Smoothing analysis of variance for general designs

Yue Cui[1], **James S. Hodges**[2*]

[1]Department of Biometrics, Abbott, Abbott Park, Illinois USA 60064

[2]Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota USA 55414

[*]*email:* hodges@ccbr.umn.edu

May 10, 2009

Summary

Hodges et al (2007) proposed smoothed analysis of variance (ANOVA) and showed how to use it to smooth interactions in balanced, single-error-term ANOVAs. Besides avoiding discontinuous choices to include or exclude effects, smoothed ANOVA addresses three practical concerns: unreplicated designs, masking in effects with many degrees of freedom, and "subgroup analysis". The present paper extends smoothed ANOVA to general designs with multiple error terms or imbalance, although certain extensions are only possible for balanced designs. While this is presented as a Bayesian analysis, we also give a smoothed ANOVA table similar to a standard ANOVA table with degrees of freedom and sum of squares.

Keywords: ANOVA, Bayesian analysis, shrinkage, smoothing, subgroup analysis

1 Introduction and motivation

Analysis of variance (ANOVA) is widely used to test the association of a response with individual factors and their combinations, i.e., interactions. Interactions are often modeled in a stepwise way in which at each step an effect stays in the model or leaves it completely. Smoothed ANOVA (SANOVA; Hodges et al 2007, henceforth HCSC) neither includes nor excludes effects, but instead smooths (or shrinks) them. In doing so, it addresses three practical concerns with standard ANOVA: unreplicated designs, masking in effects with many degrees of freedom, and "subgroup analysis". In a simulation study and a real data analysis, HCSC showed advantages of using smoothed ANOVA. Zhang et al (2007) used the SANOVA framework to smooth spatially-referenced factors in two- and higher-way ANOVAs, again balanced with a single error term. The present paper extends smoothed ANOVA from balanced, single-error-term designs to designs with arbitrary random effects and imbalance. HCSC proposed a SANOVA table as a bookkeeping device for variation in and smoothing of the dependent variable y; this paper extends the SANOVA table to general designs, although one aspect of the extension appears to require balance.

Following Hodges & Sargent (2001), HCSC proposed priors on the fitted degrees of freedom (DF) of individual effects as an indirect but intuitive way to specify priors on smoothing parameters. They also showed how to condition the priors on the DF in one or more effects so they satisfy sum constraints. This was possible only because the balanced, single-error-term ANOVA gives a natural decomposition of total DF in the fit into DF for individual effects. Cui et al (2009) extended Hodges & Sargent's (2001) definition of DF to attribute DF to individual effects in great generality. The present paper uses this extended definition to specify priors on the DF of effects in smoothing general ANOVAs.

This paper is organized as follows. Section 2 defines notation, introduces the smoothed ANOVA table and shows how to put prior distributions on DF to specify priors on smoothing parameters. The source and destination of smoothed-out DF and SS can be derived for a large class of balanced designs; Section 3 shows how. Section 4 demonstrates some properties of smoothed ANOVA and priors on DF using data from a balanced study having one between- and one within-subject effect. Section 4 concludes.

2 Theory and derivations

2.1 Notation

We use a hybrid parameterization: a one-column-per-DF parameterization for each fixed effect (FE) and a one-column-per-level exchangeable parameterization for each random effect (RE). We use the terms FE and RE in their more traditional senses (e.g., Scheffé 1959, p. 238): a FE is a factor with fixed number of levels and the interest lies in estimation of each of its levels; a RE is a factor with many levels, from which only a sample of levels can be investigated, and the interest lies in estimating variation in the population of levels. The one-column-per-DF parameterization for FE is often used in ANOVA because it assigns each effect a number of columns equal to the DF used in standard ANOVA F-tests. For RE, as traditionally understood, one-column-per-level exchangeable parameterization is the convention. Section 3 shows an example to make this more concrete. We assume the following model

$$y = X_1\theta_1 + X_2\theta_2 + \epsilon, \tag{1}$$

where $X_1 \in R^{n\times p}$ and $X_2 \in R^{n\times L}$ are respectively the design matrices for effects that will not be smoothed and for effects that will be smoothed, including REs. Partition the columns in X_2 as $[X_{21}, \cdots, X_{2l}]$, $l \leq L$ and conformably partition $\theta_2 = [\theta_{21}, \cdots, \theta_{2l}]$. The parameters $\theta_1 \in R^p$

and the first $l' \leq l$ clusters of parameters in $\theta_2 \in R^L$ are FEs while the other $l - l'$ clusters of parameters in θ_2 are REs. Then if we use an orthogonal parameterization for FEs, which gives each column one DF, in a balanced ANOVA design and with a suitable scaling of X_1 and X_2, $X_1' X_1 = nI_p$ and for $j,\ j_1,\ j_2 \leq l'$, $X_{2j_1}' X_{2j_2} = 0$ if $j_1 \neq j_2$, $X_{2j}' X_{2j} = nI$, and $X_1' X_{2j} = 0$. This parameterization is the same as in HCSC (Section 2). The j^{th} cluster of smoothed θ_{2j} has prior distribution $\theta_{2j}|\Gamma_{2j} \sim N(0, \Gamma_{2j})$, and the $\theta_{2j}|\Gamma_{2j}$ are assumed independent of each other. The normally distributed error ϵ has mean 0 and covariance matrix Γ_0. Here, Γ_0 and Γ_{2j} are nonnegative-definite symmetric covariance matrices and are not necessarily proportional to the identity matrix or even diagonal. If X_{2j} corresponds to a RE, e.g., with 5 levels, then to get an exchangeable parameterization for this effect, X_{2j} could be a binary matrix, i.e., all entries are 0 or 1, with 5 columns, and Γ_{2j} could be proportional to the identity matrix of dimension 5, or alternatively have a 2-parameter compound symmetry structure, etc.

The partition of X_2's columns and θ_2 need not correspond to effects. For example, an effect's 5 columns in X_2 could be partitioned into as many as 5 clusters, or the columns representing all of the two-way interactions could be grouped into a single cluster. However, it will often be useful to specify clusters corresponding to effects in the ANOVA.

An alternative parameterization would be an exchangeable parameterization for every effect, as proposed by Peter McCullagh (personal communication). For a 2 by 3 factorial design, this would use a 2 column and 3 column binary matrix for the two main effects and a 6 column binary matrix for the interaction. This parameterization satisfies infinite exchangeability for the levels of the two fixed effects, which makes it a sensible model according to McCullagh (2002), so that, for example, a contrast of two levels of a factor is not impacted by deletion or addition of other levels of the factor. McCullagh's mathematical formulation of a statistical model is partly motivated by a concern that a parameter's meaning should not be affected by the sample size or experimental design, with the analysis primarily oriented toward predictions of future observations. For our purposes, we lean to Besag's (2002) view of McCullagh's motivating concerns. First, this specification is over-parameterized and thus inestimable even when no term is smoothed and the model reduces to an ordinary ANOVA. Also, in computing the degrees of freedom in the model's fit, this parameterization allocates degrees of freedom among effects in a non-intuitive way (Cui et al 2009), and we consider this allocation important for interpretation and for placing prior distributions on variances of RE and variances used to smooth FE. Also, the benefits of the exchangeable parameterization, such as consistency in morphisms on experimental units (including permutation and marginalization), appear to be relevant to designed experiments, where units are regarded as exchangeable, more than to observational studies. However, we aim to produce an analysis for general designs, including observational studies.

2.2 Deriving DF and SS

As in a standard ANOVA, a smoothed ANOVA procedure gives a bookkeeping table of DF and SS. We use Hodges and Sargent's (2001) definition of DF, as extended in Cui et al (2009). This definition is identical to that of Ruppert et al (2003, section 8.3) when both are applicable. The formula to define effect-specific SS is given below. (In the preceding section's terms, here the clusters are the same as the ANOVA's effects.) DF and SS as defined here are similar in that they are both consistent with definitions in HCSC and are functions of the smoothing parameters Γ_0 and Γ_2. Since DF and SS depend on unknowns, in a Bayesian analysis they have posterior distributions, and we can use summary descriptions of these posteriors, such as the posterior mean or median. We prefer the posterior mean for reasons given in HCSC, mainly that the expectation of a sum is the sum of the expectations so that the ANOVA's bookkeeping function is preserved.

