Pity the Poor Weatherman

Despite satellites, supercomputers, and billions of observations, weather forecasting skill is improving only slowly, often too slowly for the public to notice

TV weather forecasters rarely look back. Yesterday's forecast is usually forgotten, especially if it was a bust. The National Weather Service (NWS) has not been so easy on itself. It has been keeping its own scorecard, which reveals just how hard it is to forecast the weather or to improve on past performance. Despite more than 30 years of steady technological advancements, weather forecasters must usually content themselves with slow improvements in forecasting skills, although those skills were often not that great to begin with. The weather, it seems, is simply not as easy to predict as TV forecasters would have you believe.

Will it rain tomorrow? The NWS has been about 85 percent correct in answering that question. That figure comes up often when forecast skill is discussed, but that kind of figure has certain failings. One is that it gives the forecaster an unfair advantage. In Los Angeles, for example, a forecast of no rain for everyday of the year would be correct 97 percent of the time. A more informative measure of skill takes account of the local climate; if the forecaster cannot do better on average than do forecasts based on climate averages, he would not be credited with having any skill.

Records of forecasting skill pitted against climatology show that precipitation forecasts of the 48 NWS Forecast Offices are better than informed guessing, but not always by much, and have improved over the years but sometimes only marginally. There have been some suggestions that progress has been nil, but Robert Glahn of the NWS in Silver Spring, Maryland, seems to have produced some generally accepted analyses showing an upward trend in skill (*1*). He recently compared forecasts of the probability of precipitation and observations of actual precipitation from 1967 through 1981. Summer was the toughest season. Forecasts for 12 to 24 hours ahead in 1967 were 24 percent better than a climatological forecast and 14 years later were only 30 percent better. Forecasts for 36 to 48 hours ahead in 1967 were only 10 percent better than climatology but were 15 percent better in 1981, about a 50 percent improvement over an admittedly poor initial showing.

Forecasting precipitation was easier during the winter, the shorter range forecasts for 1967 being about 37 percent better than climatology and the longer range ones about 15 percent better. Longer range winter forecasts were also easier to improve over this period, the 36- to 48-hour forecasts improving from 15 percent to 30 percent better than climatology. Overall, for forecasts up to 48 hours for all seasons across the country, forecasts began about 21 percent better than climatology and finished about 29 percent better.

The prediction of extreme events is probably the public's prime interest in weather forecasting.

To explain this less than stunning performance, forecasters point to the particularly complex nature of precipitation. A raindrop hitting the ground is the end product of a long series of physical processes, errors in any one of which affect the final outcome. And these processes can operate across distances that are so small that meteorologists find it difficult or impossible to keep track of them, even with the fastest computers.

Jerome Charba and William Klein of the NWS in Silver Spring have pointed out (*2*) that improvements in the NWS's computer forecasting model probably played a significant role in the improvements forecasters achieved in the 1970's. But that model had to sum up all the weather in a 190-kilometer square at a single point. Within that square local thunderstorms, rainstorms along fronts, or showers scattered through a storm system could bring localized rain that might easily be missed by the model or a human forecaster.

The tendency of precipitation to be more broadly and uniformly distributed during the winter than during the summer could thus explain the 50 percent higher skill during the cool season than during the warm season. The forecast model improves on the human forecaster more in the longer than in the shorter range, which perhaps explains the greater improvement during the 1970's at the longer ranges.

Knowing whether or not it will rain tomorrow is helpful, but will it be a sprinkle or a torrent? The prediction of extreme events is probably the public's prime interest in weather forecasting and one of the toughest forecasting jobs. In terms of temperature extremes, forecasters have made significant progress. The mean temperature error of forecasts up to 36 hours ahead has fallen from about 2.2°C in 1966 to about 1.9°C in 1981—slow improvement but hardly crucial to everyday life. However, the proportion of errors greater than 5.6°C has dropped steadily during that interval from more than 6 percent to 3.5 percent (*2*). That same rate of decline has held since 1949 at Salt Lake City, the longest record of that sort available. The decline means, for example, that fewer cold fronts are arriving unheralded.

