

## ERROR ANALYSIS OF DYNAMICAL MODELS IN EPIDEMIOLOGY

G.P. POKHARIYAL and A.J. RODRIGUES

Department of Mathematics and Institute of Computer Science

University of Nairobi, P.O. Box 30197, Nairobi, Kenya

(Received October 27, 1990 and in revised form December 7, 1992)

**ABSTRACT.** In this paper we conduct the error analysis of two models used to simulate the disease profiles in plant pathogen epidemics. The role of the various model parameters is discussed in relation to accuracy.

**KEY WORDS AND PHRASES.** Sigmoid curve, error analysis, latency period, asymptote, stifling effect, piecewise continuous, simulation.

**1992 AMS SUBJECT CLASSIFICATION CODES.** 92 B-F

### 1. INTRODUCTION

In the study of epidemiology various mathematical techniques have been used to understand the dynamics of the growth of infection and related phenomena. However, the mathematical basis of some of the concepts needed streamlining. By considering the interplay between pathogen and healthy plant tissues a model for growth of infection was constructed by Pokhariyal [1].

Malice & Kryscio [2] formulated a simple stochastic model and showed that the process describing infectious contacts between individuals has a major effect on the comparison of different models. Braker [3] considered simple compartmental models for infectious disease with exposed and infective periods of fixed length. He studied the threshold phenomena and the stability of endemic equilibria. A model for the evolution of viruses proposed by Rishe & Lipkind [4] allows the computer assisted determination of ancestry among viruses.

We first discuss the role of the various parameters that describe the disease profile. The initial inoculum, denoted by  $x(0)$ , which is usually very small. The latency period, denoted by  $p$ , which is the time taken for newly infected tissues to become infectious. The constant of proportionality, denoted by  $r$ , whose dimension is the reciprocal of time. The asymptote of the curve, denoted by  $a$ , where  $x(0) < a \leq 1$ .

The analysis of the observed rate of infection  $dx(t)/dt$  indicates that at the critical point (denoted by  $t_c$ ) it goes through a maximum value (denoted by  $m$ ) as shown in [1]. This factor is

important because it gives a precise measure for determining the degree of accuracy of the simulated profile denoted by  $x^*(t)$  with respect to the sigmoid profile  $x(t)$ . It therefore, enables us to compare various simulation models from the point of accuracy and conduct the error analysis.

The integral equation for  $x(t)$  is implicit, as shown by Pokhariyal & Rodrigues [5], therefore one has to use numerical solutions, that need simulations. The implications of the simulated models are discussed with respect to accuracy.

## 2. MODEL DEFINITIONS AND ANALYSIS

In the study of the theoretical model [5], we noticed that the proportion of infected tissues  $x(t)$  can be said to increase at a rate  $dx(t)/dt$ , which is affected by the factors:

- (i)  $x(t)-x(0)$ , the proportion of plant tissue infected by the progeny of the parent inoculum  $x(0)$ .
- (ii)  $r$ , a positive time constant, that can be interpreted as an indicator of the average state of the disease.
- (iii)  $a-x(t)$ , which represents the proportion of healthy tissues that is effectively susceptible, at any time.

The model equations for the three stages are:

latency stage

$$dx(t)/dt=0, \quad 0 \leq t \leq p, \quad (2.1)$$

growth stage

$$dx(t)/dt=r\{x(t)-x(0)\}\{a-x(t)\}, \quad t > p, \quad r > 0, \quad x(0) < a \leq 1 \quad (2.2)$$

and asymptotic stage

$$dx(t)/dt \rightarrow 0, \quad \text{as } t \rightarrow \infty.$$

Combining the three time stages, through the proportionalities, into one equation, we have

$$dx(t)/dt = r\{x(t)-x(0)\}\{a-x(t)\}, \quad 0 \leq t < \infty, \quad r > 0, \quad x(0) < a \leq 1, \quad (2.3)$$

which describes the dynamics of the epidemic for all  $t$ . The implicit solution of (2.3) can be written as:

$$x(t) = a - \{a - x(0)\} \exp \left[ -r \int_0^t \{x(t) - x(0)\} dt \right] \\ 0 \leq t < \infty, \quad r > 0, \quad x(0) < a \leq 1 \quad (2.4)$$

It is noticed that for all practical purposes, the infected proportion  $x(t)$  does not change after the time  $p+I$  and can be assumed to differ by a small positive number  $\delta$  from the asymptote  $a$ . Thus,  $a-x(t) < \delta$ , for  $t \geq p+I$ , where  $I$  denotes the growth period of the epidemic.

