Mathematical Biosciences 243 (2013) 137–146

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Modelling mortality of a stored grain insect pest with fumigation: Probit, logistic or Cauchy model?

Mingren Shi [a,b,*], Michael Renton [a,b,c]

[a] School of Plant Biology, University of Western Australia, 35, Stirling Highway, Crawley, WA 6009, Australia
[b] Cooperative Research Centre for National Plant Biosecurity, Australia
[c] CSIRO Ecosystem Sciences, Underwood Avenue, Floreat, WA 6014, Australia

ARTICLE INFO

Article history:
Received 16 June 2012
Received in revised form 27 January 2013
Accepted 6 February 2013
Available online 5 March 2013

Keywords:
Phosphine mortality estimation
Probit model
Logistic model
Cauchy model
Lesser grain borer
Pest control simulation

ABSTRACT

Computer simulation models can provide a relatively fast, safe and inexpensive means to judge and weigh the merits of various pest control management options. However, the usefulness of such simulation models relies on the accurate estimation of important model parameters, such as the pest mortality under different treatments and conditions. Recently, an individual-based simulation model of population dynamics and resistance evolution has been developed for the stored grain insect pest *Rhyzopertha dominica*, based on experimental results showing that alleles at two different loci are involved in resistance to the grain fumigant phosphine. In this paper, we describe how we used three generalized linear models, *probit*, *logistic* and *Cauchy* models, each employing two- and four-parameter sub-models, to fit experimental data sets for five genotypes for which detailed mortality data was already available. Instead of the usual statistical iterative *maximum likelihood estimation*, a direct algebraic approach, *generalized inverse matrix technique*, was used to estimate the mortality model parameters. As this technique needs to perturb the observed mortality proportions if the proportions include 0 or 1, a *golden section search approach* was used to find the optimal perturbation in terms of minimum least squares (L_2) error. The results show that the estimates using the probit model were the most accurate in terms of L_2 errors between observed and predicted mortality values. These errors with the probit model ranged from 0.049% to 5.3%, from 0.381% to 8.1% with the logistic model and from 8.3% to 48.2% with the Cauchy model. Meanwhile, the generalized inverse matrix technique achieved similar results to the maximum likelihood estimation ones, but is less time consuming and computationally demanding. We also describe how we constructed a two-parameter model to estimate the mortalities for each of the remaining four genotypes based on realistic genetic assumptions.

1. Introduction

The lesser grain borer, *Rhyzopertha dominica*, is a very destructive primary pest of stored grains. Fumigation with phosphine (PH_3) is a key component in the management of controlling pest infestations worldwide. However heavy reliance on PH_3 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 means to judge and weigh the merits of various management options for controlling populations and avoiding or delaying resistance evolution in pests such as *R. dominica*. But the usefulness of simulation models such as these relies on the accurate estimation of key model parameters, which should be based on reliable experimental data as much as possible.

In previously published simulation modelling research on this important topic of phosphine resistance in stored grain insect pests, accurate survivorship of different genotypes was not explicitly included in the model because adequate data were not available [1], and thus a simple single gene model was used. However, recent fumigant response analyses of PH_3 resistance in *R. dominica* in Australia have indicated the existence of two resistance phenotypes, which are labelled Weak and Strong Resistance [2–4]. The genetic linkage analysis undertaken by Schlipalius et al. [5,6] revealed that two loci confer strong resistance, thus motivating us to construct a more detailed and realistic two-locus individual-based simulation model of resistance [7–10]. In our two-locus model, for simplicity we assume that there are two possible alleles (resistance or susceptibility) at each of the two loci, meaning nine genotypes in total. As phosphine concentration and time of

* Corresponding author at: School of Plant Biology, University of Western Australia, 35, Stirling Highway, Crawley, WA 6009, Australia. Tel.: +61 8 6488 1992; fax: +61 8 6488 1108.
E-mail addresses: shi.mingren@gmail.com (M. Shi), mrenton@cyllene.uwa.edu.au (M. Renton).

0025-5564/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2013.02.005

exposure are both important in determining the intensity of response to the fumigant, the ability to estimate mortality for the different genotypes at a range of concentrations and exposure times based on experimental data is critical for the accuracy of this new two-locus simulation model.