The DF (Cui et al 2009) of the nonsmoothed FE X_1 is the rank of X_1, while the DF of a smoothed effect X_{2j} or error ϵ is the "fraction of variation" contributed by X_{2j} or ϵ out of variation that is not accounted for by the nonsmoothed effects X_1, i.e.,

$$\begin{aligned} DF(X_{2j}) &= tr\{X_{2j}\Gamma_{2j}X_{2j}'[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+}\} \\ DF(\epsilon) &= tr\{\Gamma_0[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+}\} \end{aligned}$$

where $\Gamma_2 = diag(\Gamma_{21}, \cdots, \Gamma_{2l})$; P_M is the orthogonal projection onto the column space of a matrix M, where the inner product of two vectors α and β is $\alpha'\beta$; and "+" denotes the Moore-Penrose generalized inverse.

Following the rationale used to derive DF (Cui et al 2009), the SS of a smoothed effect X_{2j} or error ϵ is the fraction of variation in y explained by X_{2j} or ϵ out of variation not accounted for by the nonsmoothed effects X_1. As in the derivation of DF, a nonsmoothed effect can be viewed as the limit of a smoothed effect, for which the prior covariance goes to infinity and imposes no constraint. For a positive scalar λ and any positive definite $\Gamma_1 \in R^{p\times p}$, the SS for, respectively, non-smoothed effects, for the j^{th} cluster of smoothed effects (FE or RE), and for error ϵ, are thus defined as:

$$\begin{aligned} SS(X_1) &= \lim_{\lambda\to+\infty} y'\{[X_1\lambda\Gamma_1X_1'+X_2\Gamma_2X_2'+\Gamma_0]^{+1/2}X_1\lambda\Gamma_1X_1' \\ &\quad [X_1\lambda\Gamma_1X_1'+X_2\Gamma_2X_2'+\Gamma_0]^{+1/2}\}y \\ &= y'P_{X_1}y \end{aligned}$$

$$
\begin{aligned}
SS(X_{2j};\Gamma_0,\Gamma_2) &= \lim_{\lambda\to+\infty} y'\{[X_1\lambda\Gamma_1X_1' + X_2\Gamma_2X_2' + \Gamma_0]^{+1/2}X_{2j}\Gamma_{2j}X_{2j}' \\
&\quad [X_1\lambda\Gamma_1X_1' + X_2\Gamma_2X_2' + \Gamma_0]^{+1/2}\}y \\
&= y'\{[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}X_{2j}\Gamma_{2j}X_{2j}'[(I-P_{X_1}) \\
&\quad (X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}\}y, \\
SS(\epsilon;\Gamma_0,\Gamma_2) &= \lim_{\lambda\to+\infty} y'\{[X_1\lambda\Gamma_1X_1' + X_2\Gamma_2X_2' + \Gamma_0]^{+1/2}\Gamma_0 \\
&\quad [X_1\lambda\Gamma_1X_1' + X_2\Gamma_2X_2' + \Gamma_0]^{+1/2}\}y \\
&= y'\{[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}\Gamma_0[(I-P_{X_1}) \\
&\quad (X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}\}y.
\end{aligned}
$$

In the limit as λ goes to $+\infty$, Γ_1's specific value does not matter.

If M is nonnegative definite, $M^{+1/2}$ denotes the symmetric and nonnegative positive square root of M as defined in Horn & Johnson (1985). They showed that for a nonnegative definite matrix, $M^{+1/2}$ always exists and is unique. It also has the same singular value decomposition as M except for replacing the diagonals in the diagonal matrix by their nonnegative square roots. The usage $[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}$ is therefore legal since the matrix in question is nonnegative definite.

The following lists some properties of SS; the Appendix gives proofs.

($SS.a$) For balanced, single-error-term designs as in HCSC, the above definition gives the same SS as in HCSC.

($SS.b$) When Γ_0, the covariance of ϵ, is positive definite, $SS(X_1)+\sum_1^J SS(X_{2j};\Gamma_0,\Gamma_2)+SS(\epsilon;\Gamma_0,\Gamma_2) = y'y = ||y||^2$, as in ordinary ANOVA.

($SS.c$) Suppose model (1) is correct, i.e., $y|\Gamma_0,\Gamma_2 \sim N(\mu, X_2\Gamma_2X_2'+\Gamma_0)$, where $\mu \in R(X_1)$, the column space of X_1, and Γ_0 is positive definite. Then $E(SS(X_1)|\Gamma_0,\Gamma_2) = tr[(X_2\Gamma_2X_2' + \Gamma_0)P_{X_1}] + \mu'\mu$ and $E(SS(X_{2j};\Gamma_0,\Gamma_2)|\Gamma_0,\Gamma_2) = tr[(X_{2j}\Gamma_{2j}X_{2j}')(I-P_{X_1})]$. So the expected SS of the unsmoothed effects X_1 come from two sources: $\mu \in R(X_1)$ and variation arising from $X_2\theta_2$ and error ϵ that falls in $R(X_1)$, the column space of X_1. By contrast, the expected SS of smoothed X_{2j} comes only from the variation in the orthogonal complement of the column space of X_1.

It may seem odd that sums of squares can be defined unambiguously for individual smoothed effects or random effects with columns X_{2j} when this is not possible in ordinary unbalanced ANOVAs (giving rise, for example, to the SAS system's four types of sums of squares). Note, however, that the sum of squares is defined above for the unsmoothed effects as a group, not for individual unsmoothed effects. Also, the sum of squares for the unsmoothed effects in X_1 is derived by first treating their coefficients θ_1 as smoothed, with smoothing covariance $\lambda_1\Gamma_1$ for positive definite Γ_1, and then letting λ_1 go to infinity. Suppose we also give the j^{th} cluster of smoothed effects, θ_{2j}, smoothing covariance $\lambda_j\Gamma_j^*$ for positive definite Γ_j^*, and then try to derive sums of squares for X_1 and X_{2j} by letting both λ_1 and λ_j go to infinity. Unless $X_1'X_{2j} = 0$, the limiting sums of squares depend on the order in which λ_1 and λ_j go to infinity, that is, the limits do not exist.

2.3 Prior distributions on smoothing parameters

HCSC Section 2.4 discussed prior distributions for smoothing balanced single-error-term ANOVA. Many of those prior distributions are applicable here as well. For cases in which $\Gamma_{2j} = \sigma_{2j}^2\Gamma_{2j}^0$ and $\Gamma_0 = \sigma_0^2\Gamma_0^0$, an obvious choice is to put priors on $\sigma_i^2, i = 0, \ldots, l$ or $\eta_i = 1/\sigma_i^2$. Apart from the familiar independent gamma priors on the η_i, the growing literature of such priors includes Gelman (2004, 2005) among others. He et al (2007) proposed re-parameterizing the σ_{2j}^2 to $\lambda = \sum_{i>0}\eta_i/\eta_0$ and $\beta = (\eta_1/\lambda, ..., \eta_l/\lambda)$, where β lies in the l-dimensional simplex, permitting a proper flat prior.

HCSC suggested priors on DF as a way to induce priors on smoothing parameters, which extend to the present case but not necessarily straightforwardly. We discuss these at length below.

HCSC (Section 2.4.2) showed how to condition prior distributions on smoothing parameters so that the sum of the DF in a group of clusters is equal to or bounded above by a constant, e.g., the sum of DF in all of the two-way interactions is fixed at some K. This can be done in the present case as well. To specify such priors, it is easiest to specify unconstrained priors for the smoothing parameters, either on the smoothing parameters themselves or on the analogous DF, and then to impose the sum constraint. HCSC (Appendix A.3, Case 2) gave an algorithm for making MCMC draws under these constraints.

The rest of this section discusses prior distributions on DF. For model (1), Cui et al (2009) proposed a method to obtain a DF-based prior on variance ratios when each of the l smoothed effects (groups of design-matrix columns) is smoothed by a single variance parameter σ_i^2 and residual error is smoothed by σ_0^2. Specifically, assume in model (1) that $\Gamma_{2j} = \sigma_j^2\Gamma_{2j}^0,\ j = 1, \cdots, l$ and $\Gamma_0 = \sigma_0^2\Gamma_0^0$, where $\sigma_j^2,\ j = 0,\ 1, \cdots,\ l$ are unknown scalars, while $\Gamma_{2j}^0, \Gamma_0^0$ are known positive definite covariance matrices. Assume further that $X_{2j} \nsubseteq R(X_1)$ for any $j = 1, \cdots, l$. Then $DF(\cdot)$ is a 1-1 mapping between $q = (DF(X_{21}), \cdots, DF(X_{2l}))'$ on q's range and $s = (log\frac{\sigma_1^2}{\sigma_0^2}, \cdots, log\frac{\sigma_l^2}{\sigma_0^2})'$, whose range is R^l, and a prior on q induces a unique prior on s. We now elaborate that idea for the general smoothed ANOVA considered here.