One of the characteristic feature of the sigmoid profile  $x(t)$  is its point of inflection, where the infection rate  $dx(t)/dt$  goes through a maximum value  $m$  at the critical time  $t_c$  (see [1]), which gives

$$x(t_c) = \{a+x(0)\}/2 \quad (2.5)$$

From (2.5) and (2.3) we get

$$\hat{m} = dx(t_c)/dt = r[\{a-x(0)\}/2]^2 \tag{2.6}$$

which gives the maximum rate of infection. Putting  $t=t_c$  into (2.4) and using (2.5), we get on simplification

$$rt_c\{a-x(0)\} = 2 \ln 2 = 1.3862 \tag{2.7}$$

This relationship can be utilized to effect a control strategy [5].

The **accuracy** of the simulated solution of (2.4), denoted by  $x^*(t)$  can be checked by analysis at the critical point  $t_c$ . The equation (2.5) can not be used as an indicator of accuracy since any simulation that goes through  $x(0)$  and  $a$  must pass through their average  $\{a+x(0)\}/2$  at a time which may be quite different from  $t_c$ . On the other hand if  $dx^*(t)/dt$  attains a maximum value that approaches  $\hat{m}$  given by(2.6), then this indicates the convergence of  $x^*(t)$  to  $x(t)$ . Thus  $\hat{m}$  can be used to judge the accuracy of simulation.

**3. TWO STAGE SIMULATION MODELS**

The simulation models are used to evaluate (2.4) by introducing functions that would approximate  $x(t)-x(0)$  in some sense and allow the development of a suitable recursive scheme to calculate the area under this curve. By considering two functions in this manner, correspondingly two models are developed.

*SIMULATION MODEL IN EPIDEMIOLOGY 1 (SME1)*

We introduce a function  $x(t-p)$ , such that

$$x(t-p)=0, 0 \leq t \leq p \tag{3.1a}$$

$$\text{graph } x(t-p) = \text{graph } x(t), \text{ delayed by } p\text{-time units.} \tag{3.1b}$$

We notice that during  $0 \leq t \leq p$ , the function  $x(t-p)$  behaves exactly like  $x(t)-x(0)$  and for very large  $t$  (when  $t \rightarrow \infty$ ) the function  $x(t-p)$  differs from  $x(t)-x(0)$  by  $x(0)$ . Thus at low initial inoculum levels (when  $x(0) \ll a$ ), the function  $x(t-p)$  is a reasonable approximator (bearing in mind that during growth state  $I$ , its acceptability as an approximator depends on whether the value of  $p$  is small as compared to  $I$ ). These deductions are purely from graphical considerations implied by (3.1)b and the graphs of Fig. 1.

Thus SME1, with  $x(t-p)$  as approximation to  $x(t)-x(0)$  in (2.3) and (2.4) respectively, may be written as:

$$dx_1(t)/dt=r_1 x_1(t-p)\{a-x_1(t)\}, 0 \leq t < \infty, r_1 > 0, x(0) < a \leq 1 \tag{3.2}$$

$$x_1(t)=a-\{a-x(0)\} \exp[-r_1 \int_0^t x_1(t-p) dt], 0 \leq t < \infty, r_1 > 0, x(0) < a \leq 1 \tag{3.3}$$

where the subscript 1 is used to distinguish the computed model profile from the theoretical model profile  $x(t)$ . On comparison, we notice that the parameters  $a$  and  $x(0)$  are unchanged. However, the time constant for the computed model has to be changed as the area function it is associated with has altered and hence we denote it by  $r_1$ .

