Phosphine Mortality Estimation for Simulation of Controlling Pest of Stored Grain: Lesser Grain Borer (Rhyzopertha dominica)

Mingren Shi, Michael Renton

Abstract—There is a world-wide need for the development of sustainable management strategies to control pest infestation and the development of phosphine (PH₃) resistance in lesser grain borer (*Rhyzopertha dominica*). Computer simulation models can provide a relatively fast, safe and inexpensive way to weigh the merits of various management options. However, the usefulness of simulation models relies on the accurate estimation of important model parameters, such as mortality. Concentration and time of exposure are both important in determining mortality in response to a toxic agent. Recent research indicated the existence of two resistance phenotypes in *R. dominica* in Australia, weak and strong, and revealed that the presence of resistance alleles at two loci confers strong resistance, thus motivating the construction of a two-locus model of resistance. Experimental data sets on purified pest strains, each corresponding to a single genotype of our two-locus model, were also available. Hence it became possible to explicitly include mortalities of the different genotypes in the model. In this paper we described how we used two generalized linear models (GLM), *probit* and *logistic* models, to fit the available experimental data sets. We used a direct algebraic approach *generalized inverse matrix technique*, rather than the traditional maximum likelihood estimation, to estimate the model parameters. The results show that both probit and logistic models fit the data sets well but the former is much better in terms of small least squares (numerical) errors. Meanwhile, the generalized inverse matrix technique achieved similar accuracy results to those from the maximum likelihood estimation, but is less time consuming and computationally demanding.

Keywords—mortality estimation, probit models, logistic model, generalized inverse matrix approach, pest control simulation

I. INTRODUCTION

THE lesser grain borer, *Rhyzopertha dominica,* is a very destructive primary pest of stored grains. Fumigation with phosphine (PH₃) is a key component in the management of the control of infestations of the pest world-wide. However heavy reliance on PH₃ has resulted in the development of strong resistance in several major pest species including *R. dominica*.

Computer simulation models can provide a relatively fast, safe and inexpensive way to weigh the merits of various management options. But the usefulness of simulation models relies on the accurate estimation of important model parameters.

In previously published modelling research "survivorship was not explicitly included in the model because adequate data were not available" [8], and thus a simple single gene model was used. However, fumigant response analyses of PH₃ resistance in *R. dominica* in Australia have now indicated two resistance phenotypes, which are labelled Weak and Strong Resistance [3]. The genetic linkage analysis undertaken by Schlipalius *et al.* [10, 11] indicated that two loci confer strong resistance. Thus we constructed a two-locus model of resistance having nine possible genotypes. The experiments in Collins *et al.* [4, 5] then performed a series of mortality rate experiments on insects that had been purified to produce strains; each corresponding to a single genotype of our two-locus model. The results of these experiments were confirmed in field trials and are the basis for the current rates used to control resistant insects in Australia. These experiments differ from others (e.g. [9]) where insects are population samples from the field that contain various mixtures of resistance genes. Hence it becomes possible now to explicitly estimate mortalities for the available strains (corresponding directly to specific genotypes).

Phosphine concentration and time of exposure are both important in determining the intensity of response to a toxic agent. In practice, a fumigation treatment needs to fix the initial concentration or dose C (mg/l) and exposure time t. The ability to estimate mortality or survival rate (1 – mortality) at a range of concentrations and exposure times based on experimental data is critical for the development of accurate simulations and management recommendations.

In this paper we described how we used two models, *probit* and *logistic* models, to fit data sets from Collins *et al* experiments (2002, 2005) [4, 5] for three strains QRD14, QRD569 and their Combined F1 (QRD14×QRD569) which corresponds to three genotypes *ss* (with both loci homozygous susceptible), *rr* (with both loci homozygous resistant) and *hh* (with both loci heterozygous) respectively [13]. We also compared the least squares errors between observed mortalities and predicted ones obtained using the two models.

II. METHOD AND MATERIAL

In statistics, the *generalized linear model* (GLM) in the form of is a flexible generalization of ordinary *least squares regression* that allows the linear model to be related to the response variable via a *link function* and the magnitude of the variance of each measurement to be a function of its predicted

M. Shi is with the University of Western Australia, Crawley, WA, 6009, and CRC for National Plant Biosecurity, Australia (phone: +61 6488-1992; fax: +61 6488-1108; e-mail: shi.mingren@gmail.com).