Forecasting extreme precipitation events has also improved in recent decades but far less dramatically. One measure of success in forecasting heavy rain is called the "threat score," a ratio involving correct forecasts and incorrect forecasts—either unforecasted heavy rain or incorrect forecasts of heavy rain—calculated so that the score ranges from 0 to 1. A threat score of 1 is a perfect forecast—forecast and observed heavy rain areas coincided—and a score of 0 means a complete failure—no heavy rain fell where forecasted. In the late 1950's, the threat score was hovering just above 0.

In 1960, a special group was formed to take advantage of the increasingly useful guidance in quantitative precipitation forecasting provided by the early computer models. The threat score rose sharply to about 0.25 for forecasts of 13 millimeters or more of rain during the first 24 hours. Then the score leveled off during the 1960's, according to analyses by Charba and Klein. Only in the 1970's, as computer models became more sophisticated, did it slowly increase to almost 0.35. In the early 1980's, however, the trend leveled off again, according to Harlan Saylor of the NWS's National Meteorological Center.

As is often the case for other aspects of the weather, the more extreme the event, the more difficult the forecasting. For 25 millimeters or more of rain in the first 24 hours, the score has only risen from about 0.17 to 0.19 in 20 years (*3*). Forecasting heavy snow (more than 10 centimeters) has been equally difficult. But, also in accordance with the usual pattern, improving forecasts at the longer ranges has been easier. While scores for forecasts of 13 millimeters of rain during the first 24 hours were leveling off, the scores for the second 24 hours continued upward, Saylor says.

In the case of even more extreme weather, forecasters have found the going to be tough. One bright spot has been an obvious jump in skill in issuing tornado watches, which followed the installation of sophisticated communication and display equipment at the National Severe Storms Forecast Center in Kansas City, Missouri. The new equipment forms the Centralized Storm Information System (*4*). The system allows the forecaster to display and even superimpose visible and infrared satellite images, radar maps, and properties derived from conventional observations made near Earth's surface. All of these help the forecaster recognize areas of potentially severe weather, especially when the observations are available as rapidly as the new system allows. Before the system's installation in early 1982, satellite images arrived 35 minutes after the satellite acquired them. Now the delay is down to 5 minutes.

Unlike the effects of other innovations in forecasting, which are usually small and masked by the effects of other forecast changes and natural variability, those of the new storm information system became obvious immediately. One forecast issued by the center is a tornado watch, which is an alert of potential tornado formation that covers an area of about 65,000 square kilometers. The percentage of watches including at least one tornado had averaged about 43 percent during the 10 years preceding 1982 and had never exceeded 46 percent. In 1982, the first year of the use of the full system, that score jumped to 52 percent. In 1983 it was 53 percent and in 1984, 54 percent. The proportion of all known tornadoes that actually struck within a watch jumped from about 30 percent to more than 40 percent. Another measure of the system's success, says forecast center director Frederick Ostby, was the capture of all reported tornadoes in the deadly 28 March 1984 Carolina outbreak within a watch issued 2.5 hours before the first tornado. That would have been impossible under the earlier system, he says.

Compared to improved skill in issuing tornado watches, an improvement that has been hailed as "dramatic," advances in the forecasting of some other kinds of severe weather have been small, even by

The Next Lap in the Race

American meteorologists are about to close the gap between their skill and that of their European colleagues in forecasting the weather 3 to 10 days ahead. They will do it by dramatically improving the sophistication of their computer forecasting model. But the Europeans have not been dillydallying while the Americans closed in. Their new model, to be implemented on 1 May, will probably put the Europeans ahead again, where they are likely to stay for some time.

The American bid for equivalency in medium-range weather forecasting depends on major improvements in a mathematical model of the atmosphere's behavior around the globe. All such models are crude, representing as they do only a select number of physical processes at widely separated points. The trick is to include as many of the significant processes in as sharp a picture of the atmosphere as the power of one's computer allows. The new model at the National Meteorological Center (NMC) near Washington, D.C., requires three times the computation time of the old model, so it could not be run until the recent installation of NMC's new CYBER 205 supercomputer.

