A Technique for Evaluating the Effectiveness of Hurricane Modification Experiments

RICHARD W. KNIGHT[1]

National Hurricane and Experimental Meteorology Laboratory, NOAA, Coral Gables, Fla. 33124

GLENN W. BRIER

Department of Atmospheric Sciences, Colorado State University, Fort Collins 80523

(Manuscript received 15 July 1977, in final form 22 October 1977)

ABSTRACT

Plans are underway to attempt to reduce the destructive force of hurricanes by artificially modifying their structure by means of cloud seeding. Since the natural variability of meteorological elements observed in hurricanes is high, the success of the project depends upon establishing a cause and effect relationship between the seeding and the hurricane's response. The small sample of mature hurricanes coupled with rigorous selection criteria make a randomized experiment impractical. Therefore, an evaluation technique based on the concept of randomization in time is developed.

1. Introduction

The seeding hypothesis used in Project STORMFURY (U. S. Department of Commerce, NOAA) is based upon definite cause-and-effect relationships that, in turn, are based upon theoretical and observational studies. Basically, the hypothesis calls for the seeding of clouds at radii greater than those of the eyewall to enhance convection outside of the eyewall region. If convection can be successfully stimulated, part of the low-level inflow that originally maintained the eyewall convection will be directed into convection at a radius greater than that of the eyewall; thus, the transport of angular momentum and water vapor to the old eyewall will be reduced. As the region of major vertical mass transport is shifted to a greater radius, the maximum winds will diminish because of conservation of momentum. The requirements, therefore, to objectively evaluate the results of a hurricane seeding are to 1) observe the entire sequence of changes that occurs in the meteorological elements to test whether this sequence corresponds to the hypothesized chain of events, in space and time, and to 2) determine the statistical significance of this correspondence.

It is well known that the natural time and space variability of meteorological elements measured in hurricanes is very high, a fact established both by land-based observations and those obtained from aircraft. This is true even when the mature storm is in an apparent steady-state condition for a period of a day or more with no obvious diurnal cycles or longer term trends. This, of course, causes serious problems in testing the results of seeding when the classical designs for comparative experiments are used. For example, even if dramatic changes in a particular element were to occur after a seeding, it would take data from many trials to show that, indeed, there must be a seeding-induced change. The process of collecting a sufficient number of treated and control cases might take years, which is an unacceptable delay.

An alternative technique, therefore, was developed for evaluation of the results of Project STORMFURY seeding experiments. The proposed statistical technique is based on a quantitative description of the hypothesized *sequence and timing of events* $A_1, A_2 \ldots A_k$, which are expected to occur during an interval l after a seeding treatment. The events A_i are physically measured quantities, such as the change from water to ice in a cloud, or a 1°C temperature drop in the eye of a storm. It becomes obvious that the likelihood of some complex sequence occurring by chance will be considerably lower than for a single-element event, such as A_1. Therefore, the technique should have the capability of providing a quantitative answer to the question: how well did the hypothetical sequence of events predict what actually happened after the seeding when compared to the variations and events during the rest of the monitoring period? Essentially, the remainder of this monitoring period acts as a control. Although spectral analysis and other classical methods have been used to examine the significant variations in time series, none of the commonly used techniques of experimental

[1] Present affiliation: NOAA, National Climatic Center, Asheville, N. C. 28801.

0021-8952/78/0222-0227$05.00

design appears to be directed at the specific problem attacked here where one or more definable discrete events are hypothesized to occur within a clearly specified interval after the application of the treatment.