M. Renton is with the University of Western Australia, Crawley, WA, 6009, CSIRO Ecosystem Sciences and CRC for National Plant Biosecurity, Australia (e-mail: mrenton@cyllene.uwa.edu.au).

value [6]. GLM includes ordinary linear regression, Poisson regression, *logistic regression* and *probit regression*. We used probit and logistic models to fit the experimental data sets.

$$Y = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k, \qquad (1)$$

We did not employ the statistical iterative approach – *maximum likelihood estimation*, but instead, used a novel algebraic direct approach – *generalized inverse matrix technique*, to estimate the model parameters [12]. This method has advantages over other methods: it is simple with only one key command, provides a more accurate estimate of parameters, and even if the coefficient matrix of the over-determined linear system is not numerically (column) full ranked it will still work and yield a solution with minimum error in the L_2 norm sense [1].

A. Probit models

The probit (meaning "probability unit") link function $\Phi(P)$ ($Y = \Phi(P)+5$) is the inverse cumulative distribution function (CDF) associated with the standard normal distribution [2, 7]:

$$P = \Phi^{-1}(Y - 5) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y-5} \exp\{-u^2/2\} du \qquad (2)$$

Note that "plus 5 to $\Phi(P)$" just makes sure all Y values are positive in practice, otherwise the parameter a in the following probit models differ by 5, with other parameters unchanged.

Using a three-parameter probit model [2], a probit plane

$$Y = (\Phi(P)+5 =) \ a + b_1 \log(t) + b_2 \log(C) \qquad (3)$$

may be fitted to the data, where t and C are respectively exposure time and concentration, and Y is the probit mortality.

In the case that the available independent data consist only of the products Ct (e.g. a range of C but a fixed constant time t), rather than separate independent values for C and t, the parameters b_1 and b_2 can be merged into a single parameter, b (two-parameter probit model):

$$Y = a + b \log(Ct) \qquad (4)$$

Whether common logarithms (base 10) or natural logarithms (base e) are used in probit models is immaterial because it only scales the estimated value of b.

An extra term $b_3 \log(t) \log(C)$ can be added to describe the interaction of the variables t and C, thus obtaining a four-parameter probit model

$$Y = a + b_1 \log(t) + b_2 \log(C) + b_3 \log(t) \log(C) \qquad (5)$$

B. Logistic models

The most typical link function for logistic models is the canonical *logit* link: (To distinguish the two models we use z instead of Y in Eq. (1))

$$z = \Psi(P) = \ln[P/(1-P)] \text{ or } P = \Psi^{-1}(z) = 1/(1+e^{-z}) \qquad (6)$$

where P is the same P shown in Eq. (2). In this case the linear predictor z is not translated by 5. The two-parameter and four-parameter logistic models corresponding to Eqs (4) and (5) are respectively,

$$z = a + b(Ct) \qquad (7)$$
$$z = a + b_1(t) + b_2(C) + b_3(Ct) \qquad (8)$$

C. Generalized inverse matrix approach

Algebraically, when any one of the above models is fitted to a data set, we have an over-determined system of linear equations *with respect to the parameters* to be estimated. For example, for the model (3), the N equations with 3 variables (a, b_1, b_2) corresponding to the data set $\{Y_i ; t_i, C_i\}_{i=1}^{N}$ are as follows:

$$Y_i = 1 \cdot a + [\log(t_i)] \cdot b_1 + [\log(C_i)] \cdot b_2 \ (i = 1,2,\ldots,N) . \qquad (9)$$

The matrix form of the above equations is $Ax = b$ where $x = (a, b_1, b_2)^T$,

$$A = \begin{bmatrix} 1 & \log t_1 & \log C_1 \\ 1 & \log t_2 & \log C_2 \\ \vdots & \vdots & \vdots \\ 1 & \log t_N & \log C_N \end{bmatrix} \quad \text{and } b = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}. \qquad (10)$$

