from their solar origins through the upper atmosphere to an observed change in the behavior of the lower atmosphere. Bruce Springer of NOAA's Space Environment Laboratory in Boulder suggested how solar sector boundary crossings might influence Wilcox's VAI, as well as more conventional measures of the weather. Springer's mechanism involves a chain of events that converts a solar-induced infall of charged particles into changes in atmospheric motions via a huge electrical current in the auroral zone called the electrojet.

The required leverage would be gained by trapping and amplifying a normally high-ranging atmospheric wave below an altitude of 60 kilometers. This wave, called planetary wave number 1, circles the globe and can play a role in U.S. dry spells and severe cold. Springer said that the effect of sector boundary crossings on atmospheric pressure can be found during the severe cold of the winter of 1979, among other times. Changes in the VAI, he says, are minor, peripheral effects of this process. Although the electrojet segment of his mechanism did not elicit great enthusiasm among his listeners, the idea of trapping a planetary wave seemed plausible, and, contrary to many sun-weather proposals, it can be checked with reliable observations that are already in hand.

Another appealing mechanism came from George Reid and Kenneth Gage of NOAA's Aeronomy Laboratory in Boulder. They suggested that solar constant variations as small as a few tenths of a percent over an 11-year cycle can influence atmospheric convection over the tropical ocean. As dramatically visualized by towering cumulus clouds and thunderheads, tropical convection plays an important role in the global transport of heat and moisture away from the strongly heated ocean surface. Reid and Gage noted that the periodic changes in solar energy received during the year, which result from the annual variation in Sun-Earth distance, do produce noticeable changes in tropical convection. Invigorated convection pushes the tropopause, the atmospheric lid over the weather system, to greater heights when Earth is closest to the sun, they say. They also found a long-term variation in tropopause height in the western Pacific that, at least during their relatively short 20-year record, is in phase with the sunspot cycle. They have not yet compared observed sea surface temperatures and troposphere heights.

The NRC panel would probably welcome more of these studies that test specific physical models and mechanisms. It went even further in calling for a shift in emphasis "from the traditional pattern of *searching for evidence* [of a correlation] to a more directed effort at *understanding the physics* of the atmosphere and the solar-terrestrial system as a whole." Sun-weather researchers have their work cut out for them.

—RICHARD A. KERR

Computer Graphics Comes to Statistics

By looking at multidimensional data on computer screens, statisticians are finding relationships that never would have been detected with standard methods

Several groups of statisticians and computer scientists have devised computer motion graphical displays of data which allow them to see multidimensional data sets and manipulate them to search for patterns. Scientists at Stanford's Linear Accelerator Center (SLAC), and at Harvard University, for example, are seeing patterns in data that never would have been picked up with standard statistical techniques.

The aim of data analysis is to discover patterns, to find nonrandom clusters of data points. Traditionally, this is done by using mathematical formulas. But, with the advent of computer motion graphics, it has become possible to look at three-dimensional projections of the data and to make use of the uniquely human ability to recognize meaningful patterns in the data.

The new method, although powerful, is still somewhat controversial, especially among conservative statisticians. Jerome Friedman, a principal designer of the SLAC system, remarks, "The old school believes that research in statistics is research in mathematics. They believe you should only use simple statistical methods where the mathematics is well worked out." By that standard, this stuff is very far out.

Peter Huber, a Harvard statistician who also has designed a computer graphics display for data analysis, cautions that there is a danger in looking at pictures rather than using standard statistical methods. "In the orthodox view, you specify all that you will do before you look at the data," he says. "You calculate the probability that things will happen and then if you see something suspicious when you look at the data, you will be fairly sure it will be significant at the 5 or 10 percent level. If you look at data long enough you will almost always see some structure. If you are ultra-orthodox you will avoid that problem by simply not looking at the data."

