Arctic, Antarctic, and Alpine Research, Vol. 42, No. 4, 2010, pp. A1–A6

Appendix. Carotenoid Pigmentation in Antarctic Heterotrophic Bacteria as a Strategy to Withstand Environmental Stresses

Mark Greenwood
Markus Dieser and
Christine Foreman

DOI: 10.1657/1938-4246-42.4.A1

Introduction

This appendix is intended to explain the statistical methods employed in more detail and to further explore the different models considered. The first section of the appendix deals generally with the additive mixed models employed, the second section explores the results from the freeze-thaw experiment, and the final section discusses the solar radiation experiment results.

Statistical Models and Methods

In the experiments conducted, each bacterial strain of interest is measured across repeated exposures (experimental treatments), either exposure to solar radiation or freeze-thaw cycles, and over the same period of time in the control group. A trial is defined as the sequence of measurements of a strain under either treatment or control conditions. Measurements for each trial depend upon initial cell concentrations (although attempts were made to start out with similar cell concentrations, cell numbers differed between the treatment and control groups for the same strain as well as between strains). Further, we assume that all measurements within a trial are correlated. At each time point where a measurement is taken, triplicate samples from a bacterial stock solution were prepared to count colony forming units (CFUs). To account for the repeated measures nature of the experiment and the different initial concentrations for each individual, we use a mixed effect model (Pinheiro and Bates, 2004) with a random effect for each trial. Other repeated measures models could have also been considered, but a random intercept models the source of similarity within the trials. A random effect for each subject also induces a correlation structure similar to compound symmetry. An additional advantage of using a random intercept is that estimates of the initial concentrations for each trial are available. This modeling framework allows comparison of different fixed effects to address the research questions of interest regarding differences due to pigmentation and control versus treatment groups, after adjusting for the random, trial specific, starting concentrations. To provide more normally distributed responses, all observations are transformed using a log base 10 transformation, with 1 added to all observations since observations range from 0 to 7.55e10.

RANDOM EFFECTS

All the models considered contain some combination of fixed and random effects. The simplest model that contains a trial-specific random effect but no group or time effects for the response variable

$$Y_{ijk} = \log 10\left(CFU_{ijk} + 1\right) \tag{A1}$$

for trial i at exposure time j, and replicate k, is

$$Y_{ijk} = \alpha + b_i + \varepsilon_{ijk}. \tag{A2}$$

The intercept, α, is the only fixed effect, suggesting no exposure time, pigmentation, or treatment differences. The random components of the model are a trial-specific random intercept [$b_i \sim N(0,\sigma_b^2)$] and the residual random variation of $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$. Maximum likelihood estimation is used to estimate these variance components as well as any additional fixed effects considered. Restricted maximum likelihood estimation (REML) is not used here since a model selection criterion is used to select across different non-nested fixed effect structures, which is not possible in a REML framework because those results are conditional on any particular set of fixed effects. For a recent, accessible introduction to mixed models see Zuur et al. (2009).

The random effect (b_i) at the trial level is used for two reasons. First, it allows for random variation in the initial CFUs at the beginning of each trial. Every attempt was made to standardize the starting concentrations, but it is not possible to make them exactly equal. Allowing the starting point for each strain to be random allows the fixed effects to consider the differences in the means over exposure time after adjusting for this random starting point for each individual trial. The second reason for the random intercept is that it induces a slightly restricted version of a compound-symmetric error structure. Compound-symmetric correlation is commonly used with repeated measures data analyses as it provides a model where all the measurements for a trial are equally correlated regardless of when they were measured, and the measurements across trials are modeled as being independent of one another. The random intercept model restricts this correlation to be positive. These assumptions are justified based on the experimental design.

For the UV exposure experiment, non-constant variance was detected in the initial model diagnostics. In order to accommodate this additional aspect of the data set, a different variance coefficient was estimated for each of the four different exposure levels. This is easily incorporated in the mixed model framework, adding three additional parameters that must be estimated. For the UV exposure results, models that assume equal variance are directly compared to those with this particular variance structure.