(This may seem restrictive, but recall that a cluster of columns X_{2l} need not be an effect.

An effect can be broken into as many clusters as the dimension of its column space, for example, HCSC Section 4 treats a 21-DF interaction as 21 clusters.)

Any prior on the vector of cluster-specific DF $q = (DF(X_{21}), \cdots, DF(X_{2l}))$ induces a joint prior on the l variance ratios $\{r_i\}_1^l$, where $r_i = \frac{\sigma_i^2}{\sigma_0^2}$. For the balanced, single-error-term smoothed ANOVAs considered in HCSC, $DF(X_{2i})$ involves only r_i, and the range of q is the Cartesian product of the ranges of the $DF(X_{2i})$. Thus HCSC put independent priors on the $DF(X_{2i})$ to obtain independent priors on the r_i. But in general smoothed ANOVA, the range of q is not necessarily regular and it is subject to constraints (property ($DF.c$), Section 2.4 of Cui et al 2009). As the ANOVA becomes more complicated, the range of q becomes subject to more constraints. For example, for the balanced design discussed in Section 4 below, the DF of the smoothed fixed effect A lies in $(0, 2)$ and the DF of the nested random effect $S|A$ lies in $(0, 26)$, but the sum of these two DF is constrained to lie in $(0, 26)$.

To avoid such complications and to speed computing, for the purpose of specifying a prior we can use an *ad hoc* approximation of DF which we call pseudo-DF. For model (1), we define the pseudo DF of X_{2j}, denoted by $DF_p(X_{2j})$, as the DF of X_{2j} in a simplified model with only nonsmoothed effects, the single smoothed effect X_{2j}, and the pure error, i.e.,

$$DF_p(X_{2j}) = tr\{X_{2j}\Gamma_{2j}X_{2j}'[(I - P_{X_1})(X_{2j}\Gamma_{2j}X_{2j}' + \Gamma_0)(I - P_{X_1})]^+\}.$$

If $\Gamma_{2j} = \sigma_j^2\Gamma_{2j}^0$ and $\Gamma_0 = \sigma_0^2\Gamma_0^0$, where $\Gamma_{2j}^0, \Gamma_0^0$ are known positive definite covariance matrices, then $DF_p(X_{2j})$ is a function only of $r_j = \frac{\sigma_j^2}{\sigma_0^2}$, and ranges from 0 to $rank((I - P_{X_1})X_{2j})$. Then $q_p = (DF_p(X_{21}), \cdots, DF_p(X_{2l}))$ defines a mapping from $r = (r_1, \cdots, r_l) \in (R^+)^l$ onto $(0, rank((I - P_{X_1})X_{21})) \times \cdots \times (0, rank((I - P_{X_1})X_{2l}))$. Now we can put unrestricted independent priors on each $DF_p(X_{2j})$ to obtain a joint prior on r. Ruppert et al (2003, Section 8.3) used the same approximation to reduce the computational burden of exact DF and reported practically no difference between the approximation and exact values for all examples they had considered. However, this is not always the case; Cui et al (2009, Section 3.2) gives an example, and Section 4 below gives another.

Usually an unbalanced design leads to a messier expression for q than does a balanced design. However, since an unbalanced design can usually be nested within a balanced design with the same set of variance parameters, DF-based priors from the larger nesting balanced design can be used as another way to construct approximate priors for the actual unbalanced design. If the imbalance is not great, this prior should differ little from a prior on DF for the imbalanced design itself. We do not consider this further here.

Section 3 below specifies another possibility, a prior on the so-called retained DF of each random effect or smoothed fixed effect which are defined below.

For the purpose of inducing priors on the variance ratios $\frac{\sigma_j^2}{\sigma_0^2}$, we can use exact DF, pseudo-DF, or retained DF to induce DF-based priors on the r_i, but after we have samples from r_i's posterior, we should only consider the exact DF as describing the smoothness of the fitted effects.

3 The destination of DF and SS smoothed out of effects

The usual ANOVA table decomposes both the DF and SS into pieces for each effect and for error. In smoothed ANOVA, θ_{2i} are shrunk toward 0 and partly removed from the fitted model and counted as error. In the one-error-term ANOVA design considered by HCSC, variation smoothed out of an effect can only be smoothed into one place: pure error. In a more complicated design, however, REs compete with pure error for the part of variation smoothed out of fixed effects, so in a general design, variation smoothed out of effects may have more than one destination. In this case, smoothed ANOVA can show how information about each error term is derived from replication and from variation smoothed out of shrunken effects and other errors. We develop these ideas using a specific balanced experiment with two error terms, a subject effect and the usual residual error, allowing explicit expressions. Three features of this example may reduce its generality: the choices of Γ_{2j}, the clear hierarchy of the error terms (REs), and the design's balance. At this section's end we discuss what is known about how much our results can be generalized.

Consider a balanced design with one between- and one within-subject fixed effect. Section 4 uses this design for an endodontics dataset, the Irrigation study. The between-subject effect A has 3 levels and nine subjects (S) were measured at each level of A. The within-subject effect B has 5 levels, giving a total of 135 observations. Let 1_k be the normalized k-vector of 1s, $1_k = \frac{1}{\sqrt{k}}(1, \cdots, 1)'$, and let H_k be a normalized orthogonal basis of the orthogonal complement of 1_k in R^k, i.e., $1_k' H_k = 0$ and $H_k' H_k = I_{k-1}$. Then model (1) becomes

$$y = X_1\theta_1 + X_2\theta_2 + \epsilon, \tag{2}$$

where $X_1 = X_{GM} = 1_3 \otimes 1_9 \otimes 1_5$ is the design matrix for the grand mean; $X_2 = [X_A, X_B, X_{AB}, X_{S|A}]$ is partitioned conforming to mean parameters $\theta_2 = [\theta_A', \ \theta_B', \ \theta_{AB}', \theta_{S|A}']'$, where A, B and AB refer to the main effects for A and B and their interaction, while $S|A$ refers to the subject random effect within levels of A. In X_2, $X_A = H_3 \otimes 1_9 \otimes 1_5$, $X_B = 1_3 \otimes 1_9 \otimes H_5$, $X_{AB} = H_3 \otimes 1_9 \otimes H_5$, and $X_{S|A} = I_3 \otimes I_9 \otimes 1_5$. Since there is no replication, the lowest-level error term is $S \times B|A$, i.e., the random effect subject-by-B within levels of A. For this example, we smooth effects in θ_2 by modelling

Effect	GM	A	B	AB	$S\|A$	$\epsilon = S \times B\|A$
DF	1	$\frac{\sigma_2^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$	$\frac{\sigma_3^2}{\sigma_3^2+\sigma_0^2} \cdot 4$	$\frac{\sigma_4^2}{\sigma_4^2+\sigma_0^2} \cdot 8$	$\frac{\sigma_5^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2 + \frac{\sigma_5^2}{\sigma_5^2+\sigma_0^2} \cdot 24$	$\frac{\sigma_0^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2 + \frac{\sigma_0^2}{\sigma_3^2+\sigma_0^2} \cdot 4 + \frac{\sigma_0^2}{\sigma_4^2+\sigma_0^2} \cdot 8 + \frac{\sigma_0^2}{\sigma_5^2+\sigma_0^2} \cdot 24 + 96$

Table 1: Effect-specific DF in the example (the Irrigation study)

each as iid normal conditional on its variance σ_i^2: $\theta_A \sim N(0, \sigma_2^2 I_2)$, $\theta_B \sim N(0, \sigma_3^2 I_4)$, $\theta_{AB} \sim N(0, \sigma_4^2 I_8)$, $\theta_{S|A} \sim N(0, \sigma_5^2 I_{27})$, $\epsilon \sim N(0, \sigma_0^2 I_{135})$, where the effects and interactions in θ_2 are independent of each other conditional on the σ_i^2. For this discussion, each effect is smoothed using a single smoothing parameter, though this is not necessary (see HCSC Section 4 and the discussion of generalizations below).

