European Journal of Plant Pathology (2006) 115:3–23
DOI 10.1007/s10658-005-1229-5

Botanical epidemiology: some key advances and its continuing role in disease management

Laurence V. Madden
Department of Plant Pathology, The Ohio State University, Wooster, OH, 44691-4096, USA
(E-mail: madden.1@osu.edu)

Accepted 13 July 2005

Key words: basic reproduction number, Bayesian decision theory, epidemic, mixed models, model, population dynamics, receiver operating characteristic (ROC) curve

Abstract

Epidemiology involves the study of the temporal, spatial, and spatio-temporal dynamics of disease in populations, and the utilization of results of experiments and surveys to describe, understand, compare, and predict epidemics. Such understanding and description of epidemics can lead directly to the development and evaluation of efficient control strategies and tactics. Mathematical and statistical models are key tools of the epidemiologist. Recent advances in statistics, including linear and nonlinear mixed models, are allowing a more appropriate matching of data type and experimental (or survey) design to the statistical model used for analysis, in order to meet the objectives of the investigator. Coupled ordinary and partial differential equations, as well as simpler growth-curve equations, are especially useful deterministic models for representing plant disease development in fields in time and space over single seasons or many years, and their use can lead to appraisal of control strategies through metrics such as the basic reproduction number, a summary parameter that may be calculated for many general epidemic scenarios. Recently, compelling arguments have been made for the use of Bayesian decision theory in developing and evaluating real-time disease prediction rules, based on measured disease or weather conditions and either empirical or mechanistic models for disease or control intervention. Through some simple calculations of predictor accuracy and (prior) probability of an epidemic (or the need for control), the success of any predictor can be quantified in terms of the estimated probability of random observations being epidemics when predicted to be epidemics or not epidemics. Overall, despite the many contributions in epidemiology over the past four decades, more effort is still needed to convince those outside of epidemiology to more fully use epidemiological results and insights into the development and evaluation of disease controls.

Introduction

In 1963, van der Plank made a most compelling case for the importance of botanical epidemiology, both for understanding plant diseases at the population scale and for determining disease management strategies (van der Plank, 1963). He also made the bold statement at the time that 'epidemiology is here to stay.' Individual disciplines enjoy 'ups and downs' of popularity, of course, and epidemiology is no exception. The tremendous growth in the discipline within plant pathology during the 1960s, 1970s, and 1980s (e.g., Campbell and Madden, 1990; Kranz, 1990; Jones, 1998; Zadoks, 2001) has been eclipsed by growth in the larger field of molecular biology over the last two decades. Nevertheless, more than 40 years after van der Plank's book (1963), botanical epidemiology is still here, and still of utmost importance in giving a sound theoretical and practical basis for disease management. This view may not always be held outside of the discipline, however, and it

remains a challenge for epidemiologists to continue to make the compelling case that epidemiology matters.

Until molecular biology or more traditional breeding results in durable resistance to all plant pathogens on all crops, coupled with the acceptance of the new cultivars by growers and the public, there will be plant disease epidemics, and many of these will result in substantial reductions in yield. There is certainly increasing use of crop GMOs around the world (James, 2003), but cultivars with very broad-acting and durable resistance have yet to be developed. Moreover, the public opinion against their use remains strong in many regions; thus, it would be naïve to expect 'super resistant' cultivars in the foreseeable future. Use of fungicides and other chemicals in a protectant or curative manner is only practical for some crops and some diseases, and there is increasing societal pressure to (drastically) reduce the use of these chemicals in many regions. Thus, a scientific basis for applying or not applying chemicals is needed, and the decision clearly involves knowledge (or prediction) of the disease dynamics under different environmental conditions. The development of resistance to fungicides and antibiotics continues, and new cultivars have a finite lifetime.

No control tactics are known that will totally eliminate epidemics in crops and forests where the pathogen is present over large areas. Biological and cultural controls may be very beneficial, depending on the pathosystem (Maloy, 1993), but variability of control efficacy may be high with the former, and grower acceptance may be low with the latter (e.g., unwillingness to rotate crops).

The public and the scientific community have been definitely reminded of the importance of epidemiology, and the research tools that epidemiologists can bring to a problem, in recent years. A few examples are given. With increasing world trade of agricultural commodities as well as international travel, the risk of pathogen invasion of new countries or regions is well recognized (NRC, 2002), and predictions of the risk of invasion involve many epidemiological characteristics of pathogens, such as survival probabilities and reproductive potential (Madden and Wheelis, 2003). Moreover, the decision to attempt to eradicate or not also involves knowledge of disease epidemiology. The cases of citrus canker in Florida, karnal bunt in Arizona, and plum pox in Pennsylvania, U.S., are three examples of disease invasions (Gildow et al., 2004; Gottwald et al., 2001; Rush et al., 2005).

New pathogens (or pathogens new to a given crop) continue to be discovered, as well as strains, races, or biotypes of previously known pathogens. The new very aggressive biotype of African cassava mosaic virus in Africa is an example of a newly evolved isolate (Legg, 1999; Strange and Scott, 2005) that is proving very difficult to control. Sudden oak death, caused by *Phytophthora ramorum*, is a newly identified disease of oak and several other plant species, which is spreading naturally and (unfortunately) with the assistance of man, in the U.S. and elsewhere (Rizzo et al., 2002).

For diseases such as sudden oak death or Asian soybean rust (newly introduced into the U.S.), there is a great need to know the extent of spread from current locations (e.g., from the point of introduction) to other locations. For any disease that is locally concentrated (e.g., around the point of a new introduction), or does not yet exist in a country or region, ethically one cannot deliberately introduce the pathogen where it does not occur in order to study spore movement and resulting disease intensity. Thus, modelling is a key research tool for understanding risks based on key epidemiological characteristics or traits of a disease (Madden and van den Bosch, 2002; Madden and Wheelis, 2003). Epidemiology as a discipline depends heavily on the tools of mathematical and statistical modelling (Campbell and Madden, 1990), so epidemiologists are, in general, quite prepared to tackle the problem of disease spread through modelling. Model parameters for these types of situations can be obtained from observations where the disease of interest does occur naturally.

Most practicing epidemiologists would strongly support van der Plank's (1963) statement that epidemiology sets the strategy for disease control, and numerous examples can be given where epidemic knowledge leads to better control (Zadoks and Schein, 1979; Fry, 1982; Maloy, 1993). Furthermore, epidemiological principles and results can also lead to *specific* control recommendations, through the process of disease forecasting or risk prediction (Hardwick, 1998; Hughes et al., 1999), as demonstrated 45 years ago (Waggoner, 1960).

However, as pointed out recently by Jeger (2004), many controls are utilized and evaluated without explicit consideration of disease dynamics in fields. Although there is great danger in basing conclusions on disease intensity measured at one time in an epidemic (especially for polycyclic diseases; see Campbell and Madden, 1990), this unfortunately happens too often. Thus, epidemiologists still need to be pro-active in working with others in developing and evaluating disease control measures.

In the remainder of this article, I discuss a few developments that I consider to be very important in the development of plant disease epidemiology. Many more topics could have been covered. I have two major themes. One deals with the advancement in our theoretical understanding of the population-dynamic processes of disease spread in space and increase in time, coupled with the improvements in relating certain models (or their parameters) to empirical results (i.e., model fitting). The other theme deals with the prediction of plant disease on a real-time basis, or prediction of the need to impose a control measure, based on principles from probability theory. Citations are deliberately sparse, and are mainly to major reviews of topics rather than to all the (many) important original papers published over the last few decades. I assume throughout that modelling and statistical data analysis are methodological foundations for understanding epidemics and utilizing any gained knowledge in disease control.

Temporal and spatial dynamics of disease

Growth curve modelling and analysis

Van der Plank (1963) used the monomolecular and logistic equations as heuristic models of monocyclic (simple interest) and polycyclic (compound interest) disease epidemics. These models continue to be the benchmarks for quantification of epidemics, especially over single growing seasons. However, plant pathologists discovered in the 1960s and 1970s that these two models did not necessarily provide an adequate description (based on statistical principles of model fitting) for many disease progress curves (Campbell and Madden, 1990). Several alternative models were proposed or developed, some of them flexible in the sense that different degrees of skewness could be represented with the same model (depending on a realized value of a shape parameter). A feature of these models is that they are all based on a single response variable (disease intensity, y) in relation to continuous time, which can be obtained as a solution for the rate of change of y with time, dy/dt [e.g., $dy/dt = r_L y(1-y)$ for the logistic model]. In some cases, the solution can be expressed as a linear model, e.g., $\text{logit}(y) = a + r_L t$, where a is a transformation of disease intensity at time 0, r_L is the per capita rate parameter, and logit(y) is a linearizing transformation of y.

A good fit of an empirical model, or even a perfect fit, to data collected over time, is not proof of any mechanism for population growth (Campbell and Madden, 1990; Zadoks, 2001). But a good fit of a particular model for several disease progress curves could lead one to hypothesize about mechanisms, and then test the hypothesis with additional data or experiments. Moreover, using a model that provides a (reasonably) good fit to data is extremely important to accurately compare epidemics; among other things, using an inappropriate model will lead to biased estimates of the rate parameter and its standard error (Neter et al., 1983).

One clear trend in botanical epidemiology is the dramatically increasing complexity of statistical models and methods that have been applied to all epidemiological data over the last few decades (e.g., Gilligan, 2002; van Maanen and Xu, 2003). This is a natural development given the fact that epidemiology is a science of populations, and populations can only be adequately characterized and compared using the methodology of statistics. Although I am sure there are some who feel that the emphasis on mathematics and statistics obscures the understanding of the biology of epidemics, I would make the opposite claim, and declare emphatically that mathematical and statistical modelling are foundations for understanding epidemics. I further believe that, with some exceptions, the use of statistical analysis is actually still inadequate in most of epidemiological research, and certainly in most of plant pathology research! Many investigators still only: measure disease at a single time, do not match the chosen form of data analysis to the type of disease intensity variable (discrete for incidence, continuous but unequal variance for severity, ordinal for many disease rating scales); do not base their analysis on the

chosen experimental design; or perform inefficient (and sometimes uninformative) analyses. An example of the latter is the still common practice of performing a separate data analysis for each assessment time during an epidemic rather than simultaneously analyzing treatments (between-subject factors) and time (within-subject factors), and their interactions. Garrett et al. (2004) and citations therein can lead the reader to some of the important recent advances in statistical data analysis of relevance in plant pathology.

