Journal of Biological Systems, Vol. 14, No. 3 (2006) 425–443

World Scientific
www.worldscientific.com

A CONTINUUM MODEL OF THE WITHIN-ANIMAL POPULATION DYNAMICS OF *E. COLI* O157

J. C. WOOD,* D. C. SPEIRS,† S. W. NAYLOR,‡ G. GETTINBY†
and I. J. MCKENDRICK*,§

*Biomathematics and Statistics Scotland, Kings Buildings
Edinburgh, EH9 3JZ, Scotland

†Department of Statistics and Modelling Science, University of Strathclyde
Glasgow, G1 1XH, Scotland

‡Scottish Agricultural College, Animal Health Group, Sir Stephen Watson Building
Bush Estate, Penicuik, EH26 0PH, Scotland
§I.McKendrick@bioss.ac.uk

Received 10 March 2004
Revised 8 November 2005

The high level of human morbidity caused by *E. coli* O157:H7 necessitates an improved understanding of the infection dynamics of this bacterium within the bovine reservoir. Until recently, a degree of uncertainty surrounded the issue of whether these bacteria colonize the bovine gut and as yet, only incomplete *in-vivo* datasets are available. Such data typically consist of bacterial counts from fecal samples. The development of a deterministic model, which has been devised to make good use of such data, is presented. A partial differential equation, which includes advection, diffusion and growth terms, is used to model the (unobserved) passage of bacteria through the bovine gut. A set of experimentally-obtained fecal count data is used to parameterize the model. Between-animal variability is found to be greater than between-strain variability, with some results adding further weight to the hypothesis that *E. coli* O157:H7 can colonize the bovine gastrointestinal tract.

Keywords: Advection-Diffusion Equation; Method of Lines; Triangular Distribution; Maximum Likelihood; *E. coli* O157.

1. Introduction

E. coli O157:H7 causes serious illness and death in human populations.[1] Since its identification in 1982, there have been increasing reports worldwide of human infection with the organism.[2] Cattle have been implicated as the source of infection in many of these cases.[3] Therefore, there is a demand to reduce the levels of infection within the bovine population. An improved understanding of *E. coli* O157:H7 population dynamics within the bovine host could aid the identification of suitable within-animal control measures. This paper uses mathematical modeling to gain insights into the *in vivo* population dynamics.

§Corresponding author.

A previous attempt to describe bacterial population levels at various sites during passage through the bovine gastrointestinal tract (GIT) involved the development of a stochastic compartment model.[4] However, development was limited by the inadequacy of existing data defining the size of the *E. coli* O157:H7 populations located at various sites along the bovine gut. Indeed, most results derived from relevant experimental studies simply consist of counts of *E. coli* O157:H7 obtained from fecal samples over a period of time. These data correspond to samples from the infection process at just one particular (spatial) point in its spatio-temporal evolution. Given the need for data in the model-building process, it is convenient to build the model around what little data is available.

Since a known fixed amount of bacteria enters an animal, generally via inoculation in experimental studies, and the amount of bacteria leaving the animal at the other end, via fecal pats, can be observed, an appealing alternative approach to modeling the *in vivo* infection process is to view the bacterial population as being represented by a density function along a one-dimensioned representation of the gut. A partial differential equation can be used to describe, deterministically, the unobservable within-animal infection dynamics of *E. coli* O157:H7 assuming constant values for the parameters along the entire gastrointestinal tract.

This paper describes the development of a simple continuum model of the passage of bacteria through the bovine gut, based on a partial differential equation, in Sec. 2.1. Both analytical and numerical solutions are sought and in Sec. 2.3, a set of detailed and extensive experimental data is used to parameterize the model for two strains of *E. coli* O157:H7. The strains differ in that only one of the strains possesses the phage encoding shiga-toxin 2. Therefore, the model can be used to explore whether the presence of this toxin enhances the duration or level of colonization of *E. coli* O157:H7 within cattle. The results are described in Sec. 3, followed by a discussion in Sec. 4.

2. Methods

2.1. *Model development*

In order to derive a model that describes the population dynamics of *E. coli* O157:H7 during transit through the bovine gut, it is necessary to adopt several simplifications. Firstly, to reduce the complexity of the model, a cylindrical tube with a constant cross-sectional area is used to represent the gut. Secondly, it is assumed that the contents of the gut are well-mixed vertically and transversely so that the population density, $n(x,t)$ varies only with time, t (days), and the distance along the gut, x (meters). Furthermore, we assume that the bacteria do not colonize or in any way adhere to the mucosal surfaces of the GIT.

The number of bacteria in the gut can change due to either the reproduction or death of existing bacteria, or by the transportation of bacteria, either advectively (with the luminal fluid, for instance) or diffusively (by molecular motions),

across the ends of the cylinder. Observations suggest that the losses to a population of *E. coli* O157:H7 due to movement through the gut exceed any gains due to reproduction, as infections appear to be transient.[5] Therefore, any potential density-dependent effects are assumed to be negligible, so that the population develops with constant *per capita* birth and death rates, and hence a constant intrinsic growth rate, β (day^{-1}).

The population of *E. coli* O157:H7 moves through the tube with a flux $F(x,t)$ at time t in the positive x direction. This flux can be broken down into two components: an advective and a diffusive flux. The advective velocity ω (meters.day^{-1}) corresponds to the rate at which bacteria are carried by the medium they inhabit within the bovine gut (for instance, the luminal fluid). The diffusive flux corresponds to the rate at which bacteria spread diffusively with coefficient γ (meters^2.day^{-1}) in both directions along the gut. It is not unrealistic to assume that diffusion occurs in both directions as Singleton[6] noted the occurrence of backflow from the duodenum to the abomasum. In fact, it has been found that backflow amounts to nearly 10% of the total peristaltic flow in sheep, and given the physiological similarities, it might be expected that a similar result holds in cattle.[7] However, boundary conditions have to be selected to ensure that implausible events do not occur, such as the diffusive flux causing bacteria to exit the tube at $x = 0$, an event which corresponds to bacteria flowing back out of the mouth of an animal. Hence, the left-hand boundary condition is defined to act as a reflecting barrier while the right-hand boundary is an absorbing barrier.