This paper describes a technique which has been developed to provide a quantitative appraisal of an experimental result obtained in these circumstances under certain basic assumptions. The first is that the probability of the sequence of events $A_1, A_2, \ldots A_k$ is uniform over the monitoring period of length T or, in other words, that the hurricane is steady state over a period of around a day or more. If a storm has a diurnal cycle or trend in its intensity, then statistical tests would be invalidated unless such factors were taken into account. Although it might be useful to evaluate the effects of seeding a developing or dissipating storm, present seeding hypotheses do not distinguish between storms that may be increasing or decreasing in intensity, but only postulate certain changes to take place after seeding a (presumably) mature storm. Furthermore, it is assumed that the seeding hypothesis specifies events $A_1, A_2, \ldots A_k$ are almost certain to occur in a limited period l after the seeding. It is important that l be much smaller than T, since the power of the statistical test will depend upon the number of intervals of length l which are contained within T. The point is that it is the time intervals within storms which are proposed as the experimental units for evaluation—not the storms. All inferences are based upon what happens during the "treated" unit compared with the remaining "control" units.

In addition, the seeding hypothesis must specify a forecast order of events in time—i.e., $A_1, A_2, \ldots A_k$. However, there is no assurance that the actual events will occur in the same order, or even occur at all! For example, if the seeding hypothesis specifies the event sequence A_1, A_2 and A_3 and the actual occurrence is A_3, A_1, then it is clear that the hypothesis tends to be negated, especially if A_3, A_1 did not occur in the specified interval l after the seeding. An objective measure to test the degree of agreement (hereafter called the B score or B) is described in the following section.

2. General description

The objective is to compute a score or index to measure the association between the hypothesized and observed sequence of events $A_1, A_2, \ldots A_k$ in a monitoring period of length T, with the forecast sequence $F(A_1), F(A_2), \ldots F(A_k)$ restricted to a period of length l when $l \ll T$. The events $A_1, A_2, \ldots A_k$ are those defined by some physical hypothesis to follow a seeding treatment (S) when monitoring is taking place. The forecasts of the events $A_1, A_2, \ldots A_k$ are designated as $F(A_1), F(A_2), \ldots F(A_k)$ and specified by the hypothesis to occur after the treatment S during the interval between t_0 and t_0+l. This forecast time length l is defined by the time span encompassed by the distribu-

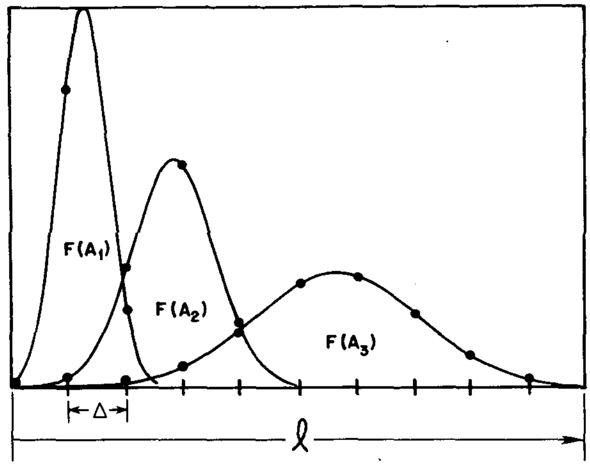

FIG. 1. Three forecast functions making up a forecast period l having a monitoring increment period of length Δ. The increment Δ is arbitrary, but should be small enough to represent the continuity in $F(A_1)$, $F(A_2)$ and $F(A_3)$. [Actual values of $F(A_1)$, $F(A_2)$, $F(A_3)$, used by the computer program, are taken from the curves at points marked by ●.]

tion curves (assumed here to be Gaussian) of the forecast functions truncated at $\pm 3\sigma$; i.e.,

$$l = (\mu_k + 3\sigma_k) - (\mu_1 - 3\sigma_1),$$

where μ_i and σ_i are the means and variances, respectively, of the normal distributions used to represent the forecast functions $F(A_i)$, as shown in Fig. 1. It is unnecessary for the forecast functions to be represented by truncated normal distributions, but for purposes of illustration, the maximum ordinate of the function $F(A_i)$ is the expected time $t(A_i)$ for event A_i to occur so that

$$t(A_i) = t_0 + \mu_i,$$

where t_0 has been specified by hypothesis as some point in time after the beginning of the seeding treatment.