It has been known for many years (Madden, 1986) that disease values collected over time in the same experimental or sampling unit (e.g., plot) are serially correlated and that the variation in disease over time within plots is different from the variation between plots. This may be in part due to the cumulative nature of disease progress curves (see pp. 521–522 in Schabenberger and Pierce, 2002, for general discussion of cumulative processes over time). Serial correlations, sometimes called temporal autocorrelations, are especially troublesome in the comparison of treatments. My recent studies now show, however, that fitting of appropriate population-growth models to disease progress data often reduces the correlation of residuals to near zero for *individual* disease progress curves, reducing the need to directly utilize cumbersome adjustments to standard errors for calculated rates (unpublished). However, in the larger setting of multiple disease progress curves, corresponding to multiple treatment factors and blocks, there will always be non-zero correlations of observations *within* the plots by the nature of the experimental design (Schabenberger and Pierce, 2002). However, the structure of the correlations and variances may be quite complex, due to the cumulative process of disease development, but simple variance-covariance models can adjust for this property. For disease progress models that can be expressed in linear form through the use of a transformation of y [e.g., logit(y)], linear mixed models provide a tremendous (and still underutilized) tool for a thorough analysis of the epidemics (Garrett et al., 2004). Most plant pathologists (including epidemiologists) are not aware of the major advances made in mixed model analysis in statistics, a field that encompasses classical ANOVA and regression, and many other topics in a unified manner (Schabenberger and Pierce, 2002; Garrett et al., 2004). Instead of estimating disease progress model parameters for each epidemic, with a follow-up analysis of variance, through mixed models one can simultaneously estimate the disease progress parameters and their appropriate standard errors based on the explicit features of the design. The former approach (e.g., estimated slope for each plot, and then an ANOVA of these slopes), still common with researchers, is known to be the least powerful approach to detect differences in treatments (Wolfinger, 1996). Through these mixed-model methods, random effects (such as locations, blocks, and possibly genotypes), and their interactions with fixed effects (treatments) can be appropriately estimated and realistic inferences made.

Many population dynamic processes can be expressed only in nonlinear form (e.g., $y = f(t; a,b)$, where $f(\bullet)$ is a nonlinear function). The recent advances in nonlinear mixed models (Garrett et al., 2004) can be applied to these situations, but the range of experimental designs is much more limited (currently), and considerably larger data sets are required to estimate and compare parameters. Nevertheless, statistically savvy and motivated epidemiologists can make considerable progress here.

Mechanistic modelling (linked differential equations)

Van der Plank (1963) clearly realized that models such as the logistic were inadequate for a biologically meaningful characterization of disease progress in time. His approach was to use a so-called differential-delay equation in order to represent polycyclic disease development. This model relates dy/dt to the *infectious* disease intensity rather than to total disease intensity, with infectious disease estimated based on assumed fixed-duration latent and infectious periods. Although the use of differential-delay equations serve as a good foundation for developing computer simulation models with fixed time steps, such equations are extremely cumbersome for mathematical analysis, making it difficult to explore implications of different biological properties of hosts and pathogens, or of different control strategies, on long-term disease development. Eventually, plant pathologists discovered the mathematical elegance of linked or coupled differential equations for characterizing disease progress (Jeger, 1986a, b; van Maanen and

Xu, 2003). The approach – which was utilized as long ago as 1911 for representing malaria epidemics (Ross, 1911) – is to use two-to-several differential equations, with some variables of interest and parameters appearing in more than one of the equations. The beauty of this approach is that new terms can be easily added, as needed, to meet the objectives of the investigator and the details of the pathosystem, and asymptotic and steady state results (such as disease persistence) can be explored quantitatively. Furthermore, even though analytical solutions cannot generally be obtained (i.e., one cannot write out y as a function of parameters and time without the use of the integral symbol), numerical solutions are now easy to obtain with many mathematical programmes such as MATHCAD and MATHEMATICA.

Statistical software such as PROC MODEL of the SAS/EST system allows direct parameter estimation of one or more parameters for these types of models (Madden et al., 1987). The approach is iterative and computationally intensive, but readily accomplished by those who have a good understanding of nonlinear models and statistics. However, unlike the case for models with analytical solutions (linear or nonlinear; see previous sub-section above), one cannot easily incorporate the features of the experimental design (e.g., split plot, etc.) into the model fitting. Rather, one generally needs to estimate parameters for each individual epidemic (e.g., each field or plot) and then perform t-tests or analysis of variance on the estimated parameters (depending on the experimental design).

A relatively simple coupled differential equation model for a polycyclic disease with no plant mortality is given by:

$$\begin{aligned} \frac{dH}{dt} &= -\beta HI \\ \frac{dL}{dt} &= \beta HI - \omega L \\ \frac{dI}{dt} &= \omega L - \mu I \\ \frac{dR}{dt} &= \mu I \end{aligned} \qquad (1)$$

where H, L, I and R are the densities of disease-free (healthy), latently infected, infectious, and post-infectious (removed) individuals (e.g., plants, leaves, roots, or even sites on leaves), $1/\omega$ is the mean latent period, $1/\mu$ is the mean infectious period, and β is the per capita transmission rate (new diseased individuals per diseased individual per healthy individual per unit time). For fungal (or oomycetes) diseases, β is the product of spore production per time unit per infectious individual, the probability that a spore comes in contact with a healthy individual, and the probability that a spore in contact with a healthy host individual causes an infection. Total disease at any time is determined as $Y = L + I + R$, and disease intensity as a proportion is given by $y = Y/(H + L + I + R)$. If initial disease intensity is very low, then at $t = 0$, initial total host density is virtually the same as initial healthy host density, H_0. The product βH_0 is analogous to van der Plank's (1963) corrected basic infection rate (new diseased individuals per diseased individual per unit time).

A fundamental result with this model is that disease will increase (i.e., an epidemic will occur) only if $\beta H_0/\mu > 1$. The expression to the left of the inequality is known as the basic reproduction number, R_0 (Diekmann and Heesterbeek, 2000). This composite parameter also indicates the final intensity of disease (after a long time) and the initial exponential rate of increase (see Segarra et al., 2001, for details). An example realization of the model in equation 1 is shown in Figure 1 for the situation with $R_0 = 2.5$. Final disease is less than 100%, and is estimated by iteratively solving $y_\infty = 1\text{-}\exp(-R_0 y_\infty)$. Control strategies are developed or evaluated by finding combinations of β, ω, and μ that give $R_0 < 1$; specific control tactics (e.g., host resistance, protectant fungicide, curative fungicide) can then be directed at reducing β, etc.

An advantage of the equation 1 formulation is the easy expansion for other situations. For instance, a simple-interest disease component (infections from resident inoculum throughout the epidemic, rather than just at the start) can be incorporated by using the πxH term, where x is the density of inoculum and π is a simple-interest rate parameter. One can consider x to be constant or to change (typically, decline over time), so that $dx/dt = \mathrm{K}x$. When x does not change, then πx is equivalent to the monocyclic rate parameter (r_M) of the monomolecular model. The πxH term is subtracted from dH/dt and added to dL/dt in equation 1. A pure simple-interest epidemic results if $\beta = 0$; otherwise, a composite of polycyclic and

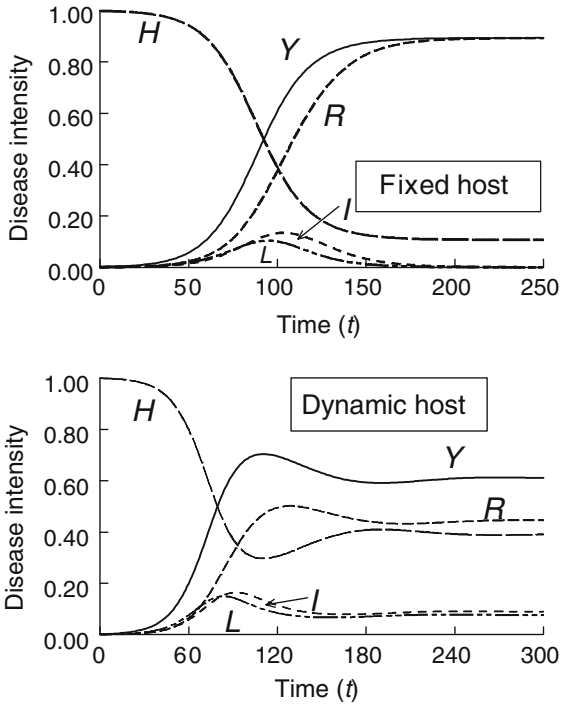

Figure 1. Density of healthy (H), latently infected (L), infectious (I), and post-infectious (R) individuals (on a proportion scale), together with total disease ($Y = L + I + R$), based on equation 1 (upper frame) and equation 2 (lower frame). Mean latent period ($1/\omega$) was 7, and mean infectious period ($1/\mu$) was 10 time units. Upper frame: $\beta H_0 = 0.25$ per time unit. Lower frame: $\beta H_0 = 0.35$ per time unit, $\eta = 0.02$, and $\pi = 0$ (no simple-interest component). Because of proportion scale, y and Y are the same here.

monocyclic processes occurs over time, very typical for root diseases (Gilligan, 2002). Host mortality can be incorporated by using a death-rate parameter η. Then ηH, ηL, ηI, and ηR are subtracted from the right hand sides of the equations for dH/dt, dL/dt, dI/dt, and dR/dt, respectively. Host growth can be incorporated in various ways. One approach is to consider just a single per capita growth rate (Ω) for disease-free individuals, and add the term Ω to the right hand side of the dH/dt equation. Suppose, further, that host size (e.g., number of citrus trees in a region) is fixed (say, at $H_{\max}$), and that new trees are only planted if others die. Then, the growth rate is also the mortality rate, and new host individuals can be expressed as $\Omega = \eta H_{\max}$; the combined growth/mortality for H can then be written as $\eta(H_{\max} - H)$.

A more general epidemic model can be written as

$$\begin{aligned}
\frac{dH}{dt} &= -\beta HI - \pi xH + \eta(H_{\max} - H) \\
\frac{dL}{dt} &= \beta HI + \pi xH - \omega L - \eta L \\
\frac{dI}{dt} &= \omega L - \mu I - \eta I \\
\frac{dR}{dt} &= \mu I - \eta R \\
\frac{dx}{dt} &= -\vartheta x
\end{aligned} \tag{2}$$

Note that in this example, total host size ($H + L + I + R$) does not change, even though there is continuous loss and addition of the host individuals (with a balance between the additions and losses). This can be seen by noting that $H_{\max} = H + L + I + R$ and adding the rates: $dH/dt + dL/dt + dI/dt + dR/dt = 0$. The model can be written in different ways to unlink the growth and mortality, to incorporate more complicated linkages, and to account for more than one disease or more than one host genotype at a time, but the example is useful to show one model formulation. When $\pi = 0$ (no simple interest component), an R_0 can be defined for many host-growth/mortality model situations. For instance, with $\pi = 0$ (no simple-interest component), $R_0 = [\beta H_{\max}/(\mu + \eta)] \cdot [\omega/(\omega + \eta)]$. An example realization of this model is shown in the lower frame of Figure 1. Note that Y ($= L + I + R$) and H oscillate a little before settling down to the steady states. The steady-state level of disease at a given R_0 is lower for the dynamic host than the fixed-host situation (equation 1); without the simple-interest component, the steady state Y is $1 - (1/R_0)$.

