Development of microcomputer applications in plant disease epidemiology

Author(s):
Blaise, Philippe

Publication Date:
1989

Permanent Link:
https://doi.org/10.3929/ethz-a-000569012

Rights / License:
In Copyright - Non-Commercial Use Permitted
DEVELOPMENT OF MICROCOMPUTER APPLICATIONS IN PLANT DISEASE EPIDEMIOLOGY

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH

for the degree of

Doctor of Technical Sciences

presented by

PHILIPPE BLAISE

Dipl. ing. agr. ETH Zürich
born 12. November 1954
citizen of France

Accepted on the recommendation of

Prof. Dr. M. S. Wolfe, examiner
Dr. C. Gessler, co-examiner

1989
CONTENTS

SUMMARY 1
ZUSAMMENFASSUNG 2

INTRODUCTION 3

Dealing with spatial data 5

GEOBASE: A simple geographical information system on a personal computer 6

Models in epidemiology 14

An extended progeny/parent ratio model: I. Theoretical development 16
An extended progeny/parent ratio model: II. Application to experimental data. 28

Computers as educational tools in epidemiology 36

APPLESCAB: A Teaching Aid on Microcomputers 38

Forecast and disease warning 44

On the search for an applicable forecasting model of downy mildew epidemics in Switzerland 45

ACKNOWLEDGEMENTS 50
Taking to computers is generally regarded as a sign of maturity in a field of study, in much the same way as experimentation with tobacco and alcohol is a sign that one's children are growing up. The analogy, if a trifle sour, has at least the merit of reminding us that there are inherent dangers in both developments, even though computers are not yet required to carry warning notices.

P.M.A. Bourke, 1970
SUMMARY

Since their emergence in the 1960's, computers have played a major role in the development of epidemiology. The increasing power and availability of microcomputers make them an ideal tool for the epidemiologist. However, the software production in this area has lagged far behind the hardware developments. This work presents four applications of microcomputers to the main domains of epidemiology: data management and representation, theoretical studies, education and forecasting.

"Geobase" is a simple geographical information system allowing management, display and preliminary analysis of spatially distributed data. It handles two kinds of data: maps and facts, where maps data describe the basis on which the facts data are located. The implementation of the program on a personal computer is presented, together with a short example showing a possible application.

"Epiphyt" is the implementation of an extension of the progeny/parent model proposed by Vanderplank in 1963. Incorporation of host growth and variable latent period allow a more realistic modelling of epidemic development while preserving the simplicity and biological meaningfulness of the original model. Host growth is driven through degree day accumulation, and the length of the latent period varies with temperature according to a beta function. The model has been tested with four sets of published epidemiological data and one set of data collected in a vineyard in Switzerland. The epidemics tested were of the diseases stripe rust and Septoria leaf blotch on wheat, apple scab and downy and powdery mildew on grapes. The disease progress curves simulated by the model reflected the epidemics in their most important stages except for the final levelling out of disease severity.

"Applescab" is a teaching tool, which simulates the development of an apple scab epidemic in an orchard. The user can test his management abilities by trying to control this epidemic with fungicide applications under different conditions.

"Vinemild" is a model for forecasting the development of downy mildew epidemics on grape. Based on literature data, it simulates the development of secondary infections according to hourly weather data and outputs the current state of the epidemic in the form of a daily report of infection and sporulation probability as well as the current severity curve. The model is composed of four parts handling respectively infection, sporulation, survival and liberation of spores. Data collected in the southern part of Switzerland in 1988 are presented and compared with the output of the model.
ZUSAMMENFASSUNG


Geobase ist ein einfaches GIS (Geographical information system), das die Manipulation, Darstellung und elementare Analyse von räumlich verteilten Daten erlaubt. Das Programm kennt zwei Arten von Daten: Karten und Fakten, wobei Karten die Unterlage beschreiben, auf welcher die Fakten lokalisiert sind. Die Implementation auf Personal Computer sowie ein kurzes Beispiel, das eine mögliche Anwendung zeigt, werden vorgestellt.


INTRODUCTION

In plant pathology, Epidemiology investigates the development of plant diseases in time, and the effect of biotic and abiotic environments on this development. Although plant disease epidemiology is now over a century old, it was only in the 1950's and 1960's that this subject developed a full role in the science of plant pathology through the provision of extremely powerful tools from the contemporary digital computer revolution.

Historically, data analysis has been the major use of computers in plant pathology (Rouse and Teng, 1984), but computers have been used by epidemiologists for two main purposes: data handling and modelling.

Data handling includes collection, storage, management, presentation and analysis. Since epidemiology is not concerned with punctual events but deals with complex systems evolving in time and space, the amount of data involved is tremendous and data manipulation is an essential part of epidemiological work. Computers are ideal tools for these manipulations: they perform automatic recording of data, allow easy storage, retrieval and transformation of large amounts of data, graphic displays and printing devices provide fast representation of abstract values, and computing power permit computations which could not be done by hand.

Modelling is the action of translating a conceptualization of the real world into a mathematical formulation, which in turn allows us to test the validity of that conceptualization. There are three main uses of mathematical models in epidemiology:
- Analysis of response of epidemic development to variation of certain factors. This increases our understanding of epidemic development and provides directions for future research.
- Simulation of epidemic development by comprehensive models consisting of many interacting components and levels, as a working and teaching aid.
- Provision of decision aids for pest management through forecast of epidemic development as well as prediction of the outcome of certain management decisions or climatic conditions.

Simulation modelling, in particular, only became practically feasible following the introduction of computers.

Until recently, a computer was a large piece of furniture, kept in a special room, with access restricted to a few pathologists with programming background or sympathetic computer specialists. Its use in phytopathology, despite the numerous advantages it offered, remained modest. The second computer revolution, which started at the end of the 1970's (and which is still going on) transformed the computer into a common tool, for people with no or little computing background. This improvement was not achieved at the cost of power: the computers standing on the desks of today are more powerful than the mainframe computer which ran EPIDEM, the first simulation model in phytopathology (Waggoner, 1969) twenty years ago (Fig.1). The evol-
ution of computing illustrates the role that personal computers will take in scientific research in the future.

The present work tries to present some possible applications of personal computers in the main domains of application in epidemiology. The tools described are the first of this kind in epidemiology, and show that results of epidemiological research can be presented in a form accessible to any phytopathologist and, in some cases, even for non-phytopathologists.

**Literature cited**


Dealing with spatial data

An important part of a scientist's work is to think of ways to represent information so that they and others can understand the underlying processes. The importance of pictures in perceiving and communicating the functional relationships in quantitative data has long been understood and the graphic capabilities of computers have significantly increased the possibilities to generate such pictures. This is especially true in the analysis of spatial data with which epidemiologists are faced as soon as they leave the two dimensional space time/severity to consider dispersal and spread of pathogens. In most cases, computers only simplify greatly the representation and analysis of spatial data, but there are instances where such analyses would not even be feasible without a computer. A study of the spread of the canker stain disease (*Ceratocystis fimbriata*) on the 8000 or more plane trees of southern Switzerland (Blaise and Gessler, 1988) would never have been started without the help of computers.

Since analysis of spatial relations and manipulation of graphics are computer intensive tasks, powerful computers are the tools of choice in this area. However, thanks to their increasing power and availability, microcomputers will become indispensable tools in this research field in the near future. From this point of view, a simple, spatially oriented data management program may introduce a new development in epidemiology.

Literature cited
Blaise, Ph. and C. Gessler. 1988. Disease mapping on PC: Application to the canker stain of plane trees in southern Switzerland. Abstr. 5th International Congress of Plant Pathology, Kyoto.
GEOBASE: A simple geographical information system on a personal computer

ABSTRACT

Spatially distributed data are often encountered in the biological sciences. Representation and analysis of such data require specific tools. A simple geographical information system is presented, which allows representation and elementary analysis of geographical coded information. It handles two kinds of data: maps and facts, where maps data describe the basis on which the facts data are located. Maps consist of objects described through a set of coordinates, while for facts a coordinate pair is associated to an unlimited number of data records containing 5 fields: a date, an element from a list, a two character code, an integer number and a real number. The input data can be displayed interactively on screen by combining logically selection criteria for each field. The facts corresponding to the selected criteria are either displayed as such, or are clustered and displayed as polygons or pies. A short example showing a possible application of the program is presented and advantages as well as limitations are discussed.

INTRODUCTION

Many applications in biological sciences are concerned with spatial data, for example geobotanical studies or population dynamics in pest management, and there is often a need for analysis of spatial distribution patterns. In plant disease epidemiology, analysis of spatial data is required for restricted scale studies such as analysis of patterns of spread of soil-borne pathogens or for large-scale studies dealing with the spread of airborne pathogens that can travel over hundreds of kilometers.

Working with data that include a spatial component is difficult since quantification is not reduced to a single value. Special tools are therefore required, that are commonly called geographical information systems (GIS). A GIS is distinguished from other data management systems by its inherent ability to maintain spatial relationships of variables as well as their attributes.

Since the appearance of the first GIS's in the 1960's, there has been a rapid increase in their number. To date, a wide variety of systems has been developed, primarily for land-use planning and natural resource management at the urban, regional, state and national levels of government, but also for applications by public utilities and private corporations (Smith et al., 1987). Although the existing systems can and have been used for biological studies, their primary use focuses rather on the geographical than on the biological side of the problems. These are mostly sophisticated, very powerful packages which in turn require trained and skilled users. They run mainly on minicomputers or mainframes, often only on special adapted hardware and a personal computer version exists for only a few of them.

