# PACIFIC SOUTHWEST Forest and Range Experiment Station

FOREST SERVICE
U.S.DEPARTMENT OF AGRICULTURE
P.O. BOX 245, BERKELEY, CALIFORNIA 94701

# PROBABILITY FIRE WEATHER FORECASTS...
## show promise in 3-year trial

Paul G. Scowcroft

*USDA Forest Service*
*Research Note PSW-201*
1970

*ABSTRACT:* Probability fire weather forecasts were compared with categorical and climatological forecasts in a trial in southern California during the 1965-1967 fire seasons. Equations were developed to express the reliability of forecasts and degree of skill shown by the forecaster. Evaluation of 336 daily reports suggests that probability forecasts were more reliable. For probability forecasting to be used operationally, fire weather forecasters will need help in the necessary computations and some decision criteria for planning.

*OXFORD:* 431.1–U551.509:U519.28.

*RETRIEVAL TERMS:* fire weather; weather forecasting; mathematical expectation.

Historically, wildland fire control agencies have based their daily prevention and presuppression activities on predicted fire danger indexes. These indexes in turn are based on U.S. Weather Bureau forecasts of expected weather conditions. Because the forecasts are not expressed in terms of probability, the user often assumes that the predicted weather condition has a 100 percent chance of occurring. This misconception can lead to costly mistakes.

Weather forecasting is filled with uncertainties. Even with data from space satellites forecasters find it almost impossible to predict exact time of onset or duration of inclement weather. They find it even more difficult to predict adverse fire weather. All weather forecasts are therefore probabilistic–even though the forecast may not mention probabilities. For the sake of clarity, we shall define probability forecasts as "numerical statements of probability of occurrence of a given critical weather event."[1]

If the fire weather forecaster states the percent probability that a certain weather condition will occur, the user should be able to make better decisions if he has some guidelines. For example, suppose a fire organization learns that the chance of moderate fire weather conditions is 90 percent and the chance of extreme conditions due to foehn (dry east) winds is 10 percent. The guidelines may indicate that the organization should prepare for extreme conditions. Had the forecast been categorical–listing only the most likely condition–however, the agency would have prepared for moderate fire weather. And it would have been caught undermanned if the foehn winds had developed. Thus, the chances of averting a large fire can be increased even when an occurrence probability as small as 10 percent is assigned to adverse weather.

Since probability forecasts provide the user with a maximum amount of information and since the public is used to precipitation probability forecasts, we assumed that foresters and firemen would benefit by this type of forecasting. We further assumed that if probability forecasting were feasible, this technique could be used in daily fire control planning by adapting a decision technique known as "mathematical expectation."[2]

In 1965, the Pacific Southwest Forest and Range Experiment Station began a joint study with the U.S. Weather Bureau to see if mathematical expectation could be used to integrate probability fire weather forecasts with daily presuppression manning plans. A total of 336 daily forecasts were issued during the study period of 1965-1967. The forecasts were made by the Weather Bureau's fire weather forecasters in Los Angeles during the fire season (July through October) for one fire danger rating area (660) on the San Bernardino National Forest in southern California.

An evaluation of these probability forecasts yielded an F score 37 percent greater than that for categorical forecasts, and about 7 percent greater than that for climatological forecasts. An F score is a combined relative measure of the validity and sharpness of a probability forecast. In climatological predictions, a fixed set of probabilities–derived from the observed long-term frequency of an event occurrence–is issued each day regardless of present or predicted weather conditions.

## FIRE DANGER RATING

The fire danger rating system used in California combines weather and fuel conditions to arrive at a Fire Load Index for any day. This index ranges from 0 to 100; zero represents the least hazardous fire conditions. Fire Load Index numbers are grouped into nine categories which correspond to nine manpower levels:

| Fire load index group | Manning plan number |
|---|---|
| 0-4 | 1 |
| 5-8 | 2 |
| 9-12 | 3 |
| 13-20 | 4 |
| 21-27 | 5 |
| 28-34 | 6 |
| 35-39 | 7 |
| 40-69 | 8 |
| 70-100 | 9 |

The frequencies–expressed as percentages by using 13 years of data–with which the plans were used in Fire Danger Rating Area 660 are shown in table 1.