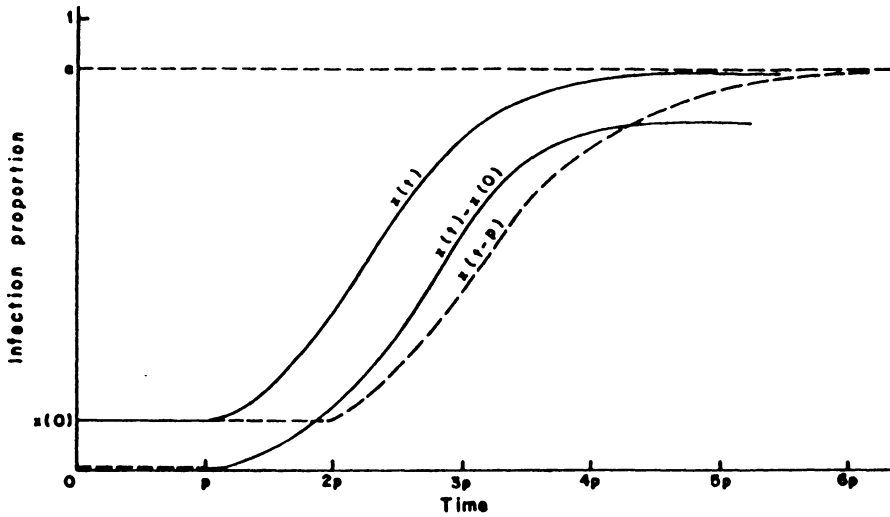


Fig. 1. In this case the infection proportion  $x(t)$ , at any time, given by  $x(t) = a - \{a - x(0)\} \exp[-r \int_0^t \{x(t) - x(0)\} dt]$  is compared with the curves  $\{x(t) - x(0)\}$  and  $x(t - p)$ .

The rate of growth of the theoretical profile  $dx(t)/dt$  at  $t=p$  is continuous, whereas the rate of growth of the computed profile  $dx_1(t)/dt$  indicates a discontinuity at  $t=p$ . A closer inspection of SME1 defined by (3.2) and (3.3) shows that in the neighbourhood of  $t=p$ ,

$$dx_1(p-\epsilon)/dt=0, \text{ for any } \epsilon>0 \tag{3.4a}$$

$$dx_1(p+\epsilon)/dt = r_1 x(0) \{a - x_1(p+\epsilon)\} \text{ for } 0 \leq \epsilon \leq p \tag{3.4b}$$

$$dx_1(p+\epsilon)/dt \rightarrow r_1 x(0) \{a - x(0)\} \text{ as } \epsilon \rightarrow 0 \tag{3.4c}$$

It can be deduced that  $x_1(p+\epsilon) > x(p+\epsilon)$ , where  $\epsilon$  is small and positive since  $x_1(p)=x(p)=x(0)$ . Thus smaller values in (3.4)b imply smaller deviations immediately after  $p$ , which provides a measure for the parametric error  $ep$  which should be very small compared to one:

$$ep_1 = r_1 x(0) \{a - x(0)\} \ll 1 \tag{3.5}$$

Furthermore, as  $\epsilon$  increases in the interval  $0 < \epsilon \leq p$ , the value of  $x_1(p+\epsilon)$  remains nearly constant, since its rate is determined by (3.4)b, a restriction that  $x(p+\epsilon)$  does not experience because it traces the exact sigmoid profile, given by (2.3). Thus for some value of  $\epsilon = \epsilon'$  say,

$$x(p+\epsilon') > x_1(p+\epsilon') \approx \text{constant of order } x(0), \text{ } 0 < \epsilon' \leq p \tag{3.6}$$

This deviation becomes more significant for large value of  $p$ . In other words, since the growth rate for  $p < t < 2p$  is nearly constant, this has a 'stifling effect' on  $x_1(t)$  during  $p < t < 2p$ . It is only when  $t > 2p$ , that  $x_1(t)$  is freed from this 'stifling effect' and able to move in accordance with the dynamics of the epidemic. Since the period for the phenomena is  $p$ , the larger its value the greater will be the deviation of  $x_1(t)$  from the true profile  $x(t)$ , during the growth stage  $p < t < p+I$ . The case  $p=I$  brings this out vividly because at  $t=p+I$ , the profile  $x(p+I)$  approaches the asymptote  $a$ , whereas  $x_1(p+I)$  is nearly of the same order as  $x(0)$ , a difference which is unacceptable (Fig. 2).