Assuming that the advection velocity and diffusion coefficient are constants, the within-animal infection dynamics of *E. coli* O157:H7 can therefore be described by the combination of a conservation equation,

$$\frac{\partial n}{\partial t} = \underbrace{\beta n}_{growth} - \underbrace{\omega \frac{\partial n}{\partial x}}_{advection} + \underbrace{\gamma \frac{\partial^2 n}{\partial x^2}}_{diffusion} \tag{2.1}$$

and the following initial and boundary conditions:

$$\begin{aligned} &\text{Initial Condition: } n(x,0) = \begin{cases} n_0 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \\ &\text{Left-hand Boundary Condition: } \omega n(0,t) - \gamma \left(\frac{\partial n}{\partial x} \right)_{x=0} = 0 \\ &\text{Right-hand Boundary Condition: } n(L,t) = 0, \end{aligned} \tag{2.2}$$

where L is the notional length of the bovine GIT. For the purposes of this model, L is assumed to equal approximately 40 meters, drawing on data in Phillipson.[7] The model is unrealistic in that it assumes constant growth, advection and diffusion parameters when these take very different values in different areas of the GIT.[8] It would not be possible to parameterize a model with constants which varied with respect to distance x, using only shedding data. However, the model can be thought of as summarizing the average behavior of each parameter over the entire GIT.

The problem is akin to that of population persistence in flowing environments such as rivers and estuaries, as studied by Speirs and Gurney.[9] However, given the transient nature of the observed infections, which imply that the *E. coli* O157:H7 populations die out over time, this analysis of persistence is of little help. Furthermore, since the distribution of *E. coli* O157:H7 within individual bovine guts is not observable, we are not interested in the form of the solutions *per se*, but rather in what they imply for the measurable flux of *E. coli* O157:H7 leaving the animal through feces. To examine these transient dynamics, it was necessary to resort to numerical methods of solution.

It is often convenient to reduce the number of unknown parameters that require estimation. However, it has been shown in Appendix A that each of the three parameters are required to accurately describe the behavior of the process in this model. As a result, a particularly efficient numerical method is required when the model is being fitted to data. A discrete-time approximation, which is described in more detail in Appendix B, is adopted, justified by the comparison outlined in Appendix C. The discrete-time approximation, which is based on a triangular distribution, produced results comparable to those obtained using the standard method of lines,[10] but was far more computationally efficient.

2.2. *Experimental data*

The set of data which is used to parameterize the model was obtained from an experimental study carried out by the University of Edinburgh and the Scottish Agricultural College.[11] The study involved a cohort of 11 calves, which were experimentally challenged at approximately 2 weeks post-weaning with a dose of 10^9 colony-forming units (cfu) of one of two different strains of *E. coli* O157:H7, labelled strains "A" and "B". Both strains were isolates from a single human case following an outbreak centered on a restaurant in Washington State, USA, with the infection ultimately being traced back to a dairy farm.[12] A series of 10 gram samples of feces were taken from each animal. For the first 10 days post-inoculation, samples were collected each day, followed by sampling every 3 to 4 days until 28 days post-inoculation or when the animal was euthanized. Each sample was suspended in 90 ml of sterile phosphate-buffered saline (PBS) and serially-diluted in 10-fold steps in PBS in order to estimate the number of cfu within a sample by direct culture.[11] Hence, the data consisted of a series of plate counts and the associated dilutions for which these counts were obtained. At each time-point sampled, this procedure was repeated up to a maximum of six times.

The issue of bacterial growth during the incubation of each sample was considered. However, this effect was disregarded, as the growth of *E. coli* O157:H7 kept under lab conditions is thought to be negligible in the short period of time between sample collection and plating.[13]

It is notable that the observed counts contain frequent zero values (Figs. 1 and 2), where observations from the same animal on congruent days, or even from

other feces samples from the same animal on the same day, show very high estimated counts. It has been established that bovine feces may be highly heterogeneous in the distribution of *E. coli* O157:H7.[11,14] The fecal samples were not homogenized before culturing, and this within-sample variability may explain the observed null counts. This heterogeneity was not, however, explicitly modeled in the parameterization exercise.

2.3. *Parameterization method*

A maximum likelihood-based approach was used to parameterize the model. Each data point (plate count, x) was assumed to be a realization from a particular Poisson distribution, i.e.

$$x \sim Po\left(\frac{\mu}{d}\right)$$

where

$$\mu = \text{actual number of cfu in 10 g of feces}$$
$$d = \text{specified dilution level.}$$

Likelihoods were derived from the Poisson distribution using the predicted values from the model, rescaled to correspond to the appropriate dilution level, where λ_{t_i} = mean cfu for a particular sample i, with observed count x_i, collected at a particular time, t. Hence, the log-likelihood to be maximized is given by:

$$F = \sum_i \{-\lambda_{t_i}/d_i + x_i \ln \lambda_{t_i}\}.$$

The numerical optimization method used is the Nelder-Mead algorithm, a downhill simplex method used for the minimization of a function of M variables. It is not very efficient in terms of the number of required function evaluations, but has the advantage that it will converge even when the initial simplex straddles two or more local minima, which is the main criterion for its adoption in this case.

The *amoeba* routine from Press *et al.*[15] is used to implement the minimization method. To avoid the problem of false convergence, the routine is re-started at a simplex near to any apparent minimum, using a method proposed by Wood[16] to eliminate the uncertainty in deciding the size of random jumps required to clear local minima. This involves stochastically perturbing the objective function, by bootstrapping the underlying data to which the model is being fitted. The idea behind this technique is that each bootstrap objective function will have the same large-scale structure as the original objective function, i.e. the same global minimum, but have different small-scale structure. It is therefore unlikely that the routine will converge to the same local minimum.

The simplex minimization routine with bootstrapping is applied to the datasets generated by each of the 11 animals in the experimental study cohort, in order to fit a model of the passage of *E. coli* O157:H7 through each individual animal.

Approximations to the variance-covariance matrix, Σ, of the parameter estimates:

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\beta, \omega, \gamma)$$

are usually calculated using Fisher's Information matrix. The use of this approximation requires the assumption of large-sample efficiency.[17] However, the small volume of data which is available makes this assumption problematic, while the evaluation of the matrix requires the numerical estimation of derivatives which, although possible, is prone to numerical error. For both of these reasons, therefore, the variance-covariance matrix of the parameter estimates is not estimated. This does, however, have consequences on our ability to interpret the results, which are discussed later.