Table 1 –*Frequency with which each manning plan actually occurred for Fire Danger Rating Area 660, San Bernardino National Forest, 1955-64*

| Manning plan | Frequency of occurrence |
|---|---|
| | *Pct.* |
| 1 | 5.07 |
| 2 | 7.52 |
| 3 | 13.32 |
| 4 | 29.07 |
| 5 | 17.76 |
| 6 | 12.40 |
| 7 | 4.90 |
| 8 | 9.51 |
| 9 | .45 |

These frequencies make up what we will call the "climatological frequency."

In this study, the fire weather forecaster issued a daily probability forecast for only three of the nine possible manning plans. This was done because his time was limited. The resulting forecast included the most likely plan and, the two close to it.

The sums of the three probabilities ranged from 0.55 to 1.00, with only two cases in the latter category. But the scoring techniques used to evaluate the forecasts[3,4,] depend on a complete probability distribution; i.e., the sum of the probabilities must equal 1. Therefore, in each forecast, the six remaining manning plans were assigned the same probability. This method of distribution seemed to be "maximally noncommittal with regard to the missing" probabilities.[5]

## EQUATIONS

Five equations were used to arrive at nondimensional numbers expressing the reliability of the forecasts and the degree of skill shown by the forecaster. Terms used in the first three equations were:

- $F$ = a relative measure of the validity and sharpness of the probability forecasts
- $N$ = total number of forecasts
- $i$ = a forecast manning plan
- $f$ = forecast probability for each manning plan
- $\phi$ = observed probability (equals 1 if the manning plan occurs, and 0 if it does not)
- $\bar{\phi}$ = average observed probability
- $k$ = probability category; i.e., probability of 0.000 to 0.049 is equivalent to $k$=1; 0.050 to 0.149 is equivalent to $k$=2, etc.

individual years, forecasts are too few to give a trustworthy estimate of reliability. But the value for the total sample is credible.

The improvement over climatological frequency was computed from equations 4 and 5 *(table 3).* The probability forecasts yielded scores—left hand term of equation 4—that were 7.29 percent greater than that for climatological forecasts—the value of C from equation 6. The amount of improvement varied slightly with each fire season, and was largely determined by resolution. Reliability—the first term on the right side of equation 4—is a penalty in equation 4: the poorer it is, the smaller the score. The gain over climatological resolution amounted to 10.23 percent. The penalty from the lack of reliability was 2.94 percent.

Table 3 —*Improvement of probability fire weather forecasts over climatological expectancy as determined from equations 4 and 5, 1965-1967*

| Fire Season | (C - F) Score | Improvement | Resolution[1] component | Improvement | Reliability component | Improvement |
|---|---|---|---|---|---|---|
| | | *Pct.* | | *Pct.* | | *Pct.* |
| 1965 | 0.0094 | 9.40 | 0.0165 | 16.50 | 0.0071 | 7.10 |
| 1966 | .0053 | 5.42 | .0094 | 9.61 | .0041 | 4.19 |
| 1967 | .0075 | 7.63 | .0114 | 11.60 | .0039 | 3.97 |
| All | .0072 | 7.29 | .0101 | 10.23 | .0029 | 2.94 |

[1]Extent to which probability forecasts are distributed toward the probability values of 0 and 1, and relative to climatological expectancy.

