(i) Sous la contrainte de monotonie, pour une matrice $\mathbf{V}^{-1} = T_m'T_m$, on a*

$$\frac{p(\beta^c|\sigma^2)}{p(\beta|\sigma^2)} = \left(\sigma\sqrt{\frac{\pi}{2}}\right)^{(-d)} \frac{\mathbf{1}_{\{\beta_1 \leq \cdots \leq \beta_K \leq \cdots \leq \beta_{K+d}\}}}{\mathbf{1}_{\{\beta_1 \leq \cdots \leq \beta_K\}}} \exp\left(\frac{\beta'(T_m'T_m)\beta - \beta^{c\prime}(T_m^{c\prime}T_m^c)\beta^c}{2\sigma^2}\right),$$

*où $T_m$ est une matrice $K \times K$ telle que* $T_m = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{pmatrix}$ *et $T_m^c$ est une matrice $(K+d) \times (K+d)$ de la même forme que $T_m$.*
*(ii) Sous la contrainte de l'unimodalité, pour une matrice $\mathbf{V}^{-1} = T_u'T_u$, on a*

$$\frac{p(\beta^c|\sigma^2)}{p(\beta|\sigma^2)} = \left(\sigma\sqrt{\frac{\pi}{2}}\right)^{(-d)} \frac{\mathbf{1}_{\{\beta_1 \leq \cdots \leq \beta_l \geq \beta_{l+1} \geq \cdots \geq \beta_{K+d}\}}}{\mathbf{1}_{\{\beta_1 \leq \cdots \leq \beta_l \geq \beta_{l+1} \geq \cdots \geq \beta_K\}}} \exp\left(\frac{\beta'(T_u'T_u)\beta - \beta^{c\prime}(T_u^{c\prime}T_u^c)\beta^c}{2\sigma^2}\right),$$

*où $T_u$ est une matrice $K \times K$ telle que* $T_u = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & -1 \end{pmatrix}$ *et $T_u^c$ est une matrice $(K+d) \times (K+d)$ de la même forme que $T_u$.*

Nous remarquons que sous contrainte de monotonie nous obtenons les cœfficients par $\beta = T_m^{-1}\zeta = \left(\zeta_1, \zeta_1 + \zeta_2, \zeta_1 + \zeta_2 + \zeta_3, \ldots, \sum_{i=1}^{K} \zeta_i\right)'$ et sous contrainte d'unimodalité $\beta = T_u^{-1}\zeta = \left(\zeta_1, \zeta_1 + \zeta_2, \zeta_1 + \zeta_2 + \zeta_3, \ldots, \sum_{i=l}^{l} \zeta_i, \sum_{i=1}^{l} \zeta_i - \zeta_{l+1}, \ldots, \sum_{i=l}^{l} \zeta_i - \sum_{i=l+1}^{K} \zeta_i\right)'$ où $\zeta \sim \mathcal{N}_K^+(0, \sigma^2\mathbb{I}_K)$ et $\mathcal{N}_K^+(\cdot,\cdot)$ est une distribution normale tronquée aux composantes positives. Concernant le choix des nœuds, nous proposons une approche entièrement bayésienne. Partant initialement d'un vecteur de nœuds $\mathbf{t}$ à modifier, pour les positions de nœuds initiales $a = t_0 < t_1 < \cdots <$

$t_\kappa < t_{\kappa+1} = b$, on pose $n_\ell = \sum_{j=1}^{n} \mathbf{1}_{\{t_\ell \le x_{(j)} < t_{\ell+1}\}}, \ell = 0, \ldots, \kappa$. Il est clair que $\sum_{\ell=0}^{\kappa} n_\ell = n$ et que $n_\ell$ représente le nombre d'observations dans l'intervalle $[t_\ell, t_{\ell+1})$. Cette idée peut se voir comme une classification des $\{x_{(j)}\}_{j=1}^{n}$ où chaque classe est de taille $n_0, n_1, \ldots, n_\kappa$. Il est facile à ce moment de remarquer que $t_1 = x_{(n_0)}, t_2 = x_{(n_0+n_1)}, \ldots, t_\kappa = x_{(\sum_{\ell=0}^{\kappa-1} n_\ell)}$. On pose $n_{\kappa+1} = \cdots = n_{n-1} = 0$ et $\mathbf{n} = (n_0, \ldots, n_\kappa, n_{\kappa+1}, \ldots, n_{n-1})'$. Dans les simulations, nous proposons de modifier la configuration des nœuds grâce au modèle Multinomial-Dirichlet suivant