3. Results

3.1. *Strain A*

The first of the strains considered, labelled "Strain A", is inoculated into five animals, identified as A_1, A_2, A_3, A_4 and A_5. The minimization routine is applied to the likelihood function for the observed counts obtained from each of these animals, in order to derive the parameter estimates that allow the model of bacterial passage through each individual animal to be fully described.

The fitted model obtained for each of the animals can be seen in Fig. 1. The solid line represents the estimated quantity of bacteria shed by the animals per day, calculated by numerical integration of the flux of bacteria into the right-hand boundary per day. Each observed count, rescaled to an equivalent daily amount by assuming an average fecal output of 8 kg per animal per day,[7] is represented on the plot by a circle, and at the time-point at which the sample corresponding to each observed count was collected, the corresponding fitted value is represented by a cross.

In order to evaluate the fit for each animal, it is necessary to measure the discrepancy between the fitted values $\hat{\boldsymbol{\lambda}}$ and the corresponding observations $\mathbf{x}$. As a Poisson likelihood was used to fit the model, an appropriate measure is the deviance, which is formed from the logarithm of a ratio of likelihoods.[18] The total deviance involves the logarithmic likelihood ratio of the null model, in which it is assumed that all observations $\mathbf{x}$ have a common mean parameter $\hat{\boldsymbol{\lambda}} = \hat{\lambda} = \bar{x}$, and the full, or saturated, model, in which there is a parameter for each data point so that the fitted model fits the set of observations exactly, i.e. $\hat{\boldsymbol{\lambda}} = \mathbf{x}$, giving the maximum possible value of the likelihood. Hence, the total deviance can be defined as:

$$-2[l(\mathbf{x}; \bar{x}) - l(\mathbf{x}; \mathbf{x})].$$

In practice, the null model is too simple and the full model is uninformative as it simply maps exactly to the data. However, this deviance provides a baseline against which the fit of an intermediate model can be measured. Therefore, the next step

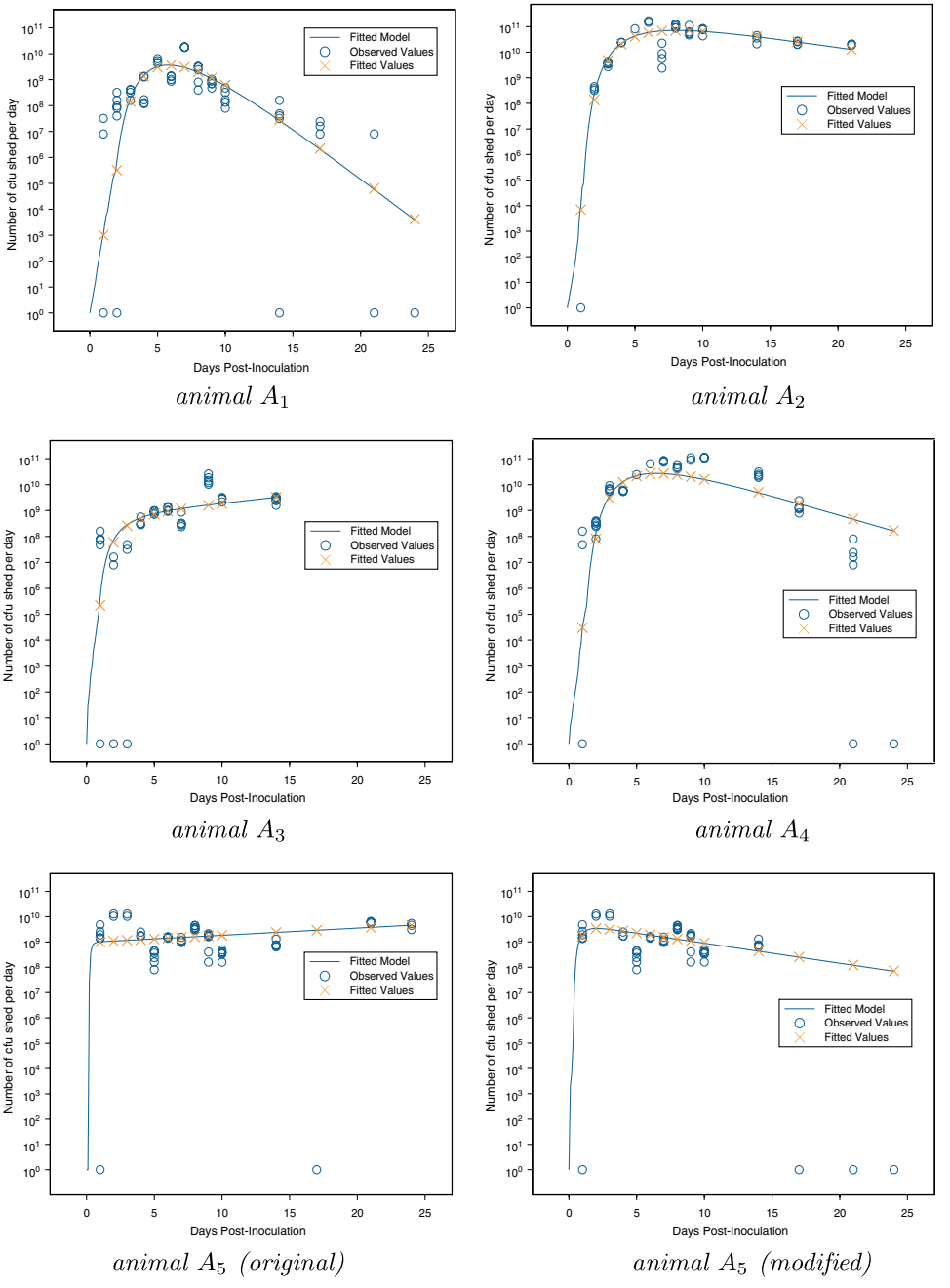


Fig. 1. Fitted model of the passage of "strain A" through individual animals following inoculation at day 0. The solid line represents the fitted bacterial distribution, the circles represent observed counts and the crosses represent the fitted values at the corresponding time-points. Circles at the 10^0 level are zero observations.

is to calculate the residual deviance, which involves the logarithmic likelihood ratio of the fitted model and the full model, and can be defined as:

$$-2[l(\mathbf{x};\hat{\boldsymbol{\lambda}}) - l(\mathbf{x};\mathbf{x})].$$

The residual deviance is analogous to the residual sum of squares,

$$\sum_i (x_i - \hat{\lambda}_i)^2$$

which is used to assess the fit of a model in normal linear regression. Therefore, the following expression:

$$1 - \frac{\text{residual deviance}}{\text{total deviance}},$$

which will be referred to as R_D^2 as it is equivalent to R^2, the coefficient of determination, is used to assess how well each of the models fits the data.