FIXED EFFECTS

Fixed effects are used in the previous mixed model structure to address differences based on pigmentation and estimate the

1523-0430/10 $7.00

mean log10-CFUs at each exposure time. Two different modeling approaches were considered for the changes over exposure time, either estimating a unique mean for each exposure time or estimating a function of exposure time. In both cases, we consider adjusting those effects based on grouping variables of interest such as pigmentation or treatment and control groups (or the interaction between them). We are generally interested in differences in the response trajectories between the groups. Those differences can take many different forms, from different linear trends for each level to groups that change quickly initially and then level off. For the freeze-thaw cycle experiment, one model that could be considered is

$$Y_{ijk} = \alpha_Y + b_i + \beta_Y \text{Cycle} + \varepsilon_{ijk} \quad \text{(A3)}$$

for the pigmented strains and

$$Y_{ijk} = \alpha_N + b_i + \beta_N \text{Cycle} + \varepsilon_{ijk} \quad \text{(A4)}$$

for the non-pigmented strains. This has a fixed linear trend effect for cycle that interacts with the pigmentation of the strains, but ignores any treatment effects. A less restrictive fixed cycle effect involves estimating a different parameter for each level of cycle for the initial two level version of the pigmentation group,

$$Y_{ijk} = \alpha_Y + b_i + \tau_{jY} + \varepsilon_{ijk} \text{ or } Y_{ijk} = \alpha_N + b_i + \tau_{jN} + \varepsilon_{ijk}, \quad \text{(A5)}$$

where $\alpha_Y + \tau_{jY}$ provides the mean at time j in the pigmented group. With 12 different freeze-thaw cycle exposure times measured and the random effect, this model requires estimation of 26 parameters, whereas the linear trend model above uses only 6 parameters since it exploits the quantitative aspect of the Cycle variable. The estimates from the two models might be very similar, but the linear trend model could be a more efficient parameterization, even with both models demonstrating an important pigmentation effect. The linear effect for Cycle might be too restrictive and the typical repeated measures model with a unique value for each level of Cycle may be overly complex. Additionally, when the control group is also considered, this full model is not estimable since no measurements were taken for cycle 1 in the control group because no change was expected to occur in control group counts in that time frame. More complicated functions of Cycle are also possible, such as various orders of polynomials. We found the best results using an alternative to polynomials using nonparametric functions of exposure in the mixed models; however, some polynomial models are also presented for comparison in the model selection results.

Wood (2006) discussed methods for incorporating spline-based smooth fixed effects in mixed models, such as those discussed above, which are called additive mixed models (AMMs). We use cubic regression splines, which are piecewise cubic polynomials with continuity conditions imposed at "knots" where the polynomials are joined. Increasing the number of knots increases both the complexity and the degrees of freedom (*df*) of the effect. A linear effect (1 *df*) is the simplest special case for the regression spline; the previous model with a linear Cycle effect is a possible result in our AMMs. Smooth effects in AMMs are defined as s(x), such as s(Cycle), which would suggest a smooth effect of Cycle in the model. While all s(x) fixed effects in each set of models start with the same maximum complexity (described through their *df*), their smoothness is estimated using additional random effects (for more details see Wood [2006] or Zuur et al. [2009]), resulting in penalized regression spline estimates of the s(x)'s. The result of this method is that the effective degrees of freedom (*edf*) are estimated for each s(x) in the model, with a lower limit of 1 *edf* (linear) and an upper limit at the number of unique points observed for x minus 1. It is possible to consider interactions between smooth terms, s(x), and a categorical or grouping variable, referred to as variable z. An interaction between s(x) and z leads to a different s(x) for each level of z and a different intercept for each level of the grouping variable as well. We estimate models that consider different functions of exposure and interactions with different groups based on treatments and pigmentation as well as simplified versions of these models to select a model(s) to interpret. All results in the paper are generated using either the nlme (Pinheiro et al., 2009) or mgcv (Wood, 2006) packages in R (R Development Core Team, 2009).