This approach of using a dynamic (but fixed total) host population size has been used in plant disease epidemiology (e.g., Madden et al., 2000), and even more so in medical epidemiology (Anderson and May, 1991) to determine whether or not an epidemic can occur (i.e., a disease invasion) as well as the persistence (or not) of disease long term. With primary infections occurring throughout the epidemic ($\pi > 0$), the concepts become a little more complicated, but there may still be a threshold (combination of parameters) that must be met for disease to persist (see review in Gilligan, 2002, and references cited therein).

Many other biological features can be incorporated in the model of equation 2. For instance,

since most plant viruses are transmitted by arthropod vectors, the rate of change in H and L does not directly depend on infectious plant individuals (I) but on infective vectors per plant (Z). Thus, the contact rate term, βHI in the first two equations of the model must be replaced by βHZ, where Z is the density of infective vectors per plant. Other components would be unchanged. There is also a need to add equations for the dynamics of the vector population, including virus-free and infective vectors. Details are given in Madden et al. (2000) and Jeger et al. (2004). Other expansions can incorporate disruptions caused by harvesting and/or planting for a multi-season time scale, as well as host responses to infection (e.g., Gilligan, 2002; Madden and van den Bosch, 2002).

The models shown so far are all deterministic. These can all be expressed in stochastic form, which is useful if one is specifically interested in heterogeneity of epidemics, small population sizes, or the epidemic outcome for individual plants or plant units. Gilligan (2002) and Gibson et al. (1999) provide more details. The mathematics definitely becomes more difficult with stochastic models.

Some spatial aspects of epidemics

There are two different threads to the characterization of the spatial component of plant disease epidemics. One thread deals with dispersal and resulting disease gradients, and the use of observed gradients to elucidate the form of the contact distribution (Campbell and Madden, 1990), the probability of a unit of inoculum at one location (ξ) coming in contact with a host individual at location s. This approach has been especially valuable for determining the rate of disease expansion from a focus, both within fields and higher spatial scales (e.g., continents) (van den Bosch et al., 1999). The contributions of van den Bosch and Zadoks (see Zadoks, 2001), Ferrandino (1993), and Aylor (1999) are especially noteworthy for aerial pathogens, and of Gilligan and colleagues (2002) for root diseases.

One of the advantages of the coupled differential equation approach of the previous section is that it can be directly expanded to account for disease at any location as well as any time. With two physical dimensions, it is now necessary to be explicit in notation about time t and location s. With two dimensions, we need to use partial derivatives rather than ordinary derivatives. Expanding equation 1, we can write the spatio-temporal model as:

$$\begin{aligned}
\frac{\partial H(t,s)}{\partial t} &= -\beta H(t,s) \int_{-\infty}^{\infty} I(t,\xi) D(s-\xi) d\xi \\
\frac{\partial L(t,s)}{\partial t} &= \beta H(t,s) \\
&\quad \times \int_{-\infty}^{\infty} I(t,\xi) D(s-\xi) d\xi - \omega L(t,s) \\
\frac{\partial I(t,s)}{\partial t} &= \omega L(t,s) - \mu I(t,s) \qquad (3) \\
\frac{\partial R(t,s)}{\partial t} &= \mu I(t,s)
\end{aligned}$$

where all parameters are as defined before, and $D(s-\xi)$ is the contact distribution, which is simply a scaled version of a disease gradient. Example contact distributions include the exponential, Pareto, Cauchy, and normal. Unlike with the simpler purely temporal model(s), the rate of decline in healthy host individuals at location s (and the rate of increase in latently infected host individuals at s) is explicitly based on the integration of the contributions of infectious individuals at all locations (all ξ values). The specific contribution at ξ to disease at s is the product of magnitude of infectious individuals at ξ multiplied by the probability that a unit of inoculum (say, spore) at ξ reaches location s (based on the contact distribution).

Both so-called wave-like and non-wave-like disease expansion is documented, where the velocity of disease expansion into new areas is constant or increases with time, and supported by the theory summarized in equation 3. The velocity of expansion (or the acceleration of expansion) is generally proportional to $\ln(R_0)$, so that there is no spread if $R_0 \leq 1$. The form of the contact distribution makes the difference in type of expansion. An example realization is shown in Figure 2 for non-wave-like expansion. The linkage of temporal population dynamics of disease and focus expansion rates is of fundamental importance because it shows (qualitatively and quantitatively) how reproduction (infection) and contact probabilities (dispersal) fully determine spatio-temporal

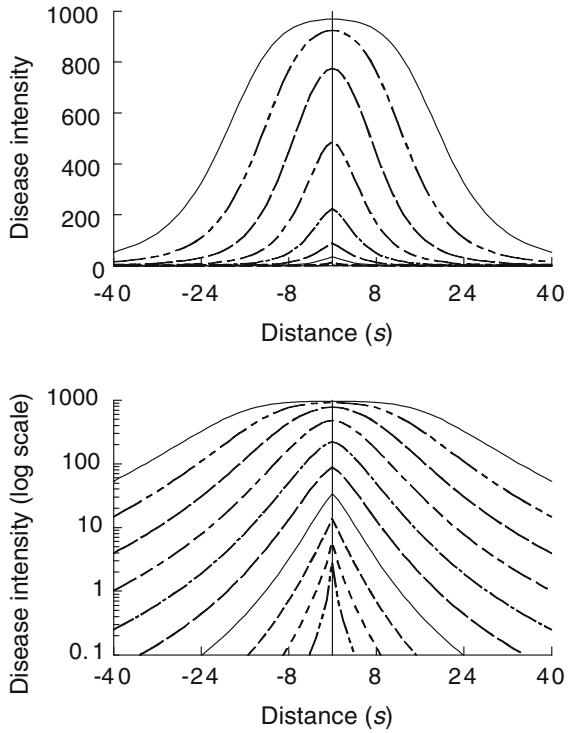

Figure 2. Density of diseased individuals ($Y = L + I + R$) vs. distance from a line source at 10-day time increments based on the numerical solution of equation 3. $H_0 = 1000$. Mean latent period ($1/\omega$) was 7, and mean infectious period ($1/\mu$) was 10 time units. $\beta H_0 = 0.4$ per time unit. A Pareto distribution was used for the contact distribution. The horizontal distance between pairs of successive curves at a single Y value (e.g., 0.1), divided by 10 gives the velocity of disease expansion.

outcomes, given a set of initial conditions. Control strategies are based, once again, on reducing R_0 to below 1, as well as reducing the scale of the contact distribution (spread parameter of the dispersal gradient) to a low value.

Equation 3 can be expanded for host growth, simple-interest dynamics, and so on, just as equation 1 was expanded to equation 2. It is (much) more difficult to work with partial differential equations than with ordinary ones, and finding numerical solutions can even be tedious. When the epidemic starts with a single focus (say, at the edge or centre of a region), then mathematical progress can be made, usually with additional assumptions (van den Bosch et al., 1990).

When there are several initial foci of infections, or unknown number and locations of initial inoculum, spatio-temporal differential-equation models, such as equation 3, are much less useful for studying epidemics because there is no single spatial starting point. With many original starting points (foci with disease at time 0), numerical solutions to equation 3 – or solutions to stochastic analogues of equation 3 (Xu and Ridout, 1998; van Maanen and Xu, 2003), – can be used to describe epidemics and explore implications of biological and physical features on disease progress, but it is more difficult to develop general principles or characterize expansion rates. Moreover, fitting a model such as equation 3 to data is generally impractical with standard statistical programmes. Thus, in epidemiology – as in ecology (Pielou, 1977) for that matter – more statistical (rather than mathematical) approaches have been generally followed to study spatial aspects of epidemics (Madden and Hughes, 1995, 2002; Hughes et al., 1997). This is the second thread of spatial characterization of epidemics. Concepts of clustering, aggregation, and regularity are utilized in terms of many different (but interrelated) statistical methods such as indices of dispersion, correlation, semivariograms, and distance statistics. This conceptual approach goes back to Cochran (1936) and Bald (1937) in plant pathology. A further advantage of the statistical approaches is that results (or concepts) are often directly useful for developing sampling plans, for either estimating disease intensity or making a decision regarding a control intervention (Madden and Hughes, 1999; Hughes et al., 2002).

The interrelationships between spatial aggregation of disease and temporal dynamics is gradually becoming more apparent. Using stochastic simulation, Xu and Ridout (1998) nicely showed how initial conditions, reproduction, and spatial contact distribution affect disease dynamics. A more theoretical approach has been to incorporate spatial properties of epidemics without explicitly using a spatial dimension (i.e., using models similar to equation 1). Models of this type are sometimes called spatially implicit, in contrast to the spatially explicit ones such as equation 3. The approach generally involves using a nonlinear function of I and/or H in the contact term, where the function depends on degree of aggregation (Zhang et al., 2000).

In recent years there has been considerable progress in bringing the two threads together (Gibson, 1997; Keeling et al., 2004), through the ingenious use of stochastic models and parameter

estimation. The results are primarily for the situation where individual plants are spatially referenced, and disease intensity is measured as a binary variable (diseased or healthy). There is still more work to do in this area, both in terms of testing the new approaches and for expanding approaches for other spatial situations (e.g., spatial referencing of just sampling units, not individual plants) and other measures of disease intensity (e.g., severity).

General thoughts on spatio-temporal disease dynamics

There is no doubt that through the expansion of models such as equations 1–3, as much detail as desired can be incorporated into models of plant disease epidemics (van Maanen and Xu, 2003). Such expansions require both knowledge of the pathosystem and knowledge of mathematics to realistically link model structure and parameterization to meaningful population-dynamic properties of the disease. Even with a model with just a few parameters, such as equation 3, mathematical insight may no longer be feasible unless starting conditions are restricted (e.g., one initial focus). Once other complicating factors are introduced, such as incorporation of environmental effects on the parameters (i.e., turning parameters into new variables that are functions of environment and new parameters), results will be limited to interpretation of numerical simulations with the model(s).