A summary of the parameters estimated for each animal along with the corresponding value of R_D^2 can be found in Tables 1 and 2. While it is relatively straightforward to obtain reasonable fits for most of the animals, it is more difficult to achieve this for animal A_5. The initial model fit predicts an increasing distribution of bacterial shedding over time, which is not expected, and as can be seen from the value of R_D^2 in Table 1, the model does not fit the observations at all well for this animal. In fact, the fitted model appears to fit the data more poorly than if a common mean parameter had been used, as in the case of the null model. A closer inspection of the observed counts in Fig. 1 can help to explain why this is the case. Two peaks can be observed in the data. There is an initial peak around day 3, which dies away slowly. No bacteria are present in any of the six samples collected on day 17, while at the next sampling point (day 21), the amount of bacteria isolated has returned to a level last seen 2 to 3 days post-inoculation. It is unclear whether the break in shedding around day 17 is simply attributable to transient shedding, or whether the animal did actually become clear of infection, but then, perhaps as a result of fecal/oral contamination, became reinfected.

If transient shedding is responsible for producing the multi-modal pattern exhibited by this dataset, then this animal corresponds to a case in which it is difficult to obtain a model that fits the data well, as the model can only successfully describe one peak in shedding. If, alternatively, reinfection did occur, it is not possible to use the data set in its current form for parameterization purposes, as the model has been developed to describe the passage of one pulse of infection through an animal. However, assuming reinfection did take place, it is reasonable to suggest that the

Table 1. Summary of parameter estimates for animal A_5, using the original set of observations and a modified set.

Animal	$\beta(\text{day}^{-1})$	$\omega(\text{m.day}^{-1})$	$\gamma(\text{m}^2.\text{day}^{-1})$	R_D^2
A_5 (original)	0.049	0.545	19.538	−0.18
A_5 (modified)	0.064	1.031	5.333	0.45

reinfection event probably occurred after day 17 (when the animal was clear of *E. coli* O157:H7) and that therefore, disregarding all positive observations following this time-point will result in a dataset that describes a single pulse of infection.

Hence, the model fitting procedure is repeated for animal A_5, but this time omitting the observations that were most likely to have arisen as a result of reinfection. The revised fitted model is plotted in Fig. 1 and the model now explains 45% of the deviance, which is a substantial improvement.

Of all the animals inoculated with strain A, the model for animal A_3 shows the poorest fit. This is also the only model to predict increasing levels of shedding over time. However, animal A_3 was necropsied on day 14 and so there are no observations available after this point. Therefore, valuable information regarding the rate at which the level of shedding diminishes is missing. The truncated dataset probably explains the relatively poor fit and the qualitatively different behavior seen in this case.

3.2. *Strain B*

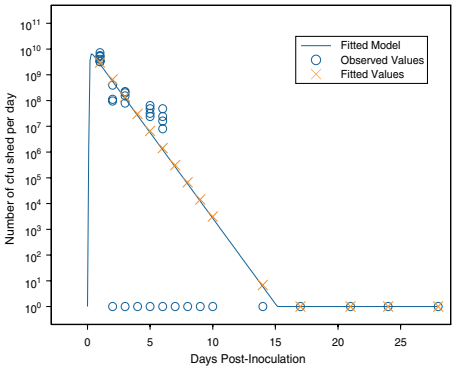
The second strain considered in this experiment differs from strain A only in that it does not possess the phage encoding shiga-toxin 2. The strain was inoculated into six animals, identified as B_1, B_2, B_3, B_4, B_5 and B_6. The parameterization method is applied to the data generated by the observation of each of the animals. A summary of the most likely parameters can be found in Table 2, and the fitted models are plotted in Fig. 2.

The only animal for which it is not possible to obtain a reasonable fit to our model is animal B_3. The explanation appears to be similar to that for animal A_5; the observations are multi-modal. In particular, there is a large drop in the shedding levels around day 21 with all but one of the samples at that time-point testing negative, followed by a large increase in shedding at the next sampling point, which is indicative of possible reinfection. Even the adoption of a similar approach to that used for animal A_5, where the fitting procedure is repeated using a dataset that omits possible reinfection observations, results in a negligible improvement to the overall fit.

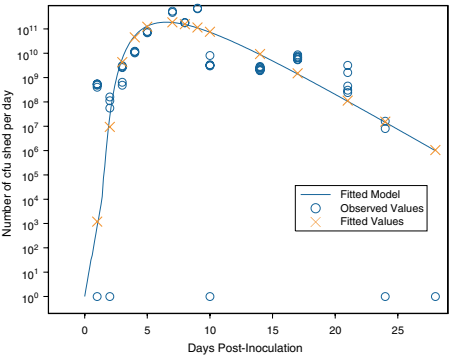
With the exception of that for animal B_6, all of the fitted models predict an eventual decline in the level of shedding over time. The high levels of bacteria recorded for animal B_6 towards the end of the sampling period suggest that reinfection may have occurred. A similar pattern of observations suggests that animal B_4 may also have become reinfected.

3.3. *Comparison of strains*

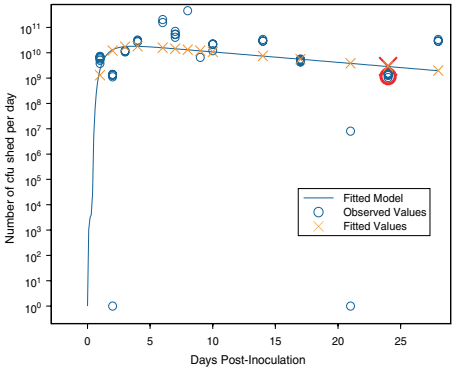
To aid comparison of the two strains, and in particular, to explore whether the presence of shiga-toxin 2 has an effect on the shedding pattern of *E. coli* O157:H7 within cattle, the estimated parameters and measures of fit for each animal are