We present here a simple GIS with limited geographical features, powerful enough to handle large data sets and with characteristics which should appeal to biologists who do not have the need for a professional GIS.
SYSTEM AND METHODS

The program was developed on an IBM PC-AT with 640 kbytes of RAM, a hard disk, a mouse and an EGA color display. It was written in Pascal, and compiled with the Turbo-Pascal compiler version 5.0 (Borland International, 1988).

The recommended configuration to run the program is an IBM PC-compatible computer with at least 512 kbytes RAM, an EGA or VGA graphic adapter with a colour display, a hard disk and a mouse. The use of an 80x87 mathematical coprocessor enhances the speed by a factor of 2 on average. The program also runs with the CGA, MCGA and Hercules graphic adapters. However, since these cards do not support colours, the use of the program in visualizing facts on screen is greatly reduced. In the same way, one can dispense with the hard disk and the mouse, but this again limits the usefulness of the program.

Limitations: Since the map files are kept on disk, their size is limited only by the available space and/or DOS limitation. The facts files (see below) are kept in memory and thus subject to the data structure limitations of the compiler. This means that the maximal amount of facts (coordinate pairs) is limited to 16384. The number of information records for each of these facts is limited only by the available disk space.

Maps can be currently printed on Postscript laser printers or plotters using the HP-GL command set. The outputs presented were produced on a TI 2108 laser printer.

IMPLEMENTATION

Geobase is an unconventional GIS in the sense that it distinguishes two kinds of data: maps and facts.

Maps represent the canvas on which the facts are located. They are composed of objects described through a name, a set of coordinate pairs and attributes describing the way these coordinates are connected together (e.g. line, closed line, filled surface). Although maps are used usually as the equivalent of geographic maps and composed of geographical objects such as borders, rivers or lakes, they can also describe any spatial structure e.g. a field, where the objects could be: border of the field, regions of the field with special soil characteristics, trees, etc.

Facts correspond to collected data and differ from map objects in three ways:
- Each fact contains an unlimited number of information records.
- The description of the localization of facts in space is (currently) limited to one coordinate pair so that the information content of a fact is associated either with a single point or with a rectangle or a circle centered on this point.
- Facts can be selected from their information content for display or clustering.

To simplify entry and data structure, the number of fields per record is fixed. Each record consists of up to 5 fields: the first is a date field, which will be used usually to keep track of either the date at which the recording was made or at which the record was entered; the second field contains an element out of a user defined list (e.g. cultivar) and the other fields contain a code (two characters), an integer number (-32768 .. +32768) and a floating point number (± 1.5*10^-45 .. 3.4*10^38).

A computer-based GIS may be viewed as having five component sub-systems (Knapp, 1978), including:
- Data encoding and input processing
- Data management
- Data retrieval
- Data manipulation and analysis
- Data display

These different sub-systems will be presented briefly, and Figure 1 shows a schematic representation of the menu tree through which the user selects the different functions.

![Schematic representation of the menu tree of Geobase. Each rectangle represents a menu or sub-menu. For clarity, lowest level menus are omitted.](image)

**Data input**: Maps are input by entering the coordinate sets defining each object. This can be tedious if many coordinates have to be entered; however, two features reduce this handicap:

- The program can import coordinates from ASCII files, so that already existing data need not be input again.
- Map files may be edited and appended at any time, i.e. a very simple map composed of just a few points can be input first and extended later.

The location of facts can be entered either by typing in the coordinates or by "clicking" on the map with a mouse. The latter is not so precise, but may be sufficient for some applications. Once the location and the first information record of a fact has been input, additional records may be added, again either by selecting the coordinates, or by "clicking" at the location of that fact on the display.

**Data management**: Special functions allow management of maps and facts data. They can be exported or imported via the ASCII format, edited or appended.
**Data retrieval:** Apart from the export of data mentioned above, data are retrieved through the graphic display: When displayed, each fact can be addressed through "clicking" on its location. The description of this fact is then displayed and the user can browse through the available records if more than one has been entered.

**Data analysis:** Simple analyses of facts can be made either on facts handled and displayed separately or on clusters of facts whereby the distance between two facts is used as a clustering criterion. The values associated with facts can be divided into classes, and then the single classes are selected for display or not. This allows queries to be answered by combining logically the different characteristics of the facts. For a survey of diseased bean fields (see example below), such a query could be: select all bean fields which in 1987 were sown with the variety X or Y, had a total disease severity of at least 10%, where plants grew to a normal height, and display them with colours corresponding to the disease severity classes. In the case of clusters, the program calculates the proportion of facts belonging to each class, within a cluster.

**Data display:** The display of facts on maps is the most important part of Geobase. The user looks at the current map and selected facts through a window defined by its centre and an associated scale. The centre of the current viewing window is selected by pointing on it or by entering its coordinates directly, and the scale is set either by selecting a previously fixed scale, entering a free scale or zooming in or out, which corresponds to halving or multiplying the current scale by a factor of two. These operations allow a precise positioning of the viewing window. In the case of maps containing many objects and thus requiring a longer time to display, classes of objects can be selectively inactivated until the desired scale and location are found. Facts are overlaid on the map as dots, characters, circles or rectangles depending on the user's selection. Clusters are displayed as polygons enclosing the facts belonging to each cluster, or as pies divided according to the number of facts in each class of the information field selected for display.

Hard copies are output in two steps: First, the size of the printed map is entered in cm, which determines the size of the viewing window on the screen (since the screen has a limited size, the window occupies the greatest possible surface by respecting the size ratio between the sides; the selected scale is thereby maintained) and then the desired map content can be built interactively, by moving the centre of the window, changing its associated scale and selecting the map objects or facts to display.

Great care has been taken to develop a user-friendly interface, so that the program can be mastered in a short time by a user without computer experience. All operations are performed through menus where the selection can be made either with a mouse or through the keyboard. All functions are performed interactively. The coordinates corresponding to the location of the mouse pointer are displayed constantly, which allows very precise work.

The following example shows a simple application of Geobase:
A study on a new disease in bean crops (hull spots) in Switzerland was initiated in 1986. Since this disease seemed to be correlated with numerous factors such as cultivar, occurrence of *Alternaria* disease or microclimate of the concerned field location, a comparative study involving over 150 fields was started to obtain more indications of the possible correlations between these factors and the incidence of the disease. Ge-
obase was chosen as a tool to keep track of the field locations and to produce status reports. Moreover, it allowed rapid testing of possible causal hypotheses.

The map used was of Switzerland and contained the information necessary for the user to locate the fields, such as borders, rivers, lakes, cities and highways.

The facts were in this case single fields, and the information recorded for each field was: date of record, cultivar, general state of the crop at harvest, accumulated disease severity level for all diseases and severity level due to the pathogen *Alternaria*. Figure 2 shows an overview (scale 1:2,500,000) of the map with the bean fields mostly in the northern part of the country. A closer look at this part is given by Figure 3A (scale 1:500,000), while Figure 3B shows the same region after clustering with a maximal distance of 4000 meters between fields. The pie shows the frequency of the different classes of total field disease severity.

**ALGORITHMS**

The structure of the map data is a cartographic structure, i.e. no spatial relationships are stored among objects. The single map objects are described through a vector model. Such a structure is commonly defined as a "spaghetti" structure. To introduce some ordering in the map files, similar structures are grouped together within a file, and information on the location of the objects is stored, thus speeding up the retrieval and display process.
Fig. 3. Printout of the northern part of the map used in a field study on bean diseases at a scale of 1:500'000. A) Bean fields are displayed as spots in a grey level corresponding to the disease severity at harvest (darker spots correspond to higher severity levels). B) Result of clustering the bean fields with a maximal distance of 4000 meters. The pies are centered in the center of gravity of the corresponding cluster. Their size is proportional to the number of fields belonging to the cluster and the different sectors correspond to the disease severity classes: 0-5%, 5-20%, 20-50%, 50-80%, 80-95% (increasing grey levels represent higher severity levels).
The facts data are stored in two separate files. The first contains the coordinates defining the location of the facts and pointers to the stored information on each fact in the second file.

During working, the coordinates of the current data set are kept in memory. To speed up access to the facts falling in the selected range of coordinates, their coordinates are sorted in ascending order of x coordinate values. To avoid resorting each time a data set is loaded, the sorted coordinates are stored in a third file. Coordinates are sorted by the quicksort algorithm (Sedgewick, 1988).

A simple algorithm has been developed to cluster the facts. The primary goal was to limit computations of distances between facts, since each such computation involves square roots and multiplications. A pseudo-code version of the clustering algorithm is given in Figure 4.

```plaintext
Function ClusterNr(Point):Integer
{ this function returns the cluster number to which "Point" belongs }
{ or zero if it does not belong yet to any cluster }
{ cluster algorithm }
For all Points P do
Begin
  If ClusterNr(P)=0 then create new cluster
  For all points R for which Rx-Px < Maxdist do
    If ClusterNr(R)=0 or ClusterNr(R) ≠ ClusterNr(P) then
      If distance between R and P < Maxdist then
        If ClusterNr(R)=0 then R belongs to ClusterNr(P)
        else Join ClusterNr(R) to ClusterNr(P)
  End
End
```

Fig. 4. Pseudo-code of the algorithm used to cluster facts in the program Geobase. Each point P is defined through a coordinate pair (Px,Py) and points are sorted after the Px coordinate. A cluster is composed of all points not further apart from the next one than the maximum distance Maxdist.