Although a model can be made indefinitely complex to represent an indefinitely complex biological system (such as plant pathosystem), such a model would violate the important principle of parsimony – keeping the model as simple as possible for the objectives of the investigator. Models, by definition, are simplifications of reality, which are useful for many purposes, including descriptions, comparisons, statistical inference, prediction, and developing understanding. Constructing models that are more complex than needed to meet the needs of the investigator – whether or not the basic biological knowledge is available for the construction of the model – is inefficient and can lead to faulty conclusions because of unrecognized (possibly erroneous) properties of the complicated model. The conclusions of Jeger (1986a) regarding the value of models with (relatively) small numbers of parameters and variables compared with large multi-variable and multi-parameter systems models is relevant here.

Thus, for many objectives, relatively simple models – such as the logistic, exponential, and monomolecular disease progress models, and empirical regression equations – will continue to be indispensable tools for the epidemiologist (Jeger, 2004). Although these models clearly are approximations, so are more complicated models. As stated by Bertrand Russell, 'Although this may seem a paradox, all exact science is dominated by the idea of approximation' (Auden and Kronenberger, 1966). Whether or not a model is too much of an approximation will always depend on the needs of the investigator.

Decision making in epidemiology

The case for disease forecasting

As stated by Gilligan (1985), 'Of the potential benefits of mathematical modelling to improving the efficiency of control of crop disease, prediction stands foremost.' Sometimes predictions or forecasts can be based explicitly on the rate parameter of a model such as the logistic or exponential (or even more complicated mechanistic model), as done with EPIPRE (Hardwick, 1998). That is, one can either use r_L to predict disease intensity some time period into the future based on either: (1) calculated r_L from previous estimates of disease in the current epidemic; or (2) predicted r_L based on environment (etc.), where the equation was developed in other studies. However, predictions need not necessarily be tied to population growth models in an explicit manner. A wide range of empirical models (often derived from regression or discriminant analysis) are utilized to simply identify conditions leading to a disease outbreak or a large reduction in yield (Madden and Ellis, 1988; Campbell and Madden, 1990). In fact, the prediction model (risk algorithm) may actually be derived without any formal statistical analysis; a good example of this is the collection of early prediction models for late blight of potato (Hardwick, 1998).

Epidemiologists continue to develop new prediction systems for plant diseases, usually used for scheduling fungicide applications, that is, for

decision-making in real time regarding the need for a control intervention (e.g., spray or not spray) (Madden and Ellis, 1988). A major development in this area over the last decade has been the application of formal Bayesian decision theory to either the construction or evaluation of the prediction systems. Growers and others (e.g., crop consultants) make numerous decisions before, during, and after growing seasons, such as when and where to plant, which cultivar to grow, whether to treat seeds with fungicide, and whether or not to apply a pesticide at any given time. Each decision can be correct or incorrect, and Bayesian decision theory provides a framework for making decisions objectively (e.g., spraying a crop) and for evaluating decisions that have been made.

The key contributions in plant pathology have been by J. Yuen, G. Hughes, and some of their colleagues (Yuen et al., 1996; Hughes et al., 1999; Yuen and Hughes, 2002; Yuen, 2003). A very recent and thorough example is Turechek and Wilcox (2005). The approach outlined below is explicitly used in the disease predictive system for Sclerotina stem rot of oilseed rape (Twengström et al., 1998), and qualitative aspects of the approach are used *implicitly* by most investigators developing and using predictive systems. I would argue, however, that a fuller utilization of the quantitative aspects of the approach will lead to better predictive systems and more efficient use of the ones already developed. The Bayesian-decision method centres on the determination of the probability of a disease outbreak (or need for a control intervention) before and after using the predictor. This approach has much in common with medical diagnostic research, where the prediction of a disease epidemic here is analogous to the diagnosis of an individual for a given disease condition. Both areas involve decisions (predictions of disease in a field or region, or the prediction of a disease status of an individual) that can only be made with some error. Plant disease prediction for crops has an additional level of uncertainty compared with medical diagnosis, since the decision is made for an entire population (e.g., a field of a given crop, or even a region where the crop is grown) rather than just for the individual.

Decision theory for disease prediction in plant pathology can be explained best with a detailed example. De Wolf et al. (2003) developed a model (their Model B) to predict major epidemics of *Fusarium* head blight of wheat in north America. The model, which is really a prediction rule in this scenario, was developed based on an analysis of 50 location-years for the disease in several parts of the U.S. The predictive system has evolved in several ways since the 2003 publication (L.V. Madden, unpublished), with many more observations analyzed as well as the development of new models, but I restrict the discussion here to the data and results of the published paper. Eighteen of the location-years were considered to be major epidemics (i.e., requiring control, if available), simply called epidemics for convenience. One can thus consider the so-called *prior* probability of an epidemic (E+) to be estimated or predicted as Prob(E+) = 18/50 = 0.36. Of course, the data set for analysis here is not necessarily representative of all locations for an indefinite period of time, but we use this calculated prior probability for now since other information was not available. Yuen (2003) discusses the use of location-dependent prior probabilities. With *Fusarium* head blight of wheat, there is a little more than a one in three chance overall that an epidemic will occur in a given location and year in the U.S. With no other information, such as measured weather variables or inoculum levels in the atmosphere or on crop debris, one would predict no epidemic – that is, one would bet against an epidemic at a given location and year, (even though one would sometimes lose the bet). This idea could be applied not just at a yearly time scale. For apple scab, one could determine the proportion of days (or weeks, for instance) where an infection period occurs in the spring. This can be considered the estimated prior probability of the need to apply a fungicide, independent of any other information (e.g., ignoring weather data).

Returning to *Fusarium* head blight, the probability of no epidemic (E−) in the De Wolf et al. data set is given by Prob(E−) = 32/50 = 0.64 = 1–0.36. For ease of calculations, it is convenient to determine the *odds* from the probability. In general, if A is some event, then odds(A) = Prob(A)/[1-Prob(A)]. If the odds are known, the probability is obtained from Prob(A) = odds(A)/[1 + odds(A)]. With the example, the odds are: odds(E+) = 0.36/[1–0.36] = 0.563, and odds(E−) = 0.64/[1–0.64] = 1.778. Note that the odds(A) = 1 when Prob(A) = 0.5. Probabilities above ½ give odds above 1, and probabilities less than ½ give odds below 1. The main symbols used in this part of the article are summarized in Table 1, for convenience.

Table 1. Some of the notation used regarding decision theory for disease prediction

Symbol	Description
Prob(E+)	Prior probability of an epidemic (or major epidemic, or for the need for a control intervention).
Prob(E−)	Prior probability of no epidemic (or for the lack of need for a control intervention)
odds(E+)	Prior odds of an epidemic: Prob(E+)/[1−Prob(E+)]
odds(E−)	Prior odds of no epidemic: Prob(E−)/[1−Prob(E−)]
Prob(P+|E+)	Probability of an actual epidemic being correctly predicted using some specified disease forecasting or predictive system; the conditional probability of a prediction of an epidemic (i.e., Z > threshold) given that an epidemic has occurred.
Prob(P−|E−)	Probability of an actual non-epidemic being correctly predicted using some specified disease forecasting or predictive system; the conditional probability of a prediction of a non-epidemic (i.e., Z < threshold) given that an epidemic has not occurred.
Prob(P+|E−)	Probability of an actual non-epidemic being incorrectly predicted to be an epidemic; [=1−Prob(P−|E−)]; the conditional probability of a prediction of an epidemic given that an epidemic has not occurred.
Prob(P−|E+)	Probability of an actual epidemic being incorrectly predicted to be a non-epidemic [=1−Prob(P+|E+)]; the conditional probability of a prediction of a non-epidemic given that an epidemic has occurred.
Z	Indicator of the risk of an epidemic (or the need for a control intervention). Can be derived with statistical or non-statistical methods.
TPP	True positive proportion (*sensitivity*); proportion of epidemics correctly predicted. An estimate of Prob(P+|E+)
TNP	True negative proportion (*specificity*); proportion of non-epidemics correctly predicted. An estimate of Prob(P−|E−)
FPP	False positive proportion; proportion of non-epidemics incorrectly predicted to be epidemics [=1−TNP]. An estimate of Prob(P+|E−).
FNP	False negative proportion; proportion of epidemics incorrectly predicted to be non-epidemics [=1−TPP]. An estimate of Prob(P−|E+)
J	A measure of accuracy [= TPP + TNP − 1 = TPP − FPP], known as Youden's index.
Prob(E+|P+)	Posterior probability of an epidemic given that one is predicted. Also known as the positive predictive value, PV(+).
Prob(E−|P−)	Posterior probability of no epidemic given that one is not predicted. Also known as negative predictive value, PV(−).
Prob(E-|P+)	Posterior probability of no epidemic given that one is predicted [=1−Prob(E+|P+)].
Prob(E+|P−)	Posterior probability of an epidemic given that one is not predicted [1−Prob(E−|P−)].
LR(+)	Likelihood ratio of a positive prediction (i.e., prediction of an epidemic), TPP/FPP.
LR(−)	Likelihood ratio of a negative prediction (i.e., prediction of a non-epidemic), FNP/TNP.
ROC	Receiver operating characteristic curve, a plot of TPP vs. FPP. Can be written mathematically as TPP = *f*(FPP).
odds(E+|P+)	Posterior odds of an epidemic given that one is predicted (see equation 4).
odds(E−|P−)	Posterior odds of a non-epidemic given that one is not predicted (see equation 5).
CR	Cost ratio, approximately equal to the cost of a false positive (C_{FP}) divided by the cost of a false negative (C_{FN})
LR*(+)	Instantaneous likelihood ratio; the slope of the tangent to the ROC curve at any (FPP, TPP) point. Also given as first derivative *f*′(FPP), of the model for the ROC curve (see equation 8 for example).

The question that arises in the context of disease forecasting or predictive systems is: can one substantially change the predicted probability of an epidemic (or the odds) based on other information? De Wolf et al. (2003) used logistic regression to develop a risk algorithm (an equation in this case) for predicting an epidemic. The following predictor was obtained:

$$Z = -3.725 + 10.5(X_1 X_2)$$

where X_1 is number of hours that temperature is between 15 and 30 °C for the 7 days prior to wheat flowering, and X_2 is the number of hours that temperature is between 15 and 30 °C and relative humidity is at least 90% for the 10 days starting at flowering, and Z is the predicted logit of the probability of an epidemic given the two weather variables. For other pathosystems, Z could represent a direct measurement of, for instance, hours of wetness, rather than being a function derived from other variables. Z could also represent an estimate of disease intensity (e.g., measured disease early in a growing season) that is used to predict final disease or crop yield (see Turechek and Wilcox, 2005; Yuen and Hughes, 2002).