With DF as defined in Cui et al (2009), Table 1 gives effect-specific DF in the fit as a function of the variances σ_i^2. If a modeler does not want to smooth the FEs A, B, or AB, which is equivalent to letting their smoothing variances σ_2^2, σ_3^2, or σ_4^2 go to infinity, then from Table 1 above, A, B, and AB will have 2, 4, and 8 degrees of freedom respectively, $DF(X_{S|A}) = \frac{\sigma_5^2}{\sigma_5^2+\sigma_0^2} \cdot 24$, and $DF(\epsilon) = \frac{\sigma_0^2}{\sigma_5^2+\sigma_0^2} \cdot 24 + 96$. It is easy to check that $DF(X_{GM}) + DF(X_A) + DF(X_B) + DF(X_{AB}) + DF(X_{S|A}) + DF(\epsilon) = 135 = dim(y)$. Direct application of Section 2.2's definition gives SS for each effect. Section 4 gives a smoothed ANOVA table for the Irrigation study.

We now describe how the DF and SS for each smoothed effect are smoothed into the two error terms in this balanced design. It is conventional in standard balanced ANOVA with a one-column-per-DF parameterization and a single error term to regard a fixed effect, say E, as the source for all variation of y that lies in $R(X_E)$, the column space of X_E, where X_E is effect E's design matrix. Thus effect E is credited with all DF attributed to variation in $R(X_E)$. However, in ANOVA with more than one error term, variation in y that lies in $R(X_E)$ comes from effect E and possibly from random effects whose design matrices are correlated with X_E, or more precisely, the column spaces of their design matrices have non-trivial intersection with $R(X_E)$. Therefore in smoothed ANOVA, effect E and those error terms share the fixed number of DF attributed to variation in y that lies in $R(X_E)$.

In our example's design, the response y is a 135-vector, so DF of the total variation in y is 135 (property (DF.a) in Cui et al 2009). Of these 135 DF, the variation that lies in $R(X_A)$ comprises 2 DF, since effect A has 2 contrasts. This 2 DF is by convention the DF of effect A in unsmoothed ANOVA. This is also true in smoothed ANOVA if effect A is not smoothed, or equivalently, if A's smoothing variance goes to infinity, so that although REs or error terms contribute to variation in $R(X_A)$, they are credited with 0 DF of A's maximum 2 DF. However, when effect A is smoothed, A only retains part of these 2 DF in the fitted model, i.e., $\frac{\sigma_2^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$. Where does the rest of the

2 DF go?

In model (1), the total (marginal) variation in y around the unsmoothed effects is $X_2\Gamma_2X_2'+\Gamma_0$, given θ_1, Γ_0 and Γ_2. Consider a given fixed effect that is being smoothed, say θ_E, with design matrix X_E and smoothing covariance Γ_E. Because we are considering balanced designs, $R(X_E)$ intersects with the column space of the nonsmoothed effects, $R(X_1)$, only at the origin. Then the total variation in y around the unsmoothed effects lying in $R(X_E)$, contributed by θ_E and possibly other effects, has total DF rank(X_E). In our example, the total variation in $R(X_A)$ around the grand mean is $P_{X_A}(X_2\Gamma_2X_2' + \Gamma_0)P_{X_A}$, and this total variation is credited with 2 DF, i.e., $tr\{P_{X_A}(X_2\Gamma_2X_2' + \Gamma_0)P_{X_A}[(I - P_{X_1})(X_2\Gamma_2X_2' + \Gamma_0)(I - P_{X_1})]^+\}$ equals 2. These 2 DF are partitioned among A, $S|A$ and $S \times B|A$ (error). Specifically, the variation arising from A is $X_A\Gamma_AX_A'$; the variation from $S|A$ that lies in $R(X_A)$ is $P_{X_A}COV_{S|A}P_{X_A}$, where $COV_{S|A} = X_{S|A}\Gamma_{S|A}X_{S|A}'$; the variation from $S\times B|A$ (error) that lies in $R(X_A)$ is $P_{X_A}COV_{S\times B|A}P_{X_A}$, where $COV_{S\times B|A} = X_{S\times B|A}\Gamma_0X_{S\times B|A}' = \Gamma_0$ because $X_{S\times B|A} = I_{135}$. The sum of the three components is the total variation in $R(X_A)$, $P_{X_A}(X_2\Gamma_2X_2' + \Gamma_0)P_{X_A}$. X_B and X_{AB} do not have a share of this variation because $P_{X_A}X_B = P_{X_A}X_{AB} = 0$.

Therefore, the DF in $R(X_A)$ attributed to effect A is $tr\{X_A\Gamma_AX_A'[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^+\}$, which reduces to $\frac{\sigma_2^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$; the part of $S|A$'s DF arising from its contribution to variation in $R(X_A)$ is $tr\{P_{X_A}COV_{S|A}P_{X_A}[(I - P_{X_1})(X_2\Gamma_2X_2' + \Gamma_0)(I - P_{X_1})]^+\}$, which reduces to $\frac{\sigma_5^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$; the part of $S\times B|A$'s (i.e., error) DF arising from its contribution to variation in $R(X_A)$ is $tr\{P_{X_A}COV_{S\times B|A}P_{X_A}[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^+\}$, which reduces to $\frac{\sigma_0^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$. The three parts add to 2, the DF conventionally attributed to A.

Define the part of a RE's DF that arises from its contribution to the total variation in $R(X_E)$ as the "DF smoothed out of X_E to this RE", because if θ_E is not smoothed, this part of DF will not be allocated to this RE. In our example, out of the total DF of $S|A$, the part $\frac{\sigma_5^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$ is termed as "DF smoothed from A to $S|A$", which becomes 0 if A is not shrunk toward 0, or equivalently if σ_2^2 goes to $+\infty$. With the balanced orthogonal parameterization of FEs, we can divide variance arising from $S|A$ into parts corresponding to variation in the column spaces of each FE and a residual, i.e., total variance from $S|A$ decomposes as

$$COV_{S|A} = P_{GM}COV_{S|A}P_{GM} + P_ACOV_{S|A}P_A + P_BCOV_{S|A}P_B + P_{AB}COV_{S|A}P_{AB} + QCOV_{S|A}Q, \quad (3)$$

where $Q = I - (P_{GM} + P_A + P_B + P_{AB}) = I_3 \otimes P_{H_9} \otimes I_5$, $P_{GM} = P_{X_{GM}}$, $P_A = P_{X_A}$, $P_B = P_{X_B}$, $P_{AB} = P_{X_{AB}}$. Thus, define the DF smoothed from A to $S|A$ as

$$DF(A \rightarrow S|A) = tr\{P_ACOV_{S|A}P_A[(I - P_{X_1})(X_2\Gamma_2X_2' + \Gamma_0)(I - P_{X_1})]^+\} = \frac{\sigma_5^2}{\sigma_2^2 + \sigma_5^2 + \sigma_0^2} \cdot 2.$$

The DF smoothed from other FEs to $S|A$ are defined similarly; the DF smoothed from the grand mean, B and the AB interaction into $S|A$, $DF(GM \rightarrow\ S|A) = DF(B \rightarrow\ S|A) = DF(AB \rightarrow S|A) = 0$ as expected, because the grand mean is not smoothed, while B and AB are within-subject effects.

