

# BAYESIAN SMOOTHING FOR MEASUREMENT ERROR PROBLEMS

Scott M. Berry<sup>1</sup>, Raymond J. Carroll<sup>2</sup> and David Ruppert<sup>3</sup>

<sup>1</sup>*Berry Consultants*

*1039 Wellington Court*

*Sycamore, IL 60178 USA*

scott@berryconsultants.com

<sup>2</sup>*Department of Statistics*

*Texas A&M University*

*College Station TX 77843-3143 USA*

carroll@stat.tamu.edu

<sup>3</sup>*School of Operations Research and Industrial Engineering*

*Cornell University*

*Ithaca, NY 14853-3801 USA*

davidr@orie.cornell.edu

**Abstract** In the presence of covariate measurement error, estimating a regression function nonparametrically is extremely difficult, the problem being related to deconvolution. In this paper we describe Bayesian approaches to modeling a flexible regression function when the predictor variable is measured with error. The regression function is modeled with smoothing splines. We provide simulations with several nonlinear regression functions. Our simulations indicate that the *frequentist* mean squared error properties of the fully Bayesian method are better than those of previously proposed frequentist methods, at least in the examples we have studied.

**Keywords:** Bayesian methods, Errors in variables model, Functional relationship, Generalized linear models, Kernel regression, Measurement error, Non-parametric methods, Structural relationship.

## 1. Introduction

In this paper we review work of [1] on a fully Bayesian approach to the problem of nonparametric regression when the independent variables are measured with error. This is known to be an extremely difficult problem in terms of global rates of convergence.

In [2] the assumption of global consistency is relaxed. They suggest two estimators: (a) a semiparametric one based on the SIMEX method of [3] which makes no assumptions about the unknown and unobserved covariates; and (b) a more parametric one that assumes that the unobserved covariates follow a mixture of normals distribution.

In [1], we proposed a new method for nonparametric function estimation when the covariate is measured with error. Our procedure can be looked at as the natural fully Bayesian extension of the techniques of [2]. It can also be looked upon as the extension to measurement error models of the MCMC technique of [7], or, viewed more broadly, the entire Bayesian formulation of smoothing splines (e.g., [10], [11]; [8], [9]).

The methodology in [1] and reviewed here is new in two respects. First, the adjustment for bias due to measurement error comes automatically from the Bayesian machinery. In contrast, other methods explicitly analyze the bias and devise a correction in a more ad hoc fashion. Second, and perhaps more importantly, the smoothing parameter selector, which also comes automatically from the Bayesian approach, is designed for the measurement error problem. Earlier work either did not propose a smoothing parameter selector or applied a smoothing parameter selector that ignores the effects of measurement error. However, measurement error has large effects on both bias and variance and a smoothing parameter that is optimal for correctly measured covariates may be far from optimal in the presence of measurement error.

In Section 2 we describe some background information on smoothing splines that is necessary for our development. In Section 3, we present our methodology. Section 4 presents simulations. The results indicate that even as a frequentist estimator our fully Bayesian method is at least competitive with that of [2], and sometimes much better. Section 5 presents a discussion of the results. Paper [1] should be consulted for more details.

## 2. Smoothing Splines

We present a brief introduction to smoothing splines. For additional information, see [10], [12], [6], [7], and [4].

Assume that  $Y_i = m(X_i) + \epsilon_i$ , where  $\epsilon_i$  has mean zero and variance  $\sigma_\epsilon^2$ . Let  $[a, b]$  be the interval for which an estimate of  $m$  is sought. Let  $g$

be the best natural cubic spline (NCS) approximator of  $m$ , that is, the NCS that minimizes  $\sum_{i=1}^n \{m(X_i) - g(X_i)\}^2$ . If  $m$  is smooth, then the error in approximating  $m$  by  $g$  typically is negligible compared to the estimation error, so we assume that  $m = g$ .