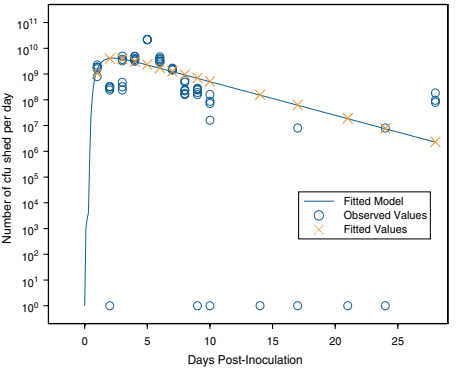
animal B_1



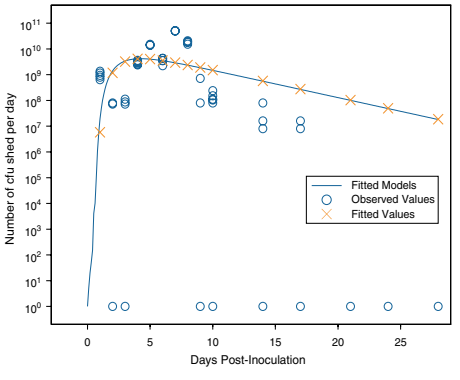
animal B_2



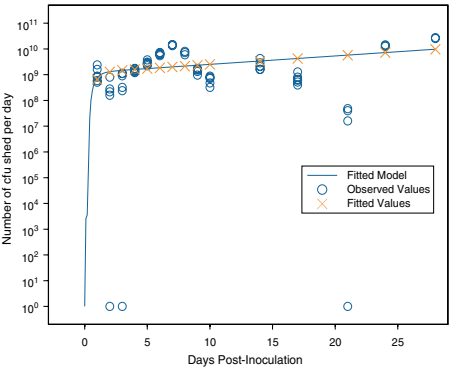
animal B_3



animal B_4



animal B_5



animal B_6

Fig. 2. Fitted model of the passage of "strain B" through individual animals following inoculation at day 0. The solid line represents the fitted bacterial distribution, the circles represent observed counts and the crosses represent the fitted values at the corresponding time-points. Circles at the 10^0 level are zero observations.

Table 2. Summary of parameter estimation results for animals inoculated with different strains.

Animal	Growth, $\beta(\text{day}^{-1})$	Advection, $\omega(\text{m.day}^{-1})$	Diffusion, $\gamma(\text{m}^2.\text{day}^{-1})$	γ/ω (m)	R_D^2
A_1	0.024	0.314	0.413	1.315	0.51
A_2	0.049	0.387	0.713	1.842	0.62
A_3	0.018	0.214	1.619	7.566	0.07
A_4	0.044	0.383	0.676	1.765	0.38
A_5	0.064	1.030	5.330	5.175	0.45
B_1	0.184	3.912	32.944	8.421	0.87
B_2	0.056	0.371	0.427	1.151	0.50
B_3	0.081	0.840	2.337	2.782	−0.06
B_4	0.061	0.927	3.774	4.071	0.27
B_5	0.040	0.522	1.628	3.119	0.05
B_6	0.041	0.657	6.670	10.152	0.23

summarized in Table 2. However, given the small number of fitted models and the inherent variability that certainly exists between animals, it will not be possible to draw any strong conclusions from these results alone.

Nevertheless, a close inspection of the individual plots (Figs. 1 and 2) does reveal some plausible between-strain differences. For instance, the predicted peak amount of bacteria shed per day by the majority of animals infected with strain A lies around the day 5 mark, whereas the peak for the animals infected with strain B lies near to day 2. Qualitative assessments of the model are described in the Appendices, where Fig. 3 illustrates that the time to reach peak flux could be decreased by either a reduction in the birth rate or an increase in the ratio γ/ω which can occur by either an increase in the diffusive coefficient γ or a decrease in the advective velocity ω. An examination of the parameter values in Table 2 suggests that the difference in the distributions is therefore likely to have been caused by the higher ratios of diffusion coefficient to advective velocity generated by animals inoculated with strain B. Nevertheless, given that there are few results available for comparison, it would be inappropriate to argue that this effect is caused by any differences in strain types, rather than some unspecified alternative factor.

Table 2 reveals that the diffusive coefficient appears to be the parameter which is subject to the most variability between animals. The growth rates between different strains may have been expected to show some variation, and the growth rate in different animals could be affected by differences in the availability of nutrients in the bowel, variety in the resident microflora which interact with the challenge bacteria, or variability in immune response. The advective velocities are prone to variation caused by a wide variety of sources such as change in diet or stress. Contrary to the results of this parameterization, it is the diffusive coefficients that might have been expected to have relatively constant values across all animals, although given the small number of animals to which the model was fitted and the lack of estimates of standard errors, it remains to be seen whether this phenomenon is genuine. This discussion implicitly assumes that the parameter estimates for each animal are largely uncorrelated. This is unlikely to be the case in practice, and some

of the peculiarities of the estimates seen in Table 3.3 may be more explicable when seen in the context of correlated (within-animal) estimates with large standard errors. However, a methodological explanation of the variability in the diffusion coefficients can be proposed based on examination of datasets to which it was not possible to obtain reasonable fits. As illustrated in Fig. 3 in the Appendices, it would appear that the diffusion coefficient is the parameter with the greatest scope to vary the shape of the shedding curve towards that observed in practice. For example, should the bacterial populations persist at high levels for a substantial length of time, an increase in the value of the diffusion coefficient will cause the shape of the curve to match, more closely, that of the observed counts over time.

4. Discussion

The development and parameterization of a novel, deterministic model of the unobserved passage of bacteria through the bovine gut is presented. The model is devised to make good use of the limited available data, and therefore uses a partial differential equation to provide a simple representation of the process.

Using this model, it is possible to obtain reasonable fits to all of the datasets considered, with one exception (animal B_3). However, in this case there is evidence that the assumption of a single pulse of infection which underlies the parameterization exercise is inappropriate. Figures 1 and 2, and Table 2, all indicate that poorer fits were obtained for those sets of observations where shedding persisted for longer periods of time at high levels. This suggests that the model also has difficulty in fitting a slow decline in the levels of shed bacteria.