Both experiments contained grouping information based on pigmentation information. Further, the freeze-thaw experiment had different treatments: exposure to stresses or not (control). The freeze-thaw experiment is discussed in detail in the section "Results for Freeze-Thaw Cycle Experiment" and the solar radiation experiment in the section "Results for the Solar Radiation Exposure Experiment." In the solar radiation experiment, fewer exposure time points were measured than in the freeze-thaw cycle experiment so the differences between AMMs and more conventional polynomial techniques are smaller. Additionally, some more conventional models are actually equivalent as, for example, there is no difference between using the cubic polynomial model and the repeated measures model since only 4 exposure times are considered.

The term "additive" can be used in two ways related to the models considered here: (1) to describe the role of s(x) in the AMM, and (2), more typically, to describe fixed effect terms in an ANOVA-type model that describe a "mean shift" across the levels of that factor, but that do not interact with any other terms in the model. After adjusting for trial specific differences in the random effect, we would expect all the groups to start at the same value for an exposure of 0. This suggests that simple ANOVA additive models that imply different intercepts for the levels and, for example, the same function of exposure, are not of interest as candidate models. We do not constrain all the models to have the same intercept, as that can cause poor fit for models with different slopes as a function of exposure, but also did not consider models that only provided different intercepts for the different groups.

MODEL SELECTION

Generating accurate hypothesis tests in mixed models and even more so in additive mixed models is a topic of much current research and discussion (Wood [2006] and more recent documentation related to the mgcv package). It is also useful only when comparing nested models, which is not the case in either experiment. For these reasons, we chose to use a model selection criterion, the

$$\text{AIC} = -2\log(\text{likelihood}) + 2p, \quad \text{(A6)}$$

originally developed in Akaike (1973), to compare models. We define p as the number of parameters that must be estimated, including the effective degrees of freedom from the additive components when they are included in the model. Selecting the model with the minimum AIC selects a model with the estimated smallest expected prediction error across data that could be obtained from the population in repeated samples like ours. It also conforms with a version of the principle of parsimony. We report ΔAICs, which are differences between the observed AICs and the model with the smallest AIC. A difference of around 2 ΔAIC units is sometimes suggested as a "large" difference between models;

TABLE A1

Model selection results for the freeze-thaw cycle experiment.

Model	Log likelihood	DF	AIC*	ΔAIC
s(Cycle)*Pigment3*Treatment	−356.78	32.33	778.21	0.00
Cycle6*Pigment3*Treatment	−348.12	44.00	784.24	6.03
Cycle5*Pigment3*Treatment	−359.33	38.00	794.66	16.45
Cycle5*Pigment3+Cycle5*Treatment	−474.51	26.00	1001.03	222.82
s(Cycle)*Pigment3+s(Cycle)*Treatment	−483.04	20.14	1006.36	228.15
Cycle*Pigment3*Treatment	−576.20	14.00	1180.40	402.19
Cycle*Pigment3+Cycle*Treatment	−589.46	10.00	1198.92	420.71
s(Cycle)*Pigment2*Treatment	−599.49	16.98	1232.94	454.73
s(Cycle)*Treatment	−606.04	12.63	1237.34	459.13
s(Cycle)*Pigment2+s(Cycle)*Treatment	−608.26	13.58	1243.68	465.47
factor(Cycle)*Pigment3	−621.34	38.00	1318.68	540.47
s(Cycle)*Pigment3	−649.85	13.32	1326.34	548.13
Cycle*Pigment2+Cycle*Treatment	−657.43	8.00	1330.87	552.66
Cycle*Pigment2*Treatment	−655.94	10.00	1331.88	553.67
factor(Cycle)*Pigment2	−697.85	26.00	1447.70	669.49
s(Cycle)*Pigment2	−714.84	9.72	1449.13	670.92
Constant mean	−740.23	3.00	1486.46	708.25

* AIC = Akaike's Information Criterion.

larger ΔAICs suggest more support for the top-ranked model (Burnham and Anderson, 2002).