Any appropriate distribution can be used for the forecast functions, and a change can be made if warranted by experience. The standard deviations and maximum ordinates may all be different. The forecast functions are "standardized" or adjusted so that the maximum ordinates sum to unity. This provides for an index I, whose value depends upon the degree of agreement between the forecast time $t(A_1), t(A_2), \ldots t(A_k)$ and observed events at $T(A_1), T(A_2), \ldots T(A_k)$, respectively. The minimum value of I is zero and occurs, for example, if events $A_1, A_2, \ldots A_k$ occur outside the interval t_0+l.

The problem now is to devise a score to determine whether a particular index I_s (which corresponds to the beginning of the sequence of events caused by seeding) is unusual with respect to the general concept of randomization in time. The following procedure is used. The hypothesized means $(\mu_1, \mu_2, \ldots \mu_k)$ and standard deviations $(\sigma_1, \sigma_2, \ldots \sigma_k)$ of the forecast functions,

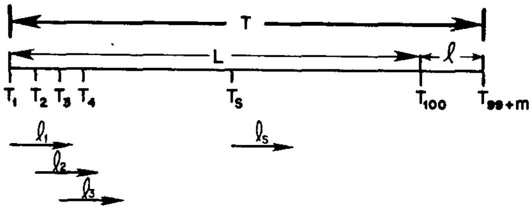

FIG. 2. The total monitoring period T, the forecast period l, and the monitoring increment numbers T_1 to T_{99+m} are depicted. Arrows at bottom of figure show how the forecast period l is moved along the monitoring period T to generate indices I_i.

$F(A_1)$, $F(A_2)$, $\ldots F(A_k)$, respectively, are specified based upon previous experience and/or results of numerical hurricane models. The forecast functions all have zero value outside of the interval l. The next step is to subtract the interval l from the total monitoring period T, leaving a period $T-l=L$ (Fig. 2). Next, the interval L is divided into a large number of parts (say, 99), so that if T_1 is the beginning of the monitoring period, then

$$\begin{aligned} T_2 &= T_1+\Delta \\ T_3 &= T_1+2\Delta \\ &\vdots \\ T_{100} &= T_1+99\Delta \\ &\vdots \\ T &= T_{100}+m\Delta, \end{aligned}$$

where m is an integer such that $m\Delta=l$. (This usually necessitates a slight adjustment in l and T.) This provides for 100 segments of length l, beginning at T_1, T_2, $\ldots T_{100}$ and defined as l_1, l_2, $\ldots l_{100}$, respectively. One of these segments l_s is specified by the seeding hypothesis to contain the events A_i. The object is to determine whether there is something unique about this interval compared with the remaining 99. This is accomplished by next computing the functions $F(A_i)$ at locations t_0, t_1, $\ldots t_m$ on the segment l where t_0 is the beginning of the segment and

$$\begin{aligned} t_1 &= t_0+\Delta \\ t_2 &= t_0+2\Delta \\ &\vdots \\ t_m &= t_0+m\Delta. \end{aligned}$$

In this example, $i=1, 2, 3$, so Table 1 is then prepared so that columns A_1, A_2, A_3 represent the observed events, zero indicates the non-occurrence of the event and unity represents the occurrence. An event may occur more than once during the monitoring period. For example, event A_3 could be a decrease of wind speed greater than 10 kt and could occur at time T_4 and T_8. The next three columns show the forecast sequences positioned so that segment l_1 is matched up with T_1 (Fig. 2). The numbers in these columns are values taken from the normal curve of each respective forecast function at m intervals of Δ time units (marked in Fig. 1). The next three columns are scores derived from the previous columns by matching the observations with the corresponding forecast as indicated. The last column is the sum of $A_1\times F(A_1)$, $A_2\times F(A_2)$, $\ldots A_k$

TABLE 1. Example* of matrix of observed and forecast events arranged for comparison and computation of scores (forecast function values are taken from points on the curves marked in Fig. 1).

