Discussion of
"Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves" by Brumback and Rice

Yuedong Wang and Grace Wahba *

December 12, 1997

In this stimulating article, Brumback and Rice (BR) propose *linear* mixed effects models for curved data using the smoothing spline structure. We wish to complement this interesting paper by offering an alternative strategy developed in Wang (1998a). We first derive an smoothing spline analysis of variance (SS ANOVA) decomposition for the progesterone example in BR. We then describe a general class of *nonparametric* mixed effects models. Comparisons of our models with those in BR are made at various places.

1 SS ANOVA Decompositions

SS ANOVA decompositions are a general technique for building nonparametric fixed (Wahba 1990, Gu and Wahba 1991, Gu and Wahba 1993a, Wahba, Wang, Gu, Klein and Klein 1995, Gu 1997) and mixed (Wang 1998a) effects models with multiple factors. They construct functional spaces with certain modular structures that parallel the classical analysis of variance decompositions.

Instead of presenting the theory behind these decompositions, we illustrate how to use this technique with the same progesterone example in BR. Ideally, such an experiment is conducted in a way that women in a particular group are randomly selected from a population for that group. Cycles of a particular woman are randomly selected from a population of all possible cycles of that woman. The inferences are usually on the populations, rather than on individual women and cycles. Nevertheless, predictions of individual women and cycles can also be derived from our models below. The goal is to build statistical models that account for variations across time, between groups, between women within a group and between cycles within a woman. Hence there are four factors: `time` (t), `group` (g), `woman` (w) and `cycle` (c). From the design of the experiment, the `time` and `group` factors are fixed;

*Yuedong Wang is Assistant Professor in the Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106. Grace Wahba is Bascom Professor, Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin 53706. This research was supported by NIH Grant R01 EY09946 and NSF Grant DMS-9704758.

the `woman` and `cycle` factors are random. Furthermore, the `woman` factor is nested within `group` and the `cycle` factor is nested within `woman`.

For simplicity of the notation, denote the time interval as $[0,1]$. Denote G as the number of groups (for generality) and $\mathcal{G}=\{1,\cdots,G\}$. For every group g, denote $\mathcal{W}_g$ as the population from which women in group g are sampled with sampling distribution $P_{w|g}$. For every woman $w\in\mathcal{W}_g$, denote $\mathcal{C}_{g,w}$ as the population from which cycles are sampled with sampling distribution $P_{c|g,w}$. Let $m(t,g,w,c)$ be the "true" mean of woman w in group g at time t during cycle c, where w and c are successively chosen according to $P_{w|g}$ and $P_{c|g,w}$. Assume observations are generated by

$$y_{tgwc} \;=\; m(t,g,w,c)+\epsilon_{tgwc}, \qquad t\in[0,1];\quad g\in\mathcal{G};\quad w\in\mathcal{W}_g;\quad c\in\mathcal{C}_{g,w}, \tag{1}$$

where $\epsilon_{tgwc}\overset{iid}{\sim}\mathrm{N}(0,\sigma^2)$ and are independent of $m(t,g,w,c)$. Both the number of observations and the design points of `time` are specific to a cycle and they are not required to be the same for different cycles.

To prepare for the SS ANOVA decomposition with all factors, let us first consider an SS ANOVA decomposition with fixed factors (`time` and `group`) only. The domains of `time` and `group` factors are $[0,1]$ and $\mathcal{G}$. We want to model the `time` effect by a cubic spline and the `group` effect by an one-way ANOVA model. Specifically, for the `time` effect, we use the reproducing kernel Hilbert space (r.k.h.s.)