$$\begin{cases} \mathbf{n}|\mathbf{q} \sim Mult(n, \mathbf{q}), \\ \mathbf{q} \sim \mathcal{D}(\alpha_0, \ldots, \alpha_{n-1}), \end{cases} \tag{4}$$

où pour $\alpha_\ell$ deux choix sont proposés $\alpha_\ell = \frac{n_\ell+1}{n}$ et $\alpha_\ell = n_\ell + 1$ et $n_\ell$ est la $\ell^e$ composante de $\mathbf{n}$ à l'étape précédente de la chaîne. Les contraintes de forme seront vérifiées à l'étape de proposition $\mathcal{P}_S^\eta$ et à l'itération $\eta$ dans le Metropolis-Hastings à sauts réversibles comme on décrit dans la suite. La proposition d'un vecteur de cœfficients $\beta^c$ s'effectue par l'insertion de $d$ cœfficients à partir de $\beta^\eta$ ou bien par la suppression de $d$ cœfficients à partir de $\beta^\eta$. Par exemple, pour proposer l'insertion de $d$ cœfficients on modifie légérement le résultat théorique de l'insertion des nœuds donné dans de Boor (2001)

$$\beta_i' = \begin{cases} \beta_i & \text{si} \quad \bar{t}_i^* < \bar{t}_l^*; \\ \omega_{i,k}(\bar{t})\beta_i + \Big(1 - \omega_{i,k}(\bar{t})\Big)\beta_{i-1} & \text{si} \quad \bar{t}_i^* = \bar{t}_l^*; \\ \beta_{i-1} & \text{si} \quad \bar{t}_i^* > \bar{t}_l^*, \end{cases} \tag{5}$$

où la fonction $\omega_{i,k}(x) = \frac{x - t_i}{t_{i+k-1} - t_i}$ si $t_i < t_{i+k-1}$ et 0 sinon et $l$ est l'indice du nouveau cœfficient inséré qu'on détermine simplement à partir de l'indice du nouveau nœud inséré $\bar{t}$, $\bar{t}_i^* = (\bar{t}_{i+1} + \cdots + \bar{t}_{i+k-1})/(k-1)$ et $(\bar{t}_{k+1}, \ldots, \bar{t}_{k+\kappa+1})' = (t_1, \ldots, \bar{t}, \ldots, t_\kappa)'$ avec $\bar{t}_1 = \cdots = \bar{t}_k = a$ et $\bar{t}_{k+\kappa+2} = \cdots = \bar{t}_{2k+\kappa+1} = b$. Nous désignons par $(l_1, \ldots, l_d)$ les indices des $d$ cœfficients qu'on propose d'insérer à l'itération $\eta + 1$. Après le calcul des cœfficients $\beta_{l_1}', \ldots, \beta_{l_d}'$ par (5), la prise en compte d'une contrainte de forme quelconque $S$ sera réalisée lors de l'étape de proposition de la façon suivante:

$(i)$ on pose $\beta' = (\beta_1', \ldots, \beta_{l_1}', \ldots, \beta_{l_d}', \ldots, \beta_{K+d}')'$;