The method of least squares is often used to generate estimators and other statistics in regression analysis [14]. If a solution minimizes

$$\sum_{i=1}^{N} \left(\tilde{Y}_i - Y_i \right)^2 = \sum_{i=1}^{N} \left([a + b_1 \log(t_i) + b_2 \log(C_i)] - Y_i \right)^2 \qquad (11)$$

where $\tilde{Y}_i = a + b_1 \log(t_i) + b_2 \log(C_i)$ is the ith predicted value, then the solution is called a *Least Squares* solution. Normally, the least squares method can be used to solve the regularized equations of $Ax = b$ i.e. $A^T Ax = A^T b$, provided that $A^T A$ is invertible. Actually, if we let A^+ be the generalized inverse of matrix A, then $A^+ b$ is such a solution [1]. Note that if A is a non-singular square matrix then $A^+ = A^{-1}$. If A is column full-ranked, then $A^T A$ is non-singular and $A^+ = (A^T A)^{-1} A^T$. But while this equation could theoretically be used to calculate A^+, it is of limited practical use for calculating A^+ numerically, because using *QR* decomposition or singular value decomposition (*SVC*) to obtain A^+ will give much smaller numerical errors than direct calculation of $(A^T A)^{-1} A^T$ [1].

D. Neatened raw data sets

Collins *et al.* in 2002 [4] observed mortalities under a range of phosphine concentrations (C: mg/l) at exposure time (t) 48 hrs for susceptible (strain QRD14 – corresponding to genotype *ss*) and strong resistant (strain QRD569 – *rr*) phenotypes and their combined F_1 progeny ((569×14)+(14×569) – *hh*). The neatened raw data are listed in Table 1. Note that if the response (kill) rate is $P = 1$ or 0 then we should give them a small perturbation (e.g. from 1 to **0.9999** or from 0 to **0.0001**). Otherwise the corresponding probit or logit value is undefined.

The $LT_{99.9}$ values (lethal time to achieve 99.9% mortality) are as follows:

C: 0.1,   0.15, 0.2,   0.3, 0.4, 0.5,   0.75, 1.0

$LT_{99.9}$: 14.02,12.74,8.509,7.144,6.55,5.628,4.233,3.74   (12)

These values are derived from the experiments of Collins *et al.* [5] who observed data for strain QRD569 which was exposed to a series of fixed concentrations from 0.1 to 1.0 mg/l for a range of exposure periods.

The two-parameter probit model (4) and logistic model (7) were used to fit the data sets for *ss* and *hh* since only one exposure time was available. The four-parameter probit model (5) and logistic model (8) were used to fit the two data sets for *rr* genotype [4, 5] since different combinations of exposure time and concentration were available, allowing their interaction to be considered and more accurate results obtained. Note that *t* (hrs) values used to fit the two data sets are: $t = 24 \times$(2, 2, 2, 2, 2, 14.02, 12.74, 8.509, 7.144, 6.55, 5.628, 4.233, 3.74) (Table 1 and [12]).

TABLE I
THE DATA OF PHOSPHINE DOSE AND THE AGGREGATE RESPONSE MORTALITY RATE FOR THE GENOTYPES *SS*, *HH*, AND *RR* AT A FIXED EXPOSURE TIME $T = 48$ HRS [4]

QRD14 (*ss*)					
Dose (mg/l)	0.001	0.0015	0.002	0.003	0.004
Mortality	0.0201	0.32	0.7047	0.9733	**1.0000**
Comb F1 (*hh*)					
Dose (mg/l)	0.0025	0.004	0.005	0.0075	0.01	0.02
Mortality	**0.0000**	0.3445	0.3940	0.8047	0.8591	0.9868
QRD569 (*rr*)					
Dose (mg/l)	0.1	0.25	0.5	1.0	1.25
Mortality	**0.0000**	0.0200	0.2254	0.5203	0.5705

III. RESULTS AND CONCLUSION

TABLE II
THE FITTED TWO PARAMETERS FOR STRAIN QRD14 (*SS*) AND COMB F1 (*HH*) AND FOUR PARAMETERS FOR QRD569 (*RR*) AND THE LEAST SQUARES ERRORS BETWEEN OBSERVED AND PREDICTED MORTALITIES OBTAINED USING PROBIT AND LOGISTIC MODELS