CONFIDENCE INTERVALS FOR NONPARAMETRIC EFFECTS

The confidence regions for the mean log-CFU reported in Figures 2, 3, and 4 in the main paper are approximate 95% Bayesian confidence intervals. They are based on using the Bayesian posterior covariance matrix for the parameters (Wood, 2006, p. 189). These are reported to perform better than the typical frequentist intervals (even when interpreted as frequentist intervals) so they are suggested as the default method for generating confidence regions. The intervals are presented mainly to allow visualization of the uncertainty in the estimation of the true mean at different exposure times. They were not formally used to choose amongst the models, but provide intervals for the likely value of the mean as a function of exposure times for the groups considered.

By re-parameterizing our models to contain an estimate for a baseline category and deviations from it for each group, we can estimate contrasts between groups as a function of exposure. Wood (2006, p. 243) discussed the details of estimating the difference between two groups in an additive mixed model. These provide estimates (and CIs) for the difference between any two groups of interest. In each experiment, we had *a priori* interests in comparing the pigmented and non-pigmented strains in their responses to the treatments. So in each experiment, we only had one planned contrast and do not adjust our intervals in any way. The estimated contrasts are displayed in Figure 4 in the main paper.

Results for Freeze-Thaw Cycle Experiment

In this experiment, the cycle, pigmentation group, and treatment were considered as fixed effects. We initially worked with two categories for pigmentation, present or absent (labeled Pigment2 in the models below). However, residual diagnostics even for the top models were problematic when ANT 11 was included with the remaining pigmented strains. Its behavior in the treatment group (and to a lesser degree in the control group) is completely different from all other strains considered in the study, suggesting that it is an outlier with respect to its "group." Instead of deleting this unusual observation, we accommodated it by creating a new group, creating three "pigmentation" categories: absent, present (not ANT 11), and present (ANT 11), labeled Pigment3 in results below. Note that ANT 11 does contain pigmentation, but at much lower concentrations than other similar carotenoids in the study. Combining the three pigmentation groups with the two treatments, there are now six levels to the combined possibilities for the overall grouping variables. The residual diagnostics were dramatically improved by allowing ANT 11 to be its own pigmentation group. We do not focus on its interpretation in as much detail as for the other groups, which have three strains per group that seem to have similar responses to the freeze-thaw exposure.

The favored model for the freeze-thaw experiment involved different additive (smooth) components for the three pigmentation groups and treatment/control levels (six different smooth functions of the number of cycles). In Appendix Table A1, results for 17 different models are presented. The models involving different orders of polynomial functions of Cycle are provided for comparison for those who are more familiar with those methods. The AMMs can provide very similar group estimates to polynomial-based models but, as the results suggest here, provide a more efficient (and easier) method of estimating the exposure effects as they interact with different groupings of the strains. All models considered the four- and six-level interactions between the two versions of pigmentation (Pigment2 and Pigment3) and treatment/control, since the four-level version was originally of interest and the six-level version was identified through residual diagnostics. Additionally, interactions between Cycle or s(Cycle) and the pigmentation and treatment effects, but not both, were also considered. Finally, removing each Cycle by group effect was considered.

Based on the ΔAIC results in Appendix Table A1, the difference between the AMM considering s(Cycle)*Pigment3*Treatment and the next top model containing a sixth order polynomial is reasonably large at 6 AIC units. This difference in AICs is partially due to the 32 *edf* used in the AMM compared to the 44 *edf* in the polynomial-based model. It may be possible to

TABLE A2

Model selection results for the solar radiation experiment.

Model	Variance	Log likelihood	DF	AIC*	ΔAIC
s(Exposure)*Pigment	Unequal	−66.65	8.21	149.74	0.00
Quadratic Exposure*Pigment	Unequal	−67.66	11.00	157.32	7.59
Cubic Exposure*Pigment	Unequal	−66.03	13.00	158.05	8.32
s(Exposure)*Pigment	Equal	−95.37	8.63	208.00	58.26
Cubic Exposure*Pigment	Equal	−95.00	10.00	210.01	60.27
Quadratic Exposure*Pigment	Equal	−97.10	8.00	210.20	60.46
Exposure*Pigment	Unequal	−103.70	9.00	225.41	75.67
s(Exposure)	Unequal	−114.03	5.20	238.46	88.73
Exposure	Unequal	−116.74	7.00	247.48	97.74
Exposure*Pigment	Equal	−119.31	6.00	250.62	100.89
s(Exposure)	Equal	−129.66	5.14	269.60	119.86
Exposure	Equal	−132.91	4.00	273.82	124.09
Constant mean	Unequal	−159.00	6.00	330.00	180.26
Constant mean	Equal	−191.06	3.00	388.12	238.39