It turns out that the chosen threshold to use for predicting an epidemic with this data set is $Z = -0.40$ (recall that Z is a logit); for a given location-year, if Z is above the threshold, then predict an epidemic (and label this P+), otherwise predict no epidemic (and label this P−). Using this rule, 15 of the 18 known epidemics were correctly predicted, giving a true positive proportion of TPP = 15/18 = 0.833. Also, 27 out of the 32 known non-epidemics were correctly predicted, giving a true negative proportion of TNP = 27/32 = 0.844. One can also calculate the proportion of known non-epidemics predicted to be epidemics, which is the false positive proportion, FPP = 5/32 = 0.156. Finally, the proportion of known epidemics predicted to be non-epidemics is the false negative proportion, FNP = 3/18 = 0.167. It can be shown that FPP = 1-TNP, and that FNP = 1-TPP. All of these calculations are based on the known (or assumed) status of each observation in the data set. Overall accuracy could be reported as (15 + 27)/50 = 0.840. However, this metric depends on the TPP and TNP values as well as the fraction of observations in each category, and can thus be a misleading indicator of model (predictor) success if the fraction of epidemics is fairly far from ½. A better overall measure of accuracy is given by Youden's index, J = TPP + TNP − 1 = TPP − FPP; J equals 1 for a perfect predictor. For the example, $J = 0.677$.

The TPP is often called the *sensitivity* of a predictor or sensitivity of a model, and is an estimate of the probability of an actual epidemic being correctly predicted, Prob(P+|E+). Likewise, TNP is often called the *specificity* of a predictor (or of a model), and is an estimate of the probability that a non-epidemic is correctly predicted, Prob(P−|E−). FPP is the estimate of the probability that an actual non-epidemic is incorrectly predicted to be an epidemic, Prob(P+|E−); FNP is the estimate of the probability that an actual epidemic is incorrectly predicted to be an non-epidemic, Prob(P−|E+). The following table summarizes the metrics and the estimates for the example.

Predicted → Actual↓	P+	P-
E+	TPP Prob(P+|E+) 0.833	FNP (= 1−TPP) Prob(P−|E+) 0.167
E−	FPP (1−TNP) Prob(P+|E−) 0.156	TNP Prob(P−|E−) 0.844

Although TPP and TNP are very similar here, this is not necessarily the case.

The effectiveness of a predictor can be expressed in another way, which is extremely useful for some calculations below. The likelihood ratio of a positive prediction (i.e., prediction of an epidemic) is estimated by: LR(+) = TPP/(1-TNP) = TPP/FPP. Furthermore, the likelihood ratio of a negative prediction (i.e., prediction of a non-epidemic) is estimated by: LR(−) = (1-TPP)/TNP = FNP/TNP. For the example, one obtains LR(+) = 5.34 and LR(−) = 0.20. An accurate predictor has, in general, large LR(+) (above 1) and small LR(−) (close to 0).

The use of a threshold of −0.4 for Z gives an overall high accuracy (high J), but this is not the only possibility. This can be seen by the TPP and TNP values over the full range of possible Z

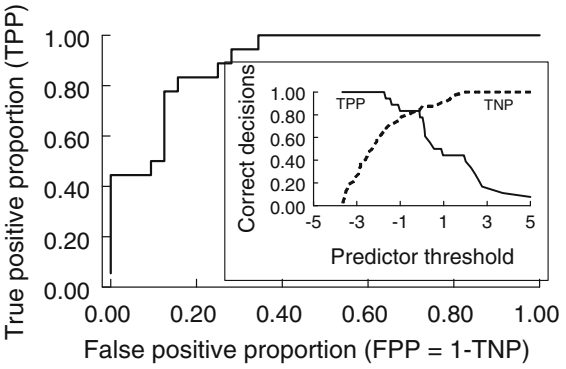

Figure 3. Main graph: An ROC curve, that is, the true positive proportion (TPP) vs. the false positive proportion (FPP), for the predictor model in De Wolf et al. (2003). Inset graph: TPP and true negative proportion (TNP = 1−FPP) vs. a full range of decision thresholds.

values, as shown by the insert of Figure 3. If one chose a low threshold (say, −3.5), then all the actual epidemics would be correctly predicted (TPP = 1), since, essentially, all observations are then predicted to be epidemics (all observations have Zs larger than −3.5). There is a major consequence of a low threshold, however, in that the actual non-epidemics are also predicted to be epidemics (TNP = 0 or FPP = 1). The J value would then be 0, a completely undesirable result. As the threshold increases above −3.5 for a predicted epidemic, TPP goes down, since it is now harder to predict an epidemic, and increasing numbers of the actual epidemics are predicted to be non-epidemics (i.e., some actual epidemics have Z values less than the threshold). However, TNP goes up with the increasing threshold, as higher numbers of non-epidemics are being correctly predicted (i.e., many of the actual non-epidemics have Z values below the threshold, as desired). Ultimately, with a very high threshold, all actual non-epidemics are correctly predicted (TNP = 1 or FPP = 0), since the threshold is higher than all the observations. However, this means that all the actual epidemics are also predicted to be non-epidemics (TPP = 0, FNP = 1), since it is then impossibly hard to predict an epidemic (i.e., all actual epidemics have Z values less than the high threshold). In between these extremes, there are thresholds around −1 to 0 where both TPP and TNP are high.

The overall performance of any predictor can be summarized with a receiver operating characteristic (ROC) curve (Metz, 1978; Linnet, 1988; Hughes et al., 1999), which is a plot of TPP vs. FPP (see Figure 3), that is, a plot of sensitivity vs. 1-specificity. The curve goes from (0,0) at the lower left corner to (1,1) at the upper right corner. The upper corner represents the *lowest* threshold tested (i.e., smallest Z value), corresponding to maximum sensitivity (high TPP) but minimum specificity (low TNP). The lower left corner represents the *highest* threshold tested (i.e., largest Z value), corresponding to minimum sensitivity and maximum specificity. If the predictor is of no value, the ROC curve will give a straight line through these two extremes, with a slope of 1. An ideal predictor will give a curve that goes vary rapidly from (0,0) up to a TPP value of 1 at an FPP barely above 0 (i.e., 0^+). The maximum J over all possible thresholds of Z for accuracy occurs at the point on the ROC curve that is closest to the upper left corner. The area under the ROC curve is an overall measure of the prediction accuracy, with a maximum of 1; for the example, the area is 0.9.

One can think of the ROC curve as representing the model TPP = f(FPP). It can be shown that the first derivative of this model [f'(FPP)] is the *instantaneous* value of LR(+) at any point FPP, that is, the tangent to the TPP:FPP curve at any FPP (Hughes and Madden, 2003). I call this likelihood ratio LR*(+). The more common calculation of LR(+), and the only one possible when the ROC curve is not available, is the straight-line slope over the interval from the point (0,0) to the point (FPP,TPP), which equals LR(+) = TPP/FPP (as indicated above).

Predictors in practice

All of the statistics shown so far deal with the success of the predictor for *known* epidemics and non-epidemics (i.e., for known status of the observations). To assess the predictor in practice, one must determine the probability that a random observation of unknown status (a particular location-year in the example) is an epidemic, given that the predictor score is positive (Z > the threshold), written as Prob(E+|P+), or the probability that a random observation of unknown status is not an epidemic, given that the predictor score is negative, written as Prob(E−|P−). Note that the conditional probabilities have been turned around from that used in developing the predictor, where the epidemic status was known; now, the prediction

status is known and not the actual status of an observation. To determine these and related probabilities for the population of interest (i.e., all location-years where the predictor is being used), one invokes Bayes' Theorem (Yuen and Hughes, 2002; Hughes and Madden, 2003), which can be most easily written (and interpreted) in terms of odds rather than probabilities.

The odds of an epidemic, given one is predicted [odds(E + |P +)] depends on the accuracy of the predictor, expressed as LR(+), and the prior odds that an observation is an epidemic [odds(E +)]. This new odds value can be written as:

$$\text{odds}(E+|P+) = LR(+)(\text{odds}(E+) \quad (4)$$

which is a simple direct product of the two terms. The left-hand side of this equation is known as the posterior odds of an epidemic (or a disease outbreak, or the need for a control intervention, etc.), given that one is predicted. The posterior odds will be high (relatively speaking) if the prior odds is high (e.g., a relatively high proportion of location-years for *Fusarium* head blight are epidemics) or if LR(+) is high (i.e., high accuracy). For the example, the posterior odds is given by: odds(E + |P +) = 5.34·0.563 = 3.0. The posterior probability can be determined by the transformation of the odds (see above), which produces Prob(E + |P +) = 3.0/(1 + 3.0) = 0.75. The posterior probability can be calculated directly, without use of prior and posterior odds (Yuen and Hughes, 2002), but the formula is cumbersome and less intuitive.

In typical usage, we would only predict an epidemic if the posterior odds is above 1 [or Prob(E + |P +) > ½]. In the example, the predicted odds of an epidemic occurring when the model predicts one is more than five times the average (or overall) odds of an epidemic when the predictor-variable information is not known (or not used). The predicted posterior probability of an epidemic when one is predicted (0.75) is a little more than double the prior probability (0.36) when no information is known (or used). Note that a predictor is only valuable if LR(+) is larger than 1 in this simple situation. When LR(+) = 1, the prior and posterior odds are the same, as well as the prior and posterior probabilities, which means that using the predictor is not giving the user any additional information about the chance of an epidemic – one is no more certain of the epidemic status of a random observation when the predictor is used compared to when it is not used.

