- In a similar way, speakers, organizers, and attendees of meetings can be electronically connected to facilitate quick and informal exchanges at all levels, largely replacing slow and costly conventional mail.

Regarding the last point, in the numerical analysis network, small meetings are often announced very early to tailor contents according to the requests from the community. This may be a more democratic way of managing meetings if insider circles can be put under pressure from the outside.

A central computer node run by the leading societies of our profession could provide further services, like maintaining and updating various databases useful to the general audience. One of them could be an on-line equivalent of a professional society directory which could be queried by sending it an electronic mail message. Also, one could establish a database of abstracts of technical reports, collected from statistics groups at universities, industrial laboratories, and journals. Full copies of reports could then be ordered from and sent by the authors or their departments, again preferably via electronic mail. An exciting aspect of this facility would be the capability for keyword search among reports and papers.

Another type of database maintained at a central node could be a collection of data sets, documented according to standards to be established. Access to the database would be provided on request from researchers trying out new methods as well as from teachers in need of data sets for their classes. One could go even farther and keep a record of short accounts of data analyses done in the past, with new accounts being incorporated as data sets are reanalyzed. This should ultimately result in an interesting history of data analysis by way of multiply analyzed data sets. A database of statistical software, or at least pointers to programs, could be tremendously useful, too.

We should point out that the proposals made above, although formulated in terms of a central node computer, could be realized in more decentralized ways as well. Schemes exist whereby copies of some or most of the databases could be kept locally in every computer, with some nodes providing updates periodically. We do not propose a particular implementation but the provision of services and capabilities of interest to the profession.

Last, we recommend that the statistical societies designate a few individuals with suitable expertise as consultants on computing issues for the profession. We hesitate to call this yet another "committee," but we sense that there are a sizeable number of statistics departments which have not yet developed sufficient expertise in computing matters and who need someone to turn to for start up help. It seems to us that some of the panel members, whose report we have at hand, would be especially qualified for this task.

Comment

David W. Scott

1. ABSTRACT

Eddy and his coauthors are to be warmly thanked for bringing together such a complete array of information for creating successful research computing environments at the departmental level. I have added a few observations, comments, and predictions directed to the very difficult task of communicating the "feel" of a good computer environment. I believe the authors' report should have a positive effect accelerating the availability of quality computing for statistical research.

David W. Scott is Professor of Mathematical Sciences, Rice University, Houston, Texas 77251. These comments were written while the author was on sabbatical leave at the Department of Statistics, Stanford University.

2. EXPANDING COMPUTER HORIZONS

Some 5 years ago in the **R** room overlooking Rice stadium, a data processing manager for an oil company addressed an assortment of academic and industrial computer experts. He remarked that for his large seismic processing operation to be successful, it required, in part, at least a 60-day backlog of jobs. Ten years ago when I graduated I believed that on-line computing was too expensive and wasteful of my time, encouraging unproductive experimentation rather than careful and thoughtful program development possible with batch processing, which of course was the only kind of computing available to me. While I cannot speak for the first gentleman, my own attitude toward computing has undergone some changes. My primary computing resource has seldom lasted longer than 2 years, beginning with an IBM 1620; in graduate

school a CDC briefly and then an IBM 370; then sequentially, a DEC 10, Prime, Hitachi, VAX (VMS), Pyramid, and VAX (Unix). Although I have used a lot of computers, I have never had much control over the choice until recently.

In the report of Eddy et al., we see that the diversity of computers and computing environments for statistical research continues to grow (surprising the authors), from isolated personal computers to decentralized workstations to departmental minicomputers to university wide mainframes as well as all possible combinations. Although the authors are quite neutral on the relative merits of these configurations, I am sure they are mystified by the conditions under which many researchers labor. While I have never seen anyone easily persuaded to switch programming languages much less operating systems or hardware, it is important for all statisticians, even causal users, to be able to share the computing experiences of other statisticians, particularly those experimenting with state of the art hardware and software. I believe departments should use the authors' report as a basis for creating a planning instrument or simply to evaluate the status quo.

3. GAINING CONTROL

The authors clearly believe there are substantial benefits to taking control of computing at the departmental level, with certain tasks reserved for supercomputers and university mainframes. The single most important ingredient for successful statistical computer research is having sufficient resources to engage in experimental computing. The clearest way to do this from a financial point of view is with a personal computer. At the other extreme, the university mainframe usually offers excellent support, long term stability of computing environment, and a wide array of software options. The cost is relatively uneven service, slow communication speeds (especially for graphics), limited opportunity to use software not designed for that machine, and expense. To obtain mainframe computing for my dissertation, I had to make a presentation to NASA to obtain approval for a $2000 reallocation of grant money. This level of difficulty remains today for many researchers. Computer center managers were caught between the conflicting demands of its users and the internal demands of its workers, for whom standard state of the art equipment is vital to maintaining their competitiveness in the job market.