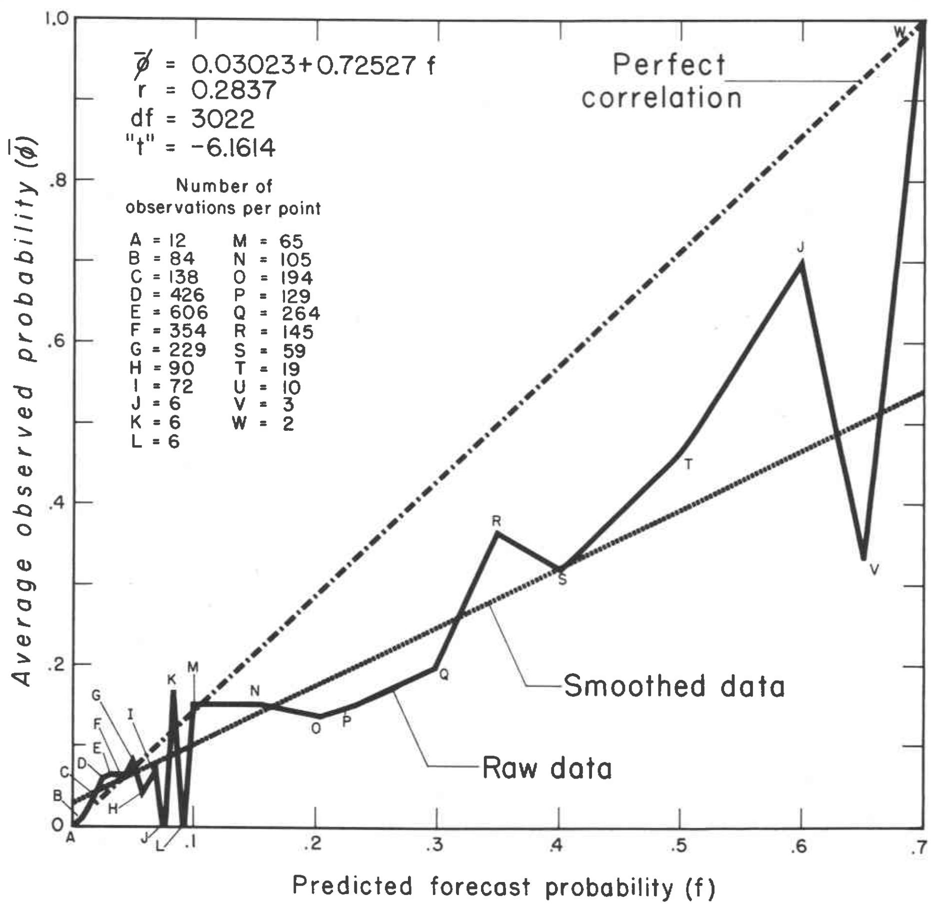


Figure 2—*Forecast probability and average observed probability for the months of July to October, 1965-1967.*

Table 2 –*F scores and their components, reliability and resolution,*[1] *and the percent improvement over categorical forecasts, by fire season and number of forecasts*

| Fire Season | Forecast | F Score[1] | Improvement | Reliability component | Improvement | Resolution[1] component | Improvement |
|---|---|---|---|---|---|---|---|
| | *No.* | | *Pct.* | | *Pct.* | | *Pct.* |
| 1965 | 106 | 0.0905 | 35.5 | 0.0044 | 32.4 | 0.0861 | 3.1 |
| 1966 | 124 | .0933 | 36.5 | .0044 | 34.2 | .0889 | 2.3 |
| 1967 | 106 | .0908 | 39.0 | .0024 | 36.1 | .0884 | 2.9 |
| All | 336 | .0916 | 37.0 | .0025 | 35.1 | .0891 | 1.9 |

[1]Extent to which probability forecasts are distributed toward the probability values of 0 and 1.