The convex hull around a cluster of facts is computed with the algorithm described by Sedgewick (1988).

Zooming is achieved through definition of viewports and associated world coordinate systems as usual in interactive graphic interfaces (Foley and Van Dam, 1983; Rogers, 1985).

**DISCUSSION**

Scientists and engineers have long understood the importance of pictures in perceiving and communicating functional relationships in quantitative data. Geographic information systems have been used successfully in several disciplines such as soil science, forestry, geology, landscape architecture and entomology. Their applications are increasing and it is estimated that approximately 4000 GIS will be operational by 1990 in North America (Tomlinson, 1984). Although the program presented here does
not pretend to compete with professional GIS, it has some features which make it useful for the biologist's daily work: its easy structure and user interface shorten the necessary training time, the map structure allows work with even small amounts of map information and its implementation on a microcomputer makes it a decentralized, affordable tool. There is a price to pay, however, for these advantages:

- The number of functions available is limited, partly because spatial relationships between map objects are not considered, a primary drawback of the vector type data model. On the other hand, this data model is more compact than tesselation type data models, and there are more algorithms available for vector models than for others (Peuquet, 1984).

- Since computation time increases with map precision and numbers of facts, working with large data sets on PCs with limited power could become very slow. However, we used Geobase for a survey of plane trees in southern Switzerland where over 8000 trees were recorded with their healthy state (Blaise and Gessler, 1988), and did not encounter real speed problems on AT-type machines. From this point of view, the maximal number of facts which the program can handle may be seen as a reasonable limit.

- The input of map data is not automated. This has not been a problem so far, but could become so if description of large maps has to be input. The possibility of importing coordinates of map objects from ASCII files should however provide a way around this obstacle.

Acknowledgements

We thank Dr. M.S. Wolfe for critical reading of the manuscript and G. Eich for kindly providing the data used in the figures. The technical assistance of B. Koller in implementing the support for Postscript laser printers is gratefully acknowledged.

REFERENCES

Blaise, Ph. and Gessler, C., 1988. Disease mapping on PC: Application to the canker stain of plane trees in southern Switzerland. Abstr. 5th International Congress of Plant Pathology, Kyoto.
Models in epidemiology

Mathematical models in epidemiology represent the formalization of hypotheses concerning the behaviour of epidemics of plant disease. They are constructed to satisfy a set of a priori assumptions, based on knowledge or intuition about the behaviour of the system. Implicit in modelling is the desire to reduce a complex system to a series of simpler hypotheses that can be tested (Gilligan, 1985). A model, therefore, is not a replica of the system it serves to represent but some approximation to it (Kranz, 1974). No review of the models developed in epidemiology will be presented here, since this has been done extensively in the recent literature (e.g. Hau, 1985, Rouse, 1985, Teng, 1985, Jeger, 1986).

Models in epidemiology have assisted in the development of theory, in the undertaking of experimental programs, and in disease management (Jeger, 1986). Jeger (1986) points out that the modeling of plant disease epidemics effectively started with the pioneering synthesis of Vanderplank (1963). Its analyses of the temporal progress of diseases provided a stimulus for much further work. Very much has been written about the few models which has been developed since then. Beside theoretical studies, the discussion has concentrated on the classification of models depending on their underlying concepts and on the way they were built. Models can be descriptive vs. explanatory, analytical vs. synthetical, mechanistic vs. functional, stochastic vs. deterministic, empirical vs. statistical... However, much more important than the class to which it belongs, the way a model behaves and the extent to which it fulfills the goals set are the criteria to judge its usefulness, and will decide upon its survival...

The scientific community is being pulled now more than even in two different directions: Basic research at an increasingly fundamental level and application of research results for solving problems. Modeling seems to offer a useful mean of tying these seemingly divergent efforts together.

Literature cited


An extended progeny/parent ratio model:  
I. Theoretical development.

ABSTRACT

A new version of Vanderplank’s epidemic model corrected for removals is presented. Incorporation of host growth and variable latent period allow a more realistic modelling of epidemic development while preserving the simplicity and biological meaningfulness of the original model. Host growth is driven through degree-day accumulation. The length of the latent period varies with temperature according to a beta function. Calculations are made through numerical integration combined with boxcar trains. The model has been implemented in a program called "Epiphyt" which runs on personal computers. Typical output produced by the program is shown.

INTRODUCTION

Among the numerous mathematical models proposed in the last decades to describe and study plant disease epidemics, the improved logistic model corrected for removals, as proposed by Vanderplank (16), remains the one with the clearest biological meaning. Although other functional models (e.g. Richard’s, Weibull’s, Gompertz’s etc.) also have an inherent biological content, the a priori assumptions that led to their build-up are often less than obvious.

The underlying hypothesis of Vanderplank’s epidemic model is that disease development is proportional to the amounts of infectious (i.e. sporulating) and available (i.e. healthy) tissues:

\[ \frac{dX(t)}{dt} = Rc \times [X(t-p) - X(t-p-i)] \times [1 - X(t)] \]  

(1)

Where \( X(t) \) is the proportion of diseased host at time \( t \) (and thus \( 1-X(t) \) the proportion of healthy host), \( Rc \) the infection rate, \( p \) the latent period of the considered pathogen and \( i \) the infectious period.

The attractive aspect of this model is that it includes in due course the essential elements influencing the development of a disease: the intrinsic multiplication potential of the pathogen, the host in the form of colonizable tissue and the past history of the epidemic expressed through the infectious tissue. Unfortunately, only a few papers are concerned with this model (e.g. 7, 5) and most references made to Vanderplank’s model (e.g. 6, 14, 10, 12) consider only the version without correction for removals (which was just a step in the development of the model), probably because of the lack of an analytical solution for the corrected model. Since this simpler form of the model is quite unrealistic, it did not find a wide application nor did it lead to further developments. Vanderplank (17) returned to the corrected model but he limited himself to a transformation of the time scale, which allowed the introduction of the progeny/parent ratio \( \alpha \). Hence, the main flaws of this model remained: absence of host growth and fixed length of the latent and infectious periods over the whole epidemic development.

We present here a new form of Vanderplank’s epidemic model corrected for removals that overcomes its major shortcomings while preserving its clarity and biological soundness. It has been implemented on a microcomputer, resulting in an interactive
tool allowing description and study of the course of epidemics with only a few known parameters.

**DEVELOPMENT OF THE MODEL**

**a) Assumptions**

The extensions to the model are based on the knowledge that host growth is an essential factor in epidemic development and that most processes in an epidemic are temperature dependent.

The following assumptions are made for the derivation of the model:

- The disease is characterized by three parameters: the latent period \( p \) which is temperature dependent and the infectious period \( i \) and the infection rate \( R_c \) (multiplication factor of the disease per time unit) which are constant.
- The host population is treated as a surface homogeneously susceptible to the disease.
- After infection, the host surface becomes diseased. After a period \( p \) it becomes infectious for a period \( i \) and is then removed.
- Infection units are dispersed uniformly over the host surface.

**b) Host growth**

Assuming an ample supply of plant nutrients and soil water, the growth of the vegetation is determined by weather conditions, among which light is almost never a limiting factor. Of the parameters driving growth rates, temperature is the most important. Therefore, the model is simplified to a form in which host growth is driven only by temperature. We assume that host growth:

- takes place only above a minimum temperature \( T_0 \)
- is linearly proportional to temperature above this level
- stops after accumulation of a fixed heat sum \( T_s \).

This results in the following equation:

\[
\Delta K(t) = \begin{cases} 
  g \left( T(t) - T_0 \right) & \text{if } \sum [T(t) - T_0] < T_s \\
  0 & \text{if } T(t) > T_0 
\end{cases}
\]

(2)

where

- \( K(t) \): Host surface at time \( t \) [L^2]
- \( T(t) \): Temperature at time \( t \) [°C]
- \( T_0 \): Minimal temperature for growth [°C]
- \( T_s \): Temperature sum after which growth stops [°C*T]
- \( g \): Norming factor [L^2/°C]

To norm the maximal host surface to 1, we set the adjustment factor \( g \) to: \( g = 1 / T_s \) and equation (2) becomes:

---

1 The program Epiphyt is not copy-protected and is available upon request from the first author. It runs on IBM PC compatible computers with any of the following graphic cards: CGA, EGA, VGA, MCGA, HGC and AT&T (Olivetti).
Initially, the host has a surface $K_0 = K(0)$, hence the host surface at any time:

$$K(t) = K_0 + \sum \Delta K(t)$$

c) The latent period

Since temperature is the most important external factor influencing the length of the latent period, the model is limited again to this factor as the driving variable of the latent period. It has been shown (1) that the beta function is well-suited to describe the relation between development of pathogenic fungi and temperature. We express therefore the temperature dependence of the latent period by:

$$p(T) = \frac{P_{\text{min}}}{f(T)}$$  \hspace{1cm} (4)

Where $P_{\text{min}}$ is the length of the latent period at the optimal temperature and $f(T)$ is a beta function of the form:

$$f(T) = a \cdot T_n^m \cdot (1 - T_n)^n$$

with $T_n = \frac{\left(T - T_{\text{min}}\right)}{\left(T_{\text{max}} - T_{\text{min}}\right)}$, $T_{\text{min}}$ and $T_{\text{max}}$ being the cardinal temperatures of the organism under consideration, which means that $p(T_{\text{min}}) = p(T_{\text{max}}) = \infty$. The two parameters $m$ and $n$ have to be determined for each host/pathogen couple.

d) Application of the model

In order to introduce host growth into Vanderplank's equation, the quantity of disease is expressed as absolute host surface instead of a proportion and equation (1) becomes:

$$\frac{dN(t)}{dt} = Rc \cdot \left[N(t-p) - N(t-p-i)\right] \cdot \left[1 - \frac{N(t)}{K(t)}\right]$$  \hspace{1cm} (5)

with $N(t)$ being the diseased and $K(t)$ the whole host surface at time $t$.