One can also determine the posterior probability that an observation is not an epidemic when an epidemic is actually predicted, Prob(E−|P +), by first calculating the posterior odds: odds(E−|P +) = odds(E−)/LR(+). Note that Prob(E−|P +) is also given by 1-Prob(E + |P +). For the example, Prob(E−|P +) = 0.25. Thus, there is still a reasonable probability (¼) that an observation is not an epidemic even when one is predicted, with the given accuracy of the model. Another valuable term is the posterior probability that an observation is not an epidemic, given that a non-epidemic is predicted, Prob(E−|P−). This can be readily determined from:

$$\text{odds}(E-|P-) = \text{odds}(E-)/LR(-) \quad (5)$$

With the example, the posterior odds is estimated by: odds(E−|P−) = 1.778/0.20 = 8.98. The posterior probability is thus: Prob(E−|P−) = 8.98/(1 + 8.98) = 0.90. In other words, the probability that there will not be an epidemic when an epidemic is not expected (= 0.9) is increased compared to the prior probability of a non-epidemic (0.64) when no other information is available. Finally, the posterior probability of an epidemic given that a non-epidemic is predicted, Prob(E + |P−), can be determined from 1-Prob(E−|P−), or by first calculating the posterior odds as: odds(E + |P−) = odds(E +)LR(−). For the example, one obtains Prob(E + |P−) = 0.10, meaning that there is only a small probability that a random unknown observation is actually an epidemic when a non-epidemic is predicted. Because of the properties of the predictor model and the prior probability of an epidemic, in the example one would have somewhat more confidence in predictions of non-epidemics than in predictions of epidemics.

The predictor is clearly of value based on the achieved likelihood ratios and the prior odds (or prior probabilities). The point of interest here is how the success of the predictor depends on the prior odds of an epidemic (which comes directly from the prevalence of epidemics). For instance, consider what would happen if epidemics were much less common than assumed originally, but one still wanted to use the developed predictor.

I call this scenario B (and the original scenario above as A). Assume the prior probability of an epidemic is Prob(E+) = 0.05, which gives: odds(E+) = 0.053, Prob(E−) = 0.95, odds(E−) = 19.0. For Fusarium head blight this is unrealistically low, but the value is used for demonstration purposes. The likelihood ratio is unchanged since this is a property of the predictor, not the prior probability. The posterior odds of an epidemic then is calculated from equation 4 as odds(E+|P+) = 5.34·0.053 = 0.28, resulting in a posterior probability of an epidemic when one is predicted of Prob(E+|P+) = 0.28/(1 + 0.28) = 0.22 (Table 2). The posterior probability of an observation not being an epidemic when one is predicted is very high, equal to Prob(E−|P+) = 1–0.22 = 0.78. The other posterior probabilities are: Prob(E−|P−) = 0.99, and Prob(E+|P−) = 0.01 (i.e., there is a very low probability that a random observation is an epidemic when one is not predicted). Using environmental data in the form of the predictor (Z), the probability that a random observation (a location-year) is an epidemic clearly increases when one is predicted (from the prior probability of 0.05 to the posterior probability of 0.22). However, there is still considerably less than a 50% chance that an observation is an epidemic when one is predicted, and there is nearly an 80% chance (posterior probability of 0.78) that an observation is a non-epidemic when one is predicted. In other words, use of the current predictor (with the selected threshold of Z for a positive prediction) would be of little value in disease management when epidemics are rare – most control interventions would be wasted since most of the predicted epidemics would turn out to be non-epidemics. This shows in general that disease forecasting may not be of direct value for rare diseases, unless one has a predictor with an extremely high overall accuracy (very large LR(+)).

With an imperfect predictor (i.e., TPP < 1, TNP < 1), there is uncertainty in any predictions of epidemics. Given that epidemics do not occur that often (hypothetically, when Prob(E+) = 0.05), the evidence must be stronger than that obtainable from the use of the predictor to conclude (at least with more than a 0.50 probability, or more than an odds of 1) that an epidemic will occur when predicted. However, if LR(+) was 20 (i.e., a much more accurate predictor), then the posterior odds would be 1.06 (when prior odds of an epidemic was 0.053), and the posterior probability would be 0.51. Under these circumstances, the use of the predictor would be of greater value in management. However, finding such accurate predictors in plant pathology may be very difficult.

There is an alternative to improving the overall prediction accuracy for rare diseases. One can use a different threshold of Z for an epidemic, which can be demonstrated with the *Fusarium* head blight results. As shown in Figure 3, TPP declines, and TNP increases, as the threshold increases. If one used a threshold (on the logit scale) of +2 (instead of −0.4), one would obtain TPP = 0.39, TNP = 0.99, FPP = 1–0.99 = 0.01, and FNP = 1–0.39 = 0.61; the likelihood ratios would be LR(+) = 39.0 and LR(−) = 0.62. I call this scenario C (see Table 2). Using a higher threshold means moving down the ROC curve (Figure 3) towards the lower left corner. By using a high threshold, almost all the known non-epidemics are correctly predicted (more specifically, almost all the *known* non-epidemics have Z values less than the new higher threshold; TNP ≈ 1), but only 40% of the *known* epidemics are correctly predicted

Table 2. Evaluation of disease predictor for *Fusarium* head blight of wheat (see De Wolf et al., 2003) under various scenarios of prior probability of an epidemic and threshold used for predicting an epidemic[a]

Scenario	Prob(E+)	Threshold	TPP	TNP	LR(+)	LR(−)	Prob(E+|P+)	Prob(E−|P−)
A[b]	0.36	−0.4	0.833	0.844	5.34	0.20	0.75	0.90
B	0.05	−0.4	0.833	0.844	5.34	0.20	0.22	0.99
C	0.05	+2.0	0.39	0.99	39.0	0.62	0.67	0.97
D	0.85	−0.4	0.833	0.844	5.34	0.20	0.97	0.47
E	0.85	−1.7	0.944	0.656	2.74	0.085	0.94	0.67

[a]See text and Table 1 for explanation of symbols and notation, as well as for terms not given in table.
[b]Scenario A is the nominal (or standard) use of the predictor as described in the article.

(TPP $\approx$ 0.4) because many of these cases have Z values below the new threshold. In other words, it is now more difficult to correctly predict a known epidemic. Although this may seem to be undesirable, the low prior probability of an epidemic (0.05) means that the posterior probability of an epidemic when one is predicted to occur is actually improved (i.e., predicted epidemics are more likely to actually be epidemics). Using the numbers in this paragraph (including a prior probability of an epidemic of 0.05), Prob(E+|P+) = 0.67 (substantially higher than the 0.22 posterior probability for scenario B), which means that in only about one third of the time, on average, would a predicted epidemic actually correspond to a non-epidemic [i.e., Prob(E−|P+) = 1–0.67 = 0.33]. When one requires stronger evidence for an epidemic [a higher threshold, giving a larger LR(+)], there is less of a chance that the prediction of an epidemic is wrong. There is a slight cost here to using the higher threshold – the posterior probability of an observation being a non-epidemic when predicted to not be an epidemic is reduced to Prob(E−|P−) = 0.97, compared with 0.99 with the nominal predictor threshold (scenario B). This is due to the increase in LR(−) compared to the original choice of threshold. Here, very little was lost in identifying non-epidemics by changing the threshold for a positive prediction (since it is, relatively speaking, easy to predict non-epidemics when epidemics are rare).

One can also consider the implication of much more common occurrence of epidemics. For instance, if the prior probability of an epidemic is Prob(E+) = 0.85 (much higher than realistic for Fusarium head blight), one obtains: odds(E+) = 5.67, Prob(E−) = 0.15, odds(E−) = 0.176. Consider the predictor used at the nominal threshold (−0.4), which gives, once again, LR(+) = 5.34 and LR(−) = 0.2. I call this scenario D (Table 2). One obtains the following posterior probabilities: Prob(E+|P+) = 0.97, Prob(E−|P+) = 1–0.97 = 0.03, Prob(E−|P−) = 0.47, and Prob(E+|P−) = 0.53. Here, the predictor works very well for predicting epidemics (there is only a 3% chance that an observation is a non-epidemic when one is predicted), but works less well for predicting the non-epidemics. Based on Prob(E+|P−), about half the observations predicted to be non-epidemics are, on average, actually epidemics. In the absence of a more accurate predictor model [that would give a combined higher LR(+) and lower LR(−)], one could move the threshold to a lower value (see Figure 3), which corresponds to a higher TPP and lower TNP (the opposite direction than used when epidemics were rare). A lower threshold means moving up the ROC curve towards the upper right corner.

With a threshold of −1.7, one obtains: TPP = 0.944, TNP = 0.656, FPP = 1–0.656 = 0.344, and FNP = 1–0.944 = 0.056; the likelihood ratios would be LR(+) = 2.74 and LR(−) = 0.085. This is scenario E (Table 2). By making it easier to predict known epidemics (i.e., lowering the threshold of Z for deciding in favour of an epidemic) when epidemics are common, one does not change the predictions of epidemics very much; that is, because of high prevalence of epidemics, Prob(E+|P+) = 0.94, only slightly less than under scenario D]. However, it is much more likely that an observation predicted to be a non-epidemic is, in fact, a non-epidemic. That is, Prob(E−|P−) = 0.67, compared to 0.47 for scenario D. But there is still a fairly high probability that a random observation is an epidemic when a non-epidemic is predicted [Prob(E+|P−) = 0.33].

The above evaluation was totally presented in terms of commonness of epidemics (or of the need to control, in general), measured by estimated prior odds, and the accuracy of the predictor for known cases, measured by likelihood ratios. The entire evaluation can be recast in terms of costs for each of the four possible decisions (true positive, true negative, false positive, and false negative), or more simply, the costs of the two incorrect decisions (false positives and false negatives) (Linnet, 1988). Hughes and Madden (2003) give a detailed account of this for regulatory problems (invasive organism risk analysis) rather than for disease forecasting. In brief, if C_{FP} and C_{FN} are the costs of a false positive and a false negative prediction, respectively, then define CR as the ratio of these: CR $\approx C_{FP}/C_{FN}$ (see Table 1). CR actually depends also on the costs of true positives and true negatives, but relative to the costs of the errors, it is quite practical to consider these other costs as nil. Then, the decision rule that minimizes the average cost of using the predictor can be shown to be:

odds(E + |P+) > CR, then predict an epidemic;

odds(E + |P+) < CR, then predict a non-epidemic.