Observed events				Forecast sequence				Product values			
Monitor increment number	A_1	A_2	A_3	Forecast increment number	$F(A_1)$	$F(A_2)$	$F(A_3)$	$A_1\times F(A_1)$	$A_2\times F(A_2)$	$A_3\times F(A_3)$	Col 1 +Col 2 +Col 3
1	0	0	0	0	0.00	0.00	0.00	0	0	0	0
2	0	0	0	1	0.42	0.01	0.00	0	0	0	0
3	1	1	0	2	0.13	0.15	0.00	0.13	0.15	0	0.28
4	0	0	1	3	0.00	0.31	0.03	0	0	0.03	0.03
5	0	0	0	4	0.00	0.09	0.08	0	0	0	0
6	0	0	0	5	0.00	0.00	0.14	0	0	0	0
7	0	1	0	6	0.00	0.00	0.15	0	0	0	0
8	0	1	1	7	0.00	0.00	0.11	0	0	0.11	0.11
9	0	0	0	8	0.00	0.00	0.04	0	0	0	0
10	0	0	0	9	0.00	0.00	0.01	0	0	0	0
11	0	0	0								
12	1	0	0								
.											
.											
.											
99	0	0	0								
.											
.											
.											
99+m	0	1	0								
											$I_1=0.42$

* This is the first of 100 tables which are used to compute 100 indices I_i.

$\times F(A_k)$, and the entry in the last line of the table is obtained by summing the last column. This is one of the 100 indices I_i that are to be computed. The values in each line of the forecast sequence in Table 1 can now be displaced one line downward so that t_0 is now matched with T_2, t_1 is matched with T_3, etc. The scores are computed as before, and these steps are continued until t_0 is lined up with T_{100}. This exhausts all the comparisons, numbering 100 in all.

Table 2 is a summary that shows the scores obtained for each of the 100 possible positions. The index I_s represents the total score that corresponds to the time when t_0 is matched up to the real time T_s, as specified by the seeding hypothesis.

Now all 100 values of the index I are compared to the index I_s that was obtained at the time of seeding T_s. The B-score is obtained by counting the number of indices that equal or exceed I_s. The B-score, therefore, takes on values between 1 and 100 and is used to determine whether there is an unusual correspondence between the actual events $A_1, A_2, \ldots A_k$ and the predictions $F(A_1), F(A_2), \ldots F(A_k)$ of the seeding hypothesis. For example, a B score of 50 would suggest no significant relationship, since one would expect by chance that 50% of the scores would be higher (or lower) than I_s when no association existed between the forecast and observed events. Thus, the B-score can be converted into a probability by division by 100 to indicate whether the null hypothesis could be rejected at some chosen confidence level. Now a judgment can be made whether there was significant association between the actual events and those specified by the seeding hypothesis. Of course, this is based upon the assumption that the probability of the sequence of events $A_1, A_2, \ldots A_k$ is uniform over the interval L and that the time of seeding T_s is *randomly* chosen in this interval. Otherwise, the accusation might be made that the experimenter had forecasting ability and introduced a bias by selecting a more favorable time for application of the treatment.

It can be shown that the same general technique can

TABLE 2. Array of indices used for B-Score computation.