The next step consists in introducing the variable latent period. As pointed out before, the second term in equation (1) represents the history of the epidemic. Therefore, in this term, $p$ is not the current latent period at time $t$, but results from the accumulation of past latent periods. To take this into account, we use the latent period $p_t$ which satisfies:

$$\int_{t-p_t}^{t} \frac{1}{p(T(t))} \, dt = 1$$

$T(t)$ being the temperature at time $t$. To obtain a simple tool to study epidemics without having to provide large sets of temperature data, we assume that the length of the latent period is constant over a day, so we need only daily weather data, or a function
T(t), which supplies a daily mean temperature. Over a year, the course of the daily mean temperature can be approximated well by a sine function of the form:

\[
T(t) = \text{Min} + \frac{1}{2} \times (\text{Max} - \text{Min}) \times (1 - \cos(\omega \times t + \phi))
\]

where

- \( \text{min} = \) minimum daily mean temperature for the year
- \( \text{max} = \) maximum
- \( t = \) time in days
- \( \omega = \frac{2}{365} \)
- \( \phi = \) phase in days (usually 20 days)

Consequently, the model defined through equations 3, 4, 5 and 6 has only one independent variable: the time \( t \). Knowledge of the following parameters is required:

- Host: \( T_0, T_s \)
- Pathogen: \( i, R_c, X(0), T_{\text{min}}, T_{\text{max}} \)
- Latent period: \( n, m \)
- Temperature: \( \text{Min}, \text{Max} \)

e) Computations

Since the model defined so far does not have a straightforward analytical solution, the use of numerical methods is required. The calculations are carried out through a combination of an Euler integration algorithm with two boxcar trains. The trains are necessary to provide the two terms with retarded arguments at any time \( t \) in order to compute the term \([X(t-p_i) - X(t-p_{i-1})]\), which Vanderplank called the 'activator'. The first boxcar train holds the latent tissue while the second holds the infectious tissue; each box consists of two compartments: one for the diseased tissue of the same age class, and one for the accumulation of fractions of the latent (or infectious) period.

The low accuracy of the Euler algorithm is not a problem here, since the required precision does not exceed \( 10^{-3} \). The data obtained with constant host surface and latent period were in agreement with the results obtained by numerical integration with a Fehlberg Runge-Kutta algorithm coupled with Hermite polynomial interpolation (Blaise, unpublished).

f) Implementation

The model was translated into the Pascal programming language and was run on an IBM AT personal computer to produce the results presented. The program is interactive, allows easy input of the parameters and displays the epidemic curves graphically. All results shown below were produced with this program and output on a Postscript laser printer (TI 2108). The parameters \( n \) and \( m \) needed for the temperature dependent latent period are estimated automatically by fitting a beta function through a set of points using nonlinear least square methods. It requires an input of at least three values of the latent period at different temperatures.

Since the model considers only latent, infectious or dead lesions, a problem arises for the representation of visible lesions. To overcome this, the program needs an additional parameter expressing the length of the incubation period which is given as a proportion of the latent period. This parameter is set by default to 0, in agreement with Vanderplank (16,17).
g) An example

To illustrate the output of the model, the course of an hypothetical epidemic obtained by using fictitious parameters is shown in Fig.1. The selected parameters were: \( i=3, \ p=3, \ Rc=0.67, \ X(0)=0.001, \ T_{min}=2^\circ C \) and \( T_{max}=30^\circ C \) for the pathogen, \( T_0=5^\circ C, \ T_s=2000^\circ C*d \) for the host, and the extremes of the daily mean temperature were set respectively to \( 0^\circ C \) and \( 20^\circ C \). The output of the model before introduction of host growth and variable latent period, which corresponds to integrating numerically Vanderplank's equation, is a sigmoid shaped curve (Fig. 1, curve A). In agreement with the calculations of Vanderplank (17, 19), the maximum disease level is not 100% but 0.797, which is the \( L \) value satisfying the equation:

\[
L = 1 - \exp(-\alpha \cdot L)
\]

where \( \alpha = Rc \times i \).

Introducing a temperature dependent latent period\(^1\) results first in a shift of the disease curve (Fig. 1B) as a consequence of a very long latent period at the beginning of the year due to low temperatures. The second effect depends on the value of \( Rc \) and on the shape of the latent period curve over the year (Fig. 2C): an effect resulting from the variable latent period can be observed only if the epidemic develops in a time interval in which the latent period varies significantly. This is the case when \( Rc \) is low and/or the shape of \( P(T) \) is wide. In our example, the epidemic reaches its maximum in 60 days, which is approximately the time when \( p_t \) is at a minimum. That is why almost no effect other than a displacement can be observed. Curve C (Fig.1) is ob-

\(^1\) The beta function was selected to produce \( p=3 \) at the optimum temperature (Fig. 2D).
tained after introduction of host growth, which is represented by curve H, without the variable latent period. In this case, the unusual shape of the disease curve is an artifact resulting from the fact that disease starts before host growth. Finally, by combining host growth and variable latent period, we obtain curve D (Fig. 1). A very important effect of the introduction of host growth is the reduction of the maximum disease

Fig. 2. Implementation of the temperature dependent latent period in the extended progeny/parent ratio model. A) Simulation of the daily mean temperature over the year through a sine function. B) An example of a beta function for a minimum latent period at 17.5 °C. C) Variation of the length of the latent period over a year resulting from the combination of curves A and D. D) Relation between the length of the latent period and the temperature, obtained through inverting the beta function (B).

level. The reason is obvious: as the epidemic reaches the critical point where the activator starts to decrease, slowing due to host growth is enough nearly to stop it. Later on, when host growth has stopped, the activator is too weak to allow disease to catch up.

The latent period and its temperature dependency are described in Fig. 2. The sinus-like course of the daily mean temperature over the year (Fig. 2A) combined with the inverse beta relation between latent period and temperature (Fig. 2D) determine the variation in the instantaneous latent period over the year (Fig. 2C). The original beta function is given in Fig. 2B.

**SENSITIVITY**

The reaction of the model to variation in some of the parameters is investigated below.
a) Variation in X(0), i & p.

To avoid interactions masking the real effect of parameter changes, host growth is not included and latent period is kept constant.

Decreasing the primary inoculum X(0) with a constant factor e.g. 10 fold, causes only a shift of the disease curve to the right (Fig. 3, curves A,B,C). Since the progeny/parent ratio is the same as in the example above, the maximum disease level remains at 0.797. Even a $10^6$ fold decrease of the primary inoculum does not change this value (Fig. 3, curve D).

Maintaining the progeny/parent ratio $R_c*i$ constant and increasing the length of the infectious period i (i=3, 4, 5) leads mainly to a decrease in speed of the epidemic (Fig. 3, curves D, E, F); however, this decrease is not proportional to the amount of increase in i, but to the proportion of increase in i.

Changing only the length of the latent period (p=3, 4, 5, 6) also causes a variation in the speed of the epidemic, which is nearly proportional to the amount of change (Fig. 3, curves F, G, H, I).

b) Variation in the yearly extremes of the daily mean temperature.

Host and pathogen parameters are set to the same values as in the example above. Varying the yearly extremes of the daily mean temperature produces changes in the daily mean temperature for the whole year. In turn, these changes influence the length of the latent period and the development of the host. Although not a very important factor, the influence of variation in extremes should not be underestimated. A minimum of 0°C and a maximum of 20°C are quite representative for most temperate climates. The corresponding epidemic is represented in Fig. 4 as well as variations for two other, unlikely, sets of extremes obtained by shifting the temperature range up or down by 5°C.