$$W_2=\{f:\ f \text{ and } f' \text{ are absolutely continuous}, \int_0^1 (f''(t))^2dt<\infty\}$$

and decompose it into $W_2=\{1\}\oplus\{t-0.5\}\oplus H_1$, where $\{1\}$ and $\{t-0.5\}$ represent linear spaces spanned by the constant 1 and linear basis $t-0.5$. H_1 is the orthogonal complement of $\mathrm{span}\{1,t-0.5\}$ in W_2. H_1 is a r.k.h.s. with reproducing kernel (r.k.) $R_1(s,t)=k_2(s)k_2(t)-k_4(s-t)$, where $k_\nu(x)=B_\nu(x)/\nu!$ and $B_\nu(\cdot)$ is the νth Bernoulli polynomial (Wang 1997). For the `group` effect, as in the classical one-way ANOVA model, we use the Euclidean G-space R^G and decompose it into $R^G=\{1\}\oplus H_2$, where $\{1\}=\{z\in R^G:\ z(1)=\cdots=z(G)\}$ and $H_2=\{z\in R^G:\ \sum_{i=1}^G z(i)=0\}$. It is easy to check that H_2 is a r.k.h.s. with r.k. $R_2(i,j)=I_{[i=j]}-1/G$. To model `time` and `group` effects jointly, we use the tensor product r.k.h.s. (Gu and Wahba 1993a)

$$\begin{aligned} & W_2\otimes R^G \\ =\; & [\{1\}\oplus\{t-0.5\}\oplus H_1]\otimes[\{1\}\oplus H_2] \\ =\; & \{1\}\oplus\{t-0.5\}\oplus H_1\oplus H_2\oplus(\{t-0.5\}\otimes H_2)\oplus(H_1\otimes H_2), \end{aligned} \tag{2}$$

where $H_1\otimes H_2$ is a r.k.h.s. with r.k.

$$R((s,i),(t,j))=R_1(s,t)R_2(i,j). \tag{3}$$

Define operators A_1 and A_2 on $[0,1]$ such that $A_1m(t,g,w,c)=\int_0^1 m(s,g,w,c)ds$ and $A_2m(t,g,w,c)=(\int_0^1 m'(s,g,w,c)ds)(t-0.5)$. For a fixed woman w in group g, define an averaging operator A_5 on $\mathcal{C}_{g,w}$ such that $A_5m(t,g,w,c)=\int_{\mathcal{C}_{g,w}} m(t,g,w,v)dP_{v|g,w}$. $A_5m(t,g,w,c)$ is the average response at time t of all cycles of woman w in group g and depends on

t, g and w only. For a fixed group g, define an averaging operator A_4 on $\mathcal{W}_g$ such that $A_4 m(t,g,w,c) = \int_{\mathcal{W}_g} A_5 m(t,g,u,c) dP_{u|g}$. $A_4 m(t,g,w,c)$ is the average response at time t of all women in group g and depends on t and g only. Define an averaging operator A_3 on $\mathcal{G}$ such that $A_3 m(t,g,w,c) = \sum_{i=1}^{G} A_4 m(t,i,w,c)/G$. $A_3 m(t,g,w,c)$ is the average response at time t of all groups and depends on t only. Note that the operator A_4 depends on g and the operator A_5 depends on g and w, which is not expressed explicitly for simplicity of the notation. The SS ANOVA decomposition for all factors is

$$
\begin{aligned}
& m \\
= \; & [A_1 + A_2 + (I - A_1 - A_2)][A_3 + (A_4 - A_3) + (A_5 - A_4) + (I - A_5)]m \\
= \; & A_1 A_3 m + A_2 A_3 m + (I - A_1 - A_2) A_3 m + \\
& A_1(A_4 - A_3)m + A_2(A_4 - A_3)m + (I - A_1 - A_2)(A_4 - A_3)m + \\
& A_1(A_5 - A_4)m + A_2(A_5 - A_4)m + (I - A_1 - A_2)(A_5 - A_4)m + \\
& A_1(I - A_5)m + A_2(I - A_5)m + (I - A_1 - A_2)(I - A_5)m \\
\triangleq \; & \mu + \alpha(t - 0.5) + a_s(t) + \beta(g) + \gamma(g)(t - 0.5) + b_s(t,g) + \delta(g,w) + \zeta(g,w)(t - 0.5) + \\
& c_s(t,g,w) + \eta(g,w,c) + \xi(g,w,c)(t - 0.5) + d_s(t,g,w,c), \qquad (4)
\end{aligned}
$$