Similarly the error term $S \times B|A$'s covariance $COV_{S\times B|A}$ decomposes as

$$\begin{aligned} COV_{S\times B|A} &= P_{GM}COV_{S\times B|A}P_{GM} + P_A COV_{S\times B|A}P_A + P_B COV_{S\times B|A}P_B + \\ & \quad P_{AB}COV_{S\times B|A}P_{AB} + P_{S|A}QCOV_{S\times B|A}QP_{S|A} + Q_{S|A}QCOV_{S\times B|A}QQ_{S|A}, \end{aligned} \tag{4}$$

where $P_{S|A} = P_{X_{S|A}} = I_3 \otimes I_9 \otimes P_{1_5}$, and $Q_{S|A} = I - P_{S|A}$, and $P_{S|A}Q$ and $Q_{S|A}Q$ are both orthogonal projections since $P_{S|A}$ and Q commute. Then the part of DF smoothed from A into $S \times B|A$ (i.e., into error) is defined as

$$\begin{aligned} DF(A \rightarrow\ S \times B|A) &= tr\{P_A COV_{S\times B|A}P_A[(I - P_{X_1})(X_2\Gamma_2X_2' + \Gamma_0)(I - P_{X_1})]^+\} \\ &= \frac{\sigma_0^2}{\sigma_2^2 + \sigma_5^2 + \sigma_0^2} \cdot\ 2, \end{aligned}$$

which would not be credited to $S \times B|A$ if σ_2^2 went to $+\infty$. DF smoothed from B and $A \times B$ into $S \times B|A$ are defined similarly. DF smoothed from $S|A$ into $S \times B|A$ is

$$\begin{aligned} DF(S|A \rightarrow\ S \times B|A) &= tr\{P_{S|A}QCOV_{S\times B|A}QP_{S|A}[(I - P_{X_1})(X_2\Gamma_2X_2' + \Gamma_0)(I - P_{X_1})]^+\} \\ &= \frac{\sigma_0^2}{\sigma_5^2 + \sigma_0^2} \cdot\ 24. \end{aligned}$$

This is also the amount of DF $S \times B|A$ would lose if σ_5^2 went to $+\infty$ and FEs were all unsmoothed.

We define the "retained" DF of $S|A$ as the part of $S|A$'s DF that is not smoothed into or out of $S|A$, i.e., $DF_r(X_{S|A}) = DF(X_{S|A}) - DF(GM, A, B, AB \rightarrow\ S|A) = \frac{\sigma_5^2}{\sigma_5^2+\sigma_0^2} \cdot\ 24$, which is the variation $QCOV_{S|A}Q$ in (3), i.e., $tr\{QCOV_{S|A}Q[(I - P_{X_1})(X_2\Gamma_2X_2' + \Gamma_0)(I - P_{X_1})]^+\}$, and is also the DF $S|A$ would have if all FEs were unsmoothed, i.e., if $\sigma_2^2, \sigma_3^2, \sigma_4^2$ are all $+\infty$. As noted earlier, a prior distribution can be placed on retained DF as an alternative to placing a prior on DF itself; Section 4 gives an example.

DF smoothed from the grand mean, B, and AB into $S \times B|A$ are $DF(GM \rightarrow\ S \times B|A) = 0$, $DF(B \rightarrow\ S \times B|A) = \frac{\sigma_0^2}{\sigma_3^2+\sigma_0^2} \cdot\ 4$, $DF(AB \rightarrow\ S \times B|A) = \frac{\sigma_0^2}{\sigma_4^2+\sigma_0^2} \cdot 8$. So the retained DF of $S\times B|A$ is $DF_r(X_{S\times B|A}) = DF(X_{S\times B|A}) - DF(A \rightarrow\ S\times B|A) - DF(B \rightarrow\ S\times B|A) - DF(AB \rightarrow S\times B|A) - DF(S|A \rightarrow S\times B|A) = 96$, which comes from the variation of $Q_{S|A}QCOV_{S\times B|A}QQ_{S|A}$ in (4), and is also the DF $S \times B|A$ would have if $\sigma_j^2 = +\infty, j = 2, 3, 4, 5$. In other words, $S \times B|A$ as the lowest error term has no DF smoothed out of it and therefore retains all of its original 96 DF.

Source	Destination: DF in the smoothed fit.						
	GM	A	B	AB	$S\|A$	$S \times B\|A$	Conventional DF
GM	1	0	0	0	0	0	1
A	0	$\frac{\sigma_2^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$	0	0	$\frac{\sigma_5^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$	$\frac{\sigma_0^2}{\sigma_2^2+\sigma_5^2+\sigma_0^2} \cdot 2$	2
B	0	0	$\frac{\sigma_3^2}{\sigma_3^2+\sigma_0^2} \cdot 4$	0	0	$\frac{\sigma_0^2}{\sigma_3^2+\sigma_0^2} \cdot 4$	4
AB	0	0	0	$\frac{\sigma_4^2}{\sigma_4^2+\sigma_0^2} \cdot 8$	0	$\frac{\sigma_0^2}{\sigma_4^2+\sigma_0^2} \cdot 8$	8
$S\|A$	–	–	–	–	$\frac{\sigma_5^2}{\sigma_5^2+\sigma_0^2} \cdot 24$	$\frac{\sigma_0^2}{\sigma_5^2+\sigma_0^2} \cdot 24$	24
$S \times B\|A$	–	–	–	–	–	96	96
DF in fit	$DF(X_{GM})$	$DF(X_A)$	$DF(X_B)$	$DF(X_{AB})$	$DF(X_{S\|A})$	$DF(X_{S\times B\|A})$	135

Table 2: A balanced design: the DF portion of the smoothed ANOVA table

Table 2 bookkeeps DF flow in the model fit. In Table 2, rows are sources of variation except for the bottom row ("DF in fit") which is the column sum; columns are final locations of variation, except for the right-most column ("Conventional DF"), which is the row sum; the entry in a row and column is the piece of the row's total DF that, for given Γ_0 and Γ_2, is retained in or smoothed into the column's effect. Here, for example, only the between-subject factor A has DF smoothed into $S|A$. The bottom row is the sum of DF retained or smoothed into the column's effect in the fitted smoothed ANOVA, and is the same as in Table 1. Retained DF is on the diagonal of Table 2. Table 2 gives us the correct DF in the fit for each effect, rationalizing our definition of DF smoothed from a FE to each error term. Besides accounting for the variation in each effect, this table clarifies the various sources of information about the components of error variation.

The exact DF in the fit, pseudo-DF DF_p (i.e., approximate DF) and retained DF_r are identical in a balanced design with orthogonal parameterization and a single error term. In this special situation, all smoothed DF goes to the error term, so in the smoothed ANOVA table (Table 2), except for the error term, a column has only one nonzero element, on the diagonal, so the DF in the fit of the column's effect is its retained DF. The pseudo-DF of X_{2j} — the DF of X_{2j} in the simplified model including all nonsmoothed effects, smoothed effect X_{2j}, and the pure error — can be obtained by taking $\sigma_i^2 = 0$ for $i \neq 0$ or j in the expression of $DF(X_{2j})$ in the original model. In a balanced design with orthogonal parameterization and a single error term, $DF(X_{2j})$ depends only on $\frac{\sigma_j^2}{\sigma_0^2}$, as derived in HCSC, so the pseudo-DF is the same as the DF in the fit.

The amount and destination of SS smoothed out from each effect can be obtained from the same decomposition of covariance, equation (3) and (4). For example, the SS smoothed out from

A to $S|A$, written as $SS(A \rightarrow\ S|A)$ is

$$y'[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}P_ACOV_{S|A}P_A[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}y,$$

and SS smoothed out from $S|A$ to $S \times B|A$ (error), $SS(S|A \rightarrow\ S \times B|A)$ is

$$y'[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2}P_{S|A}QCOV_{S\times B|A}QP_{S|A}\cdot$$
$$[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)\cdot(I-P_{X_1})]^{+1/2}y.$$

As noted, three features of this example may reduce its generality: the Γ_{2j} that were used, the nesting of $S \times B|A$ within $S|A$ and $S|A$ within A, and the design's balance. As seen in equations (3) and (4), flow of DF and SS depends on decomposition of variance from RE and the error, which does not involve the specific value of Γ_{2j} of any FE. Thus for fixed effects (A, B, AB), the above results generalize immediately to $\Gamma_{2j} = \sigma_{2j}^2\Gamma_{2j}^0$ for known positive-definite Γ_{2j}^0, although the resulting expressions are not as explicit as those given above. However, it is not clear to which $\Gamma_{S|A}$ and $\Gamma_{S\times B|A}$ our results can be extended.

The nesting of $S \times B|A$ within $S|A$ permits the decomposition of $COV_{S\times B|A}$ in (4) and hence allows unambiguous declaration of the nesting effect ($S|A$) as the source and the nested effect ($S \times B|A$) as the destination. In crossed random effects, which have no such hierarchical relationship, DF and SS in the fit can be defined unambiguously, but it is unclear that variation can be partitioned as in (4) except by convention.