Shifting the temperature up, that means producing a warm climate, allows host growth to start earlier and to attain faster its maximum while in a cold climate, produced through shifting the temperature range down, maximum development is not reached due to a late start of growth and a slow development. The main effect on the epidemic development comes from the shift of the start of the growth of the host. In
the case of high temperatures, the advance of the host is important enough to maintain the disease level under 10% until the host stops growing. Conversely, cold temperatures do not influence the epidemic very much (at least for this set of parameters), except at the beginning.

c) Variation in host growth parameters

Through variation of the host parameters, it is possible to adapt the development of the host surface to particular crop conditions of interest. A displacement of the zero development temperature from 2°C to 5°C produces, under an identical temperature pattern, a shift of the start of the host curve and a small alteration of the slope of this
curve (Fig. 5, curves B, C). The result is first a small displacement of the epidemic curve and, far more important, a decrease of the maximum disease level reached (Fig. 5, curves D, E). This effect is even stronger if the required heat sum is reduced from 2000 °C*d to 1000 °C*d while keeping the zero-development temperature at 5°C (Fig. 5, curves D, F).

d) Variation in the beta function

Fig. 6 shows the model output (curves A, B, C) for three possible shapes of the beta function (Fig. 7, A, B, C). To avoid disturbance effects, host growth was not included. The parameters selected were $i=3$, $R_c=0.67$, $X(0)=0.001$, $T_{min}=0$ °C, $T_{max}=30$ °C, $Min=0$ °C and $Max=20$ °C. The corresponding beta functions are represented in Fig. 7. The resulting effect of changing the beta function is, as expected, similar to changes produced by varying the length of the latent period. However, curve A (Fig. 6) shows the effect which can result from an important lengthening of the latency period due to unfavourable temperature conditions during the epidemic. In this case, the effect is compensated by decreasing temperatures after day 200, which produce a decrease of the latency period and thus allows the epidemic to catch up.
DISCUSSION

Mathematical models of plant pathogen epidemics have an important role to play in helping us to understand the empirical world (6). The way in which epidemiological data should be evaluated and which model should be applied have been extensively discussed in the literature (e.g. 3). Pro and cons of extreme synthesis can be found. A frequent approach uses functional growth models, i.e. a single equation or a set of equations that can be manipulated and fitted to data sets. The major criterion for the choice of a specific functional growth model is often its better fit to the existing data, so the most favoured is often the equation with most parameters, such as the polynomial equation (20).

Conversely, Vanderplank (17) stresses the point that biological comprehension and the significance of the parameters are more important than the best fit. Abstracting from real data and starting from the logistic equation, he developed an extremely simple model with only a few parameters, that were all biologically sensible. However, the flaws common to most of the functional epidemic models remained and until now, no real attempt has been made to improve this model.

Waggoner (20) attempted to integrate the effect of varying temperature in the logistic equation by making \( r \) temperature dependent. However, the logistic model is highly synthetic, and the parameter \( r \) is a synthesis of many biological parameters, so that varying \( r \) with the temperature cannot account for all biological variation due to temperature.

Since the epidemic model of Vanderplank contains explicitly the parameters most relevant for epidemic development, we are able to apply an effect of the temperature specific to the different components. The most important temperature effect is certainly its influence on the length of the latent period; this was already recognized by Zadoks (22), who introduced a temperature dependent latent period in EPISIM.

By introducing a temperature dependent \( r \), Waggoner assumed either a linear (19) or a parabolic (20) relation; although the parabolic relation is quite realistic, problems arise at the pathogen's cardinal temperatures, where \( r \) becomes negative. Zadoks (22) derived a relation between temperature and latent period for stripe rust of wheat from field observations. However, the function he used to describe this dependency was not adequate to handle temperatures below 4°C or over 19°C, so that he introduced arbitrarily a constant latent period outside this range. The beta function, while in good agreement with experimental data, offers the advantage that its range of definition can be chosen within any limits. To be able to reflect the effect of past latent periods, we use fractions of latent period, and we allow for the addition of temperature effects. The hypothesis of additivity of temperature equivalents was already used as a useful approximation by Zadoks (22) and was shown to be valid for rates of development of *Alternaria solani* (21).

Kushalappa and Ludwig (9) developed a method to correct the logistic equation for host growth. They used either the monomolecular, the logistic or the Gompertz growth model while Waggoner (20) used the logistic model. Yet these models, despite the realistic appearance of the curves, are less adequate for representing host development since plant growth is not limited by an inherent maximal possible extension. The use of degree-day accumulation assumes that there is a linear relationship between development rates and temperature, which is an oversimplified assumption. Nevertheless, it has led to some useful applications in modelling the growth development of some crops (e.g. 2, 4). The introduction of an improved host growth model would permit a
better adjustment to different environmental conditions; on the other hand, this could lead to a loss in flexibility of the model.

Since temperature plays an important role in our model, special care should be taken in providing these data. However, obtaining meteorological data corresponding to existing disease data is usually very difficult. Waggoner (19, 20) avoided this problem by generating daily temperature data through a sine function oscillating around an annual mean temperature. We use here a similar function, except for the use of different parameters to specify the limits. In the normal case, the "standard" year is quite realistic, but if real data are available, it is of course possible to use them instead of the sine function.

Although the practical applications of the model will be discussed in a subsequent paper, some interesting aspects of this model need to be emphasized:

- Vanderplank stated that the maximal disease level reached by an epidemic depends only on the progeny/parent ratio (17) and the primary inoculum (18). According to the first results produced by our model, host growth can have some effect on this parameter (Fig. 5).
- From the original equation [1], the epidemic can start only if \( X(t-p) \) is different from \( X(t-p-i) \) so that the activator is non-zero; from a biological point of view, if \( p \) and \( i \) are kept constant diseases starting from overwintering inoculum, i.e. pathogens which do not sporulate in the winter, will never give rise to an epidemic since the activator is equal to zero. A temperature-dependent latent period, which is very long at the beginning of the season, eliminates this inconsistency.
- Most authors, except those using explicitly the incubation period, set implicitly the incubation period equal to 0 by comparing real disease values to calculated values. The possibility to select the length of the incubation period allows an exact comparison of real and model data.

In spite of the increased realism of the model, there are still some important factors lacking:

- Humidity and rain are, for some pathogens, critical factors which can have a decisive influence on the development of the epidemic
- The influence of temperature on the length of the infectious period can be important (15, 13), and may play some role in some epidemics.
- The pustules or lesions of many leaf pathogens often sporulate over extended periods although the bulk of spores tends to be produced in the early phase of the infectious period (11).

Kranz wrote in 1974 (8): "it will not be long before Vanderplank's model, though readily accepted for some epidemics, face the fate of all models; they will be adapted, amended, expanded or replaced by others". Unfortunately this prediction came only partially true and we hope that our work will awaken new interest for this model since there is probably still much to learn from it.

**LITERATURE CITED**


An extended progeny/parent ratio model:
II. Application to experimental data.

ABSTRACT

A previously published model based on Vanderplank's progeny/parent ratio equation was tested using "Epiphyt", its implementation on a microcomputer, with four sets of published epidemiological data and one set of data collected in a vineyard in Switzerland. The epidemics tested were of the diseases stripe rust and Septoria leaf blotch on wheat, apple scab and downy and powdery mildew on grapes. The disease progress curves simulated by the model reflected the epidemics in their most important stages except for the final levelling out of disease severity. Advantages of such a model based solely on the parameters initial inoculum, daily progeny/parent ratio and length of latent period, are discussed.

INTRODUCTION

The past decade has seen an increasing number of publications describing and using new models of plant disease epidemics. These models tend to be increasingly complex and often have to be considered as black box systems where only input and output are obvious to the user. A quarter of a century ago, Vanderplank (17) proposed an easily understandable model based on only three essential parameters: the initial inoculum, the latent period and a daily multiplication factor which he later called progeny/parent ratio (18). Using the length of the infectious period as a unit for the independent variable, time, allowed him to compare real field epidemics at a theoretical level. The model remained at a theoretical level and was never used to simulate real epidemics. In a previous paper (3) an adaptation of the Vanderplankian model was presented with the latent period as a function of temperature, incubation period as fraction of the latent period and a simple host growth model included in the disease severity calculation. The independent variable time was in units of days. The new model retained the easily understandable concepts of the original progeny/parent ratio model. The implementation of the model on a microcomputer with the program "Epiphyt" allows simulations of epidemics by input of calculable and measureable parameters. This paper compares actual disease severity data with the output of "Epiphyt".

MATERIAL AND METHODS

Model: The theoretical basis of the model, its construction and requirements were presented previously (3).

Disease severity data: Four sets of data used to test the model were taken from older publications for stripe rust on wheat (19) in the Netherlands, leaf blotch on wheat (4) in Switzerland, apple scab (1) in Germany and powdery mildew on grapes (9, 10) in California (Tab.1). The data on downy mildew on grape were collected in Southern Switzerland in 1988. The original data points of the epidemics are presented in the corresponding figures in addition to the simulated epidemics.
It is well known that a scab epidemic never reaches more than a few percent of the leaf surface present on an apple tree. Analytis (1) transformed his data correspondingly so that the asymptote of the effective maximum possible attack was approximately 3.6% of the whole leaf area. The figure of 100% disease severity of the model would in this case correspond to an effective attack of 3.6% of the whole leaf area.

Weather: The only weather data required for the model are the yearly maximum and minimum daily average temperature. If any information on these parameters was given with the published epidemic data it was used as input for the model, otherwise we used an approximation of the long-term yearly maximum and minimum daily average for the given location (Tab. 1).