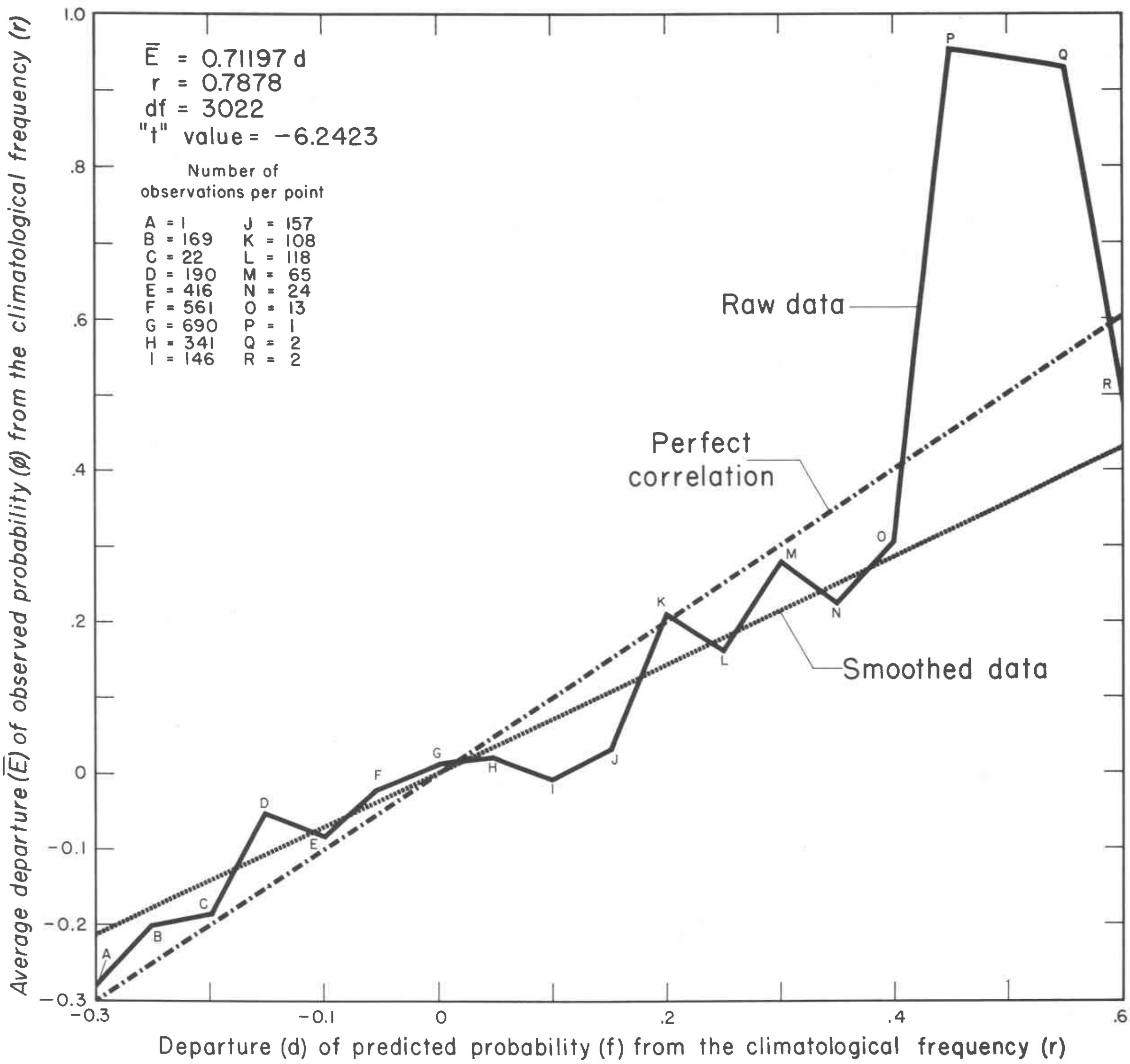


Figure 1 –*Departure of forecast probabilities from climatological expectancy versus average departure of observed frequency of occurrence from climatological expectancy, months July to October, 1965-1967.*

M = number of forecasts in each probability caregory k.