Experiments by Collins et al. [11,12] and Daglish [13] provided results that help to quantify the expected mortality of these nine resistance genotypes. These were conducted on insects that had been purified to produce strains. The data sets from the experiments of Collins et al. contain three strains QRD14, QRD569 and their Combined F1 (QRD14 × QRD569) and those from Daglish's experiments contain QRD14 (the same as Collins'), QRD369 and their hybrid (QRD14 × QRD369). Each of the five strains corresponds to a single genotype of the nine possible genotypes in our two-locus model (Table 1), whereas in most previous studies (e.g. [14]), datasets were obtained from population samples from the field likely to contain various mixtures of resistance genes. Hence these new results provide the means to accurately estimate mortalities for the available five strains, which is the first phase of our mortality modelling. Moreover, those estimates can in turn be used to construct a model to estimate mortalities for the remaining four genotypes, based on reasonable genetic assumptions [8], which is the second phase.

Our previous papers [7,8] presented some discussion of phosphine mortality estimation. However, these focussed on the numerical algorithm used for model fitting, with only a limited use of experimental data for illustration [7], and presented preliminary modelling results based on simple probit models only [8]. Moreover, there was a limitation in the numerical treatment in the two papers. The kill rates in some of the experiments included one and zero [11]; to enable the probit least squares approach to be used, these two values need to be changed from 1 to $1-\varepsilon$ or from 0 to ε where ε is a small perturbation, otherwise, their link function values (see below) are undefined. But in the previous two papers [7,8] the choice of ε was arbitrary and fixed and certainly not optimal in terms of minimum least square error. This limitation motivated us to conduct this current more comprehensive study on the best way to model mortality phosphine mortality for different genotypes.

In this paper we describe how we used three models, *probit*, *logistic* and *Cauchy* models to fit the available data sets using either C (concentration or dose) and t (exposure time) themselves or $\log(C)$ and $\log(t)$ as the independent model variables, and compared the relative accuracy of probit, logistic and Cauchy models for mortality estimation. We also tested and compared two-parameter and four-parameter sub-models for each of the three models. We also show how we developed an approach for identifying an optimal perturbation value when mortality was 0 or 1 based on the *golden section search* method.

2. Materials and methods

2.1. Two-locus model with nine genotypes

To parameterise the mortality component of our simulation model, we needed to develop empirical mortality models for each genotype. Since there are two loci in the model [7–10], with two possible alleles on each of the loci, there are nine genotypes in total (Table 1).

2.2. Three generalised linear models of mortality

In statistics, the *generalised linear model* (GLM) in the form of

$$Y = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k \tag{1}$$

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* for Y and the magnitude of the variance of each measurement to be a function of its predicted value [15]. GLMs applicable to binomial mortality data include *probit regression* (with a *probit* link function), *logistic* regression (with the canonical *logit* link) and *Cauchy regression* (with the *tangent* link) which we used to fit the experimental data sets.

2.2.1. Probit model

The probit (="probability unit") link function $\Phi(P)$ ($Y = \Phi(P) + 5$) is the inverse cumulative distribution function (CDF) associated with the standard normal distribution [16,17]:

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

where P is the actual mortality (proportion that died, $0 \leqslant P \leqslant 1$) and Y is the probit transformed mortality. Note that adding five to $\Phi(P)$ just ensures all Y values are positive in practice, and simply means the parameter a is transformed by a constant value of five compared to an alternative link function where this is not applied.

Using a three-parameter probit model [16], a probit plane

$$Y = a + b_1m(t) + b_2m(C) \tag{3}$$

may be fitted to the data, where t and C are respectively exposure time and concentration, and Y is the probit mortality. We considered two choices for the function m: the logarithmic function, i.e. $m(t) = \log(t)$ and $m(C) = \log(C)$ (whether common logarithms (base 10) or natural logarithms (base e) are used is immaterial), or the identity function, i.e. $m(t) = t$ and $m(C) = C$.

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

$$Y = a + bm(Ct) \tag{4}$$

Table 1
The identifiers of the nine genotypes (*ss*,*sh*,...,*rr*) in the two-locus model, and the correspondence of genotypes and the five strains for which experimental mortality data was available [*s* – homogeneous ("homo") susceptible ("suscept"); *r* – homogeneous resistant; *h* – heterozygous ("hetero")].

2^{nd} gene / 1^{st} gene	***s*** homo suscept	***h*** heterozygous	***r*** homo resist
s homo suscept	***ss*** Both homo susceptible **(QRD14)**	***sh*** 1st homo suscpt & 2nd *heterozygous*	***sr*** 1st homo suscpt & 2nd homo resist
h heterozygous	***hs*** 1st hetero & 2nd homo suscpt **(QRD14×QRD369)**	***hh*** Both heterozygous **(QRD14×QRD569)**	***hr*** 1st hetero & 2nd homo resist
r homo resist	***rs*** 1st homo resist & 2nd homo suscpt **(QRD369)**	***rh*** 1st homo resist &2nd heterozygous	***rr*** Both homo resistant **(QRD569)**