Computer Science Departments led the revolt against computer centers when it became clear that minicomputers could be supported by researchers in a single department and would not require full time operators. Many other departments followed suit, although it is clear from this paper that many have not. Today, minicomputers, workstations, and microcomputers offer even small departments the opportunity to control research computer expenses by limiting operating costs to maintenance costs (assuming purchase and installation are covered by grants and university sources); see the authors' Appendices II and III. But the risk and headaches of dealing with one's own machine can reduce the attractiveness of this option. The authors give a valuable and realistic description of what responsibility a department is assuming and what support resources are necessary.

Researchers become attached to their computing facilities. After a few months of exclusive use, it is easy to believe that a good computing environment is "whatever you've got." Thus I feel the authors' italicized statement that "standardization is more important to long range effectiveness than either price or performance" may be misinterpreted as favoring the status quo. I believe the correct emphasis is on ease of communication among the array of CPUs and output devices in the system. Computing that is an inconvenience does not further the goal of statistical research. Many departments have adequate CPU power in principle, but find that most users gravitate to the more easily accessible central "server" rather than walk down the hall to a theoretically more powerful but isolated single-user CPU. That sort of bad planning or bad luck can be demoralizing since few departments have the resources to absorb such errors. As to the authors' definition of "standard," duplicating computers in the local computer science or electrical engineering department or the central facility can have distinct advantages, such as extra volume price discounts.

4. LOOKING INTO THE FUTURE

Research computing became exciting 5 years ago when maintenance costs of minicomputers could be covered by pooling the computational budgets of a homogeneous small group of researchers. Today, workstations are rapidly evolving to the point where the same statement can be made for *individual* researchers. I expect that universities supporting mainframes exclusively will find distributed workstations more cost-effective within 4 years. The model is clearly illustrated in Appendix II: a minicomputer works as a "server," supporting as many as a dozen workstations. It should be recalled that most workstations themselves can support an extra terminal or two. The advantage of this model is that additional CPUs (workstations) can be added to the server network at small marginal cost, providing positive benefits to everyone connected to the server.

It is exciting to watch the convergence of micro-

computers, workstations, and minicomputers. My own feeling is that microcomputers have a slight advantage in that they will add minicomputer features when the price of the machine will bear it while minicomputer manufacturers have to face the prospect of (for them) radical price reductions. Furthermore, microcomputers have a tradition of providing remarkably powerful software at prices attractive to individuals. Microcomputers are rapidly approaching the threshold where they may substitute for a minicomputer's nonserver functions. Mainframe and minicomputer software vendors moving into the microcomputer and workstation environment face great uncertainty about how to price their packages in addition to shedding the "mainframe feel" for a workstation environment.

The authors highlight the support issue. This includes annual maintenance expense of course, but the primary headache is system support at the university level for the diverse set of hardware. We seem to be making a full circle. Whereas once we had fully centralized computing support, researchers led by computer scientists were eager to escape that style of computing. Now with large inventories of microcomputers and networks of minicomputers and workstations, some centralized expertise once again is appropriate and necessary. While the fate of computer centers was uncertain a few years ago, it now seems clear that they can serve a useful university wide role of advisor for computer purchases and central provider of software and hardware maintenance.

The irony of this circle is that the new supercomputer centers would appear to be a return to an older style of computing. Access to supercomputers is critical for many identified problems. But the relatively narrow bandwidth of these machines limits their utility for problems in graphics. Furthermore the computational advantage between supercomputers and some workstations is not large and the time lag narrowing. Careful planning is required by granting agencies to assure that adequate resources for nonsupercomputing equipment remain available even when unexpected budget shocks are felt. Relatively modestly priced hardware is available that permits a comfortable experimental and developmental environment before moving to a supercomputer.

5. FINAL REMARKS

While the number of statisticians proudly proclaiming their ignorance of computing has diminished, a surprisingly large group of apathetic folks remain. I hope this report will stimulate them to get involved. Many surprising attitudes remain. For example, some researchers find electronic mail as intrusive as the telephone. I can't imagine having to return to the endless cycle of returned phone calls on simple matters. Another finding in the survey that particularly surprised me was the current lack of use of symbolic algebra software. Finally it is difficult to divorce use of computers for statistical research from other day to day requirements. Do the authors believe these can or should be separated?

ACKNOWLEDGMENT

This research was supported in part by Contract N00014-85-K-0100 from the Office of Naval Research.

Comment

Prem K. Goel

First of all, I must congratulate the workshop members for producing such a comprehensive report on the use of computers in statistical research and the editors of *Statistical Science* for publishing it and facilitating its dissemination. This report will be extremely helpful in convincing decision makers about the need for a "dedicated" computing facility and to statistics researchers in making the right choices. The report correctly emphasizes the role of computers in ongoing statistical research and the opportunity for cross-fertilization with other sciences. Because of computers, statistical research has come full circle. It is now fashionable in the research community to talk about developing useful and powerful data analytic methodology rather than just proving asymptotic properties of procedures which would not have been implemented on large data sets, in any case, without access to cheap and fast computing power. However, only a very small proportion of the research

Prem K. Goel is Professor of Statistics and Director of the Statistical Consulting Service at The Ohio State University. He served as Director of the Statistics and Probability Program at the National Science Foundation during 1982–1983. His mailing address is Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, Ohio 43210.