EISENSMITH, Scott P.: Acquisition, storage, and retrieval of epidemological
data in plant pathology

When studying plant disease epidemics, large environmental and biological data
bases are accumulated. This creates a need for efficient methods for
organizing, storing, and summarizing the data. Computers are suited for storing
and manipulation large, heterogeneous data bases, and computer facilities are
available in many parts of the world. Powerful statistical packages (1, 4, 12,
14) are available but do not meet all the needs of the plant pathologist.
Although computer programs have been written for specific teaching (3, 9, 10),
extension (7, 8, 13), and research purposes (11), little has been done to
provide computer software for handling general epidemiological data. The
computer program EPISTAR was developed to fill this need and reduce the time
researchers must spend in data organization and retrieval. This program was
developed in conjuntion with Drs. Rosemary Loria, Brian Olson and Alan Jones at
Michigan State University. Our approach and several of its advantages are
outlined below.

EPidemiological Information STorage And Retrieval (EPISTAR) is a computer-based
system designed to assist researchers in collection, reduction, storage, and
analysis of host, pathogen, and environmental data. EPISTAR consists of:

o   data collection forms for organizing data into a standard format

o   an interacitve FORTRAN V computer program that will display, summarize, and
    perform coincidence analysis (11) on data

o   a user's manual explaining how to use the system

o   a programmer's manual which discusses program modification and
    transportability.

The EPISTAR computer program is well commented and easily modified to allow for
new data processing commands or additional data types. For example, at Michigan
State University, algorithms developed at Cornell University (5) that allow
transcription of hourly values from hygrothermograph or De Wit leaf wetness (M.
De Wit, Hengelo, Holland) charts have been added to our version of EPISTAR.

The program will accomodate both environmental and biological data, collected
by automatic or manual systems. Currently EPISTAR will handle ten environmental
parameters (air temperature, barometric pressure, leaf wetness, rainfall, rela-
tive humidity, soil moisture, solar radiation, soil temperature, wind
direction, and wind speed) with up to 24 observations per day for each parame-
ter. All output is in the same units as the original data.

Biological data types processed by EPISTAR include: BURKARD spore trap (Burkard
Scientific Sales Ltd., Rickmansworth, Hertfordshire, England) catches, Rotorod
spore trap (Ted Brown Associates, Los Altos Hills, CA 94022) catches, water
trap catches, disease ratings, and stages of host phenology. Burkard spore trap
catches may be recorded up to 24 times per day, while other biological data
types may be recorded, at most, once daily. Data may be taken from multiple
environmental sensors, spore traps, or plots in the same time period.

The computer program performs three basic operations: displays stored data for
examination, calculates summary statistics, and searches for possible
relationships among variables. The display command allows for convenient access

to portions of the original data and facilitates the creation of subfiles for plotting or statistical analyses. To obtain the data, the user specifies the data type(s) and the date range. Then the program searches the data file and prints out the results of the search. The summary command provides various statistics that are useful in analyzing epidemiological information. Each biological and environmental factor is summarized on a daily basis. Summary calculations include minimum, maximum, and average values; number of time periods above or below user-specified thresholds; and number of missing values. The user may select subsets of data for summarization, i.e. data from an environmental sensor, plot, plant, shoot, shoot type, or leaf. The coincidence analysis command permits the location of time periods within the data file with conditions that match user-specified values. The conditions are specified with the FORTRAN relational operator codes 'LT', 'LE', 'EQ', 'NE', 'GT' and 'GE' in reference to a user-supplied value. Series of comparisons may be joined together with the logical operators 'AND' or 'OR'.

Coincidence analysis may be used on all environmental data and Burkard spore trap catches, and is particularly useful in detecting associations between environmental conditions and biological events. For example, it may be desirable to locate time periods with greater than 0,02 cm of rainfall, temperatures less than 23 C, and a spore trap catch of at least 10 spores/hour.

Once the EPISTAR program is accessed, all the information necessary to run the program is requested from the user. The user may choose to display, summarize, perform coincidence analysis, change the placement of the output, select a new data file, or terminate the program by entering the appropriate mnemonic command.

EPISTAR accesses the data file and writes out information about the data file -- including the data collection date range, collector, collection location, and the data types present in the file. If the user chooses to display, summarize, or perform coincidence analysis; the desired data types and probe, plot, or trap number(s) must be selected from those listed as available. If the data are to be summarized, appropriate thresholds should be provided for environmental data.

Relational operators, relational values, and possibly a logical operator must be chooses for coincidence analyses. When the EPISTAR program has completed the data processing, it will again ask the user which procedure is desired -- display, summary, coincidence analysis, new data file, or program termination. User responses are checked for errors, and invalid responses are rejected with a diagnostic message. No knowledge of FORTRAN programming is necessary to use EPISTAR.