A *smoothing spline* is defined as the minimizer over  $g$  of the penalized sum of squares

$$S(g) = \sum_{i=1}^n \{Y_i - g(X_i)\}^2 + \alpha \int_a^b \{g''(x)\}^2 dx, \quad (1)$$

for  $\alpha > 0$ . This minimizer is a NCS with knots at the distinct  $X_i$  values. The integral term of (1) is a roughness penalty and  $\alpha$ , is the smoothing parameter. Let  $\mathbf{g} = \{g(X_1), g(X_2), \dots, g(X_n)\}^\top$ . The penalty term can be written as  $\alpha \int_a^b \{g''(x)\}^2 dx = \alpha \mathbf{g}^\top \mathbf{K} \mathbf{g}$ , where  $\mathbf{K}$  is an  $n \times n$  dimensional matrix of rank  $n - 2$ , defined in [4]. The smoothing spline minimizing  $S(g)$  is  $\hat{\mathbf{g}} = \mathbf{A}(\alpha) \mathbf{Y}$ , where  $\mathbf{A}(\alpha) = (\mathbf{I} + \alpha \mathbf{K})^{-1}$  and  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ . The vector  $\hat{\mathbf{g}}$  uniquely defines the smoothing spline.

The Bayesian approach to smoothing splines gives the vector  $\mathbf{g}$  a prior density proportional to the ‘‘partially improper’’ Gaussian process:

$$(\alpha/2\sigma_\epsilon^2)^{M/2} \exp \left\{ -(\alpha/\sigma_\epsilon^2) \mathbf{g}^\top \mathbf{K} \mathbf{g} \right\}, \quad (2)$$

where  $M = n - 2$  and  $\mathbf{K}$  is defined as above. Although both  $\mathbf{K}$  and  $\mathbf{g}$  depend on the knot locations, since  $\mathbf{g}^\top \mathbf{K} \mathbf{g} = \int_a^b \{g''(x)\}^2 dx$  this prior is independent of the knot locations. If the observations  $Y_i$  are independent and normally distributed with mean  $g(X_i)$  and variance  $\sigma_\epsilon^2$  then the posterior distribution for  $\mathbf{g}$  is multivariate normal with mean  $\hat{\mathbf{g}} = \mathbf{A}(\alpha) \mathbf{y}$ , and covariance matrix  $\sigma_\epsilon^2 \mathbf{A}(\alpha)$ .

### 3. General Model

We consider the following measurement error model,

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (3)$$

where the  $\epsilon_i$  are independent normal random variables with mean 0 and variance  $\sigma_\epsilon^2$ . The  $X$ 's are not observable, i.e., they are latent variables, but  $W$  that are surrogates for the  $X$ s are observed;

$$W_{ij} = X_i + U_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (4)$$

where the  $U_{ij}$  are independent normal errors with mean 0 and variance  $\sigma_u^2$ .

Model (4) is more general than it first looks. It can be interpreted as stating that a *known* function of the observed covariates ( $W$ ) is the

same function of the latent variables, plus independent, homoscedastic normally distributed measurement errors. We have written (4) as if the function were the identity function, but it could be anything, e.g., the logarithm. The reason for this generality is that in (3), the function  $m(\cdot)$  is unknown, so that for example if  $m_*(v) = m\{\exp(v)\}$ , then  $m_*\{\log(x)\} = m(x)$ .

Denote  $\boldsymbol{\theta} = (\mathbf{g}, \mathbf{X}, \sigma_\epsilon^2, \sigma_u^2, \alpha)$ . The posterior density is

$$f(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{W}) \propto f(\mathbf{Y}|\mathbf{g}, \mathbf{X}, \sigma_\epsilon^2) f(\mathbf{W}|\mathbf{X}, \sigma_u^2) f(\mathbf{g}|\alpha) f(\sigma_u^2) f(\sigma_\epsilon^2) f(\mathbf{X}) f(\alpha). \quad (5)$$

### 3.1 Fully Bayesian Approach

In this section the fully Bayesian approach to this problem is developed. Prior distributions are placed on all parameters, including the structural parameters  $(\mu_x, \sigma_x^2)$  and the variance components  $(\sigma_\epsilon^2, \sigma_u^2)$ , and the joint posterior distribution is calculated. One of the benefits of this approach is that observations of the smoothing spline are generated from the posterior and therefore we estimate the entire posterior distribution of  $g$ , not just its mode. Thus, various forms of “error bars” are straightforward to calculate. These credible sets take account of the measurement error of the independent variables and the use of a data-based smoothing parameter.

