A Critique of the Superposed Epoch Analysis Method: Its Application to Solar–Weather Relations

MARION W. HAURWITZ AND GLENN W. BRIER

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

(Manuscript received 27 October 1980, in final form 7 July 1981)

ABSTRACT

Superposed epoch analyses, based on solar sector boundary crossings as key times and the Vorticity Area Index as the response variable, are tested for significance using both parametric and randomization techniques. We conclude from a comparison of these techniques that the randomization procedure leads to markedly different results from those obtained from parametric tests. In particular, the results are strongly affected by the modest skewness of the Vorticity Area Index distribution.

1. Introduction

Solar–weather relationships that have been claimed are usually viewed with skepticism since there are no accepted hypotheses on how solar influences could penetrate into the lower stratosphere or troposphere. As a result, the evidence for such relationships has depended upon statistical associations between solar events and weather phenomena. The validity of the methods used and the conclusions reached have often been questioned on purely statistical grounds. An important technique that has been used in many of these investigations is the superposed epoch method. When properly used, this method can be a valuable investigative tool, although it is relatively unknown in the mathematical-statistical literature and to theoreticians specializing in time series.

This paper presents a critique of the superposed epoch method, pointing out some of the pitfalls to avoid and offering some guidelines that should help strengthen the credibility of such analyses. Particular emphasis is placed on randomization tests in which superposed epoch "key" times are selected by Monte Carlo techniques to simulate conditions consistent with the null hypothesis. In the experiment described here, results obtained from 1000 such "nonsense" key time samples are compared with those based on a sample of "real" key times.

As a vehicle for discussing the analysis method and related data considerations, we have selected solar sector boundary crossings as key times and superposed about these times the Vorticity Area Index as our weather-related index.

2. The superposed epoch and data considerations

The superposed epoch is a row-column array in which the "response" index values filling any row are data pertaining to a single key event. Thus the number of rows is the sample size of such events. The columns line up the index values in fixed time relation to the key times; column averages comprise the "superposed epoch analysis." By this averaging method, any fluctuations in the response index that are locked in time relative to the key time column are preserved in the average, whereas fluctuations shifting in time from row to row are averaged out.

Before constructing a superposed epoch and analyzing it, consideration should be given to the two types of data that enter such an analysis: the sample of key times and the response index which is superposed about these times. Pertinent to both are considerations of after-the-fact selectivity. In the absence of theoretical guidance or of results from previous analyses of independent data, a typical approach is to search among available subsamples of interesting key-time events and among different indices, going through an exploratory superposed epoch analysis and, perhaps, a significance test of each likely combination. After a variety of such analyses has been conducted, the strongest from among them is selected and, based on this, a relationship is proposed and an *ex post facto* definition given of key event characteristics and response index. Also, a significance level for the selected result is often quoted. Such exploratory analyses are often necessary and desirable, but use of their results is only appropriate for defining a

0027-0644/81/102074-06$05.50

© 1981 American Meteorological Society

Unauthenticated | Downloaded 07/22/24 01:09 AM UTC

specific hypothesis to be tested on independent data, and for defining the data characteristics; any significance level ascribed to the selected exploratory result itself must be of dubious validity.

Thus when confronted with a non-familiar superposed epoch result, the questions should be asked: To what degree have the data been culled and hypotheses adjusted *a posteriori* to maximize the strength of an apparent relation? Has a predefined hypothesis been tested on at least one independent sample? Then, since naturally occurring sequential key-event samples are unlikely to have similar patterns of autocorrelation, a further consideration should be that of experiment replication. Only if several independent samples are available and yield consistent results, can one assume that the correlation patterns are not crucially affecting the results; only then can one reliably predict the behavior of other such naturally occurring samples.

Although there may be little theoretical basis for anticipating a relationship between a given type of key event and a given type of index, the characteristics of the two types of data should be considered. Is the key event one of general physical interest or importance, for example, in terms of energy considerations or other known relations involving it? Is the response index well-constructed to measure the physical quantity it is designed to measure? Is that quality itself one of general interest or is it more nearly an *ad hoc* construction?

Other important questions involving the weather-related response index are: What is the degree of serial correlation in the index? Are there abrupt inhomogeneities in the data? Are there long-term trends or seasonal variations? Are there other periodicities in the time series? Is the index approximately normally distributed? What would be the effect of departure from normality?

3. Problems with testing superposed epoch results: Parametric techniques

After post-selectivity or related multiplicity problems have been resolved, there remains a variety of pitfalls associated with the application of parametric testing methods to superposed epoch analyses. (These methods involve the use of tabled probability values derived for distributions with known parameters.)

If the response index is not approximately normal, the typically used parametric tests are not valid. Examples are different forms of the Student's t-test and various "error bar" tests, conducted on column subsample means. Other deficiencies than those related to non-normality also apply to such tests, and will be discussed.