The basic equation[3] is:

$$F = \frac{1}{N} \sum_{i=1}^{N} (f_i - \phi_i)^2$$

Sanders[4] expressed this equation in terms of 11 probability categories (k), which represented probabilities of 0.0, 0.1, 0.2, etc.

$$F = \frac{1}{N} \sum_{k=1}^{11} \sum_{i=1}^{M} (f_k - \phi_{ki})^2$$

Sanders further expanded equation 2 to get equation 3:

$$F = \frac{1}{N} \sum_{k=1}^{11} M_k (f_k - \bar{\phi}_k)^2 + \frac{1}{N} \sum_{k=1}^{11} M_k \bar{\phi}_k (1 - \bar{\phi}_k)^2$$

The first term on the right side of equation 3 represents the *reliability* of the forecasts–extent to which the probability forecasts agreed or disagreed with what actually occurred. The second term is a measure of the forecast *resolution* –the extent to which the probability forecasts were distributed toward the probability values of 0 and 1.

Sanders[4] modified equation 3 to show the amount of improvement of probability forecasts over climatological frequency (r) (climatological frequency was the percent of times each manning plan had actually occurred over the past 10 years). The new equation was expressed in terms of 21 probability departure categories (p), which represented probabilities of –1.0, –0.9, . . ., 0.0, . . ., +0.9, +1.0.

$$C - F = \frac{1}{N} \sum_{p=1}^{21} M_p \bar{E}_p^2 - \sum_{p=1}^{21} M_p (d_p - \bar{E}_p)^2$$

in which

- C = the climatological control score
- r = climatological frequency
- d = departure of predicted probability (f) from the climatological frequency (r); $d = (f - r)$
- p = probability departure category; i.e., a value of d ranging from –0.76 to –0.85 would be represented by category –0.8
- E = observed departure of relative frequency ($\phi$) from the climatological frequency (r); ($E = \phi - r$)
- $\bar{E}$ = average observed departure.

In equation 4, reliability and resolution as measured *relative* to climatological likelihood are expressed by the right and left hand terms, respectively. Expressing the amount of improvement as a percent produced

$$\frac{(C - F)}{C} \times 100 = \frac{\left(\sum_{p=1}^{21} M_p \bar{E}_p^2 - \sum_{p=1}^{21} M_p (d_p - \bar{E}_p)^2\right)}{\sum_{i=1}^{N} E_i^2} \times 100$$

in which

$$C = \frac{1}{N} \sum_{i=1}^{N} E_i^2$$

## COMPUTATION

Using equation 3, we computed the F scores for the sample data. One way of interpreting these scores was to view them relative to the theoretical minimum and maximum F values. The minimum value which corresponds to perfect forecasting would be 0. This value would be achieved when all forecast probabilities are 0 or 1, and every forecast is correct. Such forecasting accuracy now or in the near future is unlikely. The significant minimum value of the F score is, therefore, somewhat greater than zero. The maximum score would occur when all the forecast probabilities were 0 *or* 1, and all forecasts were incorrect. It would be equal to 0.222. The extreme values for reliability are 0 and 0.125. The minimum value for the resolution term is also 0, but the maximum value is 0.097.

By using these theoretical limits, it is possible to draw three conclusions: (1) an F score of zero indicates perfect forecasting; (2) perfect reliability will be shown by a value of zero, and (3) zero indicates maximum resolution.

Further interpretation of these scores was attempted by comparing them with the scores obtained by using present categorical forecasting methods, in which only one plan is indicated *(table 2).* Each categorical forecast was assumed to have a 100 percent probability value associated with it, because the user would probably interpret the forecast that way.

The comparison revealed that the probability forecasts were about 35 percent more reliable than present forecasts. However, the amount of improvement for the resolution was only about 2 percent. Reliability improves with successive years, but is least for the total sample. This tendency suggests that in

The departure ($d_k$) of forecast probability from climatological expectancy was plotted against the average departure ($\bar{E}_k$) of observed frequency of occurrence from the climatological frequency for the sample period *(fig. 1.)*. The ideal relation between the two variables could be represented by a straight line with a slope of 1. Using a least squares regression analysis made it possible to compare the sample with the idealized population. This was done by using a t-test for slope differences. The analysis indicated that slope of the sample population was significantly different from that of the true population. We concluded that the probability forecasts were less than perfect.