where μ is the grand mean; $\alpha(t-0.5)$ is the linear main effect of `time`; $a_s(t)$ is the smooth main effect of `time`; $\beta(g)$ is the main effect of `group`; $\gamma(g)(t-0.5)$ is the linear interaction between `group` and `time`; $b_s(t,g)$ is the smooth interaction between `group` and `time`; $\delta(g,w)$ is the main effect of `woman`; $\zeta(g,w)(t-0.5)$ is the linear interaction between `woman` and `time`; $c_s(t,g,w)$ is the smooth interaction between `woman` and `time`; $\eta(g,w,c)$ is the main effect of `cycle`; $\xi(g,w,c)(t-0.5)$ is the linear interaction between `cycle` and `time`; $d_s(t,g,w,c)$ is the smooth interaction between `cycle` and `time`. The first six components in (4) are fixed because they depend on the fixed factors `time` and `group` only. These six components belong to the six subspaces in (2) where orthogonality between spaces provides natural side conditions for identifiability. The last six components in (4) are random because they depend on random factors `woman` and `cycle`. One simple way to model the random effects is to assume that $(\delta(g,w), \zeta(g,w)) \sim \mathrm{N}((0,0), \sigma^2 D_g)$, $(\eta(g,w,c), \xi(g,w,c)) \sim \mathrm{N}((0,0), \sigma^2 E_g)$, and $c_s(t,g,w)$ and $d_s(t,g,w,c)$ are realizations of two independent stochastic processes on $[0,1] \times \mathcal{G}$ with mean 0 and covariance function $\sigma_1^2 R$ and $\sigma_2^2 R$ respectively, where R is given in (3). If only one cycle is sampled for a woman, which is true for conceptive group, `woman` effect is confounded with the `cycle` effect. The corresponding components of `cycle` effect are excluded.

SS ANOVA decomposition (4) is a natural extension of the classical ANOVA decomposition for nested designs (Scheffé 1959) (p.248). Another way to derive essentially the same model as (4) without the `cycle` factor is given in section 4 of Wang (1998a) (Model C). See Wang (1998a) for more examples. One advantage of the SS ANOVA decomposition (4) is that all components have the the same interpretations as in the classical ANOVA as main effects and interactions. Whether a component is fixed or random is determined naturally by the design and/or treatment structure. All these components can be estimated and tested (Gu and Wahba 1993b, Wahba et al. 1995, Wang 1998a), which is very useful for building models. Combining components we can calculate the main effect of `time` as $\alpha(t-0.5)+a_s(t)$;

the interaction between `group` and `time` as $\gamma(g)(t-0.5)+b_s(t,g)$; the interaction between `woman` and `time` as $\zeta(g,u)(t-0.5)+c_s(t,g,u)$; and the interaction between `cycle` and `time` as $\xi(g,u,v)(t-0.5)+d_s(t,g,u,v)$. It is possible to test if a linear model for a main effect or an interaction is appropriate by testing the corresponding smooth component equals to zero. Another advantage of the SS ANOVA decomposition is that one can build several nested models based on this decomposition by deleting components successively (Wang 1998a).

SS ANOVA decompositions for crossed samples of curves are easier and are thus not presented here. We have four factors in the progesterone example: continuous `time` and discrete `group`, `woman` and `cycle`. General SS ANOVA decompositions can be defined for an arbitrary number of factors, each factor takes values on an arbitrary domain (e.g. discrete, continuous, circle, sphere, Euclidean space and tensor products of them), each factor may be fixed or random, and some factors may be nested within other factors. They can be applied to fit longitudinal data, spatial data, spatial-temporal data and spherical-temporal data (see Wang (1998a) for details).