If there is serial correlation in the response index, estimates of the number of degrees of freedom, used for example in a two-sample t-test involving the superposed epoch background, may be too large and will give overly optimistic significance levels. Also, in the case of the two-sample t-test, the pooled estimate of the variance, based on the assumption of equal population variances, will be suspect. Error bar tests not only involve the normality assumption but also, since they are based on standard deviations from individual column subsample means, provide no quantitative determination of the significance of intercolumn differences.

In the discussion to follow we will repeatedly refer to key events, response events and the event interval. The first refers to the sample of key times, in our case, the sample of solar sector boundary crossings (SSB's). The second refers to individual features, peaks or dips, in the superposed response index, the Vorticity Area Index (VAI). The event interval, also involving the response index, is defined in the following section.

A serious problem, intractable by parametric testing techniques, involves the event interval in the superposed epoch. This interval is the group of contiguous columns in the superposed epoch within which the investigator expects a predefined type of response event to occur. Thus this interval should be defined before the analysis is made. Often in practice, the superposed epoch is constructed and the results are scanned for the most outstanding feature. This feature is then taken to be the response event and is tested as if its exact location had been predefined. Actually, the feature could have occurred anywhere within a "reasonable" time interval relative to key time and it would have been post-selected as the response to be tested. Thus, the probability for chance occurrence of the response is dependent on the width of the event interval and on the degree of serial correlation in the response index. If such correlation were negligible, the binomial theorem could be invoked to determine the number of independent chances for the feature to occur within the event interval; however, given the typically strong autocorrelation in geophysical time series, use of the binomial theorem is not appropriate.

4. Testing of superposed epoch results: Randomization techniques

In randomization testing of results, effects due to population characteristics of the response index do not enter, since these affect the "real" superposed epoch and superposed epochs based on randomly selected key times in the same way; all that is of concern is the relative occurrence frequency in the randomized superposed epochs of results as good as, or better than, the real result. This relative frequency is then the estimated probability defining the

Unauthenticated | Downloaded 07/22/24 01:09 AM UTC

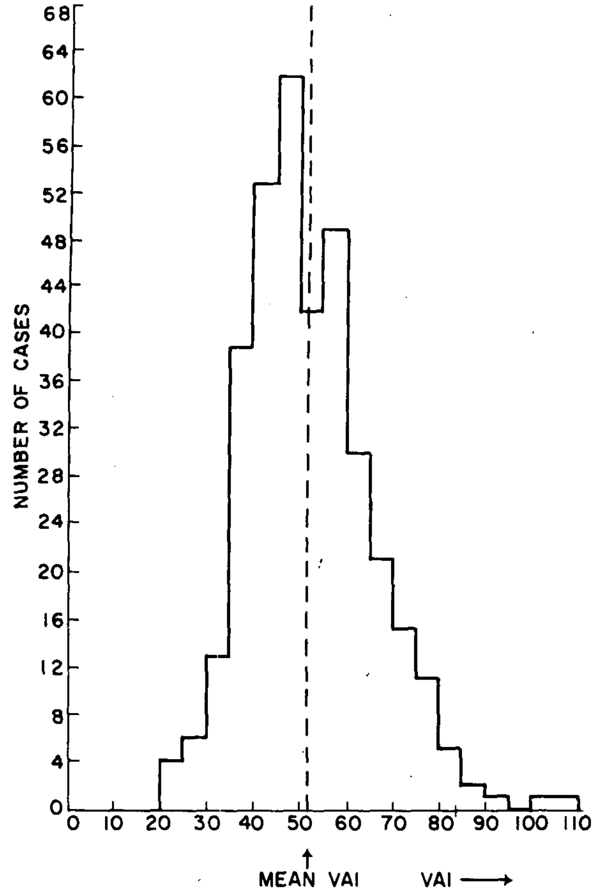

FIG. 1. Frequency distribution of VAI values. Random selection of values from the 1964–70 interval.

significance level of the real result. Thus randomization accomplishes the following: 1) eliminates the normality requirement on the index variable (or any requirement on distribution characteristics); 2) eliminates the equal-variance requirement, described for the two-sample *t*-test; and 3) eliminates the problem with serial correlation in the response index. Further, randomization provides a method for handling problems associated with the event interval, described above. In the present randomization study, analyses involving a variety of event interval widths are conducted, and the consequences are discussed and graphically displayed.

5. Rationale for present work

The germinal studies of Wilcox *et al.* (1974, 1976), in which they show a strong decrease in mean Vorticity Area Index (VAI) one day after passage past the earth of solar sector boundaries (SSB's) have been followed by a number of analyses by other workers who variously accept or reject the reality of the effect (Hines and Halevy, 1977; Williams and Gerety, 1978; Shapiro, 1979). The original Wilcox result is the most dramatic-appearing relationship thus far obtained between a solar activity variable and a neutral atmosphere variable, and as such recommends itself as an interesting and controversial example on which to base a critique of parametric and Monte Carlo techniques as applied to superposed epochs.