The method is as follows. Without loss of generality, we replace  $\alpha/\sigma_\epsilon^2$  by  $\gamma$ . The prior distributions for  $\sigma_\epsilon^2$  and  $\sigma_u^2$  are inverse-gamma distributions and the prior distribution for  $\gamma$  is a gamma distribution:  $\sigma_\epsilon^2 \sim IG(A_\epsilon, B_\epsilon)$ ,  $\sigma_u^2 \sim IG(A_U, B_U)$ , and  $\gamma \sim G(A_\gamma, B_\gamma)$ . We use the following definitions of the inverse-gamma and gamma distributions:

$$f(x|A, B) = \frac{1}{\Gamma(A)B^A x^{A+1}} \exp\left(-\frac{1}{Bx}\right) I_{(0, \infty)}(x).$$

$$f(x|A, B) = \frac{1}{\Gamma(A)B^A} x^{A-1} \exp\left(-\frac{x}{B}\right) I_{(0, \infty)}(x).$$

For the simulations and examples in this paper we use a hierarchical Bayes approach. A normal distribution with mean  $\mu_x$  and variance  $\sigma_x^2$  is used, where  $\mu_x \sim \text{Normal}(d_x, t_x^2)$  and  $\sigma_x^2 \sim IG(A_x, B_x)$ .

The hyperparameters that are fixed a priori and thus are “tuning constants” are denoted by Roman fonts. They are  $A_Y, B_Y, A_U, B_U, A_\gamma, B_\gamma, d_x, t_x^2, A_x$ , and  $B_x$ .

Assuming the hierarchical normal structure for  $[X]$ , the joint posterior is proportional to

$$\begin{aligned}
 & \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n \{Y_i - g(X_i)\}^2 - \frac{1}{2\sigma_u^2} \sum_{i=1}^n \sum_{j=1}^{m_i} (W_{ij} - X_i)^2 \right. \\
 & \left. - \frac{1}{2\sigma_x^2} \sum_{i=1}^n (X_i - \mu_x)^2 - \frac{1}{2t_x^2} (\mu_x - d_x)^2 \right\} \\
 & \times \exp \left\{ -(\gamma/2) \mathbf{g}^\top \mathbf{K} \mathbf{g} - \frac{1}{B_\epsilon \sigma_\epsilon^2} - \frac{1}{B_U \sigma_u^2} - \frac{\gamma}{B_\gamma} - \frac{1}{B_x \sigma_x^2} \right\} \quad (6) \\
 & \times \sigma_\epsilon^{-2(n/2+A_\epsilon+1)} \sigma_u^{-2(\frac{1}{2} \sum_{i=1}^n m_i + A_U + 1)} \sigma_x^{-2(n/2+A_x+1)} \gamma^{(A_\gamma + M/2 - 1)},
 \end{aligned}$$

where  $M = n - 2$  as in (2). The sampling is done using a successive substitution algorithm ([5]).

The complete conditional distributions for the parameters are

$$\mathbf{g} | \mathbf{X}, \gamma, \sigma_\epsilon^2, \mathbf{Y}, \mathbf{W} \sim \text{Normal} \{ \mathbf{A}(\sigma_\epsilon^2 \gamma) \mathbf{y}, \sigma_\epsilon^2 \mathbf{A}(\sigma_\epsilon^2 \gamma) \};$$

$$\begin{aligned}
 X_i | \mathbf{W}_i, \mathbf{g}, \sigma^2, \sigma_u^2, \mathbf{Y}, \mathbf{W} & \propto \exp \left[ -\frac{1}{2\sigma_u^2} \sum_{j=1}^{m_i} (W_{ij} - X_i)^2 \right. \\
 & \left. - \frac{1}{2\sigma_\epsilon^2} \{Y_i - g(X_i)\}^2 - \frac{1}{2\sigma_x^2} (X_i - \mu_x)^2 \right]; \quad (7)
 \end{aligned}$$

$$\sigma_\epsilon^2 | \mathbf{g}, \mathbf{X}, \mathbf{Y}, \mathbf{W} \sim IG(A_\epsilon + n/2, [1/B_\epsilon + (1/2) \sum_{i=1}^n \{Y_i - g(X_i)\}^2]^{-1});$$