In SS ANOVA decomposition (4), the first six components represent a smoothing `group` mean ($s_{g(i)}$ in BR). Components 7 to 9 measure `woman` effect (subject departure, $s_{w(i)}$ in BR) and the last three components measure the `cycle` effect (cycle departure, $s_{c(i)}$ in BR). It is easy to see that model (10) in BR includes all components in (4) except $a_s(t)$. However, $b_s(t,g)$ is treated as random effect and $\delta(g,w)$, $\zeta(g,w)$, $\eta(g,w,c)$, $\xi(g,w,c)$ are treated as fixed effects. As a consequence, the observations from different women in a group are correlated and there is a large number of parameters.

2 Nonparametric Mixed Effects Models

One of the most useful features of the mixed effects models is their ability to model the mean structure (fixed effects) and the covariance structure (random effects and random errors) simultaneously. As demonstrated in Wang (1998a), the fixed effects usually represent a function on some domain and the random effects represent realizations of a stochastic process on another domain. Flexible models can be derived by modeling the function for fixed effects nonparametrically and by modeling the covariance function of the random effects.

SS ANOVA decompositions provide *one* way to model both the fixed and the random effects. Just like writing the classical ANOVA decompositions in forms of linear models, we may present the SS ANOVA decompositions in a general form as (Wang 1998a)

$$\boldsymbol{y} = \boldsymbol{f} + Z\boldsymbol{b} + \boldsymbol{\epsilon}, \tag{5}$$

where $\boldsymbol{y}$ is the observation vector, Z is the design matrix for the random effects $\boldsymbol{b}$, $\boldsymbol{\epsilon}$ is the random error vector, $\boldsymbol{b} \sim \mathrm{N}(0,\sigma^2 D)$, $\boldsymbol{\epsilon} \sim \mathrm{N}(0,\sigma^2\Lambda)$ and is independent of $\boldsymbol{b}$. $\boldsymbol{f}$ is a vector of a deterministic function f evaluated at design points. $f \in \mathcal{H}$, where $\mathcal{H}$ is a r.k.h.s. on an arbitrary domain $\mathcal{T}_1$. For the decomposition (4), $\mathcal{T}_1 = [0,1]\times\mathcal{G}$ and $\mathcal{H} = W_2 \otimes R^G$.

Model (5) is called a *nonparametric* mixed effects model. Similar to a classical linear model, it is a general model and is not limited to models generated by SS ANOVA decompositions. Linear mixed effects models including the models (10) and (16) in BR are special cases of (5). SS ANOVA models with fixed factors are also special cases of (5). Several methods have been proposed to model the covariance function of the random effects by using smoothing spline structure (Anderson and Jones (1995) and BR), eigenfunction expansion (Rice and Silverman 1991) and specific stochastic processes (Zeger and Diggle 1994). The resulting models usually can be written in a form of (5).

Suppose $\mathcal{H} = \mathcal{H}^0 \oplus \mathcal{H}^1 \cdots \oplus \mathcal{H}^q$, where $\mathcal{H}^0$ is a finite dimensional space containing functions which are not going to be penalized, and the $\mathcal{H}^1, \cdots, \mathcal{H}^q$ are orthogonal subspaces (Gu and Wahba 1993a). For the decomposition (4) with the fixed effects belong to the subspaces in (2), $\mathcal{H}^0 = \{1\} \oplus \{t - 0.5\}$, $\mathcal{H}^1 = H_1$, $\mathcal{H}^2 = H_2$, $\mathcal{H}^3 = \{t - 0.5\} \otimes H_2$, $\mathcal{H}^4 = H_1 \otimes H_2$. Denote $t_1, \cdots, t_n \in \mathcal{T}_1$ as the design points. Suppose $\mathcal{H}^0 = \text{span}\{\phi_1, \cdots, \phi_M\}$ and denote $T = \{\phi_\nu(t_i)\}_{i=1}^{n}{}_{\nu=1}^{M}$. For $k = 1, \cdots, q$, let R^k be the r.k. of $\mathcal{H}^k$ and denote $\Sigma_k = \{R^k(t_i, t_j)\}_{i,j=1,n}$. Let P_k be the orthogonal projector in $\mathcal{H}$ onto $\mathcal{H}^k$. The function f is estimated by maximizing the penalized likelihood