* AIC = Akaike's Information Criterion.

consider constraining some of the effects for the polynomial model to lower dimensional polynomials to get performance closer to the spline-based model; the ability of the AMM to provide linear and other curving effects are highlighted in both the simplicity of fitting this model and its performance relative to a polynomial-based model with similar degrees of freedom (a quartic polynomial model provides 32 *df* but is over 50 points worse on AIC). While this is not a formal test to compare the models, the efficiency of this model in characterizing the differences between the pigmented and non-pigmented groups leads to overwhelming support for it compared to the other models considered.

The favored AMM contained the following random effect estimates: the random intercept was

$$b_i \sim N\left(0, \hat{\sigma}_b^2 = 0.383\right) \tag{A7}$$

and the residual random variation was

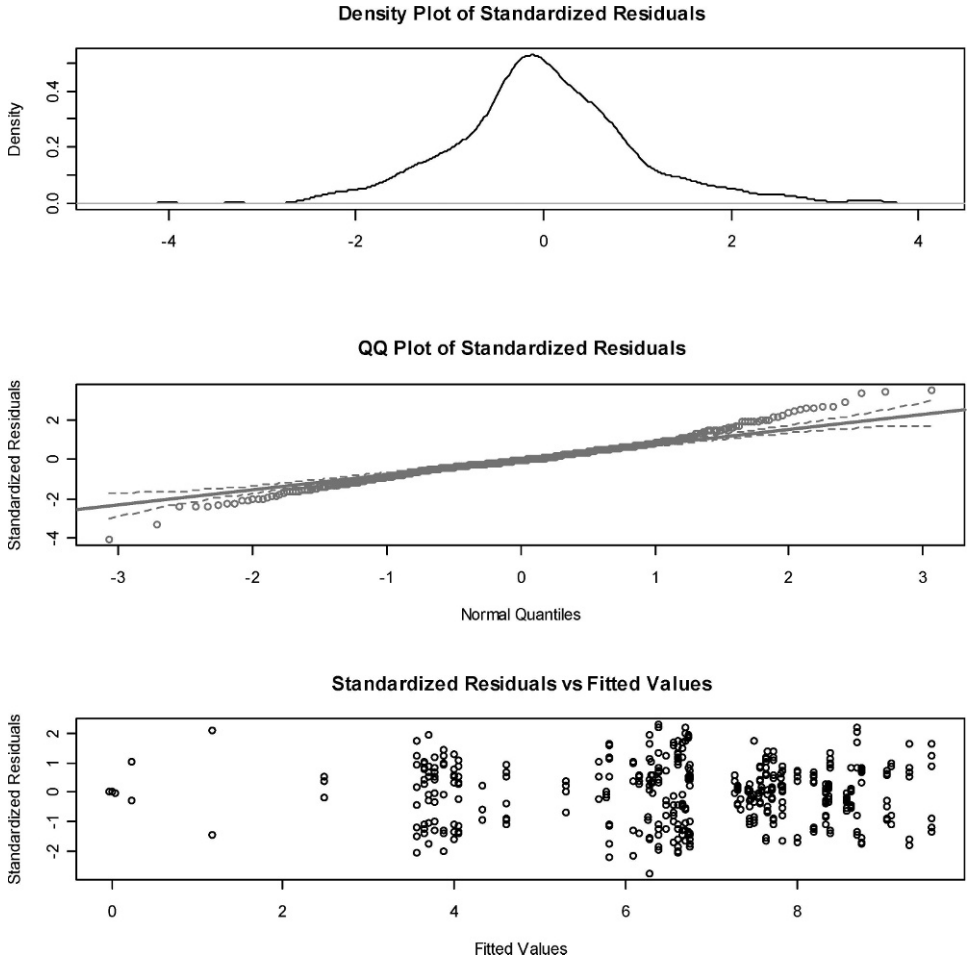

FIGURE A1. Residual diagnostic plots from the top model for the freeze-thaw experiment.