$$\gamma | \mathbf{g}, \mathbf{X} \sim G(A_\gamma + \frac{M}{2}, [1/B_\gamma + \frac{1}{2} \mathbf{g}^\top \mathbf{K} \mathbf{g}]^{-1});$$

$$\mu_x | \mathbf{X} \sim \text{Normal} \{ (n\bar{X}t_x + d_x \sigma_x^2) / (nt_x^2 + \sigma_x^2), \sigma_x^2 t_x^2 / (nt_x^2 + \sigma_x^2) \};$$

$$\sigma_x^2 | \mathbf{X} \sim IG[A_x + n/2, \{B_x^{-1} + (1/2) \sum_{i=1}^n (X_i - \mu_x)^2\}^{-1}];$$

$$\sigma_u^2 | \mathbf{X} \sim IG(A_U + (1/2) \sum_{i=1}^n m_i,$$

$$[1/B_U + (1/2) \sum_{i=1}^n \sum_{j=1}^{m_i} (W_{ij} - X_i)^2]^{-1}).$$

Observations from each of the complete conditionals are drawn iteratively in the order presented above. The generation of an observation of  $\mathbf{g}$  is computationally difficult for smoothing splines because they have  $n$  knots. Because the values of  $\mathbf{X}$ ,  $\mathbf{K}$  (for smoothing splines) and  $\gamma$  are continually changing in the algorithm the matrix  $\mathbf{A}(\sigma_\epsilon^2\gamma)$  (which is  $n \times n$ ) and its inverse must be recomputed for each iteration of the MCMC algorithm. Reference [7] discusses an algorithm for generating observations of  $\mathbf{g}$  in  $O(n)$  operations.

The complete conditionals for the  $X_i$ 's require a Metropolis-Hastings step. This is done by generating a candidate observation of  $X_i$  from a normal distribution with a mean of the current value of  $X_i$  and a standard deviation of  $2\sigma_u^{(i)}/\sqrt{m_i}$ , where  $\sigma_u^{(i)}$  is the current value of  $\sigma_u$  in the MCMC algorithm. Using  $\bar{W}_i$  as an estimate of  $X_i$ , without the information in the regression function, has a standard error of  $\sigma_u^{(i)}/\sqrt{m_i}$ . Using the rule-of-thumb of a candidate value having a standard deviation twice the standard deviation of the marginal posterior provides a conservative candidate distribution for  $X_i$ . In terms of efficiency of the Metropolis-Hastings step, in our experience it is better to overestimate this standard deviation than to underestimate it. The evaluation of the complete conditional for  $X_i$  is computationally straightforward.

Generating observations from each of the other complete conditionals is straightforward and fast. Because the position of  $\mathbf{X}$  changes throughout the algorithm, when using smoothing splines, we keep track of the value of  $g$  at a uniformly distributed grid of points. For each realization of  $g$  in the sampler the value of  $g$  for each grid point is recorded. This enables us to keep track of pointwise moments and percentiles.

Having observations from the joint posterior distribution provides a powerful tool for inference. The pointwise mean curve is a natural estimate of the regression mean function  $m$ . Pointwise credible intervals can also be calculated very easily from the observations of  $\mathbf{g}$ . Functions (linear or non-linear) of the regression function can also be estimated, along with standard errors. This is the approach used by [11] in non-measurement error cases and [7] in non-measurement error semi-parametric models. The work in [11] predated the revolution in Bayesian computations and she treats the smoothing parameter as fixed. Reference [7] uses the Gibbs sampler to adjust the credible sets for uncertainty in the variance components that define the smoothing parameter. In the present work, use of the Gibbs sampler also adjusts the credible sets for measurement error.

For an example we consider simulated data. The data consist of 100  $X$ 's generated from a standard normal distribution. The responses,  $Y$ , were generated from a normal distribution with a standard deviation of

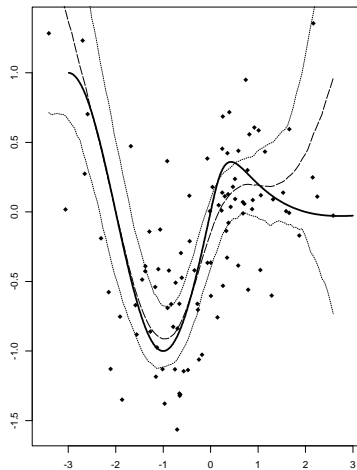


Figure 1. An example of the fully Bayesian Spline. The solid curve is the true regression function. The dashed middle curve is the mean of the posterior of the regression function. The dotted error bars represent the piecewise 90% credible intervals.