Position of t_0 relative to T_i	I_i
1	I_1
2	I_2
3	I_3
4	I_4
5	I_5
.	.
.	.
.	.
s	I_s
.	.
.	.
99	I_{99}
100	I_{100}

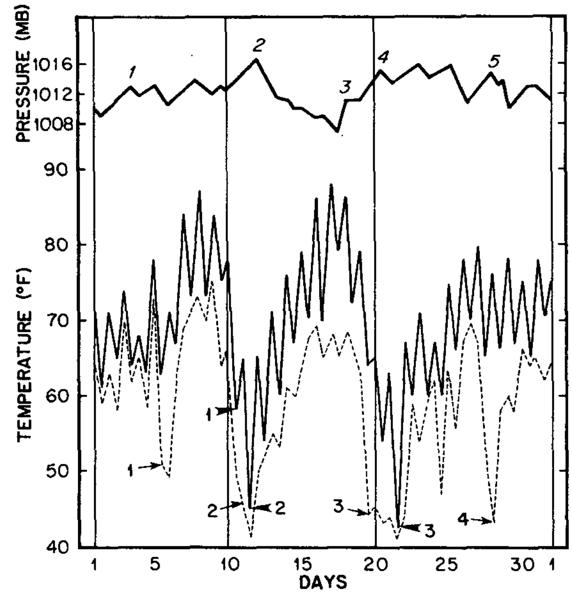

FIG. 3. Pressure, temperature and dew point observed at a hypothetical station during a month. The times when the chosen criteria are satisfied are marked with numbers. Heavy vertical lines indicate frontal passages on the 10th and 20th at 0000 GMT.

be applied to storms treated more than once during a monitoring period. However, some power might be lost unless the monitoring period was increased, since the sensitivity of the test depends upon the ratio of the treated experimental units to the total number of units. This problem could be mitigated (perhaps considerably) if the hypothesis specified events $A_1, A_2, \ldots A_k$ to follow the first treatment and different types of events $B_1, B_2, \ldots B_k$ to follow the second seeding.

3. Example of the B-score computation from a hypothetical case

In this section, the use of the B-score is demonstrated by applying it to a hypothetical weather situation. Two cold fronts have passed a station during a month. The score is used to test whether either of these fronts had significantly affected the observed weather after their passage.

Three elements are chosen for monitoring that can be easily measured and that can provide a good indication that the event (cold frontal passage) changed the weather as hypothesized. The three elements are the temperature, dew point and pressure measured at the station. A one-month record of these is shown in Fig. 3. Table 3 lists the change criteria chosen as indicators of significant weather change and the appropriate forecast functions. In this situation, the forecast functions were chosen by looking at past observations of frontal passages at the station, and then basing the forecast functions on this prior record of observations. Together, the three change-criteria make up a sequence of events that, when observed, indicates the occurrence of the

TABLE 3. Change criteria and forecast function parameters used with cold frontal passages at a hypothetical station.

Measured element	Criterion	Times of occurrence after a significant cold frontal passage (hours) Mean	Standard deviation
A Temperature	Fall of 15°F	9	3
B Dew point	Fall of 20°F	20	5
C Pressure	Rise of 4 mb	40	10

hypothesized effect, or, in this case, a significant change in the weather caused by a cold frontal passage.

After choosing appropriate forecast functions, the total observation period is divided into two monitoring periods so that the cold frontal passages occurring on the 10th and 20th of the month can be dealt with independently. We do this so that the test of significance of one frontal passage will not be in any way obscured by the effects of another. The next step is to compute the B score parameters—that is, the values for the monitoring period T, the forecast function period l, the monitoring interval Δ and the number of increments m (Table 4). As noted in Section 2, slight adjustments usually have to be made in T and l so that m and Δ are integers. The number of integers computed for comparison purposes does not have to be 100; adjustments can be made if the data require it. There are two reasons for making this adjustment. First, if one or more of the forecast functions are relatively "sharp," an adjustment may be made to obtain adequate resolution of the particular function. Fig. 1 can be used to illustrate this point. In the figure, the curve $F(A_1)$ is relatively sharp. Since $m\Delta=l$, as m becomes smaller Δ becomes larger, which could result in $F(A_1)$ being missed, or being poorly represented in the computer program. Another reason for changing Δ may be convenience—that is, to make it conform to the measuring increment of the observational data. From the definitions of the B-score parameters, the following equation can be derived that shows the relationship between the m and the number of indices:

$$m=\frac{1}{(T/l)-1}\times[(\text{number of indices})-1].$$

TABLE 4. B score parameters for two hypothetical experiments.