The posterior odds is calculated from equation 4, based on the prior odds and accuracy of the predictor (the likelihood ratio). The lower the CR, the lower the posterior odds needed to predict an epidemic. A low CR would occur when the costs of false positives (such as the cost of spraying a crop where the fungicide application is not needed) is relatively low compared to the costs of false negatives (such as the yield loss due to the disease that occurs because a needed fungicide spray was not used). As shown by Hughes and Madden (2003), the optimum threshold of Z to use for minimizing costs of using a predictor can be determined based on CR. In particular, one can write:

$$\mathrm{LR}^{*}(+)_{\mathrm{opt}} = \mathrm{CR}/\mathrm{odds}(\mathrm{E}+) \tag{6}$$

where the left-hand side is the optimum LR*(+) in which to make epidemic predictions (a function of TPP and FPP, which are properties of the predictor). Equation 6 is easy to calculate. The right hand side is based on commonness of epidemics and the relative costs of predictor errors (simply as a ratio, so that absolute values of the costs are not needed), but does not involve the accuracy of the predictor. It should be re-emphasized that the predicted LR*(+) here is the 'instantaneous' change in TPP with change in FPP (slope of the tangent to the TPP:FPP curve at FPP), given as f' (FPP). To translate equation 6 into an exact predicted combination of TPP and TNP (or FPP = 1-TNP), one must first have a specific model for the ROC, TPP = f(FPP), in order to obtain f' (FPP) at the optimum point (Hughes and Madden, 2003). An example is equation 4 in Hughes and Madden (2003). The fit of this model to the ROC curve in Figure 3 using nonlinear least squares results in the following equation:

$$\mathrm{TPP} = \left(1 + e^{-4.42}\left(\mathrm{FPP}^{-2.37} - 1\right)\right)^{-1/2.37} \tag{7}$$

As shown in Figure 4, the model provides a good fit to the TPP values. The first derivative of equation 7 is given as:

$$f'(\mathrm{FPP}) = \frac{\left(1 + e^{-4.42}\left(\mathrm{FPP}^{-2.37} - 1\right)\right)^{-1/2.37} e^{-4.42}\mathrm{FPP}^{-2.37}}{\mathrm{FPP}\left(1 + e^{-4.42}\left(\mathrm{FPP}^{-2.37} - 1\right)\right)} \tag{8}$$

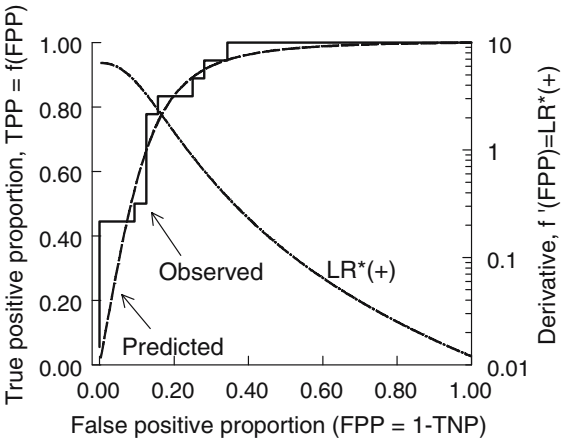

Figure 4. Left-hand axis: An ROC curve (true positive proportion, TPP, vs. the false positive proportion, FPP) for the predictor model in De Wolf et al. (2003), together with predicted TPP from equation. Right-hand axis: derivative of equations 7 and 8 vs. FPP, the instantaneous likelihood ratio [LR*(+)].

which is based on equation 5 in Hughes and Madden (2003). As required because of the shape of the ROC curve, f' (FPP) = LR*(+) declines with increasing FPP (Figure 4; right-hand axis). Mathematically, one can solve the f' (FPP) equation for FPP, and then obtain TPP based on f(FPP) (equation 7). From this combination of TPP and FPP (which gives TNP and FNP), one can determine the standard LR(+) (= TPP/FPP) and use equation 4 for risk assessment.

The use of equations 6 and 7 can be demonstrated with the *Fusarium* head blight data. Previously, an epidemic was predicted in practice if odds(E+|P+) was greater than 1. This is equivalent to specifying that the cost ratio, CR, equals 1, where both types of errors are equally costly. Using odds(E+) = 0.563, as stated previously, one finds from equation 6 that $\mathrm{LR}^{*}(+)_{\mathrm{opt}} = 1.78$ when CR = 1. Graphically, one finds this value of f' (FPP) in Figure 4, and then determines the corresponding FPP and TPP at this value. One can see that FPP $\approx$ 0.18 and TPP $\approx$ 0.81 at f' (FPP) = 1.78 in the graph, which are similar to the

values used in the nominal situation described in Table 1 (with a Z threshold of −0.4; scenario A). (There will be some discrepancy because equation 7 is not a perfect fit to the ROC curve.) With the TPP and FPP values here, LR(+) ≈ 4.5, and the posterior odds of an epidemic when one is predicted is 4.5·0.563 = 2.5, giving Prob(E+|P+) = 0.71 (close to the value found at slightly different TPP and TNP values in Table 1). It is important to emphasize that the posterior odds of an epidemic (or non-epidemic) are actually calculated with LR(+), *not* with the instantaneous rate LR*(+).

Now consider the situation in which a false negative decision is four times as costly as a false positive decision ($CR = C_{FP}/C_{FN} = 1/4 = 0.25$). With the nominal prior odds of an epidemic, one finds that $LR^*(+)_{opt} = 0.25/0.563 = 0.444$. From Figure 4, one can see that this derivative occurs at FPP ≈ 0.32 (or TNP ≈ 0.68) and TPP ≈ 0.94. An increased TPP and decreased TNP compared to the nominal situation (with CR = 1) is higher up the ROC curve (towards the right-hand corner), which means that an epidemic is predicted to occur at a lower Z threshold (Figure 3). That is, there is a less stringent criterion to predict an epidemic. Using the listed sensitivities and specificities here, one obtains LR(+) = 0.94/0.32 = 2.94, and LR(−) = (1–0.94)/0.68 = 0.088. The posterior odds of an epidemic when one is predicted then is odds(E+|P+) = 2.94·0.563 = 1.66, giving a posterior probability of Prob(E+|P+) = 1.66/(1 + 1.66) = 0.62. It can be shown that the posterior probability of a non-epidemic when a non-epidemic is predicted is Prob(E−|P−) = 0.95. One is more certain about the true epidemic status of a random observation when non-epidemics are predicted compared to when epidemics are predicted. Also, Prob(E+|P+) is lower here than when CR = 1, but this reduction in certainty of epidemics is required to minimize the average cost of making predictions.

In general, as demonstrated in the previous paragraph, as CR declines at a given prior probability of an epidemic, one moves up the ROC curve towards the right-hand corner (higher TPP and FPP; lower TNP and FNP), which means that a lower Z threshold is used for predictions of epidemics. Loosely speaking, with high cost of false negative decisions, one would not want to make too many of these errors (i.e., one would want a low FNP). Conversely, as CR increases (e.g., false positives are more costly than false negatives), one moves down the ROC curve towards the left-hand corner (lower TPP and FPP; higher TNP and FNP), which means that a higher Z threshold is used for predictions of epidemics. Loosely speaking, with high cost of false positives, one would not want to make too many of these errors (i.e., one would want a low FPP). The approach outlined here can also be coupled with consideration of different prior probabilities of epidemics, as presented previously in this section. It is quite possible for a given pathosystem that there are combinations of prior probabilities and costs of false predictions that one would always assume that an epidemic will occur or always assume that an epidemic will not occur. The decision-theory approach provides the formal mechanism for evaluating these scenarios.

The analyses discussed here are just the beginnings of the possibilities for applying risk assessment to disease prediction (Yuen and Hughes, 2002; Yuen, 2003). In fact, only the initial aspects of this approach have been formally applied to *Fusarium* head blight forecasting at this stage; costs of decisions and consideration of prior probabilities of epidemics for this pathosystem will be addressed more formally after a more accurate prediction system is developed for known epidemics and non-epidemics. Other areas requiring research for plant diseases in general include: having more than a dichotomy of decisions (such as spray, do not spray, and wait-and-see what happens); dealing with more than a dichotomy of predictions (such as predicting the degree of expected damage from a disease, predicting spraying once or weekly for the rest of the season); dealing with more than a dichotomy of measured outcomes (such as intensity of disease on a continuum for different predictions, for either binary predictions or a continuum of predictions); and dealing with multiple diseases or pests.

Discussion

Botanical epidemiology has advanced on many fronts in the years since van der Plank's first book was published in 1963, and the discipline continues to be a foundation for understanding and predicting diseases at the population scale. I have

chosen to outline just two out of many possible broad topics where substantial advances have been, and continue to be, made. Many of the speakers and poster presentations at the 9th International Workshop on Plant Disease Epidemiology reported in this special issue of the *European Journal of Plant Pathology* have dealt with other valuable topics.

The use of growth-curve and mechanistic population dynamic models, especially the coupled (ordinary and partial) differential equations outlined in this article, provide a flexible and powerful methodology for representing the temporal, spatial, and spatio-temporal dynamics of diseases, and provides the framework to elucidate general thresholds for epidemic occurrences (disease invasion), long-term persistence of disease, velocity at which disease expands from foci, and the initial rates of disease increase over time. Many of these qualitative and quantitative properties of epidemics can be summarized by the basic reproduction number (R_0).

The coupled differential equations (or other model formulations, such as stochastic difference equations) can be made extremely complicated, and care should be taken to keep the principle of model parsimony in mind when modelling epidemics! Strategies for disease control can be readily explored by finding the combination of disease properties (e.g., latent and infectious periods, transmission rate) that result in a R_0 less than 1. This approach is less useful for real-time prediction of epidemics, or for determining the need to intervene with a control measure, because precise estimates of parameters under specific environmental conditions are often not available. Usually, simpler prediction equations (such as regression equations, discriminant functions, or *ad hoc* rules) are used to actually make predictions. This use of more descriptive equations can be further justified by the fact that even very complicated mechanistic models often result in simple exponential-type population increase when disease intensity is not high (see Segarra et al., 2001).

The incorporation of probabilistic decision theory into disease predictions (whether these come from empirical rules or equations, or even population-dynamic models) is the second topic I covered where key advances have been made recently (Yuen and Hughes, 2002). Although most plant pathologists, including epidemiologists, clearly are not yet thinking formally or explicitly in terms of prior and posterior odds, and likelihood ratios, the concepts follow directly from intuitive understandings of how prediction rules are applied when there is some inaccuracy in the predictors and when epidemics and non-epidemics (or the need to intervene or not with a control method) are not equally common in a given area. The next generation of advances in this area will deal more with the costs of decisions (correct or incorrect), and in addressing some of the biological and environmental interactions in more complicated pathosystems, possibly involving multiple diseases (and crops simultaneously) (McRoberts et al., 2003).

Whether one is working with elaborate population-dynamic models or testing prediction rules for decision making, I accept as an axiom that appropriate statistical methods be used to fit models to data, compare results within and between studies, and test hypotheses. Developments in linear and nonlinear mixed models, for instance, are drastically improving the matching of the statistical methods to the data and experimental design, and intended inferences of the investigator (Schabenberger and Pierce, 2002). There is also growing evidence that Bayesian methods are also useful for developing prediction equations (Mila and Carriquiry, 2004), and not just in evaluating the performance of predictors (as demonstrated in this article). Except for the most quantitative of the botanical epidemiologists, more effort is still needed to encourage plant pathologists to keep abreast of developments in statistics and utilize the appropriate old and new techniques (Garrett et al., 2004).