The continuum model assumes constant parameter values throughout the gut, and hence, these rates are applied to the internal bacterial population as a whole. The model would be robust when fitted to data arising from a situation where the parameters take different values at different points in the gut as long as the change applies to all bacteria at that point. For instance, if bacteria passing through a particular section of gut moved with a velocity that differed from the rate found in the rest of the gut, an equivalent temporal distribution of shed bacteria could be obtained from the model fitting a weighted average of the two velocity rates to the entire gut. However, the model will not fit well to data arising from a situation where different sets of parameters are applied to subsets of the bacterial population. The difficulties noted above in fitting models to observations obtained from "high-shedding" animals suggest that these observations have been obtained from a situation in which the assumptions of the model do not hold. Therefore, it is plausible that a subset of the bacterial population within these animals is subjected to a different set of parameter values than that of the majority population at some point during its passage through the gut. This is consistent with the occurrence of colonization, as reported by Naylor *et al.*,[11] which would correspond to a sub-population becoming subject to negligible advection at some point in the continuum. In future experiments, direct swabbing of the potential

site of colonization[19] will allow the identification and quantification of such subpopulations. Models such as that described in this paper will still have relevance in describing the population dynamics of non- or negligibly-colonizing strains of bacteria, or of the population dynamics in animals which are refractory to colonization.

The collection of bacterial data externally from an animal, for instance from fecal pats, is cheaper, more convenient and more ethical than the alternative, which inevitably involves surgical intervention in the animal to access the gastrointestinal tract. In so far as pat data easily provides a dynamic picture of shedding within individual animals, it is also more informative. The continuum model was specifically designed to ensure that information gathered externally from an animal would be sufficient for parameterization purposes. The data used for this purpose was the most detailed and extensive set produced to date. Therefore, an attempt was made to compare the within-animal infection dynamics of the different strains used in the experimental study. Unfortunately, the small number of datasets for which parameterization was possible did not allow any general conclusions to be drawn, beyond the observation that between-animal variability was greater than between-strain variability. This inability to generate sufficient evidence to evaluate biological hypotheses, despite the use of a large dataset and a specially-developed model, highlights the need for further research in this area, and in particular, a requirement to conduct more extensive field and experimental studies. Future parameterization models should also explicitly allow for the effect of heterogeneity in the feces sample, including the possibility of observing a zero count from a sample collected from a shedding animal, perhaps through the use of a zero-inflated count model.[20] It will also be important to develop a robust estimator for the variance-covariance matrix of the parameter estimates, facilitating better understanding of the interplay between different parameter estimates, and allowing the identification of animals with statistically significant differences in shedding behavior.

Acknowledgments

We wish to thank Glenn Marion and Chris Low for helpful comments on the manuscript. JCW gratefully acknowledges the support of a BBSRC studentship. IJM acknowledges support from the Scottish Executive Environment and Rural Affairs Department (SEERAD), project BSS/028/99 and the Wellcome Trust International Partnership Research Award for Veterinary Epidemiology. SWN acknowledges the help and support of Dr. D. Gally of the University of Edinburgh, holder of a DEFRA Veterinary Fellowship Award. The experimental work also benefited from the support of SEERAD.

Appendix A: Reducing the Dimensions

An attempt can be made to reduce the number of unknown parameters for which estimates are required. While it is possible that the advective flux can be inferred

from the mean length of time it takes a dose of *E. coli* O157:H7 to pass through the bovine gut, it is difficult to identify any mechanism to estimate the intrinsic growth rate and the diffusion coefficient, short of fitting the results of the model to shedding data. Therefore the model is rewritten in a form which allows these two parameters to appear independently.

The length of the digestive tract, L, can be estimated from the literature,[7] providing a natural scale for space, x. Similarly, the characteristic passage time through the gut can be denoted by L/ω, which provides a suitable time scale. Therefore, defining

$$y \equiv \frac{n}{n_0}, \quad T \equiv \frac{t\,\omega}{L}, \quad X \equiv \frac{x}{L}, \quad B \equiv \frac{\beta L}{\omega}, \quad K \equiv \frac{\gamma}{\omega L}, \tag{A.1}$$

yields the following normalized model with unit passage time through the gut, $T = 1$:

$$\frac{\partial y}{\partial T} = By - \frac{\partial y}{\partial X} + K\frac{\partial^2 y}{\partial X^2}, \tag{A.2}$$

with boundary conditions:

$$y(0,T) - K\left(\frac{\partial y}{\partial X}\right)_{X=0} = 0, \qquad y(1,T) = 0 \quad \forall\, T, \tag{A.3}$$

and initial condition:

$$y(X,0) = \begin{cases} 1 & \text{if } X = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{A.4}$$

A standard approach to solving such systems is to use the method of lines, in which discretization in space is independent of discretization in time.[10] This approach leads to a set of ordinary differential equations which are solved as initial value problems.

The calculation of a numerical solution for the rescaled model enables the range of behavior that can be exhibited by the model to be investigated. Figure 3 illustrates the effect of varying the normalized diffusion parameter, K, and the growth rate parameter, B, on the amount of *E. coli* O157:H7 shed from the animal per unit time (i.e. the terminal flux). Comparison of the three frames appears to indicate that the shape of the temporal distribution of the flux is most strongly affected by the level of diffusion. For a value of $K = 1$, used in the uppermost frame, the flux of bacteria leaving the gut has a right-skewed distribution with respect to time. However, the lower frames show that as diffusion levels decrease, the shedding of *E. coli* O157:H7 becomes more concentrated in time, leading to a more symmetrical distribution of the flux, whose mode lies closer to the mean passage time through the gut, $T = 1$.

By contrast, the growth rate, B, has a weaker effect on the temporal pattern of the flux, acting in the main as a scaling parameter. However, reductions in B do appear to result in small decreases in the time to reach the peak flux. It can be seen that, even when local growth rates are positive, the rate of terminal flux decays over time.