	Experiment 1 (h)	2 (h)
Original monitoring period	360	384
Original forecast function period	70	70
Adjusted monitoring period (T)	366	366
Adjusted forecast function period (l)	69	69
Monitoring interval (Δ)	3	3
Number of increments (m)	23	23

Since T and l are constants (both may be adjusted slightly so that m is an integer) for a given experiment, m is proportional to the number of comparisons made.

For the first experiment, the original monitoring period was 15 days or 360 h. This was adjusted to 366 h or 15.3 days. The forecast function period was adjusted from 70 to 69 h. Similar adjustments were made to the second experiment (Table 4). Next, the observed data are checked for changes that meet the chosen criteria (Fig. 3). These times are then converted into a corresponding increment number ranging from 1 to $99+m$. (For the general case, increment numbers range from 1 to $A+m$ where A is some reasonably large number—usually $\geqslant 25$.) Similarly, the times of the cold frontal passages are converted, and this information is loaded into the computer program, which computes the B-score for each time that a frontal passage occurred. The probabilities obtained are then used to accept or reject the null hypothesis.

In this case, the null hypothesis H_0 states that the cold frontal passages at the station produced no significant change in the subsequent observed weather. Thus, any observed changes could have been well within the bounds of the natural variations that are observed when there are no frontal passages. The B-score probabilities are used to reject H_0 at a chosen level of significance.

The observational data occurring on the 10th show excellent correspondence with the hypothesized sequence of events that is expected to occur after a significant cold front (Fig. 3). The temperature and dew point dropped rapidly through the chosen criteria, and the pressure rose slowly through its threshold value as required by the hypothesized sequence of events. On the other hand, the events occurring after the cold front passed on the 20th show poorer correspondence with the hypothesis. The pressure rose simultaneously with the frontal passage. The temperature dropped as required, but was considerably late, and the dew point criterion was not met at all. The B-score probabilities obtained for the two cold fronts were 0.03 and 0.08, respectively. Therefore, in the case of the cold front on the 10th, H_0 can be rejected at the 3% level, which is strong evidence that indeed there was a significant change in the weather, as we have defined it, after the frontal passage. In the second case, an examination of the raw data appears to indicate a frontal effect. However, based on the data observed during the rest of the month, there is one chance in 12 that the observed sequence of events could have occurred with no frontal passage. Therefore, in the second case H_0 would probably be accepted.

The above example demonstrates the ability of the technique to assess statistical significance of the relationship between forecast and observed phenomena that have a large degree of natural variability. Had the forecast functions been used separately, their power to reject the null hypothesis would have been minimal because each change-criterion was met one or more times when there was no frontal passage. However, when combined into a *sequence of events* the method becomes a powerful tool for evaluating hypothesized cause-and-effect relationships.

4. Summary and conclusions

If hurricanes can be successfully modified, the resultant changes are expected to be similar in magnitude to naturally occurring changes. This makes it somewhat difficult to obtain convincing evidence of seeding-induced changes with standard statistical methods. Therefore, the evaluation of hurricane modification experiments has relied and continues to rely on cause-and-effect relationships. The technique described in this paper is based upon the sequence and timing of more than one event and should provide a powerful statistical analysis tool which will greatly supplement the physical analysis approach. It gives a quantitative assessment as to whether a specified series of events occurred as hypothesized.

Currently, there is little intuition about what the forecast functions should be. However, after continuous long-term monitoring ($\sim$30 h) of specified parameters is accomplished in a few more seeded and unseeded storms, we should have a much better idea about which of the meteorological elements provide evidence of seeding-induced change and what their respective forecast functions should be. With this information, the B score technique should prove to be an excellent means of attaching statistical significance to the results of hurricane seeding experiments.