Finally, although balance is not necessary in obtaining effect-specific DF or SS, it appears to be important in deriving the flow of DF and SS smoothed out of effects. In unsmoothed ANOVA of unbalanced designs with RE, due to collinearity there is no unique way of specifying denominator DF for F-tests and no unique way to specify the "original" DF of a RE. For SS, even when there is no RE, there is no unique way to define an effect's "original" SS, e.g., SAS's Type I to Type IV SS (Chapter 12, SAS online Doc^{TM}: Version 8). So although in smoothed ANOVA we know the DF or SS of each effect in the model fit, there is no easy way to determine how much is smoothed out or into a given effect.

4 Irrigation study analyzed

This section shows some analyses of the Irrigation study (Kinsey, 2006) which has the design used in Section 3 and data given in Table 3. All of the capabilities of SANOVA displayed in HCSC's

Section 4 could also be displayed here, for example, subgroup analysis, or fixing the DF in the fit for a group of effects and requiring those effects to compete for those DF. However, doing so would have no novelty, so we restrict our attention to those aspects of SANOVA developed in this paper.

	$A1$ (NAOCL)					$A2$ (NAOCL/EDTA)					$A3$ (NAOCL/MTAD)				
patient	$B1$	$B2$	$B3$	$B4$	$B5$	$B1$	$B2$	$B3$	$B4$	$B5$	$B1$	$B2$	$B3$	$B4$	$B5$
$S1$	12.751	5.170	5.701	2.347	11.627	10.901	0.591	3.143	5.205	20.044	11.062	6.196	8.430	3.457	5.515
$S2$	10.387	3.160	1.764	8.293	8.415	6.146	5.672	2.337	4.232	8.745	8.990	4.961	11.146	6.324	1.810
$S3$	13.170	4.596	6.879	5.531	1.568	12.833	4.335	0.898	4.723	17.187	9.346	21.181	10.793	21.039	0.717
$S4$	18.166	2.446	11.838	12.353	16.706	17.255	7.896	10.637	4.268	1.988	1.112	13.427	6.163	10.563	16.135
$S5$	8.497	4.295	3.646	6.101	15.159	1.899	0.869	7.254	1.650	3.844	12.130	5.294	4.348	0.035	3.215
$S6$	2.023	0.129	0.337	3.597	1.272	5.191	2.324	2.443	6.139	4.396	9.402	1.240	2.245	5.793	1.630
$S7$	5.861	0.753	0.673	2.253	2.787	8.893	0.282	2.040	3.406	1.872	4.175	0.608	0.410	0.110	1.314
$S8$	11.839	1.424	2.454	5.707	14.193	6.454	13.143	0.641	0.629	1.489	4.730	8.759	10.234	1.338	7.606
$S9$	3.312	3.772	6.042	15.531	9.262	1.244	11.461	0.950	2.503	16.525	9.929	6.801	9.118	5.947	7.718

Table 3: Irrigation data

The Irrigation study compared three solutions for rinsing ("irrigating") a tooth's prepared root canal before sealing it. The outcome measure was the strength in mega-Pascals (MPa) of the bond between the remaining root structure and a so-called post inserted into the canal and used as the foundation of a crown. The between-subject factor A was irrigant, with three levels: NAOCL, NAOCL/EDTA, and NAOCL/MTAD. The within-subject factor B was segment of the tooth root, with levels 1, 2, 3, 4, and 5 in order from the top of the tooth's root (cervix) to the bottom (apex). Markov Chain Monte Carlo (MCMC) was used to sample from the posteriors; the Appendix derives the necessary posterior distributions. Table 4 summarizes the posterior means of the variances σ_5^2 and σ_0^2 and of various effects' DF, under three priors: flat on DF, flat on retained DF (DF_r), and flat on pseudo-DF (DF_p), called π, π_r and π_p, respectively. Each prior was flat on the legal range of $DF,\ DF_r$, or DF_p, as appropriate. The induced distributions on the variance ratios $r_j = \frac{\sigma_j^2}{\sigma_0^2}$ are

$$\begin{aligned}
\pi(r_2, r_3, r_4, r_5) &= \frac{1}{50}\left[\frac{48}{(r_2 + r_5 + 1)^2(r_5 + 1)} + \frac{4}{(r_2 + r_5 + 1)^3}\right] \cdot \frac{1}{(r_3 + 1)^2} \cdot \frac{1}{(r_4 + 1)^2}, \\
\pi_r(r_2, r_3, r_4, r_5) &= \frac{1}{(r_2 + r_5 + 1)^2(r_5 + 1)} \cdot \frac{1}{(r_3 + 1)^2} \cdot \frac{1}{(r_4 + 1)^2}, \\
\pi_p(r_2, r_3, r_4, r_5) &= \frac{1}{(r_2 + 1)^2} \cdot \frac{1}{(r_5 + 1)^2} \cdot \frac{1}{(r_3 + 1)^2} \cdot \frac{1}{(r_4 + 1)^2}\ .
\end{aligned}$$

Clearly $r_3,\ r_4$, and (r_2, r_5) are independent of each other in all three priors, and the marginal of r_i is $\frac{1}{(r_i+1)^2},\ i = 3,\ 4,\ 5$ in all three priors. (The integrals for r_5 can be done explicitly for π and π_r;

	$E(\sigma_5^2\|y)$	$E(\sigma_0^2\|Y)$	$E[DF(A)\|y]$	$E[DF(B)\|y]$	$E[DF(AB)\|y]$	$E[DF(S\|A)\|y]$
range	—	—	(0,2)	(0,4)	(0,8)	(0,26)
prior on DF	18.771	19.801	0.708	2.290	3.253	11.196
prior on DF_r	18.725	19.792	0.711	2.302	3.247	11.213
prior on DF_p	19.031	19.759	0.565	2.320	3.265	11.390

Table 4: Summary of posterior estimates for smoothing variances of random effects and DF of fixed effects and random effects in the Irrigation study.

we omit the details.) However, the three marginal priors on r_2 are different:

$$\begin{aligned}\pi(r_2) &= \frac{1}{50}[\frac{2}{(r_2+1)^2} + 48(\frac{log(r_2+1)}{r_2^2} - \frac{1}{r_2(r_2+1)})],\\ \pi_r(r_2) &= \frac{log(r_2+1)}{r_2^2} - \frac{1}{r_2(r_2+1)},\\ \pi_p(r_2) &= \frac{1}{(r_2+1)^2}.\end{aligned}$$

The variance ratios r_2 and r_5 are correlated in priors induced by flat priors on DF and DF_r, but independent in the prior induced by DF_p. The joint priors on $(s_2, s_5) = (log(r_2), log(r_5))$ from DF and DF_r are displayed in Figure 1; they look almost identical due to the similarity of their expressions. Figure 2 plots the three marginal priors of $s_2 = log(r_2)$; we can see that DF and DF_r induce almost identical marginal priors on s_2 and thus on r_2. Results from all three posteriors are close (Table 4); those from the flat priors on DF and DF_r are almost identical. Although the three priors give similar exact posterior mean $DF(A)$ in Table 4, their (approximate) pseudo-DF range from 0.9 to 1, noticeably different from the exact values of 0.565 to 0.7. This is a counterexample to Ruppert et al's observation (2003, Section 8.3) that the exact DF and the pseudo-DF were almost the same in all examples that they had considered. DF and pseudo-DF diverge because $R(X_A)$ is a subspace of $R(X_{S|A})$, and this collinearity is almost certainly the "culprit". Cui et al (2009), Section 3.2 gives another counterexample in which DF and pseudo-DF diverge, again because of similarly severe collinearity.

Table 5 displays the smoothed ANOVA table from the actual data using the flat prior on DF π, showing posterior mean DF and SS. The four smoothed effects are smoothed substantially, each retaining less than half of its original DF, except for B. $S \times B|A$ has about 20 extra DF smoothed into it, mostly from $S|A$.

Source	Destination in the smoothed fit.						
	GM	A	B	AB	$S\|A$	$S \times B\|A$	Total DF of effect
GM	1 (5387.9)	0	0	0	0	0	1 (5387.9)
A	0	0.7 (9.3)	0	0	0.6 (7.6)	0.7 (9.5)	2 (26.4)
B	0	0	2.3 (149.9)	0	0	1.7 (112.7)	4 (262.7)
AB	0	0	0	3.3 (116.6)	0	4.7 (170.2)	8 (286.8)
$S\|A$	–	–	–	–	10.7 (437.9)	13.3 (547.4)	24 (985.3)
$S \times B\|A$	–	–	–	–	–	96 (1865.3)	96 (1865.3)
net DF/SS	1 (5387.9)	0.7 (9.3)	2.3 (149.9)	3.3 (116.6)	11.2 (445.5)	116.5 (2705.1)	135 (8814.5)

Table 5: Irrigation study: the smoothed ANOVA table. In each cell, the value is "$E(DF|y)$ $(E(SS|y))$"

5 Discussion

This paper presented a generalization of smoothed ANOVA as proposed in HCSC. Section 2 extended their theory to general ANOVA designs, gave a smoothed ANOVA table, and showed how to specify DF-induced priors on REs and smoothed FEs.