<table>
<thead>
<tr>
<th>Disease</th>
<th>Min1</th>
<th>Max1</th>
<th>T0</th>
<th>T1</th>
<th>T2</th>
<th>P4</th>
<th>i^5</th>
<th>IPH6</th>
<th>r7</th>
<th>Re8</th>
<th>X08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stripe Rust/Wheat</td>
<td>1</td>
<td>17</td>
<td>2</td>
<td>700</td>
<td>10 (16)</td>
<td>20 (14)</td>
<td>0.9</td>
<td>0.2</td>
<td>1.505</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Leaf blotch/Wheat</td>
<td>-5</td>
<td>20</td>
<td>2</td>
<td>700</td>
<td>12 (19)</td>
<td>9 (4,20)</td>
<td>0.5</td>
<td>0.777</td>
<td>0.388</td>
<td>0.255</td>
<td></td>
</tr>
<tr>
<td>Powdery mildew/Grape</td>
<td>5</td>
<td>24</td>
<td>9</td>
<td>1150 (10)</td>
<td>6 (5)</td>
<td>4 (5,9)</td>
<td>1.0</td>
<td>0.08</td>
<td>0.47</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Downy mildew/Grape</td>
<td>2</td>
<td>25</td>
<td>10</td>
<td>1150</td>
<td>5 (7)</td>
<td>1 (7)</td>
<td>1.0</td>
<td>0.28</td>
<td>4.65</td>
<td>5.5E-7</td>
<td></td>
</tr>
<tr>
<td>Scab/Apple</td>
<td>0</td>
<td>19</td>
<td>4.4</td>
<td>900</td>
<td>9 (1)</td>
<td>8</td>
<td>0.8</td>
<td>0.12</td>
<td>0.56</td>
<td>0.0155</td>
<td></td>
</tr>
</tbody>
</table>

1 Yearly minimum and maximum average daily temperature.  
2 Minimal average daily temperature for host growth.  
3 Temperature sum above the host min. needed for the maximum development of the host.  
4 Latency period in days used to calculate Re. In parentheses source used to determine the beta function.  
5 Infectious period used to calculate Rc and for the simulation.  
6 Incubation period expressed as fraction of the instantaneous latent period.  
7 Apparent infection rate calculated from the epidemic data used to calculate Rc.  
8 Daily progeny/parent ratio calculated with the formula \( R_c = r \times \exp[(i + p) \times r] / (\exp(i \times r) - 1) \).  
9 Initial inoculum

Initial inoculum: Data on the amount of disease (primary inoculum) on the 1st of January is usually not available. Often, because of the kind of overwintering, it does not exist. The initial disease level was therefore adjusted so that the simulated epidemic curve passed through the first recorded field data point (Tab. 1).

Progeny/parent ratio: In epidemics caused by aerially dispersed pathogens the number of daughter lesions a single lesion can produce is an essential parameter (14, 15, 16). We calculated this according to the following formula (17, 18):

\[ R_c = r \times \exp[(i + p) \times r] / (\exp(i \times r) - 1) \]

where \( R_c \) is the multiplication factor per time unit, \( r \) the apparent infection rate, \( i \) the length of the infectious period and \( p \) the length of the latent period. \( r \) was measured from the epidemic data in the exponential growth phase of the respective epidemic (Tab. 1).

Latent period: For each epidemic, a beta function was forced from independent literature data describing the relation between length of the latent period and the temperature (Tab. 1).
Incubation period: The period between infection and appearance of the corresponding symptoms was assumed to follow the same relation with temperature as the latent period. It was therefore expressed as a proportion of the latent period and estimated from general plant pathology literature.

Infectious period: Since precise values corresponding to our needs were rarely available, the value of this parameter was taken from general plant pathology literature for leaf blotch, powdery and downy mildew and apple scab. It was estimated for stripe rust from data from van den Bosch et al. (16).

Host: The host growth curve is fixed by two parameters: Minimum development temperature (MDT) and sum of degree-days above MDT needed for maximum development. These parameters were taken either from the literature (Apple, 2) or calculated from published data on the respective host growth. The model assumes that all host tissue present is always susceptible and that this susceptibility remains at a constant level.

RESULTS and DISCUSSION

The model outputs (Fig. 1-5) reproduce the corresponding epidemics in their most essential aspects: start of the visible increase and steepness of the explosive phase of the epidemic. However, although data on the final stage of disease development are mostly missing, the model seems to fail at this point.

The levelling off at a certain severity in the final stage of a real epidemic can be ascribed mostly to changes in the host growth and/or susceptibility, or to significant weather changes: the stripe rust epidemic described by Zadoks (19) stopped before the substrate was exhausted, either because of a change of weather or of a change in the host-parasite compatibility (19); epidemics of apple scab stop because susceptible apple tissue is available only for a limited time.
The prediction of the model for the last phase of the epidemic is, in all cases, presented much steeper and higher than in reality: it even predicts a complete destruction of all tissue for the wheat diseases (Fig. 1, 2) and grape downy mildew (Fig. 4).

This effect can be attributed either to the lack of flexibility of the model towards weather changes, or to the simplicity of the incorporated model of host growth. Since weather is reduced to a daily mean temperature, all other factors can be accounted for only indirectly and only during the time window used to calculate $R_c$. Moreover, the course of temperature during the year is fixed. These limitations lead to inaccuracy of the model in the case of conditions deviating from the mean for long periods such as drought periods or cold spells. A simple means of increasing precision (e.g. sensu sensitivity to years deviating from the long term averages), would be to input real weather data from the particular site and year considered.

The inadequacy of the host growth model is particularly clear in the case of the apple scab epidemic (Fig. 5). The host growth model assumes that the tissue suscept-
tible to apple scab increased linearly with the temperature until a maximum was reached. In reality the host tissue available for infection (leaves without ontogenic resistance) would be better described by a bell-shaped curve with a maximum around May/June. Such a curve would flatten the disease severity curve in its upper part. The Septoria leaf blotch epidemic (Fig. 2) described by Brönnimann (4) never exceeded 60% of attack because the host was then ripe and the leaves died of natural causes; again the incorporated model of host growth does not account for death of host tissue.

The complex relationships between host, pathogen and environment are condensed into the three essential variables of the disease model, Rc, p and i. The success of such a model in mimicking reality is bound up with the extent to which these variables can retain and express the information they contain. The progeny/parent ratio is certainly the most obscure: the number of daughter lesions produced by a parent lesion depends on the instantaneous environmental conditions as well as on dispersal and the host-pathogen interaction. The model assumes that the progeny/parent ratio per day (Rc) is constant over the whole year, in other words a lesion sporulating under unfavourable conditions produces potentially the same number of daughter lesions per day as a lesion sporulating in favourable conditions. In reality the Rc value may drop to zero during a drought period, and compensate to unknown values during unusually favourable weather conditions, such as constant high humidity or heavy rainfall. However, because Rc was calculated from the increase rate of a logistic equation fitted to the epidemic data (apparent infection rate, r) using the average latent period and infectious period length (15), it already incorporates the effect of the average particular environmental conditions of that epidemic in itself.

An illustration of this compensating effect is given in Figure 4: The model, even using the long-term average weather data, simulates accurately the extraordinary epidemic of downy mildew on grape from 1988 in southern Switzerland. The epidemic progressed with a daily r value of 0.28 between the 4th of June (severity of 0.002% of leaf surface attacked) and the 13th of July (severity 51%). From this value we calculated an Rc value of 4.65 with i = 1 and p = 5 days. After this period rain and average humidity returned to more normal conditions and Rc probably dropped to much lower values.
Important long-lasting changes in the course of the epidemic outside the time window used to calculate $R_c$, will not be reflected by the model. The potential progeny/parent ratio should therefore be measured experimentally and then modelled with biological parameters influenced by independent variables such as rain and leaf wetness, instead of being calculated from the epidemic data. Unfortunately such measurements have been rarely made, mostly due to experimental difficulties. Van den Bosch et al. (16) encountered the same difficulties in determining the "gross reproduction" which is essentially the same parameter as the $R_c^\ast i$ mentioned in this paper. For diseases having a longer infectious period than the latency period it was necessary to determine this value indirectly by calculating it from $r$ during the exponential phase using $p$ and $i$.

The total number of daughter lesions that a mother lesion can produce in an uninfected field is the product of the length of the infectious period and the number of daughter lesions per day. The number of lesions per day is intimately related to the sporulation intensity and in reality is not a constant value during the infectious time for an individual lesion, but follows rather a gamma function (16). The shape of the gamma function describing the relative spore production of a lesion per time unit, and the length of the infectious period depend on the current temperature (16), but the total spore production is somewhat constant. As soon as the different generations of lesions overlap, a block function (15) is a sufficient approximation of such a distribution. However, the surface covered by the block function has to be equal to the integral of the gamma function, which leads to shorter infectious periods than the maxima that can be measured.

The third essential parameter, the length of the latency period, is better understood. The relation between the length of the latent period and temperature is known for many pathogens and we were able to incorporate it from data collected independently of the epidemics.

The fundamental condition for the start of any epidemic is the presence of initial inoculum. In our case, the words initial inoculum represent the number of latent lesions arising from infections occurring on the first of January, expressed as the fraction of a minimum overwintering host surface ($K_0$, (3)). The Septoria epidemic
presented was a special case, as it was initiated by a single artificial infection resulting in a high disease severity early in the season (20% on April 11) followed by a dilution effect due to host growth. The good correlation between the real epidemic and the simulation of this particular case is a confirmation of the suitability of our approach in resolving the problem of initial inoculum.

A current trend in computer applications is the WYSIWYG concept (What You See Is What You Get). In the case of epidemic modelling, to account for this trend, we have to represent the visible fraction of the diseased host tissue by setting the incubation period to a non-zero value.

Conclusions

The evaluation of a model may be divided, according to Teng (12), into verification and validation. Verification, which consists in checking that the model performs in a reasonable way, was described previously (3). Validation of a model must prove its correctness and that of its underlying concepts. Statistical criteria are often used for this purpose. Another approach permits a subjective validation (12) by using the "Turing test" (13). It consists of submitting a set of data from the literature to a competent person together with the output of the model, and to ask him/her to differentiate between them (11). This is regarded as a severely critical test (12). To observe the rules strictly, the real data should be presented in the same form as the model output (in our case as a connected line). In the five cases presented the simulation can hardly be distinguished from reality and the requirements for a subjective validation can be considered to be fulfilled.