Huber believes that the ultra-orthodox statisticians are veering too far in the direction of safety. With his newly developed graphical displays, he has already noticed some heretofore undetected anomalies in published data analyses. For example, Huber's colleagues David Donoho and Mathis Thoma discovered an unusual structure in data representing the fuel consumption of cars of various years. When they looked carefully at that structure, they discovered there had been a change in the definition of horsepower during the years the data were gathered. This change in definition, although present in the description of the data, was buried and had been overlooked by other analysts.

The advent of computer motion graphics in statistical analysis took place 10 years ago at SLAC when John Tukey, a Princeton statistician with a long-standing enthusiasm for looking at data to see patterns, took a sabbatical at Stanford. It was Tukey's mission to study how to use computer graphics in statistics; at SLAC he had the makings of a system—albeit an extremely expensive one. William Miller, who now is president of SRI, had put together a graphics facility whose hardware alone cost $150,000 in late 1960 dollars. The graphics facility was hooked up, by way of an expensive fast link, to a $6-million SLAC computer.

Using this system, Tukey and Mary Ann Fisherkeller of Stanford designed, in 4 months, the first graphical display of statistics data. They called their system PRIM-9, for Picturing Rotation, Isolation, and Meshing up to nine dimensions.

0036-8075/82/0903-0919$01.00/0

"The original PRIM-9 was very well received but it was not emulated because of the cost of hardware," says Friedman. For a while there were not many new ideas on how to publish results using PRIM-9 because the patterns in the data were displayed in three dimensions. The SLAC group made a film of the system which they sent out on request and which was seen, says Friedman, "by many more people than would have read a publication." Stanford statistician Rupert G. Miller and physician G. M. Reavan of the Veterans Administration Hospital in Palo Alto used PRIM-9 to analyze diabetes data and had an artist draw what was on the computer screen.

What helps make the graphics technique work is the idea of "projection pursuit," a concept developed by Friedman and Tukey. "In watching people use the original PRIM-9, we asked, What are people looking for?" Friedman remarks. "We came across the notion that they are looking for nonuniformities in the distribution of points, tendencies of the points cluster. We came up with a structure index that is related to entropy—we estimate the entropy of a picture. We then can use a numerical optimization algorithm to get the best picture."

The advantage of the projection pursuit algorithm is that it can quickly sift through all possible pictures to find the best one. The human eye is much slower but is a better pattern recognizer. So the SLAC scientists decided to combine the two—to have the projection pursuit algorithm find what it calculated as the best picture and then have the human observer fine tune the picture.