Model and fitted parameters		
Strain	Probit	Logistic
QRD14	$Y = 14.0963 + 8.4248[\log(Ct)]$	$z = -6.4461 + 70.5897(Ct)$
CombF1	$Y = 7.6101 + 4.7740[\log(Ct)]$	$z = -3.0425 + 8.5558(Ct)$
QRD569	$Y = -11.8492 + 10.0363 \log(t)$	$z = -4.9769 + 0.0193(t)$
	$-3.4563 \log(C)$	$-4.6877(C) + 0.1741(Ct)$
	$+ 3.6357 \log(t) \log(C)$	
Least squares error		
QRD14	0.00054	0.02924
CombF1	0.03368	0.14878
QRD569	0.00325	0.03357

The fitted parameters and least squares errors between observed and predicted mortalities (see Eq. (11)) obtained from the two models are listed in Table 2. Also mortality curves (against doses) for the three strains are plotted in Figs 1- 3.

It can be seen from Table 2 and Figs 1 and 2 that the sigmoid curves have the same shape for the *ss* and *hh* genotype insects and the predicted mortalities at the experimental doses obtained using probit and logistic models are both close to the observed values.
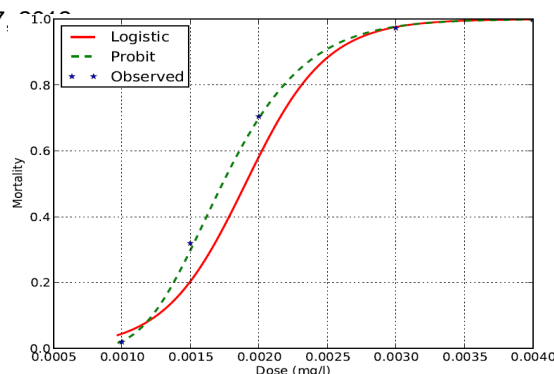

Fig. 1 Observed mortalities and predicted mortalities by probit and logistic models for the strain QRD14 (*ss*)
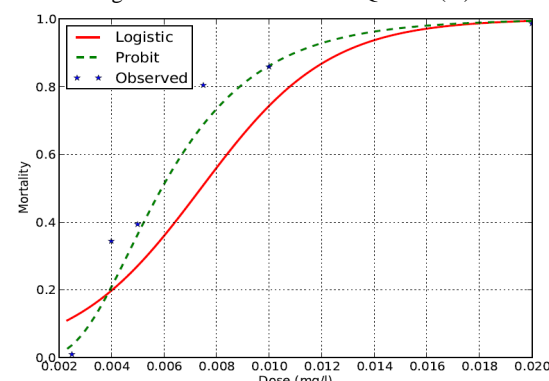

Fig. 2 Observed mortalities and predicted mortalities by probit and logistic models for the strain Comb F1 (*hh*)

The mortality curves for the *rr* insects show that the predicted values from the probit model are closer to the observed ones than those from logistic model.

The least squares errors (Table 2) from the logistic models are all more than those from the probit models; about 4 times for the *hh* beetles, 10 times for the *rr* beetle and 54 times for the *ss* beetles.

The two probit lines for QRD14 are close to each other (Fig. 4). But it can be seen from comparing the least squares errors that the generalized inverse matrix approach has smaller numerical error (for the predicted probit values) in the sense of formula (11): 0.2214 compared with 0.3850 (maximum likelihood) for the *ss* beetles (also for the other two genotype beetles).
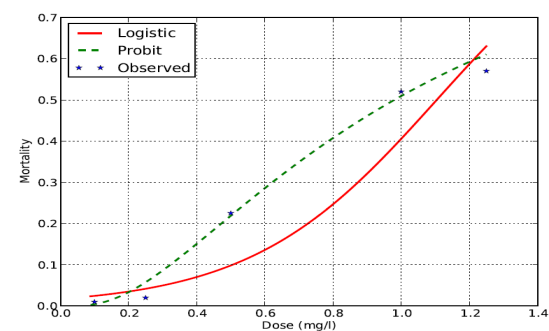