$$\min_{f \in \mathcal{H}} \left\{ (\boldsymbol{y} - \boldsymbol{f})' W (\boldsymbol{y} - \boldsymbol{f}) + \sum_{k=1}^{q} \lambda_k \|P_k f\|^2 \right\}. \tag{6}$$

where $W^{-1} = \text{Cov}(\boldsymbol{y})/\sigma^2 = ZDZ' + \Lambda$. The norms on $\mathcal{H}^k$ are the tensor product norms induced by the norms on the component subspaces (see Wang, Wahba, Gu, Klein and Klein (1997b) for an example). The random effects $\boldsymbol{b}$ are estimated by the posterior mean.

A Bayesian model may be constructed for (5) by specifying a prior of f as

$$F(t) = \sum_{\nu=1}^{M} \theta_\nu \phi_\nu(t) + \sigma \sum_{k=1}^{q} Z_k(t)/\sqrt{\lambda_k}, \qquad t \in \mathcal{T}_1,$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_M)' \sim \text{N}(0, aI)$, $Z_k(t)$ are independent zero mean Gaussian stochastic processes, independent of $\boldsymbol{\theta}$, with covariance $\text{E}Z_k(s)Z_k(t) = R^k(s,t)$. Suppose that observations are generated by

$$\boldsymbol{y} = \boldsymbol{F} + Z\boldsymbol{b} + \boldsymbol{\epsilon}, \tag{7}$$

where $\boldsymbol{F}$ is a vector of F evaluated at design points. Then it can be shown that with the variance a approach to infinity, the estimates of the function and the random effects equal to the posterior means of the Bayesian model (Wang 1998a). Posterior covariances are used to construct Bayesian confidence intervals for various components. Bayesian confidence intervals have good frequency properties viewed as across the function and they behave similarly to the bootstrap confidence intervals based on parametric resampling (Wang and Wahba 1995). Smoothing parameters and covariance parameters are estimated jointly by the generalized maximum likelihood (GML) method based on the Bayesian model. Extensions of the cross validation, generalized cross validation, unbiased risk methods may also be used. However, they are inferior to the GML method with moderate sample sizes (Wang 1998b, Wang, Guo and Brown 1997a).

The nonparametric mixed effects model (5) is *not* a linear mixed effects model since its mean function is nonparametric. Nevertheless, using the connection between a smoothing spline model and a linear mixed effects model, it can be related to the following linear mixed effects model:

$$\boldsymbol{y} = T\boldsymbol{d} + \sum_{k=1}^{q} \boldsymbol{u}_k + Z\boldsymbol{b} + \boldsymbol{\epsilon} = T\boldsymbol{d} + (I \ \cdots \ I \ Z) \begin{pmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_q \\ \boldsymbol{b} \end{pmatrix} + \boldsymbol{\epsilon}, \tag{8}$$

where $\boldsymbol{u}_k \sim \mathrm{N}(0, \sigma^2\Sigma_k/\lambda_k)$, $\boldsymbol{b} \sim \mathrm{N}(0, \sigma^2 D)$, $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2\Lambda)$, and they are mutually independent. More precisely, the estimates of the function f evaluated at design points and the prediction of $\boldsymbol{y}$ in the nonparametric mixed effect model (5) are the same as the BLUP estimates of the corresponding elements in the linear mixed effects model (8); the GML estimates of the smoothing parameters and the covariance parameters in (5) are the same as the REML estimates of the variance components in (8); and posterior covariances of the elements in the Bayesian model (7) are the same as covariances of the corresponding elements in (8). This connection suggests a way of fitting nonparametric mixed effects models using the existing programs for linear mixed effects models such as SAS procedure `proc mixed`. See Wang (1998a) for details.