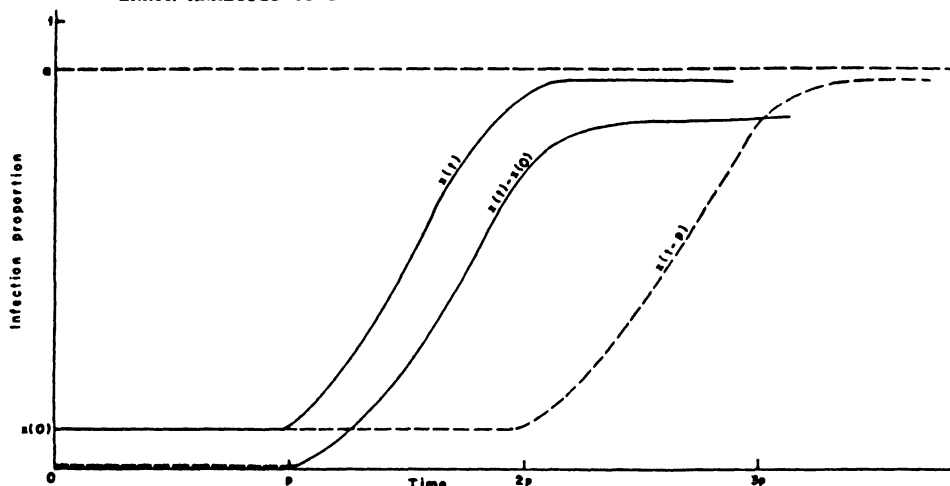


Fig. 2. In this case, the infection proportion  $x(t)$ , at any time, given by  $x(t) = a - \{a - x(0)\} \exp[-r \int_0^t \{x(t) - x(0)\} dt]$  is compared with the curves  $\{x(t) - x(0)\}$  and  $x(t - p)$  for the situation when  $p$  is large and  $p = I$ .

Thus, in general, the consequence of increasing the 'stifling period'  $p$  is that the computed profile  $x_1(t)$  would asymptote towards  $a$ , at time increasingly greater than  $p+I$ .

**SIMULATION MODEL IN EPIDEMIOLOGY 2 (SME2)**

The inaccuracy of SME1 in matching the true profile confirms the inadequacy of the function  $x(t-p)$  as a approximation of  $x(t)-x(0)$  during the growth stage. To overcome this problem, we define a picewise continuous function  $g(t)$  (see [5]), such that

$$g(t)=0, 0 \leq t < p \tag{3.7a}$$

$$g(t)=x(t-h), h < p, t \geq p \tag{3.7b}$$

where  $h$  represents the 'time shift'. However, in the functional sense  $x(t-h)$  is defined as  $x(t-h)=0, 0 \leq t < h$  and  $\text{graph } x(t-h) = \text{graph } x(t)$ , delayed by  $h$  time units. Thus from purely graphical considerations, we have

$$g(t) \rightarrow x(t)-x(0) \text{ as } h \rightarrow 0, x(0) \ll a \tag{3.7c}$$

The error of such an approximation can be made as small as we wish by choosing a smaller value of  $h$ . Hence SME2, with  $g(t)$  as an approximation to  $x(t)-x(0)$  in (2.3) and (2.4) respectively, may be written as

$$dx_2(t)/dt = r_2 g(t) \{a - x_2(t)\}, 0 \leq t < \infty, r_2 > 0, x(0) < a \leq 1 \tag{3.8}$$

$$x_2(t) = a - \{a - x(0)\} \exp[-r_2 \int_0^t g(t) dt], 0 \leq t < \infty, r_2 > 0, x(0) < a \leq 1 \tag{3.9}$$

Where the subscript 2 defines the profile of SME 2.