$(ii)$ $\widetilde{\beta}_{l_1} \sim \mathcal{U}_{\{S(\beta', l_1) \cap [\beta_{l_1}' \pm \epsilon']\}}$ où $S(\beta', l_1) = \{\widetilde{\beta}_{l_1} : (\beta_1', \ldots, \beta_{l_1-1}', \widetilde{\beta}_{l_1}, \ldots, \beta_{K+d}')' \in S\}$;
$\widetilde{\beta}_{l_2} \sim \mathcal{U}_{\{S(\beta', l_2) \cap [\beta_{l_2}' \pm \epsilon']\}}$ où $S(\beta', l_2) = \{\widetilde{\beta}_{l_2} : (\beta_1', \ldots, \beta_{l_2-1}', \widetilde{\beta}_{l_2}, \ldots, \beta_{K+d}')' \in S\}$;
$\vdots$
$\widetilde{\beta}_{l_d} \sim \mathcal{U}_{\{S(\beta', l_d) \cap [\beta_{l_d}' \pm \epsilon']\}}$ où $S(\beta', l_d) = \{\widetilde{\beta}_{l_d} : (\beta_1', \ldots, \beta_{l_d-1}', \widetilde{\beta}_{l_d}, \ldots, \beta_{K+d}')' \in S\}$;

$(iii)$ $\beta^c = (\beta_1', \ldots, \widetilde{\beta}_{l_1}, \ldots, \widetilde{\beta}_{l_2}, \ldots, \widetilde{\beta}_{l_d}, \ldots, \beta_{K+d}')'$,

où la constante $\epsilon'$ permet de contrôler la variance de la proposition. La valeur initiale $\beta^1$ est choisie selon un examen visuel des données $\boldsymbol{y} = (y_1, \ldots, y_n)'$. On propose d'estimer $f$ par le mode a posteriori. En effet, contrairement à l'espérance a posteriori, le mode a posteriori vérifie nécessairement les contraintes de forme. Les simulations réalisées à partir de l'algorithme de type Metropolis-Hastings à sauts réversibles permettent de calculer le mode a posteriori et l'espérance a posteriori et de les comparer par une étude numérique de simulation.

# References

C. Abraham and K. Khadraoui. Bayesian regression with b-splines under shape and smoothness constraints. *Soumis*, pages 1–24, 2012.

G. Claeskens, T. Krivobokova, and J. D. Opsomer. Asymptotic properties of penalized spline estimators. *Biometrika*, 96:529–544, 2009.

C. de Boor. B(asic)-spline basics. *Computer Aided Geometric Design (CAGD) Handbook*, 1: 1–34, 2001.

D. Denison, B. K. Mallick, and A. F. M. Smith. Automatic bayesian curve fitting. *Journal of the Royal Statistical Society: Series B*, 60:333–350, 1998.

I. DiMatteo, C. R. Genovese, and R. E. Kass. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88:1055–1071, 2001.

P. H. C. Eilers and B. D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–102, 1996.

T. J. Hastie. Pseudosplines. *Journal of the Royal Statistical Society*, 58:379–396, 1996.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of American Statistical Association*, 90: 773–795, 1995.

G. Kauermann, T. Krivobokova, and L. Fahrmeir. Some asymptotic results on generalized spline smoothing. *Journal of the Royal Statistical Society*, 71:487–503, 2009.

C. Kelly and J. Rice. Monotone smoothing with application to dose response curves and the assessment of synergism. *Biometrics*, 46:1071–1085, 1990.

F. Leitenstorfer and G. Tutz. Knot selection by boosting techniques. *Computational statistics and data analysis*, 51:4605–4621, 2007.

F. O'Sullivan. A statistical perspective on iii-posed inverse problems. *Statistical Science*, 1: 502–518, 1986.

R. L. Parker and J. A. Rice. Discussion of some aspects of the spline smoothing approach to non-parametric regression curve fitting by silverman. *Journal of the Royal Statistical Society*, 47:40–42, 1985.

D. Ruppert. Selecting the number of knots for penelized splines. *Journal of computational and graphical statistics*, 11:735–757, 2002.

C. J. Stone, M. H. Hansen, C. Kooperberg, and Y. K. Truong. Polynomial splines and their tensor products in extended linear modeling. *The annals of statistics*, 25:689–705, 1997.

M. P. Wand. Smoothing and mixed models. *Computational statistics*, 18:223–249, 2003.