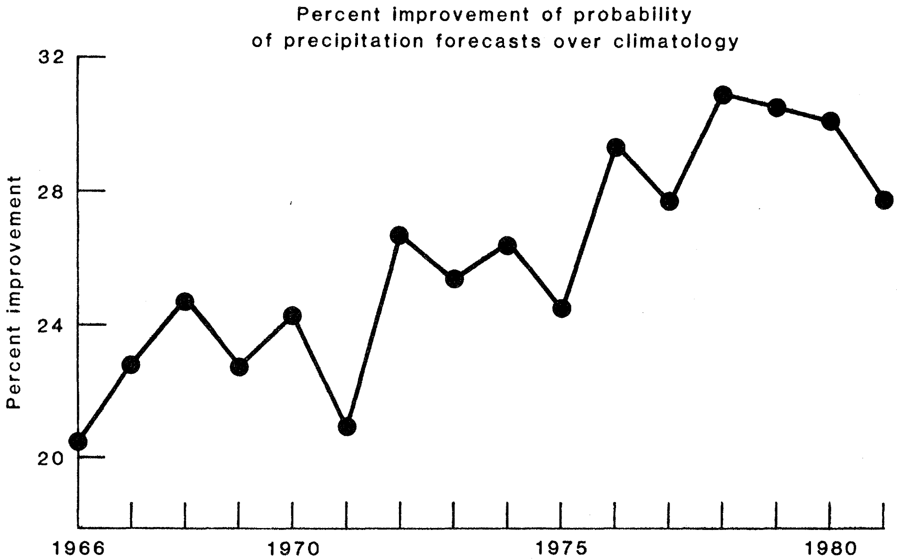
One model improvement is the division of the atmosphere into 18 layers, instead of the previous 12 layers, to sharpen the portrayal of vertical atmospheric features. Another is that the model no longer operates in continuous darkness—the sun shines, the land warms, and land, air, and clouds exchange energy through radiation emission and absorption. The land can also now warm the overlying air. Both changes are based on work by researchers at the Geophysical Fluid Dynamics Laboratory in Princeton, New Jersey. And instead of the model's Rockies being a broad lump no more than 1500 meters high, their profile is more finely detailed and rises as high as 2500 meters.

According to tests of the model in February and March, the improvements in vertical resolution and physical processes extended the usefulness of the model's forecasts from about 5 to about 6 days. In the weather forecasting business, that is a huge step ahead. Still, even if this success holds throughout the year, it just brings the Americans abreast of the Europeans.

In the meantime, researchers at the European Center for Medium-Range Weather Forecasting have developed a new model to run on their recently acquired Cray XMP-22, which can operate three to five times faster than the NMC machine. They are using that computing capacity to improve the model's horizontal resolution from about 170 kilometers, which is a bit better than that of the NMC model, to 100 kilometers, which is almost the equal of the resolution over North America in the NMC's short-range model. The other major change was an improvement in the way the model simulates rising air driven by surface warming and by condensation of water vapor.

Some of the results were predictable. Fronts in the model became sharper, airflow around mountains strengthened, intense storms were more accurately portrayed, and large-scale, convection-driven circulation in the tropics and subtropics was improved. The huge investment of computer power in increased resolution also seems to have paid off in an improved forecasting of overall weather patterns, according to David Burridge of the European Center. The improvement is particularly obvious in the summer over Europe, for reasons that are not so obvious. Overall, the changes will extend useful forecasts by about 12 to 24 hours.

The race will continue, and, as in the past, the prize will go to those who have the biggest, swiftest computer. That is the Europeans for now and probably for some time to come. The NMC operated its previous computer for 9 years. The European Center moved up to the next-generation computer after only 5 years.—**R.A.K.**



A long, slow climb for weather forecasters

Forecasts of the probability of precipitation made by National Weather Service Forecast Offices have improved but relatively slowly. Skill is measured as the percent improvement over forecasts based solely on the frequency of precipitation in the past, that is, climatology. Looking on the bright side, such skill has increased by about 50 percent since 1966. More critically, forecasters are still doing only 30 percent better than climatology. [*Source: National Weather Service*]

the necessarily generous standards of forecasters. Errors in the forecasting of hurricane positions 24 hours ahead, for example, fell from 230 kilometers in 1954 to 198 kilometers in 1980, due in large part to the advent of satellite-determined storm motions (*5*). But in the mid-1970's, as the limits of satellite positioning were approached and other types of observations from remote tropical oceans began to degrade for various reasons, the error decline stopped. There has been no improvement for 10 years, according to Charles Neumann of the National Hurricane Center in Coral Gables, Florida.