Acknowledgements

Salaries and research support were provided by state and federal funds to the Ohio Agricultural Research and Development Center, Ohio State University, USA.

References

Anderson RM and May RM (1991) Infectious Diseases of Humans; Dynamics and Control, Oxford Science Publication, Oxford.

Auden WH and Kronenberger L (1966) The Viking Book of Aphorisms, Viking, New York.

Aylor DE (1999) Biophysical scaling and the passive dispersal of fungus spores: relationship to integrated pest management strategies. Agricultural and Forest Meteorology 97: 275–292.

Bald JG (1937) Investigations on 'spotted wilt' of tomatoes III: infection in field plots. Bulletin 106. Melbourne, Australia: Council for Scientific and Industrial Research.

Campbell CL and Madden LV (1990) Introduction to Plant Disease Epidemiology, John Wiley & Sons, New York

Cochran WG (1936) The statistical analysis of field counts of diseased plants. Supplement to Journal of Royal Statistical Society 3: 49–67.

DeWolf ED, Madden LV and Lipps PE (2003) Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data. Phytopathology 93: 428–435

Diekmann O and Heesterbeek JAP (2000) Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation, John Wiley and Son, Ltd., Chichester, U.K, 303

Ferrandino FJ (1993) epidemic waves: I. Focus expansion within a linear planting. Phytopathology 83: 795–802

Fry WE (1982) Principles of Plant Disease Management, Academic Press, New York

Garrett KA, Madden LV, Hughes G and Pfender WF (2004) New applications of statistical tools in plant pathology. Phytopathology 94: 999–1003

Gibson GJ (1997) Investigating mechanisms of spatiotemporal epidemic spread using stochastic models. Phytopathology 87: 139–146

Gibson GJ, Gilligan CA and Kleczkowski A. (1999) Predicting variability in biological control of a plant-pathogen system using stochastic models. Proceedings Royal Society London B 266: 1743–1753

Gildow F, Damsteegt B, Stone A, Schneider W, Luster D and Levy L (2004) Plum pox in North America: identification of aphid vectors and a potential role for fruit in virus spread. Phytopathology 94: 868–874

Gilligan CA (1985). Introduction. In: Gilligan CA) Advances in Plant Pathology, Vol. 3; Mathematical Modelling of Crop Disease (pp 1–10) Academic Press, London

Gilligan CA (2002) An epidemiological framework for disease management. Advances in Botanical Research 38: 1–64

Gottwald TR, Hughes G, Graham JH, Sun X and Riley T (2001) The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. Phytopathology 91: 30–34

Hardwick NV (1998). Disease forecasting. In: Jones DG (eds) The Epidemiology of Plant Diseases (pp 207–230) Kluwer Academic Publishers, Dordrecht

Hughes G and Madden LV (2003) Evaluating predictive models with application in regulatory policy for invasive weeds. Agricultural Systems 76: 755–774

Hughes G, McRoberts N, Madden LV and Nelson SC (1997) Validating mathematical models of plant disease progress in space and time. IMA Journal of Mathematics Applied in Medicine and Biology 14: 85–112

Hughes G, McRoberts N and Burnett FJ (1999) Decision-making and diagnosis in disease management. Plant Pathology 48: 147–153

Hughes G, Gottwald TR and Yamamura K (2002) Survey methods for assessment of citrus tristeza virus incidence in urban citrus populations. Plant Disease 86: 367–372

James C (2003) Global Status of Commercialized Transgenic Crops: 2003, International Service for the Acquisition of Agri-Biotech Applications (ISAAA), Manila

Jeger MJ (1986a). The potential of analytic compared with simulation approaches to modeling in plant disease epidemiology. In: Leonard KJ and Fry WE (eds) Plant Disease Epidemiology, Vol. 1: Population Dynamics and Management (pp 255–281) Macmillan Publishing Company, NY

Jeger MJ (1986b) Asymptotic behaviour and threshold criteria in model plant disease epidemics. Plant Pathology 35: 355–361

Jeger MJ (2004) Analysis of disease progress as a basis for evaluating disease management practices. Annual Review of Phytopathology 42: 61–82

Jeger MJ, Holt J, vanden Bosch F and Madden LV (2004) Epidemiology of insect-transmitted plant viruses: modeling disease dynamics and control interventions. Physiological Entomology 29: 1–14

Jones DG(.) (1998) The Epidemiology of Plant Diseases, Kluwer Academic Publishers, Dordrecht, 460

Keeling MJ, Brooks SP and Gilligan CA (2004) Using conservation of pattern to estimate spatial parameters from a single snapshot. Proceedings National Academy of Science USA 101: 9155–9160

Kranz J(.) (1990) Epidemics of Plant Diseases: Mathematical Analysis and Modeling, 2nd edn. Springer-Verlag, Berlin

Legg JP (1999) Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. Crop Protection 18: 627–637

Linnet K (1988) A review of the methodology for assessing diagnostic tests. Clinical Chemistry 34: 1379–1386

Madden LV (1986). Statistical analysis and comparison of disease progress curves. Pages 55–84. In: Leonard K and Fry WE (eds) Plant Disease Epidemiology (pp 372) Macmillan Publishing Co, New York

Madden LV and Hughes G (1995) Plant disease incidence: distributions, heterogeneity, and temporal analysis. Annual Review of Phytopathology 33: 529–564

Madden LV and Hughes G (1999) Sampling for plant disease incidence. Phytopathology 89: 1088–1103

Madden LV and Hughes G (2002). Plant epidemics, models, and analysis. In: El-Shaarawi A and Piegorsch W (eds) Encyclopedia of Environmetrics (pp 1557–15623) John Wiley & Sons

Madden LV and Ellis MA (1988). How to develop disease forecasters. In: Kranz J and Rotem J (eds) Experimental Techniques in Plant Disease Epidemiology (pp 191–208) Springer-Verlag Publishers

Madden LV and vanden Bosch F (2002) A population-dynamics approach to assess the threat of plant pathogens as biological weapons against annual crops. BioScience 52: 65–74

Madden LV and Wheelis M (2003) The threat of plant pathogens as weapons against U.S. crops. Annual Review of Phytopathology 41: 155–176

Madden LV, Pirone TP and Raccah B (1987) Temporal analysis of two viruses increasing in the same tobacco fields. Phytopathology 77: 974–980

Madden LV, Jeger MJ and vanden Bosch F (2000) A theoretical assessment of the effects of vector-virus transmission mechanism on plant virus disease epidemics. Phytopathology 90: 576–594

Maloy OC (1993) Plant Disease Control, John Wiley & Sons, New York

McRoberts N, Hughes G and Savary S (2003) Integrated approaches to understanding and control of diseases and pests in field crops. Australasian Plant Pathology 32: 167–180

Metz CE (1978) Basic principles of ROC analysis. Seminars in Nuclear Medicine 8: 283–298

Mila AL and Carriquiry AL (2004) Bayesan analysis in plant pathology. Phytopathology 94: 1027–1030

Neter J, Wasserman W and Kutner MH (1983) Applied Linear Regression Models, Richard, D. Irwin, Inc, Homewood, IL, 547

NRC (2002). Predicting Invasions of Non-indigenous Plants and Plant Pests. National Research Council, National Academy of Sciences, National Academy Press, Washington, DC

Pielou EC (1977) Mathematical Ecology, John Wiley & Sons, New York

Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW and Koike ST (2002) *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. Plant Disease 86: 205–214

Ross R (1911) The Prevention of Malaria, 2nd edn. Murray, London

Rush CM, Stein JM, Bowden RL, Riemenschneider R, Boratynski T and Royer MH (2005) Status of Karnal bunt of wheat in the United States 1996 to 2004. Plant Disease 89: 212–223

Schabenberger O and Pierce FJ (2002) Contemporary Statistical Models for the Plant and Soil Sciences, CRC Press, Boca Raton, Florida

Segarra J, Jeger MJ and vanden Bosch F (2001) Epidemic dynamics and patterns of plant diseases. Phytopathology 91: 1001–1010

Strange RN and Scott PR (2005) Plant disease: a threat to global food security. Annual Review of Phytopathology 43: (in press)

Turechek WW and Wilcox WW (2005) Evaluating predictors of apple scab with receiver operating characteristic (ROC) curve analysis. Phytopathology 95: 679–691

Twengström E, Sigvald R, Svensson C and Yuen J (1998) Forecasting sclerotinia stem rot in spring sown oilseed rape. Crop Protection 17: 405–411

vanden Bosch F, Metz JAJ and Diekmann O (1990) The velocity of population expansion. Journal of Mathematical Biology 28: 529–565

vanden Bosch F, Metz JAJ and Zadoks JC (1999) Pandemics of focal plant disease, a model. Phytopathology 89: 495–505

vanMaanen A and Xu X-M (2003) Modelling plant disease epidemics. European Journal of Plant Pathology 109: 669–682

vander Plank JE (1963) Plant Diseases: Epidemics and Control, Academic Press, NY, 349

Waggoner PE (1960). Forecasting epidemics. In: Horsfall JG and Dimond AE (eds) Plant Pathology, AN Advanced Treatise (pp 291–313III) Academic Press, New York

Wolfinger RD (1996) Heterogeneous variance–covariance structures for repeated measures. Journal of Agricultural, Biological, and Environmental Statistics 1: 205–230

Xu X-M and Ridout MS (1998) Effects of initial epidemic conditions, sporulation rate, and spore dispersal gradient on the spatio-temporal dynamics of plant disease epidemics. Phytopathology 88: 1000–1012

Yuen J (2003) Bayesian approaches to plant disease forecasting. Online. Plant Health Progress doi:10.1094/PHP-2003-1113-06-RV. [http://www.plantmanagementnetwork.org/php]

Yuen JE and Hughes G (2002) Bayesian analysis of plant disease prediction. Plant Pathology 51: 407–412

Yuen JE, Twengström E and Sigvald R (1996) Calibration and verification of risk algorithms using logistic regression. European Journal of Plant Pathology 102: 847–854

Zadoks JC (2001) Plant disease epidemiology in the twentieth century: a picture by means of selected controversies. Plant Disease 85: 808–816

Zadoks JC and Schein RD (1979) Epidemiology and Plant Disease Management, Oxford University Press, Oxford, 427

Zhang X-S, Holt J and Colvin J (2000) A general model of plant-virus disease infection incorporating vector aggregation. Plant Pathology 49: 435–444