We want to draw a distinction between a smoothing spline (nonparametric mixed effects) model and its related linear mixed effects model. A spline model is defined on all points in a domain (e.g. [0,1]) and often belongs to an infinite dimensional space (e.g. W_2). The linear mixed effects model, on the other hand, is only defined on the design points with a finite number of parameters. They provide the same predictions at the design points only. More importantly, the interpretations of these two models are different: a spline function is deterministic and observations are independent; the linear mixed effects model has random effects and observations are thus correlated. It is not clear which model ((9) or (10)) in BR is believed to be the true model that generates observations. If (9) is the true model, observations are independent and model (10) is used merely for calculation. If (10) in BR is the true model, observations are correlated and (9) serves as a motivation.

3 Conclusion

SS ANOVA technique combines flexibility of the smoothing spline models with power of the classical ANOVA. For both crossed and nested designs, they can be used to build flexible mixed effects models for multiple fixed and random factors on arbitrary domains. Components in a model constructed by these decompositions have the standard interpretation as main effects and interactions and are fixed or random based on the design and/or treatment structure.

SS ANOVA decompositions and many other methods give rise to the general form of the nonparametric mixed effects model. Therefore calculations of the estimates and inference can be dealt with in a unified fashion. Relationship between a nonparameteric mixed effects model and a linear mixed effects model allow us to calculate these estimates using the SAS procedure `proc mixed`.

With different assumptions and goals, self-modeling models may also be used to fit curved data (Wang and Brown 1996).

References

Anderson, S. J. and Jones, R. H. (1995). Smoothing splines for longitudinal data, *Statistics in Medicine* **14**: 1235–1248.

Gu, C. (1997). Discussion of "Polynomial splines and their tensor products in extended linear models" by Stone, Hansen, Kooperberg and Truong, *Annals of Statistics* **25**: 1432–1443.

Gu, C. and Wahba, G. (1991). Comments to "Multivariate adaptive regression splines", by J. Friedman, *Annals of Statistics* **19**: 115–123.

Gu, C. and Wahba, G. (1993a). Semiparametric ANOVA with tensor product thin plate spline, *Journal of the Royal Statistical Society B* **55**: 353–368.

Gu, C. and Wahba, G. (1993b). Smoothing spline ANOVA with component-wise Bayesian confidence intervals, *Journal of Computational and Graphical Statistics* **2**: 97–117.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves, *Journal of the Royal Statistical Society B* **53**: 233–243.

Scheffé, H. (1959). *The Analysis of Variance*, Wiley, New York.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59.

Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy, *Annals of Statistics* **23**: 1865–1895.

Wang, Y. (1997). GRKPACK: fitting smoothing spline analysis of variance models to data from exponential families, *Communications in Statistics: Simulation and Computation* **26**: 765–782.

Wang, Y. (1998a). Mixed-effects smoothing spline ANOVA, *Journal of the Royal Statistical Society B* **60**.

Wang, Y. (1998b). Smoothing spline models with correlated random errors, *Journal of the American Statistical Association* **93**, March issue.

Wang, Y. and Brown, M. B. (1996). A flexible model for human circadian rhythms, *Biometrics* **52**: 588–596.

Wang, Y. and Wahba, G. (1995). Bootstrap confidence intervals for smoothing splines and their comparison to Bayesian confidence intervals, *J. Statist. Comput. Simul.* **51**: 263–279.

Wang, Y., Guo, W. and Brown, M. B. (1997a). Spline smoothing for bivariate data with applications to association between hormones, Submitted. Available at `http://www.pstat.ucsb.edu/~yuedong`.

Wang, Y., Wahba, G., Gu, C., Klein, R. and Klein, B. (1997b). Using smoothing spline ANOVA to examine the relation of risk factors to the incidence and progression of diabetic retinopathy, *Statistics in Medicine* **16**: 1357–1376.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, *Biometrics* **50**: 689–699.