It can be argued that the Vanderplankian approach leads to models that are descriptive rather than explanatory (6) but the tested model, because of its extensions, can provide insights into the fundamentals of epidemics and their interaction with the weather so that its explanatory value cannot be denied.

With increasing use of computers in epidemiology, the number of models and their complexity will increase even further. One should therefore keep in mind that complexity is not always a synonym for reality, nor does it necessarily increase our comprehension of the real world.

LITERATURE CITED

5. Delp, C.J. 1953. Some environmental factors which influence the development of the grape powdery mildew fungus, Ucinula necator (Schw.) Burr. - Ph. D. diss.
Computers as educational tools in epidemiology

Computers now feature conspicuously in biological education at all levels (Murphy, 1986). The possibility to use computers in teaching biology was recognized very early. One of the first teaching programs was developed by Crosby in 1961 and was used to train students in genetics.

In epidemiology, the introduction of computers for teaching purposes came much later, and most programs in use are simulations. A computer simulation is based on a mathematical model, and students develop understanding of this model by changing input parameters and observing the effects on output (Tawney, 1979). The advantages of simulation in teaching epidemiology are obvious: a whole season can be simulated in a few minutes, parameters can be varied individually at will, and wrong decisions do not lead to economic damage.

The first simulation model used in teaching epidemiology was developed in the form of a game, APPLES CAB, where the player's role was a pest manager in an apple orchard (Arneson et al., 1979). This game was followed by others e.g. SPUDCROP, LATE BLIGHT or GAMERUST (Teng and Rouse, 1984). These games simulate disease progress given certain environmental conditions and, although they are usually based on relatively simple models, they capture enough of the stochastic nature of an epidemic to be useful. However, the success of these games remained limited, because their use required access to mini or mainframe computers and their interactive capabilities were very restricted. APPLES CAB was the first teaching game in epidemiology to be transferred to a microcomputer (Blaise et al., 1987, 1988) and was followed recently by other microcomputer based educational programs: Resistan helps in the understanding of selection of fungicide-resistant plant pathogens in response to application of fungicides (Arneson, 1988) while CORN-III simulates corn growth and northern leaf blight epidemics caused by Exserohilum turcicum (Bowen and Pedersen, 1989).

Literature cited
APPLESCAB: A Teaching Aid on Microcomputers

INTRODUCTION

Of the many uses for computers in plant pathology, epidemiologic models offer an exciting challenge for teaching. With a realistic model, students can learn through trial and error, exercising their management abilities without restriction in making decisions; there is no risk of crop loss or environmental damage resulting from a bad decision. An attractive program should be easy to use, self-explanatory, interactive, accessible at any time, and fun. Unfortunately, little has been done to extend this possibility since the introduction of APPLESCAB (1). Most programs run on mainframes (or minicomputers) to which access is often limited and sometimes formidable for pathology students. We present here a microcomputer version of APPLESCAB, enhanced with graphics and self-explanatory at any level.

THE MODEL

The structure of the model has not been changed, and a more detailed description can be found in Arneson et al (1).

The APPLESCAB simulator has four major submodels:

Weather generator. This part supplies weather data to the other parts and offers two possibilities: 1) "canned" weather, which is read from an external weather file and provides reproducible weather patterns, and 2) simulated weather which is provided by the weather simulator developed by Bruhn (2). 1-, 2- and 3-day forecasts are generated with an error factor.

Tree growth. This part provides the growth stages of the trees, total area of the leaves and leaf area susceptible to apple scab.

Disease development. Disease development is simulated through a time-varying distributed delay with attrition, in an approach similar to the one used by Kranz et al. (3).

Fungicide attenuation. This submodel handles the disappearance of fungicide residues on fruit and leaf surfaces due to fungicide degradation, rainfall and tree growth.

THE PROGRAM

Characteristics. The program, originally written in FORTRAN IV was translated into Pascal, which is currently the most popular structured programming language on microcomputers. The Turbo-Pascal compiler (Version 3.0, Borland International, 4585 Scotts Valley Drive, Scotts Valley, CA 95066) was used. This is a very inexpensive and rapid compiler that generates fast, compact code and has become the de facto

---

1 The APPLESCAB program for personal computers is not copy-protected and will be distributed as shareware. Copies of the program may be requested through the authors. Updates and source may be requested from Ph. Blaise, Institut für Phytomedizin, ETH-Zentrum/LFW, CH-8092, Zürich, Switzerland.
standard on most personal computers. The graphic windowing was generated with the Turbo Graphix Toolbox (Borland International). The use of overlays allows APPLES-CAB to be run with only 256K bytes of RAM (random access memory). The current version runs on the IBM PC, XT (or compatible) with a graphics card (CGC); in addition, the high-resolution mode (640 X 400 pixels) of the Olivetti M24 PC (AT&T 6300) is supported.

**Installation.** The program comes ready to run. All default values such as prices and costs, running mode, and names of fungicides and apple varieties can be changed.

**Sample run**

**Initialisation.** The game allows the user to choose the starting conditions, including weather regime, levels of inoculum and of fungicide resistance, and apple cultivar (Fig. 1). The user also selects the frequency of reports on the state of ascospore maturation, i.e., never, weekly, daily, or on request; costs are calculated according to the
choice and presented in the final financial report. The daily fungicide cover report can be set "off" or "on".

Run. On a day-by-day basis the player can, as in reality, manage the fungicide schedule according to personal knowledge and observations (weather forecast, eventual ascospore report and fungicide cover). The actual state of the system (tree growth, weather conditions, and disease level) is displayed for each day of simulation; disease progression is graphically displayed from the beginning of the season and updated daily (Fig. 2). Ascospore and fungicide reports are displayed according to the initial choice (Fig. 3), and infection periods, calculated according to the Mills table, are signalled to the user as they occur. At any moment the user can perform one of four functions: advance a day, advance directly to the end of the season, spray, or (if the option "on request" has been selected) ask for an ascospore report. Help can be requested at any time to get information on fungicides, biology of the pathogen, etc. At the end of the season, the user gets an overview of the weather and scab epidemic (Fig. 4), a report of scab on fruits and the fungicide treatments, and a financial statement (Fig. 5), with which success as a disease manager can be evaluated.

Fig. 3. The display during the game with fungicide cover and "On request" ascospore report.

Fig. 4. The first display at the end of the season. The top graph shows disease development with arrows indicating dates and type of fungicide sprays. The middle graph shows mean daily temperature and the bottom graph shows rainfall.
Thanks to the incorporated graphics, the player can see not only the actual conditions but also the development of scab up to that point. A decision has to be made daily on whether to advance a day or perform an action. The response to the decision is immediate. The simulation can be extended over many seasons, in which case the amount of primary inoculum each season is determined by the epidemic of the previous season. The final proportion of the fungal population resistant to systemic fungicides is carried over.

**Autoreset mode.** In automatic reset mode, the program returns to the beginning if no key has been pressed during a certain period. We have used this feature in exhibitions where people just walk through and, out of curiosity, start a run without finishing it.

**Execution time** Depending on the processor used, a daily step requires 0.1 sec and 0.4 seconds to compute and a whole season without intervention (obtained by advancing directly to the end of the season) takes 12-48 seconds (Tab. 1). The time needed for computation is short enough to allow several runs in a session. Most of the time is needed for decision making, which is unlimited.

**DISCUSSION**

Models have been and will continue to be widely used as training and instruction aids, with an increasing number being run on personal computers (4,6). The disadvantages of most epidemicologic simulation models requiring mainframes (7) are overcome in the version of APPLES C A B presented here. Graphic display gives a good overview and allows one to follow the dynamic development of the disease influenced by the interventions. This corresponds to the request made by Zadoks (9) that a dynamic rather than a static approach be taken to decision making.

Residue information, as presented in the game, does not represent the real world, where little or no information is available to the grower. Similarly, ascospore maturity information is available on a weekly basis, at best, in the real world, and assessments of the disease level in the orchard require a considerable effort on the part of the grower unless scouting or management services are used. The described version of
APPLESCAB overcomes these inconsistencies by allowing the user to decide whether residue information is provided and to select the frequency of ascospore reports. Reality is best reflected by eliminating all reports, but these reports can be very useful for learning purposes. The obligatory day-by-day step eliminates the risk of overlooking an important event. Moreover, the unlimited reflection time permits discussions of events and decisions. The ease of use allows concentration on the content rather than on the operation of the program.

The use of the Pascal programming language, which is gaining in popularity among scientific programmers (5), facilitates insight into the program structure, and changes are therefore easier to implement. Weather and help data have been separated from the program so that they can be replaced as desired.

The original mainframe version of APPLESCAB, which is one of the most widely distributed simulation programs in plant pathology, is implemented in more than 50 institutions. The wide distribution of personal computers and the reduced format of a disk should increase this number exponentially. For the first time, a simulation program is available in four languages - English, German, French and Italian. This facilitates its use even further. We can agree with Teng and Rouse’s (8) optimistic outlook that as microcomputers become more powerful and less expensive, epidemic simulations will gain more importance in plant pathology and that many of the large epidemic simulation models will be implemented on microcomputers (5).