We notice that in the vicinity of  $t=p$

$$dx_2(p-\epsilon)/dt = 0, \text{ for } \epsilon > 0 \tag{3.10a}$$

$$dx_2(p+\epsilon)/dt = r_2 x(0) \{a - x_2(p+\epsilon)\}, \text{ for } 0 \leq \epsilon \leq h \tag{3.10b}$$

$$dx_2(p+\epsilon)/dt \rightarrow r_2 x(0) \{a - x(0)\}, \text{ as } \epsilon \rightarrow 0 \tag{3.10c}$$

These equations indicate that the derivative function is discontinuous at  $t=p$ , as in the case of SME1. However, the corresponding 'stifling period is now equal to  $h$  (time shift). Thus decreasing  $h$  reduces the impact of the stifling effect and yields

increasingly accurate profiles provided that the product of parameters in (3.10)c is small compared to one i.e the parametric error:

$$ep_2 = r_2 x(0)\{a-x(0)\} \ll 1 \quad (3.11)$$

#### 4. SIMULATION GUIDELINES AND DISCUSSION

The following guidelines are observed to simulate the theoretical profile  $x(t)$ , SME1 profile  $x_1(t)$  and SME2 profile  $x_2(t)$ , that are denoted by  $x^*(t)$ ,  $x_1^*(t)$  and  $x_2^*(t)$  respectively.

(i) We use Simpson's rule, with a step length  $h < p$ , to evaluate the area under the curve in (3.3) and (3.9).

(ii) We check that  $dx_1(p+\epsilon)/dt$  and  $dx_2(p+\epsilon)/dt$ , are much smaller than 1 at the start of the growth stage. This would ensure good simulation accuracy since the respective parametric errors  $ep_1$  and  $ep_2$  are  $\ll 1$ .

(iii) We find  $\hat{m}_1$  and  $\hat{m}_2$ , the maximum growth rates of  $x_1^*(t)$  and  $x_2^*(t)$  by computing the values using (3.2) and (3.8) respectively and then picking out the maximum value in each case. These values are then compared with the value given by (2.6), which would then give an indication of overall accuracy of  $x_1^*(t)$  and  $x_2^*(t)$ .

From various simulations that have been carried out by Vander Plank [6] and in [5], we notice that SME2 is relatively more accurate as compared to SME1 and VPRM (Vander Plank's removal model). The accuracy of SME2 is greatly improved when the step length is reduced and the long growth stage  $I$  attained with SME1 shrinks drastically with SME2. This is due to the significant reduction of the stifling effect. Thus smaller values of  $h$  give more accurate profiles that have shorter growth stages and the simulation  $x_2^*(t)$  converge to the exact profile  $x(t)$  as  $h$  approaches zero. The benefit of these mathematically complex models over the simpler model discussed in [6] is improved accuracy of the simulation provided the parametric error condition is satisfied which in general holds since the level of initial inoculum  $x(0) \ll 1$ .

#### REFERENCES

1. POKHARIYAL, G.P., A model for the growth of infection, J. Theor. Bio., **119** (1986), 181-188.
2. MALICE, M.P. and KRYSCIO, R.J., On the role of variable incubation periods in simple epidemic model, IMA J. Math. Appl. Med. Bio. **6** (1989), 233-242.
3. BRAUER, F., Epidemic models in population of varying size, Lect. Notes Biomath **81** (1989), 109-123.
4. RISHE, N. and LIPKIND, M., Modelling of virus inter-relationships for computer analysis of inhibition data, Maths. Comput. Modelling **13** (1990), 39-48.
5. POKHARIYAL, G.P. and RODRIGUES, A.J., An accurate epidemiological model, Applied Maths & Computation (1992), 1-12.
6. VAN DER PLANK, J.E., Host-Pathogen Interactions in Plant Disease, Academic Press, New York, 1982.