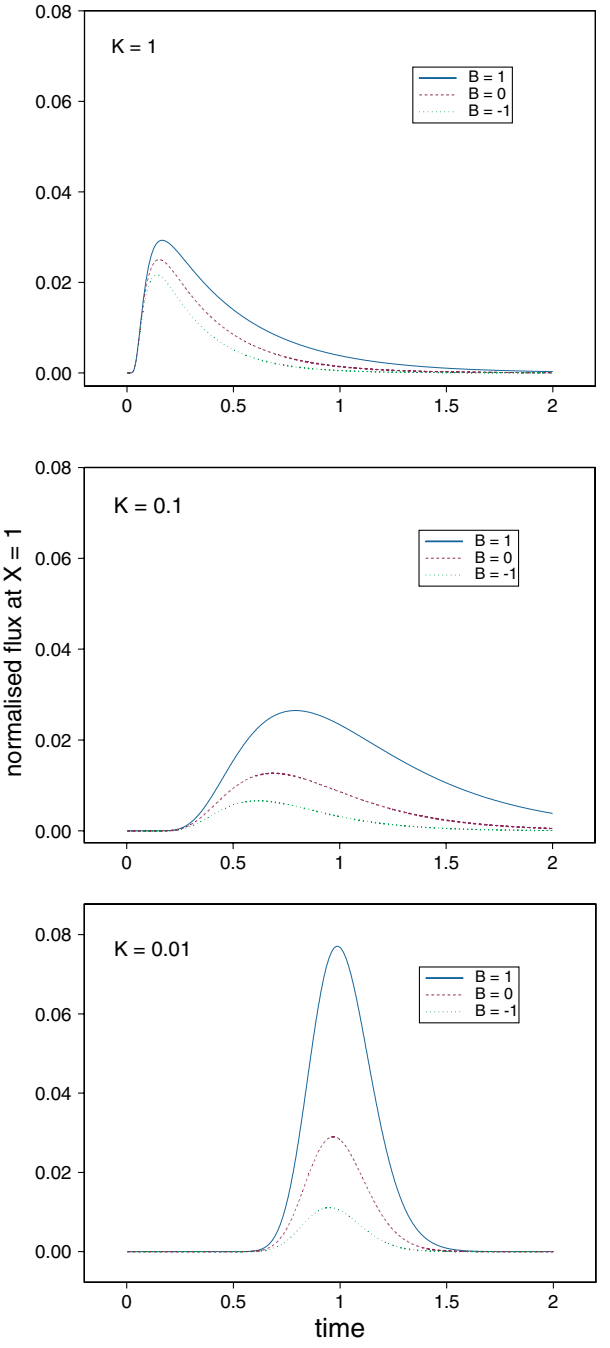


Fig. 3. Plots of the normalized terminal flux calculated for different parameter values. Each frame shows results for a different value of the normalized diffusion coefficient, K, while the lines within each frame show the effect of changing the intrinsic rate of increase, B.

Table 3. Mean residence times (in units relative to the characteristic passage time) of *E. coli* O157:H7 in the digestive tract for different values of the growth and diffusion parameters, B and K, respectively.

		B		
		−1	0	1
	1	0.308	0.371	0.481
K	0.1	0.784	0.905	1.107
	0.01	0.976	0.995	1.015

It is clear from Fig. 3 that the peak flux occurs prior to the mean passage time, for $K > 0$. This earlier than expected peak is due to the diffusive component, which dominates movement through the absorbing right-hand boundary, and since the overall population growth is negative, the amount leaving the animal is at a maximum early in the process, even though the mean location of the population is not yet near the boundary.

The mean residence times of bacteria in the gastrointestinal tract are calculated from the terminal fluxes obtained for different values of the parameters, B and K, used in Fig. 3. As can be seen from Table 3, the mean residence time of bacteria in the gut is not the same as the time required to advect the length of the gut, with the mean residence time depending on both B and K. This difference is unfortunate as it suggests that the advective velocity should be retained in the equation to be solved and that therefore a model of this system should incorporate each of the three unknown parameters.

Appendix B: An Efficient Discrete-Time Approach

In order to fit the model to data, an efficient discrete-time approximation, based on a triangular distribution, was adopted. Having discretized space using m nodes spaced Δx apart, the method essentially involves approximating the distribution of bacteria after some time increment Δt, following an initial point release at the ith node, x_i, by a triangular distribution displaced a distance δ from x_i. Since in this problem, the distribution around the initial node after one time step is independent of the starting position, it is convenient to express this as a dispersal distribution, D_j, which gives the fraction of the population which has been displaced j nodes from the source,[21] i.e.:

$$D_j = \begin{cases} (1 + (j\Delta x - \delta)/\alpha)\Phi & \text{if } \delta - \alpha < j\Delta x < \delta \\ (1 - (j\Delta x - \delta)/\alpha)\Phi & \text{if } \delta < j\Delta x < \delta + \alpha \\ 0 & \text{otherwise} , \end{cases}$$

where α is half the width of the distribution, and Φ is a normalization parameter such that:

$$\sum_j D_j = 1. \tag{B.1}$$

The distribution, D_j, can only be nonzero for values of j such that:

$$(\delta - \alpha)/\Delta x < j < (\delta + \alpha)/\Delta x, \tag{B.2}$$

and Eq. (B.1) can easily be obtained numerically by bisection on Φ, for any δ and any $\alpha > \Delta x$. The restriction on α occurs since the triangle must be sufficiently wide to ensure that it spans at least two nodes in order to guarantee normalization with Φ in the range 0–1.

To incorporate growth, the population at each node is multiplied by the net reproductive rate used in discrete-time models, equivalent to a multiplier of

$$\exp(\beta \Delta t)$$

where β is the intrinsic growth rate used in the continuous-time model (see Eq. (2.1)). These growth and dispersal procedures are applied to each node at the end of each time step.

However, in some instances the triangular distribution determines that dispersers should arrive at nodes lying outwith the domain $\{1, 2, \ldots, m\}$. Therefore, in order to use this triangular distribution in a system of finite length, it is necessary to approximate the boundary conditions. This is done using the method of images.[22] Hence, for a reflecting boundary located halfway between two nodes (i.e. $x = 0$), the part of the distribution which falls outside the domain is "folded" back and added to the weight associated with the matching cell that is equidistant from the boundary on the interior of the domain. For the absorbing boundary at $x = b$ the process is the same, except that the weights associated with nodes outside the boundary are subtracted from the corresponding nodes in the interior. Hence, as the nodes are numbered from 1 to m, and taking into account the reflecting boundary on the left-hand side and the absorbing boundary on the right-hand side, any destination node k lying outside the domain $\{1, 2, \ldots, m\}$ is mapped as follows:

$$k \to \begin{cases} -(k-1) & \text{if } k < 1 \\ -k + 2m + 1 & \text{if } k > m. \end{cases}$$

Appendix C: Comparison of Numerical Methods

Gurney and Nisbet[21] have shown that where a population is initially located at a single cell, successive iterations of the triangular distribution will converge to a normal distribution in space, approximating the diffusion equation. This result allows a comparison to be made between two numerical methods (triangular distribution and method of lines).