The form of output depends on the kind of processing and on the type(s) of data selected. Output from the program may be printed at a computer terminal, or written onto a disk file for use with statistical analysis (1, 12) and plotting (2, 6) programs. In display output, each line represents one day, and begins with the date; data type code; and the probe, plot, or trap number. This is followed by the raw data, with decimal places inserted according to the standard formats.

Each line of summary output begins with the date and the day of the year, followed by the summarized data values for the date. Summary values are calculated from data sets with missing data, but the number of missing values is indicated in the output. If all data values are missing, missing value indicators are given instead of the summarized values. Each data type is separated by a column of stars. Each line of coincidence analysis begins with the date and day of the year.

Those time periods, represented in the title as the first through the 24th hours, for which the user-specified requirements have been satisfied, are indicated with stars. The total number of time periods for which condition where satisfied for a day, and for a particular time period, are given as the sums of the rows and columns in the table, respectively.

The EPISTAR system facilitates collection, storage, and analysis of epidemiological data, and may serve as a model for other standardized epidemiological data processing systems. Data forms allow for immediate organization of the information into computercompatible format, thus facilitating the use of other computer software for data analysis on plotting.

Subsets of the master data file may be easily selected by date range and data type. Multiple observations for a single day may be summarized into daily values. Those data meeting user-specified conditions can be quickly located within the data file.

The computer program has been written in a modular format to facilitate modification, including the addition of new data types or processing routines. We hope it may thus serve as a framework for additional epidemiologically-oriented computer software.

Literature Cited

1. ANONYMOUS: Michigan State University STAT system: User's guide VIII. Computer Laboratory, Michigan State University, East Lansing, MI 48824, 1974

2. ANONYMOUS: Statistical plotting on-line command system: user's guide. Pest Management Technical Report 13. Dept. of Entomology, Michigan State University, East Lansing, MI 48824, 54 pp., 1977

3. ARNESON, P.A., OREN, T.R., LORIA, R., JENKINS, J.J., GOODMAN, E.D. and COOPER, W.E.: APPLESCAB: A pest management game. Dept. of Entomology, Michigan State University, East Lansing, MI 48824, 1978

4. BARR, A.J., GOODNIGHT, J.H., SALL, J.P. and HELWIG, J.T.: A user's guide to SAS-76 Institute, P.O.Box 10066, Raleigh, NC 27605, 1976

5. BLUME, M.C., SEEM, R.C. and BARNARD, J.: Two computer programs used in the analysis of rectangular and circular charts from continously recording weather instruments. Search 9 (1) Plant Pathology Dept. No. 5, N.Y. Agric. Expt. Stn., Geneva, NY. 11 pp., 1978

6. BOBZIN, J., EISENSMITH, S.P. and JONES, A.L.: Plotting of environmental and biological data: User's guide. Pest Management Technical Report 16, Dept. of Botany and Plant Pathology, Michigan State University, East Lansing, MI 48824, 82 pp., 1978

7. COTTER, H.V.T., MAC HARDY, W.E. and WARREN, J.A.: Storing, retrieving, and reporting plant disease occurence data using a computer data base management system. Plant Dis. Rept. 63 (2), 117-121, 1979

8. CROFT, B.A., HOWES, J.L. and WELCH, S.: A computer based extension pest management delivery system. Envir. Ent. 5 (1), 20-34, 1976 DIMOFF, K. and KILMER, N.: SAMPLE user's manual. Dept. of Entomology, Michigan State University, East Lansing, MI 48824, 21 pp., 1978

9. LOGAN, P.: Orchard: A serious predator-prey interaction game. Dept. of Entomology, Michigan State University, East Lansing, MI 48824, 26 pp., 1978

10. MOGK, M.: Automatic data processing in analysis of epidemics. Pages 55-77 in: J. Kranz, ed. Epidemics of Plant Disease. Springer-Verlag, New York, NY. 170 pp., 1974

11. NIE, H.C., HULL, H., JENKINS, J.G., STEINBRENNER, K. and BENT, D.H.: Statistical Package for the Social Sciences. Mc Graw-Hill, Co. Inc., New York, NY. 675 pp., 1978

12. PENNYPACKER, S.P., KNOBLE, H.D. and LONGENECKER, J.L.: Plant disease survey information (PDSI): a computer data storage and retrieval system. (Abst. No. 192), Proc. Am. Phytopahtol. Soc. 3:246, 1976

13. RYAN, T.A., JOINER, B.L. and RYAN, B.F.: MINITAB: A Student Handbook. Duxbury Press Div. of Wadsworth Pub. Co., North Scituate, MA. 341 pp., 1976