Fig. 3 Observed mortalities and predicted mortalities by probit and logistic models for the strain QRD569 (*rr*) at 48 hrs exposure time
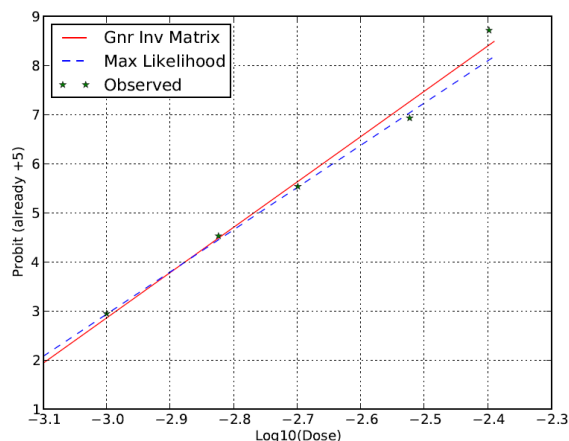
Fig. 4 Probit lines obtained using probit model with generalized inverse matrix approach (Gnr Inv Matrix) and maximum likelihood estimation

To sum up, both probit and logistic models fit the data sets well but the former is much better in terms of small least squares (numerical) errors. Meanwhile, the generalized inverse matrix technique and the maximum likelihood estimation achieved similar accuracy, but the former is a direct algebraic approach and easy to use while the latter is an iterative approach that is more time consuming and computationally demanding.

REFERENCES

[1] Ben-Israel A, Greville TNE (2003) Generalized Inverses: Theory and Applications, 2nd ed. New York, NY: Springer
[2] Bliss CI (1940) The relation between exposure time, concentration and toxicity in experiments on insecticides. Annals of the Entomological Society of America 33(4): pp. 721-766
[3] Collins PJ (1998) Resistance to grain protectants and fumigants in insect pests of stored products in Australia. In: *Proc First Australian Postharvest Technical Conference* (ed by Banks HJ, *et al*) Canberra, Australia, pp 55–57
[4] Collins PJ, Daglish GJ, Bengston M, Lambkin TM, Pavic H (2002) Genetics of resistance to phosphine in *Rhyzopertha dominica* (Coleoptera: Bostrichidae). Journal of Economic Entomology 95(4): 862–869
[5] Collins PJ, Daglish, GJ, Pavic, H, Kopittke, RA (2005) Response of mixed-age cultures of phosphine resistant and susceptible strains of lesser grain borer, *Rhyzopertha dominica*, to phosphine at a range of concentrations and exposure periods. Journal of Stored Products Research 41: 373–385
[6] Dobson AJ, Barnett AG (2008) Introduction to Generalized Linear Models (3rd ed.), Boca Raton, FL: Chapman and Hall/CRC
[7] Finney DJ (1971) Probit Analysis (3rd ed.) Cambridge University Press
[8] Hagstrum, D.W., Flinn, P.W. (1990) Simulations comparing insect species differences in response to wheat storage conditions and management practices. Journal of Economic Entomology 83 (6) 2469-2475.
[9] Pimentel MAG, Faroni LA, Totola MR, Guedes RNC (2007). Phosphine resistance, respiration rate and fitness consequences in stored-product insects. Pest Manag Sci 63: 876-881
[10] Schlipalius DI, Collins PJ, Mau Y, Ebert PR (2006) New tools for management of phosphine resistance. Outlooks on Pest Management 17: 52–56
[11] Schlipalius DI, Chen W, Collins PJ, Nguyen T, Reilly PEB, Ebert PR (2008) Gene interactions constrain the course of evolution of phosphine resistance in the lesser grain borer, *Rhyzopertha dominica*. Heredity 100: 506–516
[12] Shi M, Renton M (2011) Numerical algorithms for estimation and calculation of parameters in modelling pest population dynamics and evolution of resistance in modelling pest population dynamics and evolution of resistance. Mathematical Biosciences, 233(2): 77-89
[13] Shi M, Renton M, Ridsdill-Smith J, Collins PJ (2002) Constructing a new individual-based model of phosphine resistance in lesser grain borer (*Rhyzopertha dominica*): do we need to include two loci rather than one? Pest Science, in press, DOI 10.1007/s10340-012-0421-6
[14] Wolberg JR (2005) Data analysis using the method of least squares: extracting the most information from experiments. Springer.