0.3 and a mean function of

$$m(x) = \frac{\sin(\pi x/2)}{1 + 2x^2\{\text{sign}(x) + 1\}}. \tag{8}$$

Each  $m_i = 2$  and the  $W_{ij}$ , for  $j = 1, 2$ , are normally distributed with a mean of  $X_i$  and a standard deviation of 0.8. A pointwise posterior mean curve is used for the estimate of  $m$ . Credible curves are calculated by interpolating the pointwise  $100(1 - \alpha)\%$  credible intervals. A burn-in time of 500 observations is used with 1000 observations from the posterior. Figure 1 shows the estimate of  $g$  and the 90% pointwise credible curves.

There are at least two possible methods for choosing the smoothing parameter for a smoothing spline. We place a prior distribution on  $\gamma$ . Reference [7] uses an identical procedure within semi-parametric models. It is worth noting that by placing a continuous density prior on  $\gamma$ , we have automatically given zero prior probability to the possibility of doing no smoothing at all. This is an automatic way of avoiding the possibility of gross undersmoothing that caused so much trouble for the methods of [2].

## 4. Simulations

### 4.1 Basic Simulations

We performed a series of simulations to compare our methods with those of [2]. In each case, 200 simulated data sets were generated, with  $X_i$  generated as independent normal random variables with mean  $\mu_x$  and variance  $\sigma_x^2$ , with two replicates ( $m_i = 2$ ), and with  $\epsilon_i$  also normally distributed. In each simulation, for the fully Bayesian method, the following prior distributions are used:  $\sigma_\epsilon^2 \sim IG(1, 1)$ ,  $\sigma_u^2 \sim IG(1, 1)$ ,  $\gamma \sim \text{Gamma}(3, 1000)$ ,  $\mu_x \sim N(0, 10^2)$ , and  $\sigma_x^2 \sim IG(1, 1)$ . These priors were selected because of their relative flexibility. They are all proper, yet they are not strong, in the sense of bringing a lot of information to the problem. We found the results insensitive to moderate modifications of these priors. The flexibility of these priors are demonstrated by their success in the different regression functions used in the simulations.

For purposes of bias and mean squared error calculations, the smoothing spline estimates of  $g$  were computed on a grid of 101 points in the interval  $[a, b]$ , the interval chosen to contain most of the distribution for  $X$ . The mean squared biases and mean squared errors were computed over this grid.

The five cases considered were the following:

Case 1: The regression function,  $m$  is given in (8), with  $n = 100$ ,  $a = -2.0$ ,  $b = 2.0$   $\sigma_\epsilon^2 = 0.3^2$ ,  $\sigma_u^2 = 0.8^2$ ,  $\mu_x = 0$  and  $\sigma_x^2 = 1$ .

Case 2: Same as Case 1 except  $n = 200$ .

Case 3: A modification of Case 1 above except that  $n = 500$ .

Case 4: Case 1 of [2], so that  $m(x) = 1000x_+^3(1-x)_+^3$ ,  $x_+ = xI(x > 0)$ , with  $n = 200$ ,  $a = 0.1$ ,  $b = 0.9$   $\sigma_\epsilon^2 = 0.0015^2$ ,  $\sigma_u^2 = (3/7)\sigma_x^2$ ,  $\mu_x = 0.5$  and  $\sigma_x^2 = 0.25^2$ .

Case 5: A modification of Case 4 of [2], so that  $m(x) = 10\sin(4\pi x)$ , with  $n = 500$ ,  $a = 0.1$ ,  $b = 0.9$   $\sigma_\epsilon^2 = 0.05^2$ ,  $\sigma_u^2 = 0.141^2$ ,  $\mu_x = 0.5$  and  $\sigma_x^2 = 0.25^2$ .