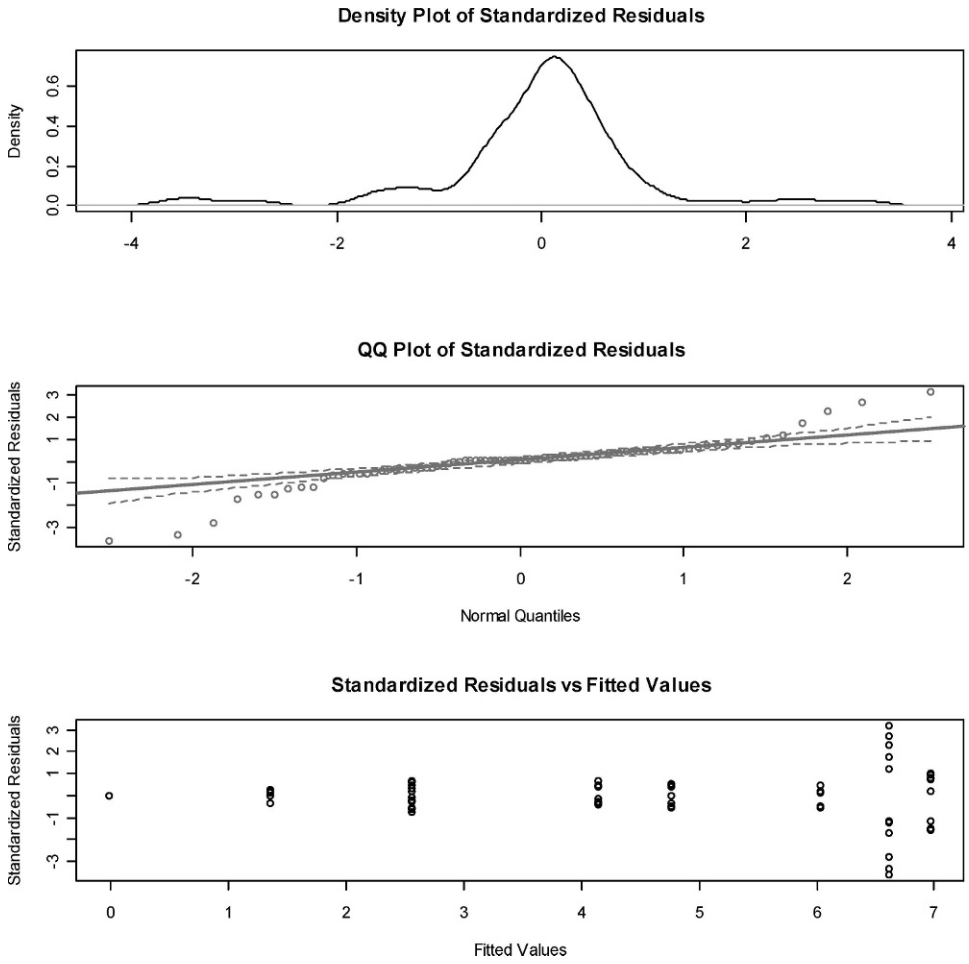

FIGURE A2. Residual diagnostic plots from the top model for the UV exposure experiment.

$$\varepsilon_{ij} \sim N\left(0, \hat{\sigma}_{\varepsilon}^2 = 0.260\right). \quad \text{(A8)}$$

This also provides an estimate of the within-trial correlation that is induced in this model of

$$\hat{\sigma}_b^2/(\hat{\sigma}_b^2 + \hat{\sigma}_{\varepsilon}^2) = 0.60. \quad \text{(A9)}$$

This suggests that there is a strong dependency of the measurements within the trials and that failing to account for this correlation would lead to considerably different results. We did not formally compare models with different random effects structures, instead we used the design of the study to dictate this component of the models. However, the size of this correlation is suggestive of the importance of the random intercept effect in these models. The fixed effects are displayed in Figure 2 in the main paper; the non-pigmented control and treatment as well as the ANT 11 treatment groups had between 5.6 and 6.7 *edf*, the ANT 11 control group had 3.5 *edf*, and the treatment and control pigmented groups (excluding ANT 11) were estimated to have linear, 1 *edf* effects.