Section 3 rationalized the flow of DF out of smoothed effects into error terms in a balanced design with nested error terms. It is unclear whether this extends in any generality to unbalanced designs. One *ad hoc* extension is to nest the unbalanced design in the minimal larger balanced design with the same set of variance smoothing parameters. After obtaining posterior information on the smoothing parameters, one could use the flow of DF and SS from the bigger design to describe DF flow in the actual design.

Also, the tidy expression of DF-induced *priors* from the Irrigation study (Section 3) depends on the balance of the design. The same trick mentioned in the preceding paragraph – nesting the unbalanced design in the minimal larger balanced design — could be used to specify simpler priors on DF for unbalanced designs, while using the actual design to make inference. Although the obvious intuition is that it makes little difference when the design not too far from being balanced, further investigation is needed into this matter.

One concern with smoothed ANOVA is collinearity among covariates. Collinearity between two nonsmoothed effects reduces to the familiar collinearity problem in ordinary regressions. The new problem is how unsmoothed FE change estimation of smoothed effects when they are collinear, and how two collinear smoothed effects influence each other. Intuition suggests that when a smoothed effect is collinear with a nonsmoothed effect, the smoothed effect will account for less variation than it otherwise would, while two correlated smoothed effects compete in an as-yet undescribed manner (Cui et al 2009). However, study of this problem is in its infancy, and as yet nothing is known about when two such effects will mask each other and perhaps lead to false-positive findings. This question points to rich opportunities for future research.

6 Appendix

Proofs for Section 2.2

($SS.a$) Assume the model in HCSC, i.e., in model (1), $X_1'X_1 = nI$, $X_{2k}'X_{2k} = nI$, $X_{2k}'X_{2k'} = 0$ if $k \neq k'$, $X_{2k}'X_1 = 0$, $\Gamma_{2k} = diag(1/\eta_{j(k_1)}, \cdots, 1/\eta_{j(k_{n_k})})$, where n_k is the dimension of X_{2k}, $\Gamma_0 = \frac{1}{\eta_0}I_n$, and the function $j(\cdot)$ is as defined in HCSC. Then under the definition in Section 2.2,

$$SS(X_{2k};\Gamma_0,\Gamma_2) = y'X_{2k}diag\left((n+\frac{\eta_{j(k_1)}}{\eta_0})^{-1},\cdots,(n+\frac{\eta_{j(k_{n_k})}}{\eta_0})^{-1}\right)X_{2k}'y.$$

It is easy to show that $X_{2k}\Gamma_{2k}X_{2k}' = \frac{X_{2k}}{\sqrt{n}}diag(n/\eta_{j(k_1)},\cdots,n/\eta_{j(k_{n_k})})\frac{X_{2k}'}{\sqrt{n}}$, and

$$[(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]^{+1/2} =$$
$$\sum_{m=1}^{l}\frac{X_{2m}}{\sqrt{n}}diag\left((\frac{n}{\eta_{j(m_1)}}+\frac{1}{\eta_0})^{-1/2},\cdots,(\frac{n}{\eta_{j(m_{n_m})}}+\frac{1}{\eta_0})^{-1/2}\right)\frac{X_{2m}'}{\sqrt{n}}+(\frac{1}{\eta_0})^{-1/2}\frac{X^c}{\sqrt{n}}\frac{X^{c\prime}}{\sqrt{n}},$$

where X^c is any set of columns forming a basis for the orthogonal complement of the column space of $[X_1, X_2]$. The conclusion follows from the definition in Section 2.2.

($SS.b$) From the definition of SS in Section 2.2,

$$\sum_{1}^{J}SS(X_{2j};\Gamma_0,\Gamma_2)+SS(\epsilon;\Gamma_0,\Gamma_2) = y'H^{+1/2}HH^{+1/2}y$$

where $H = [(I-P_{X_1})(X_2\Gamma_2X_2'+\Gamma_0)(I-P_{X_1})]$. Since for any nonnegative definite matrix H, $H^{+1/2}HH^{+1/2} = P_H$, then under the assumption that Γ_0 is positive definite, the column space of H is the column space of $I-P_{X_1}$. Therefore, $P_H = I-P_{X_1}$, and $\sum_1^J SS(X_{2j};\Gamma_0,\Gamma_2)+SS(\epsilon;\Gamma_0,\Gamma_2) = y'(I-P_{X_1})y$. Together with $SS(X_1) = y'P_{X_1}y$ the conclusion follows.

($SS.c$) Under the assumption in ($SS.c$), $E(y|\Gamma_0,\Gamma_2) = \mu$ and $Var(y|\Gamma_0,\Gamma_2) = X_2\Gamma_2X_2'+\Gamma_0$. Then

$$\begin{aligned}E(SS(X_1)|\Gamma_0,\Gamma_2) &= E(y'P_{X_1}y|\Gamma_0,\Gamma_2)\\ &= tr\left((X_2\Gamma_2X_2')P_{X_1}\right)+\mu'P_{X_1}\mu\\ &= tr\left((X_2\Gamma_2X_2')P_{X_1}\right)+\mu'\mu\end{aligned}$$

since $\mu \in R(X_1)$. Also,

$$\begin{aligned} E(SS(X_{2j})|\Gamma_0, \Gamma_2) &= E\left(y' H^{+1/2}(X_{2j}\Gamma_{2j}X_{2j}')H^{+1/2}y|\Gamma_0, \Gamma_2\right) \\ &= tr\left(H^{+1/2}(X_{2j}\Gamma_{2j}X_{2j}')H^{+1/2}(X_2\Gamma_2X_2' + \Gamma_0)\right) + \mu' H^{+1/2}(X_{2j}\Gamma_{2j}X_{2j}')H^{+1/2}\mu \\ &= tr\left(H^{+1/2}(X_{2j}\Gamma_{2j}X_{2j}')H^{+1/2}H\right) + \mu' H^{+1/2}(X_{2j}\Gamma_{2j}X_{2j}')H^{+1/2}\mu \\ &= tr\left((X_{2j}\Gamma_{2j}X_{2j}')(I - P_{X_1})\right) \end{aligned}$$

since $H^{+1/2}HH^{1/2} = I - P_{X_1}$ as shown in the proof for $(SS.b)$, and $H^{+1/2}$ is symmetric and $N(H^{+1/2}) \supseteq R(X_1)$, where $N(H^{+1/2})$ is the null space of $H^{+1/2}$.

Distributions for MCMC in Section 4

This is a derivation of posteriors of the mean-structure parameter θ and the smoothing variances $\sigma_i^2, i = 0, 2, \cdots, 5$ in the Irrigation example. The joint posterior of θ and $\{\sigma_i^2\}$ is

$$f(\theta, \sigma_i^2|\, Y) \propto \pi(\sigma_0^2, \sigma_2^2, \cdots, \sigma_5^2)|\Sigma|^{-1/2} exp(-\frac{1}{2}(Y - X\theta)'\Sigma^{-1}(Y - X\theta)).$$

where $Y = \binom{y}{0_{41}}$, $\Sigma = diag(\Gamma_0, \Gamma_2)$, $\Gamma_0 = \sigma_0^2 I_{135}$, $\Gamma_2 = diag(\sigma_2^2 I_2,\ \sigma_3^2 I_4,\ \sigma_4^2 I_8,\ \sigma_5^2 I_{27})$, and

$$X = \begin{pmatrix} X_1 & X_2 \\ 0 & I_{41} \end{pmatrix}.$$

Integrate out θ and change variables from $\{\sigma_i^2\}$ to $(\eta_0, \mathbf{r})$, where $\eta_0 = 1/\sigma_0^2$ and $r_i = \frac{\sigma_i^2}{\sigma_0^2}$, $i = 2, \cdots, 5$. This gives the joint marginal posterior of $(\eta_0,\ \mathbf{r})$:

$$f(\eta_0, r|y) \propto\ \pi(\eta_0, r)\eta_0^{\frac{135-1}{2}} exp(-\frac{1}{2}\eta_0 W(r))(r_5 + 1)^{-12}(r_2 + r_5 + 1)^{-1}(r_3 + 1)^{-2}(r_4 + 1)^{-4},$$

where $\eta_0 W(r) = Y'\Sigma^{-1}Y - Y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y = \eta_0 y'[(I - P_{X_1})(X_2\Gamma_2X_2' + \Gamma_0)(I - P_{X_1})]^+ y$. If η_0 has a gamma prior independently of r, $\pi(\eta_0, r) \propto \pi(r)\eta_0^{\alpha-1} exp(-\lambda\eta_0)$, then the marginal posterior of $(r_2, \cdots, r_5)$ is

$$f(r|Y) \propto \pi(r)(2\lambda + W(r))^{-(\frac{135-1}{2}+\alpha)}(r_5 + 1)^{-12}(r_2 + r_5 + 1)^{-1}(r_3 + 1)^{-2}(r_4 + 1)^{-4},$$

and the conditional posterior of η_0 given r is a gamma distribution,

$$f(\eta_0|r, Y) \propto \eta_0^{(\frac{135-1}{2}+\alpha-1)} exp(-\eta_0(\lambda + \frac{W(r)}{2})).$$

Following the derivation in HCSC Section 2.2, the posterior of θ is a multivariate-t on $\nu = 135-1+2\alpha$ DF, with center $\hat{\theta} = \Sigma_r^{-1}X_d'y$ and dispersion matrix $\frac{2\lambda+W(r)}{\nu}\Sigma_r^{-1}$, where $X_d = [X_1,\ X_2]$, $\Sigma_r = X_d'X_d + \begin{pmatrix} 0 & 0 \\ 0 & \Gamma_2^{-1}/\eta_0 \end{pmatrix}$. Note that Σ_r is a function of $(r_2, \cdots, r_5)$ and not the individual η_j.

Acknowledgement

Prof. Peter McCullagh of the University of Chicago generously provided very useful comments despite deep disagreements about our formulation.

References

Besag J (2002). Discussion of McCullagh (2002). *The Annals of Stat.*, 30, 1267-1277.

Cui Y, Hodges JS, Kong X & Carlin BP (2009). Partitioning degrees of freedom in hierarchical and other richly-parameterized models. *Technometrics*, accepted.

Gelman A (2004). Parameterization and Bayesian modeling. *J. Amer. Stat. Assoc*, 99:537-545.

Gelman A (2005). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:1-19.

He Y, Hodges JS, Carlin BP (2007). Re-considering the variance parameterization in multiple precision models. *Bayesian Analysis*, 2:529- 556.

Hodges JS, Cui Y, Sargent DJ & Carlin BP (2007). Smoothing balanced single-error-term analysis of variance. *Technometrics*, 49, 12-25.

Hodges JS & Sargent DJ (2001). Counting degrees of freedom in hierarchical and other richly-parameterized models. *Biometrika*, 88, 367- 379.

Kinsey L (2006). The effect of the intracanal irrigant MTAD on bond strength of resin cements within the canal. MS thesis, Division of Endodontics, School of Dentistry, University of Minnesota.

McCullagh P (2002). What is a statistical model? (with discussion) *Annals of Stat.*, 30, 1225-1310.

Ruppert D, Wand MP & Carroll RJ (2003). *Semiparametric Regression*. Cambridge Univ Press.

SAS online Doc^{TM}: Version 8 (1999). SAS Institute Inc., Cary, NC, USA,
URL *http://v8doc.sas.com/sashtml/*

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

Schott JR (1997). *Matrix Analysis for Statistics*. New York: Wiley.

Zhang Y, Hodges JS & Banerjee S (2007). Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing. Research Report 2007-031, Division of Biostatistics, University of Minnesota. Tentatively accepted by Annals of Applied Statistics.

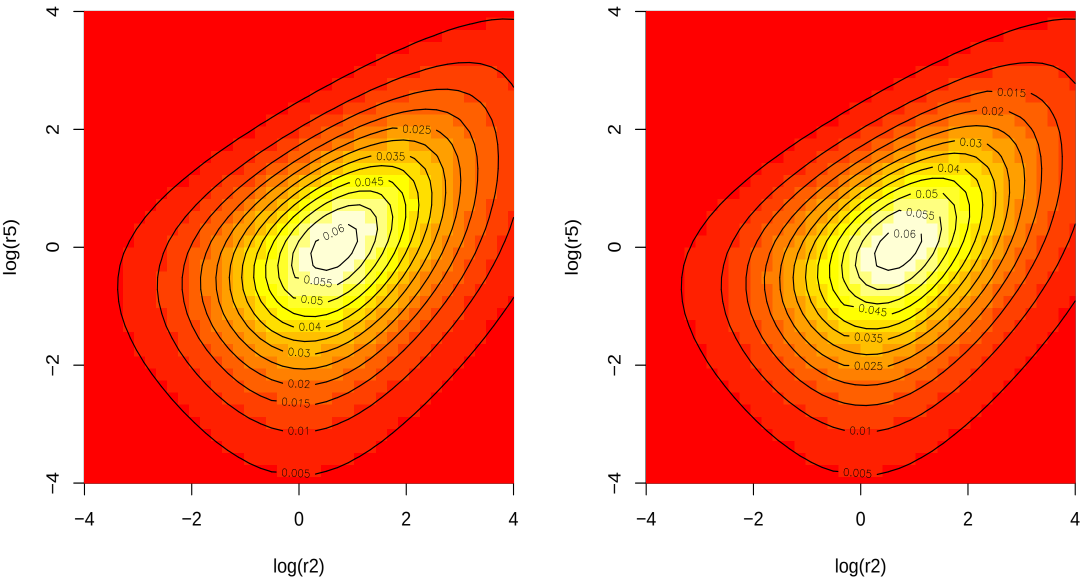

Figure 1: For the Irrigation example, prior contour plot on $(log(r_2), log(r_5))$, left from DF q, right from retained DF q_r.

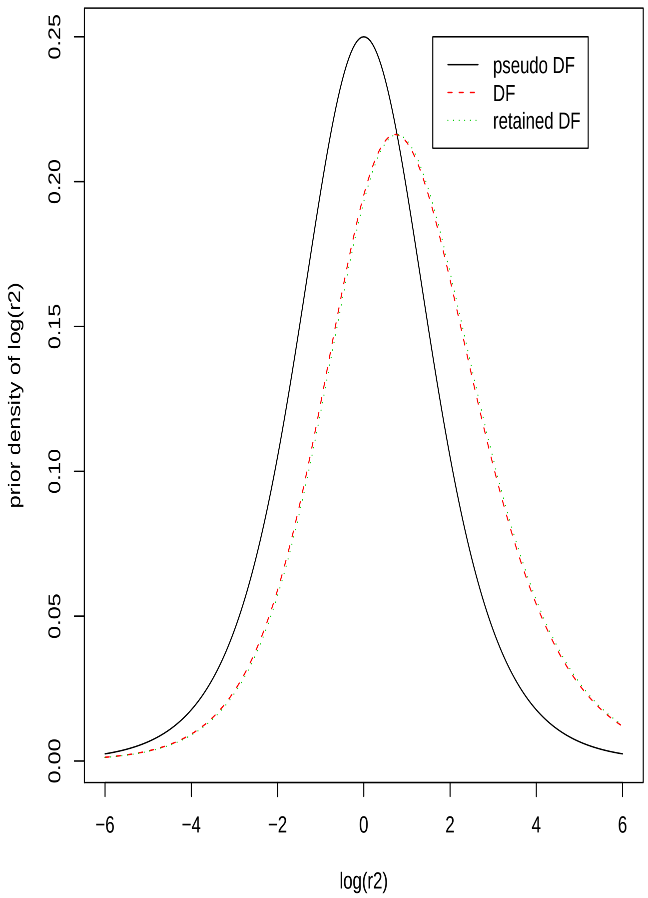

Figure 2: For the Irrigation example, marginal prior of $s_2 = log(r_2)$. The densities for DF q and retained DF q_r are almost exactly superimposed and thus visually indistinguishable.