But were the forecasters over- or under-confident? Sanders[4] explained that when the magnitude of the departures ($d_k$) of forecast probability exceeds that of the departures ($\bar{E}_k$) of relative frequency of occurrence, then the forecaster is leaning toward overconfidence. This appears to be the case in this study. Sanders has found that this characteristic tends to disappear with time as the forecaster gains experience with the probability format.

Figure 2 illustrates the graphic relationship between forecast probability ($f_k$) and average observed frequency of occurrence $\bar{\phi}_{ki}$. Overconfidence by forecasters is apparent. The sample population differs significantly from the true population. Reliability seems to be good at $f_k$ values less than 0.20, but most of these values were assumed, not predicted.

## CONCLUSIONS

We had hoped that the improvement of probability over climatological forecasts would be large. Such was not the case, but an explanation is possible. Nine fire danger rating classes were assigned probabilities. The classes were of uneven width, some being very narrow, i.e., five index points for manning plan No. 7; and others very wide, i.e., 31 index points for manning plan No. 9. The forecaster easily distinguished the wider classes. But he had more difficulty in identifying the narrower classes in terms of probability; their inclusion served to decrease the probability that could be assigned to each class. The combination of the two effects probably resulted in the small improvement over climatology.

It would be more realistic from the forecaster's viewpoint to start with four or five classes of more-or-less uniform width and then expand as justified by his forecasting skill. This would obviously necessitate a reappraisal of fire prevention and presuppression guidelines. But if monetary gains could be realized and natural resources preserved by the integration of probability forecasts into fire planning, then the effort would be well worthwhile.

In summary, "probability forecasts permit the forecasters to state their best estimates of the probability of occurrence of all classes, and thus provide the fire control organization with the maximum weather information."[6] The probability fire weather forecasts evaluated in this study showed an improvement over present categorical forecasts.

To avoid optimism, I should point out that despite these findings, the issuance and use of probability forecasts very much depends upon (a) the amount of time the forecaster has to make his forecasts, and (b) the existence of a decision criteria for using probabilities in daily fire prevention and presuppression planning. Fire weather forecasters are now pressed for time; they are responsible for large and diverse forecast areas. Should probability forecasts be adopted, they would have to use computers to handle the workload. Some work has been done in developing decision criteria, but to date, nothing promising has resulted. But if these two difficulties can be overcome, I believe wildland fire control agencies would be justified in seriously considering the adoption of probability forecasting.

### NOTES


[1]Root, Halbert E. *The value of weather forecasts.* U.S. Weather Bureau, Region IV, Tech. Note 3, 1 p. 1965.

[2]Miller, D. W., and Starr, M. K. *Executive decisions and operations research.* 446 p. Englewood Cliffs, N.J.: Prentice-Hall. 1960.

[3]Brier, Glen W. *Verification of forecasts expressed in terms of probability.* Mon. Weather Rev. 78(1): 1-3. 1950.

[4]Sanders, Frederick. *On subjective probability forecasts.* J. Appl. Meteorol. 2(2): 191-201. 1963.

[5]Jaynes, E. F. *Information theory and statistical mechanics.* Phys. Rev. 106(4): 620-630. 1957.

[6]Root, Halbert E., U.S. Environmental Science Services Administration. (Personal communication, July 30, 1969.)



**The Author**

**PAUL G. SCOWCROFT** was formerly assigned to fire management systems research, with headquarters at the Station's Forest Fire Laboratory, Riverside, Calif. He is now on the staff of the Station's Institute of Pacific Islands Forestry, Honolulu, Hawaii. He earned B.S. (1964) and M.S. (1966) degrees in forestry at the University of Missouri. He joined the Forest Service in 1965.