COMPUTERS AND THE FIELD RESEARCH STATION

J. Robert Stottlemyer

Recently the National Science Foundation sponsored a workshop on data management at biological field stations, the first of what probably will be a series of sessions dealing specifically with the questions of data cataloging, data administration, computers and software options for data management, and mechanisms and incentives for intersite exchange of information. The workshop The workshop objective was to consider these topics as they relate primarily to the "small" and "medium-sized" field research station. Such stations commonly are associated with Long Term Ecological Research sites, Experimental Ecological Reserves, university field stations, and a few national parks. Results of the workshop will be published this fall.

A topic shown much interest at the workshop and one debated often by researchers is the expanding interest in and use of automated data handling techniques. The driving factor behind this interest, beyond the intense advertising particularly on the part of micro-computer manufacturers, is the ever-changing need for more sophisticated and responsive data bases and their management systems. Increasingly, data bases are needed for maintenance of research "memory", to bring in multiple data sets to address complex regional issues, for the production of more rigorous data sets with improved quality control, the stimulation of new hypotheses and their testing, and to provide easy access to commonly used data. Such needs do not, in themselves, dictate the need for computer application, but they raise the question of whether such should or should not be used.

There is always the possibility of technology overkill in applying computer methodology to field station data bases. For example, many data sets simply do not need to be computer based. Also technology overkill appears more probable when research programs have a quick start-up, such as often results with large federally-funded programs that try to "find the answer" for a newly emerged problem. Some attribute the shortcomings of the International Biological Program to this, especially when weighed against the less grand but highly respected concurrent ecosystem research conducted at places such as Hubbard Brook Experimental Forest, among others. As a rule these latter programs started out small and evolved slowly and methodically. Regardless, if one examines the ongoing research programs at sites such as Coweeta, H. J. Andrews and Hubbard Brook experimental forests, one finds many differences in approach to data base and data base management systems. The basic reason is probably the differing philosophies of the principal investigators and the differing states of evolution in the research programs.

This appears a logical point to risk the ire of the more knowledgeable on the topic of computers and field station data bases, and make some specific observations and suggestions regarding

computer application which I think reflect to some degree a concensus among users. Data bases, their management, and data exchange are not considered here in detail as they are, in themselves, equally complex topics.

The most difficult question to assess objectively and the one that must be answered first is what are my data base and data base management needs? It is the needs of the user that dictate the subsequent choices in software, which in turn dictate the hardware choice. While this sounds intuitively obvious, in many instances this sequence of steps is not followed. Quite often it is exactly the opposite even when the investment is a $5-million mainframe. The small research station director can even less afford this mistake. While equipment cost is becoming a lesser concern, the topics mentioned above—data base needs, data management needs, and data exchange objectives—are the key questions to answer before determining software and hardware needs. More specifically, data entry, manipulation, and analysis needs must be specified. So too are data documentation and cataloging needs. Data documentation refers to the conditions under which the data were collected, by whom, and when did procedures change, etc. Presently, such information is rarely readily accessible in data bases if it is there at all. The question of data documentation is perhaps a "threshold" question for there is currently no good software on the market for micro- or mini-computers to fully utilize or analyze a documented data base, let alone merge raw data with other available information. Data base management software systems (DBMS)—that is, programs that can collate and merge information from various available sources—are relatively rare, expensive, and until now only available on mainframes. However, SAS, which is one popular DBMS, probably will become available for mini-computers in the not too distant future.

Statistical processing systems (SPS) and/or statistical computation systems (SCS) are readily available for all micros and minis. These generally are not too expensive (typical price about $300 - $500), and many versions, each with their individual strengths, are on the market especially for the more popular computer makes. These will effectively cooate and merge raw data files. They also will do most standard statistical testing and trend analyses. Most also have relatively crude graphics capabilities which can do much to reduce the amount of "what if" file manipulation. Most major micros and minis have sophisticated graphics software systems available which, when combined with a decent plotter, can produce "publication quality" end products. There are problems and pit falls in the graphics area however. Many cheap plotters on the market are simply not up to the physical task of producing good copy. Also, it is often difficult if not impossible without access to a programmer to get the results from your statistical analysis package into final graph form using the graphics software system. But to jump into hardware briefly a good single color plotter can be purchased for less than $2000.

Here I must confess to an already apparent bias. For those with access to micros and minis I strongly lean to using off-the-shelf or packaged software. The array of programs available for the major computer brands is truly formidable. One must do a little shopping, since most of it is available from relatively small vendors, but the trend is favoring the scientist and other professionals more and more. Initially, the small- and medium-size computer market was primarily the business world, but especially with the advent of the "personal computer" more and more software is becoming available for the academician, researcher, etc. And the emphasis clearly is "user friendliness." That is, the less intimidating the software the more marketable it is. This is not to say that access to programming skills is not desirable, for it is! But before any small research facility manager devotes a position to a programmer the options should be really carefully studied.