Records of skill in forecasting other kinds of severe weather are too short to permit any conclusions about trends, but they do suggest that forecasters have plenty of room for improvement. According to statistics compiled by Kristine Campbell of the NWS in Silver Spring (*6*), only about 10 percent of the 1280 flash flood, high wind, and severe winter weather events of 1981 through 1983 occurred without the issuance of a watch or warning. Of the total warnings issued, about 78 percent proved to be well founded because actual severe weather was reported. But about 42 percent of the warnings were issued when the severe weather was in progress somewhere in the warning area.

Distilling all of this information into a final grade for the NWS and meteorology in general over recent decades of modern forecasting has proved controversial. There is the problem of an incomplete record. Only within the past few years has the NWS attempted to develop a comprehensive national program that would tell how it was doing in forecasting a wide variety of weather, especially severe events. Everyone agrees that too many of the longest records that have been verified against actual weather are limited to only one or a few cities, or, like rain versus no-rain forecasts, are not informative enough to allow solid conclusions.

Where the verification record is sufficient, forecasters and researchers agree that progress has generally been slow. When new technology is first applied to a problem ripe for improvement, the progress can be relatively rapid, especially where performance had been relatively poor. Computer modeling and quantitative precipitation forecasting or satellite imaging and hurricane position forecasting are examples. But even in such cases, the effects of technological breakthroughs are limited by the morass of complexities inherent in the weather. To realize even slow continued progress, forecasters must achieve repeated major improvements in their techniques.

The main driving force behind forecasting progress has been the continual improvement of computer models of the atmosphere. After 40 years of development, these forecasting tools devour tens of thousands of observations a day and demand the fastest computers in the world. Progress in forecasting by computer modeling is easily verified; it has been considerable and steady (*7*). The frustrating part has been that models for the most part turn out rather vague maps of future circulation patterns that must be converted into detailed forecasts of the weather elements that people really care about such as wind, rain, snow, and temperature. Progress in forecasting this true weather has been slower and spottier, especially in the crucial first 24 hours. Modeling does promise further improvements, such as the 36 percent increase in the quantitative precipitation threat score achieved by the Regional Analysis and Forecast System now being implemented by the NWS (*8*).

How good should forecasts be after all the computers, satellites, and research? The answer seems to depend on whether you view the glass as half full or half empty, as one researcher puts it. Those in the front lines of daily weather forecasting see the glass as half full. They acknowledge the slow progress but note the huge obstacles presented by the scale and complexity of the atmosphere. In that light, a 50 percent improvement in skill represents considerable progress even though it still limits the forecaster to doing only 20 to 30 percent better than informed guessing. A few meteorologists, such as Colin Ramage of the University of Hawaii, express surprise that there has not been more improvement and wonder why. Were weather satellites oversold? Are the science of meteorology and the art of forecasting kept too separate for either's good? Most meteorologists conclude that, although such influences may play a role, weather forecasting is simply one of the toughest problems that science and technology will ever face.—**Richard A. Kerr**

Next: Have weather satellites met their promise by this, their 25th anniversary year?

References

1. H. R. Glahn, *Trends in Skill and Accuracy of National Weather Service PoP Forecasts* (NOAA Technical Memorandum NWS TDL 73, Department of Commerce, Washington, D.C., 1984); *Bull. Am. Meterol. Soc.*, in press.
2. P. D. Polger, *National Weather Service Public Forecast Verification Summary April 1978 to March 1982* (NOAA Technical Memorandum NWS FCST 28, Department of Commerce, Washington, D.C., 1983).
3. J. P. Charba and W. H. Klein, *Bull. Am. Meteorol. Soc.* **61**, 1546 (1980).
4. F. P. Ostby, in *Recent Advances in Civil Space Remote Sensing* (Society of Photo-Optical Instrumentation Engineers, Bellingham, Wash., 1984), p. 78.
5. C. J. Neumann, *Bull. Am. Meteorol. Soc.* **62**, 1473 (1981).
6. A. K. Campbell, *National Weather Service 1982 and 1983 Watch/Warning Verification: Flash Flood, Winter Storm, and High Wind* (NOAA Technical Memorandum NWS FCST-29, Department of Commerce, Washington, D.C., in press).
7. R. A. Kerr, *Science* **220**, 39 (1983).
8. ———, *ibid.* **228**, 40 (1985).