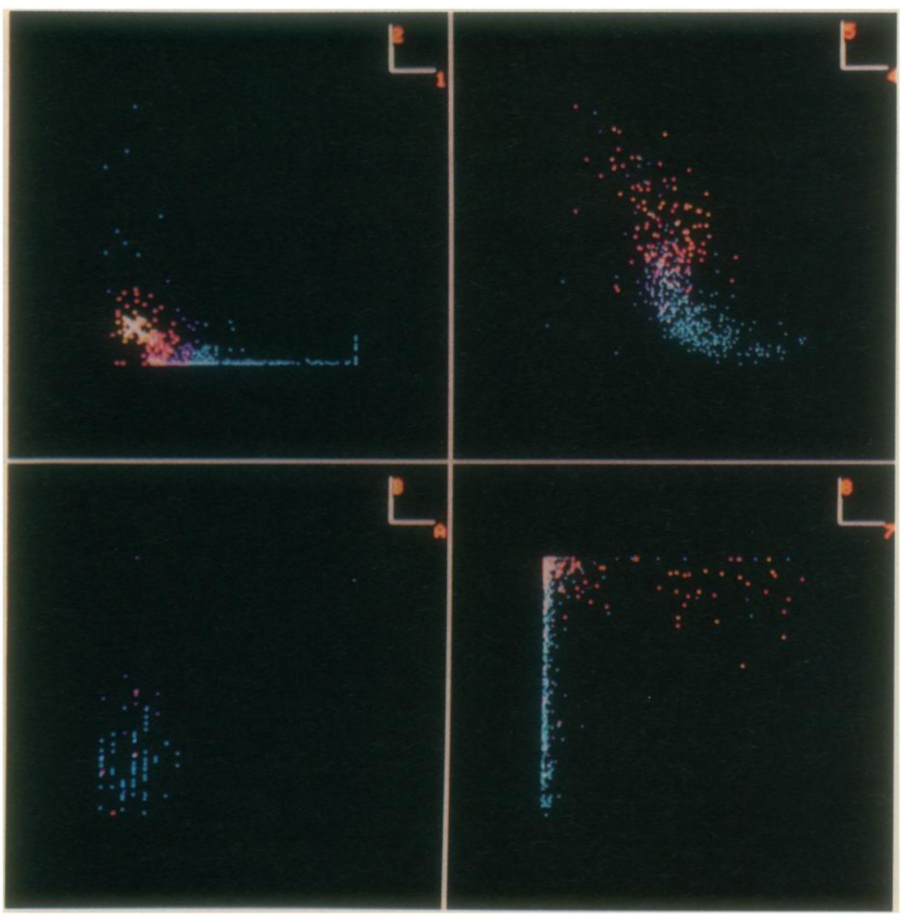
About 1½ years ago, Friedman, SLAC statistician Werner Stuetzle, and their associates began to build the system they have today. "What prompted us was that the original PRIM-9 was directed toward cluster analysis but not toward other sorts of statistical analysis that are very important, such as regression and classification," says Friedman. "In 1979, Werner and I came up with a way to use projection pursuit to treat those problems. That rekindled our interest," he remarks. At about the same time Huber at Harvard set up his graphics system, which makes use of standard graphics programs used by Harvard chemists.

When Friedman and Stuetzle returned to the problem of computer graphics in statistical analysis after their 8-year hiatus, they came back to a totally different computing environment. Before about 1976, graphics were very expensive and did not employ color. Then, when the cost of computer memory dropped, it occurred to computer scientists that they could use standard color television technology for computer graphics. They could make a big memory with a bit in the memory for every point on the screen. An electron beam would continuously scan the screen and every time the beam hit a point the memory would tell it whether the point should light up and, if so, what color it should be. The SLAC system, for example, has eight bits of color information at each point so, by combining these primary colors, it is possible to get 256 different colors. The researchers can then label points of interest in a data set with color in order to keep track of them.

The SLAC group built its own special-purpose hardware for its statistics displays at a cost of $40,000. They store their files in SLAC's large computer. "The thing that's revolutionary is the amount of computing power that we can devote to a single user," says Friedman. "That little box has as much computing power as many typical university computing centers. It is inexpensive enough that a reasonable place can buy several."

The SLAC system is not yet ready for large-scale use by statisticians, however. The researchers still must program a projection pursuit algorithm onto the device to aid users in finding the best patterns. And they also need to program in a memory for patterns that were observed so that if a person finds a particularly intriguing pattern, he can have the computer immediately go back to it.

But the hard steps already are taken. The device works, it is inexpensive, and it is powerful. "Over the next few years these computer graphics devices will lead to a totally new style of data analysis," Huber predicts.—Gina Kolata



The 506 dots on the computer screen represent 506 census districts in metropolitan Boston. There are 14 variables measured for each district, representing such things as average value of homes in the district, crime rate, student-teacher ratio in the public schools, commuting distance to Boston, and the percentage of blacks in the district. What shows up on the computer screen is a projection of these 14-dimensional points on particular three-dimensional spaces. The researchers can move the three-dimensional space around the cloud of points so the points will be projected at different angles. The four pictures represent four different projections of the census data. In the upper left picture, an x *is arbitrarily placed in the midst of the data points. The points are then colored to indicate their distance from the* x. *The points retain the same colors in the other pictures. Thus if there is an interesting structure in one view, it can be followed through these color codings in other views of the same data.*