To summarize the issues concerning the definition of data needs and software, one must decide on what capabilities he feels his facility truly needs, what is the user group, and what is the present and future budget situation? To repeat, the determination of data needs is the first question to answer before software and hardware needs are addressed.

With the questions of data needs and software carefully thought out, the array of hardware options becomes much more manageable. However, there still remain some major considerations and a series of minor questions. One of the first decisions is whether to purchase a system outright or contract with outside services. For field research stations the purchase or lease of a large computer is neither necessary nor practical. If access to such computational capacity is truly needed the purchase/contract question is reduced to either purchasing or renting a micro or "dumb" terminal or contracting with an outside service in total. With the price of micros and "dumb" terminals rapidly falling, the leasing of such equipment is increasingly a poor choice economically. This is especially true for the micro, since it cannot only be used as a terminal for front-ending a larger mini or mainframe system, but when not serving that purpose it can do a variety of other tasks such as word processing, preliminary file analysis and manipulation, etc.

If extensive access to a mainframe is not needed and the present array of micros on the market is not really adequate than a mini might be the clear option. Several research stations I am familiar with, each having about 8 - 10 scientists, have gone with a mini to meet their total file and analysis needs. These have been equipped with peripherals such as hard disk, word processing, and publication quality graphics to round out the system, making it suitable for station administrative needs also. The major lines of minis have an extensive assortment of packaged software systems available. They are also usually multiple-user systems. And as mentioned earlier SAS, a DBMS, will pro-

bably soon become available for minis. Since one can easily invest $25 - $75 thousand or more in such a system, leasing remains a viable option. But, while it is dangerous to make any generalization in this very dynamic industry, it appears most research stations that determine they need mini capability are purchasing such systems outright.

The capabilities of micro computers are perhaps the hardest to call since the competition is very keen, and we are in the stage where many companies are entering the market and many will fold in the years to come. While the technology employed in micros is at least 10 years old, the emphasis here clearly is on "user friendliness." At times it can be almost sickening. But there also have been some significant hardware improvements over the last year or two. Most of the major companies now offer 16 or 16/32 bit micro processors, which greatly reduce file analysis times. Most also have at least one model with 256k or 512k Random Access Memory (RAM), and "on line" solid disk storage can exceed 30 million characters. Solid disk storage also has opened the door to limited "multiple-user" capabilities on micros. This means that while one person is using the micro for analysis, up to two other terminals (or micros) can be simultaneously entering data for other files. Cost is not a major consideration in micros anymore, and as earlier mentioned they can be used to "front end" a mainframe with the proper handshake software.

There are a few additional points to consider regarding hardware. While the bigger vendor is not always the better, size is a factor to consider, particularly in the volatile micro and mini arena. Vendor size and stability is also a concern in servicing equipment especially in remote areas. It is a significant factor in off-the-shelf software availability. The large micro and mini hardware manufacturers are not emphasizing the production of software. This is mostly done by small specialty firms. Since it is virtually impossible to prohibit the mass duplication of most software, the small software manufacturer has but a limited time to market a new system and must therefore place emphasis on software for popular makes.

If one's objective is inter-site exchange of data a whole series of questions are relevant to computer networks. And while handshake software can be purchased off-the-shelf for the more popular computer makes and custom made for virtually any other, getting computers to "talk" to one another is generally not a task for the impatient or easily intimidated. This is especially true when the communication is to be between a single-user and multiple-user system.

To conclude, this discussion is by no means a treatise of the topic. While I am certainly not free of bias, my objective here was to present a sequence of questions and considerations that I feel are important in assessing the possible application of computers to the operation of a small to medium sized field research

station. It is very evident that the demands on the field research station for more rigorous data sets with better quality control and much improved access will intensify. The reasons are two-fold. One is the fact that presently most assessments of regional or national issues use information of very limited resolution. There is a certain "depth" or level of data resolution which is never penetrated. Second, the quality of data used is generally not a consideration. Typical regional and national problem assessments pay much attention to summarized research results, but very little, if any, attention to variation in data sets, mean variance, or the differing conditions under which the data were collected (data documentation). There is considerable hazard in this approach, particularly as the results of such analyses are carried beyond the assessment stage and are applied to correct the problem.

J. ROBERT STOTTLEMYER is Leader of the NPS Great Lakes Area Resource Studies Unit, and Adjunct Professor of Biological Sciences at Michigan Technological University, Houghton.

PROGRAM FOR THE TRIENNIAL CONFERENCE

Program announcements have been sent to all members and prospective participants for the Triennial Conference, to be held in Washington, DC, 18-20 October 1982. The announcement contains a program sequence, a letter containing addition details from President Tommy Gilbert, several blank Pre-Registration forms and a return envelope. Recipients are urged to return to Pre-Registration form as soon as possible—this will allow better pre-Conference planning. The extra Pre-Registration forms have been enclosed for passing on to interested acquaintances. Additional forms are available on request. See center-fold for a copy of the Conference Program, which is now complete (the mailoutannouncement contained three "[being arranged]" items).