**LITERATURE CITED**


Forecast and disease warning

Disease warning has been one of the first preoccupations of epidemiologists. Starting from simple rules of thumb, disease warning has evolved to sophisticated warning systems involving automatic weather recording stations coupled with plant growth simulators, simulation models of epidemic development and models predicting yield losses. Predictive systems are especially useful for explosive diseases such as those caused by peronosporaceous fungi like downy mildew of vine (*Plasmopara viticola*) and potato leaf blight (*Phytophthora infestans*) however, computer based disease warning systems have been developed for numerous types of diseases e.g. *Alternaria solani* (Madden et al., 1978) or *Cercospora apii* (Berger, 1969).

A plant disease forecasting system produces a forecast according to a given input (usually weather data). Depending on the size of the system and on the input required, the transfer of information has been implemented in different ways e.g. small electronic devices in the field, postal services, radio, or on-line connection with a mainframe computer. Until now, personal computers have not been used for this purpose, but their increasing presence on farms (Holt, 1985) may well lead to the development of decentralized forecasting systems adapted to the specific needs of the single farmer and with input/output as well as prediction occurring on-site.

Predictive models may be divided into two categories (Krause and Massie, 1975): disease prediction and infection prediction. Krause and Massie divide further both disease and infection prediction systems in empirical and fundamental systems. Most of the disease forecast models developed are empirical disease prediction models, this means that they predict occurrence of disease symptoms, based on the comparison of historical disease and weather records for a specific area. The model presented below is, according to the above definition, a fundamental disease prediction model: based on data obtained experimentally in the laboratory, it gives the probability of disease occurrence for a given set of weather data.

**Literature cited**


On the search for an applicable forecasting model of downy mildew epidemics in Switzerland

ABSTRACT

Evaluation of currently available forecast models for downy mildew epidemics on grape with past weather data from Switzerland did not lead to outputs reflecting the disease situation. Therefore, we initiated the development of a new forecast model. In a first step, an epidemic model based on literature data was built which simulates the development of secondary infections. Based on hourly weather data, it outputs the current state of the epidemic in form of a daily report of infection and sporulation probability, and as the current severity curve. The model is composed of four parts handling respectively infection, sporulation, survival and liberation of spores. The model was translated into the Pascal programming language and runs on personal computers. For the fine tuning of the parameters, detailed epidemiological field data were collected for the first time in a vineyard in the southern part of Switzerland in 1988 where the epidemic was extremely severe. The collected data are presented and compared to the output of the model.

INTRODUCTION

In the region of Zürich, specific control measures against downy mildew are unnecessary in 60% of all years. A forecasting system that avoids treatments in these years corresponds to the needs of IPM.

Besides accuracy, we wanted the system to meet the following requirements: biological soundness of the driving parameters and variables, high flexibility since vineyards vary greatly in Switzerland, and easy implementation of new results. The final implementation should allow the decentralized use of the forecasting system by non-programmers.

In the light of these requirements, two prognosis models currently in use were evaluated. The first model (Strizyk, 1983) does not take into account the biology of the pathogen and failed under the conditions tested to rank correctly the disease severities for the years 1960-1981 (Blaise and Gessler, 1988). The second model (Hill, 1988), although based on biologically meaningful parameters, is limited to a summing of temperature dependent factors which can be read in a look-up table and thus cannot fulfill the mentioned requirements.

Therefore, we initiated the development of a new forecast model. To respond to the requirements, a descriptive model was chosen and implemented on a microcomputer.

THE VINE MILD MODEL

Based on available literature data (Blaeser, 1979), a descriptive model divided into sporulation, survival of zoosporangia (on sporangiophores and detached), infection and host growth was developed. A symbolic representation of the model is given in Fig. 1.

In the following, the main components will be briefly described. The numbers in parentheses are the default values of the model:
Sporulation occurs only in the dark (≥ 4 hr) and requires a minimum temperature (13 °C) and a minimum relative humidity (> 98%)

Survival time (in days) of spores shortens with the increase of the water saturation deficit (Sd) and differs between zoosporangia still attached on the sporangiophores (Za) and those already released (Zr):

\[ Z_a = 9.27 - 1.12 \times S_d + 0.04 \times S_d^2 \]
\[ Z_r = 5.67 - 0.47 \times S_d + 0.02 \times S_d^2 \]

Infections occurs only in the presence of a water film and after a minimum number of degree-hours (50 °C*Hr for wet applied zoosporangia, 71 °C*Hr if dry applied).

Host growth is a simple function fitted to field data collected in southern Switzerland (Fig. 2). It is not yet implemented.

The dissociation of the zoosporangia from the sporangiophores is due only to rain (≥ 5mm).

A fundamental property of the Vinemild model is that it does not work with absolute quantities. The units used are arbitrary units which correspond to the probability for an event to occur. This means for example that the model does not compute the amount of new infections for a given day, but only the probability that infections will occur.
IMPLEMENTATION OF THE VINEMILD MODEL

The model was translated into the Pascal programming language and runs on IBM-compatible personal computers. It is menu-driven and can therefore be used by non-programmers. The weather data can be input either with the built-in editor or read from external files. Once weather data have been entered, the season is run and the program displays four items: the probability for sporulation, the potential for infections and zoosporangia on zoosporangiophores, and the probability of infection. The user receives a status report for any day selected and has also the possibility to influence the model through a feedback feature which allows correction of any difference between the model output and real observations. An additional option allows the display of the current "epidemic level", e.g. the sum of the infections which potentially occurred up to the current day.

PRELIMINARY RESULTS

For the validation and fine-tuning of the model parameters, detailed epidemiological field data were collected for the first time in a vineyard in the southern part of Switzerland in 1988 (Fig. 2). Hourly weather data were registered through an automatic station and included temperature, rain and relative humidity (Fig. 3). The corresponding model output is shown in Fig. 4. The collected epidemiological data were lesion number per leaf, number of leaf infected and size of lesions. Development of the leaf area per plant was also recorded.

According to our observations, the model appeared to predict the first infections too early. However, it is probable that the first infection from oospores has not been detected. On the other hand, the explosion of the epidemic in the second half of June was correctly simulated some days before the emergence of the first symptoms.
DISCUSSION

The accuracy of forecasts depends mostly on two factors:
- The accuracy of the model in simulating the development of the epidemic
- The ability of the system to make the correct appreciation of the past and current epidemiological components in order to help in decision making.

The former criterion is an intrinsic property of the model and can be tested on past data. The output of the model run with the weather data collected shows a good agreement with the development of the extremely severe epidemic of 1988 and although this remains to be confirmed, seems to be able to simulate accurately downy mildew epidemics.

The second criterion depends mostly on experience so that adaptation to the particular conditions of the target vineyard will always be necessary.

Despite these encouraging results, the model still lacks some important features:
- The presence of leaf wetness, which is not required as input is still determined according to a humidity level or the occurrence of rain. An attempt to integrate the formula proposed by Pedro and Gillespie (1982) did not succeed.
- There is no "real" simulation of the epidemic in the sense that the model does not calculate a disease severity. The major consequence of this is that the current epidemiological situation cannot be fully appreciated and so limits seriously the fulfillment of the second criterion cited above. We are currently trying to remedy to this problem by integrating a synthetic epidemic model as proposed by Vanderplank (1963) modified through integration of host growth and variable latent period (Blaise and Gessler, 1989).
- Finally, oospore maturation is not taken into account. Currently, the model assumes that oospores are always present at the beginning of the season, and that their germination depends upon the same conditions as the zoosporangia. Since other workers are looking closely at these phenomena, we hope to be able to implement their results and so to improve the model in this domain.
Fig. 4. Screen display from the Vinemild program at the end of the season. Weather data from Cugnasco, Ticino, 1988. Bottom: daily probability for production of zoosporangia; lower middle: cumulative potential number of ripe living zoosporangia ready to be released; upper middle: cumulative potential number of dispersed living zoosporangia ready to infect; top: daily probability of the dispersed living zoosporangia to infect the host.

LITERATURE CITED

ACKNOWLEDGEMENTS

I wish to thank all the people who helped me in any way, either directly or indirectly during my work. I am especially grateful to the following people:

C. Gessler. He gave me the opportunity to work on a topic of my choice. He always found the time for invaluable discussions and his comments were always very supportive.

H. Kern. Together with G. Defago and C. Gessler, they were my masters in phytopathology and infected me with their virus-like interest for this science.

M.S. Wolfe for being my examiner and for the critical review of my "English" manuscript.

B. Koller. For the Postscript support and the subtle complicity...

G. Eich, B. Büeler, M. Stanga and G. Medici for providing some of the data used in this work.
CURRICULUM VITAE

November 12, 1954  born in La Ferté s/s Jouarre, France
1959 - 1965  Primary school in Cormeilles en Parisis, France
1966 - 1973  Secondary school in Pontoise, France
1974 - 1977  Student at the faculty of Mathematics and Physics at the Swiss Federal Institute of Technology (ETHZ) in Zürich
1978 - 1983  Student at the faculty of Agronomy at the Swiss Federal Institute of Technology (ETHZ) in Zürich
1983  Diploma in Agronomy
1984 - 1989  Scientific collaborator at the Institut for Phytomedicine of the Swiss Federal Institute of Technology (ETHZ) in Zürich (Head: H. Kern, V. Delucchi, M.S. Wolfe)