The convergence result indicates that the variance (with respect to space) of a population diffusing from a point source over a period of Δt is $2\gamma \Delta t$ for a diffusion coefficient of γ, while the mean displacement is $\omega \Delta t$. Thus, by calculating the mean

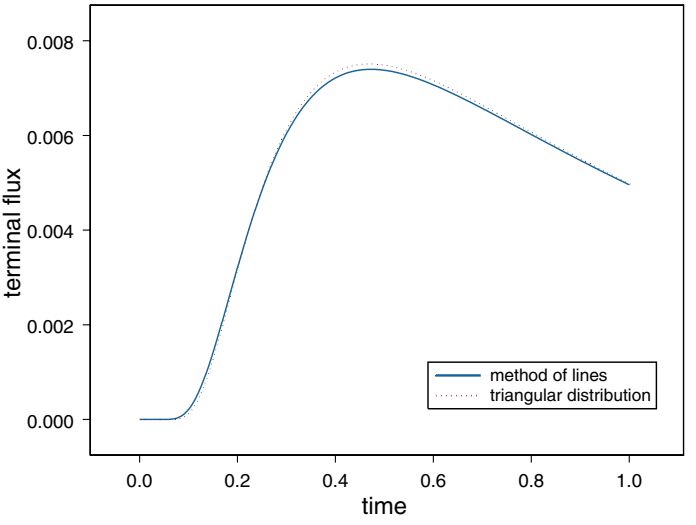


Fig. 4. The dots indicate the terminal flux obtained from iterating the triangular distribution with 100 nodes and $\Delta t = 0.01$ with $\alpha = 0.20022$ and $\delta = 0.0002$. These parameters imply an advective velocity of 0.2 and a diffusion coefficient of 0.333, and the solid line shows the solution to the equivalent continuous-time formulation using the method of lines.

and variance of the triangular distribution, estimates of the advective velocity, ω, and diffusion coefficient, γ, of the corresponding continuous-time model can be obtained as follows:

$$\hat{\omega} = \frac{1}{\Delta t} \sum_j D_j j \Delta x \tag{C.1}$$

$$\hat{\gamma} = \frac{1}{2\Delta t} \left[\sum_j D_j (j\Delta x)^2 - \left(\sum_j D_j j \Delta x \right)^2 \right]. \tag{C.2}$$

Figure 4 shows a comparison between the two numerical methods to illustrate that the more computationally efficient discrete-time method produces acceptable results. First, the terminal flux obtained from iterating the triangular distribution with 100 nodes, $\Delta t = 0.01$ and parameters $\alpha = 0.20022$ and $\delta = 0.0002$ is plotted. Equations (C.1) and (C.2) are used to derive the parameters for the corresponding continuous-time model and hence, the terminal flux is obtained using the method of lines. It is clear from Fig. 4 that the results are virtually identical for all times. Hence, the discrete-time method involving the triangular distribution is preferred, given that it is the more efficient algorithm, which is particularly important for computationally intensive tasks, such as parameter estimation.

References

1. Douglas A, Kurien A, Seasonality and other epidemiological features of haemolytic uraemic syndrome and *E. coli* O157:H7 isolates in Scotland, *Scott Med J* **42**:166–171, 1997.
2. Riley L, Remis R, Helgerson S *et al*, Hemorrhagic colitis associated with a rare *Escherichia coli* serotype, *New Engl J Med* **308**:681–685, 1983.
3. Chapman P, Siddons C, Wright D, Norman P, Fox J, Crick E, Cattle as a source of verotoxigenic *Escherichia coli* O157*, Vet Rec* **132**:323–324, 1992.
4. Wood JC, Mathematical modelling of *E. coli* O157:H7 infection dynamics within the bovine reservoir, PhD thesis, University of Strathclyde, 2002.
5. Besser T, Hancock D, Pritchett L, McRae E, Rice D, Tarr P, Duration of detection of fecal excretion of *E. coli* O157:H7 in cattle, *J Infect Dis* **175**:726–729, 1997.
6. Singleton A, The electromagnetic measurement of the flow of digesta through the duodenum of the goat and the sheep, *J Physiol* **155**:134–147, 1961.
7. Phillipson A, *Dukes' Physiology of Domestic Animals*, 9th ed., Cornell University Press, Ithaca, NY, Ch. 22, 1977.
8. Wood JC, McKendrick IJ, Gettinby G, A simulation model for the study of the within-animal infection dynamics of *E. coli* 0157, *Prev Vet Med* **74**:180–193, 2006.
9. Speirs D, Gurney W, Population persistence in rivers and estuaries, *Ecology* **82**: 1219–1237, 2001.
10. Schiesser W, *The Numerical Method of Lines Integration of Partial Differential Equations*, Academic Press, San Diego, 1991.
11. Naylor S, Low C, Besser T, Mahajan A, Gunn G, Pearce M, McKendrick IJ, Smith D, Gally D, Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic *E. coli* O157:H7 in the bovine host, *Infect Immun* **71**:1505–1512, 2003.
12. Ostroff S, Griffin P, Tauxe R, Shipman L, Greene K, Wells J, Lewis J, Blake P, Kobayashi J, A statewide outbreak of *E. coli* O157:H7 infections in Washington state, *Am J Epidemiol* **132**:239–247, 1990.
13. Naylor S, personal communication 2002.
14. Robinson SE, Brown PE, Wright EJ, Bennett M, Hart CA, French NP, Heterogeneous distributions of *Escherichia coli* O157 within naturally infected bovine faecal pats, *FEMS Microbiol Lett* **244**:291–296, 2005.
15. Press W, Teukolsky S, Vetterling W, Flannery B, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge, 1992.
16. Wood S, Minimising model fitting objectives that contain spurious local minima by bootstrap restarting, *Biometrics* **57**:240–244, 2001.
17. Silvey S, *Statistical Inference*, Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1975.
18. McCullagh P, Nelder J, *Generalized Linear Models*, Monographs on Statistics and Applied Probability, Chapman and Hall, New York, 1983.
19. Rice DH, Sheng HQQ, Wynia SA, Hovde CJ, Rectoanal mucosal swab culture is more sensitive than fecal culture and distinguishes *Escherichia coli* O157:H7-colonized cattle and those transiently shedding the same organism, *J Clin Microbiol* **41**:4924–4929, 2003.
20. Hall DB, Zero-inflated Poisson and binomial regression with random effects: a case study, *Biometrics* **56**:1030–1039, 2000.
21. Gurney W, Nisbet R, *Ecological Dynamics*, Oxford University Press, Oxford, 1998.
22. Smyth W, *Static and Dynamic Electricity*, McGraw-Hill, New York, Ch. 3, 1968.