Appendix Figure A1 displays standardized residuals for the top AMM. The first panel displays a nonparametric estimate of the distribution of the residuals, the second panel plots those same standardized residuals versus what would be expected from a normal distribution, and the bottom panel displays the residuals versus fitted values. The distribution of the residuals is relatively symmetric although it is slightly heavier tailed than a normal distribution. This can have some impact on the performance of the inferential techniques, but these deviations are relatively minor and the sample size is large enough that these minor differences with a normal distribution should have little impact due to the central limit theorem. The variance is relatively constant in the bottom panel, and no systematic deviation of the residuals from the estimated model is easily detected.

Results for the Solar Radiation Exposure Experiment

The solar radiation experiment differs from the freeze-thaw experiment in two important ways: the measured number of exposure times is much smaller, only 4 time points, and the control observations are not considered in this analysis since we demonstrated that all bacterial strains in the control group increased in CFUs over a 24 hr period. The favored model here is also an AMM with an interaction between a smooth function of exposure time and pigmentation present/absent groups. Diagnostics are more favorable here and do not suggest systematic deviation for any strain from the others in its group, so we do not separate ANT 11 from the other pigmented strains in the study. But the diagnostics did suggest a problem with all the models, demonstrating variability that was larger at the two middle exposure times and smaller at the beginning and end time points. To adjust for this, different variances were estimated for each exposure time. The top model was

$$Y_{ijk} = \alpha_N + b_i + s(\text{time})_N + \varepsilon_{ijk} \quad \text{(A10)}$$

for the non-pigmented strains, and

$$Y_{ijk} = \alpha_Y + b_i + s(\text{time})_Y + \varepsilon_{ijk} \quad \text{(A11)}$$

for the pigmented strains, with different variances for each exposure time j. The *edf* for the non-pigmented and pigmented time effects were 2.8 and 1.4, respectively. The random effect was estimated to be

$$b_i \sim N\left(0, \hat{\sigma}_b^2 = 0.117\right), \quad \text{(A12)}$$

and the residual random variation was

$$\varepsilon_{ij} \sim N\left(0, \hat{\sigma}_\varepsilon^2 = 0.0035\right), \quad \text{(A13)}$$

which is the variance estimate for exposure time zero. The variance at two hours of UV exposure was estimated to be 0.554, at six hours of exposure, 1.73, and at 12 hours of exposure, 0.428. The non-constant variance aspects of the model are also visible in the plots of the estimated effects and the inferred precision at each exposure time.

The second- and third-ranked models involve quadratic and cubic polynomials that provide similar results to the favored AMM, but use the model degrees of freedom less efficiently than the AMM. Results only from the AMM are reported, but the inference for the groups are similar with either method.

Residual diagnostic plots are displayed in Appendix Figure A2 for the top model. The distribution of the residuals is slightly heavy tailed relative to a normal distribution, but the deviation is relatively minor. There is also a negligible asymmetry to the residual distribution with the negative residuals having a slightly longer tail than for the higher values. The residual diagnostics initially did point to a problem with unequal variances that this model addresses. The structure of non-constant variance is a little unusual, in that it increases as a function of the estimated values and then decreases for the largest values, but was accommodated using a different variance for each exposure level. A small amount of changing variance is still visible in this plot that was not accounted for by the model, but these results are much better than for the models with constant variance.

Additional References Cited

Burnham, K. P., and Anderson, D. R., 2002: *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach.* Second edition. New York: Springer.

R Development Core Team, 2009, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M., 2009: *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.