The methods compared were the following:

- naive smoothing spline fit ignoring measurement error;
- fully Bayesian approach;
- Structural method ([2], 5 knots).
- Structural method, 15 knots.

Table 1 presents summary results for mean squared error. The striking feature of this table is that our Bayesian estimator has at least as good

Mean Squared Error $\times 10^2$								
Method	Case 1	2	3	4	5	6	7	8
Naive	6.91	5.57	5.38	11.55	3.79	5.77	5.84	16.48
Bayes	<b>2.84</b>	<b>1.56</b>	<b>1.47</b>	<b>1.95</b>	1/03	<b>2.69</b>	<b>2.49</b>	<b>7.41</b>
Structural(5)	8.17	3.82	1.73	2.17	2.03	7.27	7.91	16.84
Structural(15)	9.90	5.40	1.85	2.37	<b>0.80</b>	6.94	9.91	20.22

Table 1. The mean squared error for the simulation. “Naive” is the naive smoothing spline, “Bayes” is the fully Bayesian method, and “Structural(m)” is the Structural regression spline of [2] with  $m$  knots. In each column, the smallest MSE values is in boldface.

*frequentist* properties as the frequentist methods. In cases 1 and 2, it clearly dominates, having less than half the mean squared error of the other methods. In cases 3, its mean squared error efficiency is 20% greater than the structural spline with 15 knots, while it is only 25% less efficient in Case 5. The improvement of the fully Bayesian method over the frequentist methods is especially large for smaller sample sizes, e.g., Cases 1 and 6 (see below) where  $n = 100$ .

Clearly, even this limited simulation suggests that our Bayesian method is at least competitive with other methods proposed previously in the literature.

## 5. Discussion

The Bayesian approach to measurement error, modeling the mismeasured variables as latent random variables, and integrating them out, is a powerful one. In this paper we have developed a Bayesian method for nonparametric regression in the presence of measurement error. By modeling a smoothing spline from a Bayesian point of view we create algorithms to calculate the posterior distribution of the regression function. The resulting estimate accounts for the effects of measurement error both upon the estimator and upon the smoothing parameter. The resulting smoothing parameter selector appears to be the first that adjusts for the effects of measurement error.

The simulations demonstrate the flexibility of the fully Bayes approach, and even its efficiency in the frequentist sense, at least in the examples we have investigated. The fully Bayes approach also enable inference on more than just the regression function.

We believe that the reason that the fully Bayesian approach works better than previous proposals is this: often one can estimate the unknown  $X_i$  significantly more accurately using *all* information in the data

about  $X_i$  rather than by using just the  $W_{ij}, j = 1, \dots, n_i$ , see [1] for details. Clearly, this leaves open the possibility that with highly nonnormal errors, or highly heteroscedastic ones, the misspecified information from our simple model will lead to bias or other deleterious behavior in the Bayesian method.

## Acknowledgments

Carroll's research was supported by a grant from the National Cancer Institute (CA-57030), and by the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-ES09106). Ruppert's research was supported by NSF Grant DMS-9804058.

## References

- [1] S. M. Berry, R. J. Carroll, and D. Ruppert. Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 96:to appear, 2001.
- [2] R. J. Carroll, J. D. Maca, and D. Ruppert. Nonparametric regression with errors in covariates. *Biometrika*, 86:541-554, 1999.
- [3] J. R. Cook and L. A. Stefanski. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89:1314-1328, 1994.
- [4] R. L. Eubank. *Nonparametric Regression and Spline Smoothing (Second Edition)*. Marcel Dekker, Inc. New York, 1999.
- [5] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398-409, 1990.
- [6] P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London, 1994.
- [7] T. Hastie and R. Tibshirani. Bayesian backfitting. *Statistical Science*, 15:193-223, 2000.
- [8] D. Nychka. Bayesian confidence intervals for a smoothing spline. *Journal of the American Statistical Association*, 83:1134-1143, 1988.
- [9] D. Nychka. The average posterior variance of a smoothing spline and a consistent estimate of the average squared error. *Annals of Statistics*, 18:415-428, 1990.
- [10] G. Wahba. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B*, 40:364-372, 1978.
- [11] G. Wahba. Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, 45:133-150, 1983.
- [12] G. Wahba. *Spline models for observational data*. SIAM Press, Providence, 1990.