Here we will be concerned with the study of the same SSB sample as was used in the original 1974 Wilcox *et al.* paper, and we will thus reanalyze their original effect. Our intent is not to add one more study to the growing list of those designed to confirm or refute the reality of the SSB-VAI relation. In particular, since the 1974 work was of an exploratory nature and involved post-selectivity of hypotheses and of data (for discussions, see Hines and Halevy, 1977; Shapiro, 1979), no valid estimate of significance, based on either parametric or randomization tests, could be given. We present such tests strictly for the purpose of comparing the two methods; the probability values we obtain are not to be interpreted as valid significance levels for any SSB-VAI relation.

6. Present analysis

As a first step in the present analysis, we obtain the frequency distribution of the response index (VAI) by selecting values randomly from the 1964–70 interval that was used by Wilcox *et al.*, (1974). Fig. 1 strongly suggests that the distribution is non-normal, with skewing toward a high-value tail. Although the degree of skewness is modest, it appears to strongly affect the randomized tests of significance, and demonstrates the inappropriateness of parametric tests. In the randomized superposed epochs, peaks and dips test quite differently, as will be shown.

The VAI values in the above distribution, and in all of the analysis to follow, have been preprocessed by forming 3 half-day running means; the central half-day is weighted by 0.5 and the preceding and following half-days entering the mean are weighted by 0.25. These means were formed to balance contributions from 0000 and 1200 GMT VAI values, since these are not determined in a consistent manner (Wilcox *et al.*, 1974). Any gaps in the data were filled with interpolated values, if such gaps were neither too wide nor too closely spaced; interpolated values also always contained a balance of 0000 and 1200 GMT values. [The sample of 54 SSB's used by Wilcox *et al.* (1974) is reduced in our analysis to 51 cases, since three cases involved gaps in the response index which could not be filled with interpolated values.]

The random selection of key times in our analysis

Unauthenticated | Downloaded 07/22/24 01:09 AM UTC

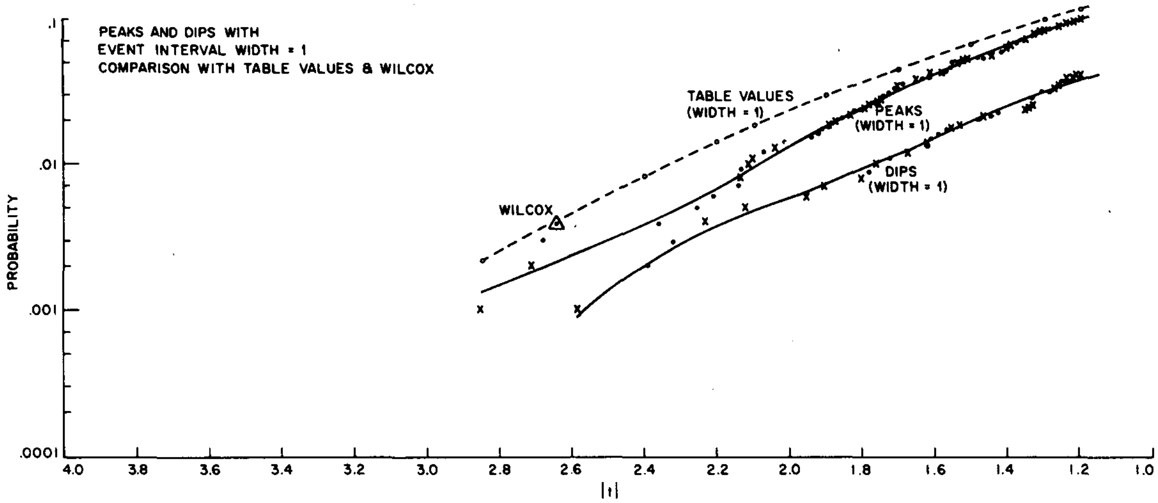

FIG. 2. Probability values versus magnitudes of Student's t. The dashed curve gives tabled probabilities based on the normality assumption, degrees of freedom = $N_E + N_B - 2$, and implied event interval width = 1 unit. The triangle point on this curve is for the Wilcox *et al.* (1974) result. The solid curves are for randomization runs (dots and crosses refer to two different runs); curves for superposed epoch peaks and dips are drawn separately; the event interval width = 1 unit.

is made in a manner consistent with restrictions placed on event selection by Wilcox *et al.* (1974). We thus restrict ourselves to the five months November, December, January, February and March of the years 1964–70, and require that no two randomly selected events be closer together than four days.

The superposed epochs in our analysis, both the "real" case (that obtained using 51 of the original 54 SSB's) and the 1000 randomized cases (each obtained using 51 randomly selected key times under the restrictions described above), are constructed by forming each superposed epoch final column mean from the average of two adjacent columns; such two-column averaging is done to introduce a time-smearing which is consistent with that typically found in average key-time-response event time relations.

We use as our test statistic the t value for the difference between the selected response-*E*vent mean (i.e., a single "final" column value) (the E-mean) and the superposed epoch *B*ackground mean (the B-mean). In the real superposed epoch case, the corresponding significance level, or probability value, is read from a standard table for Student's t probabilities; it should be reemphasized

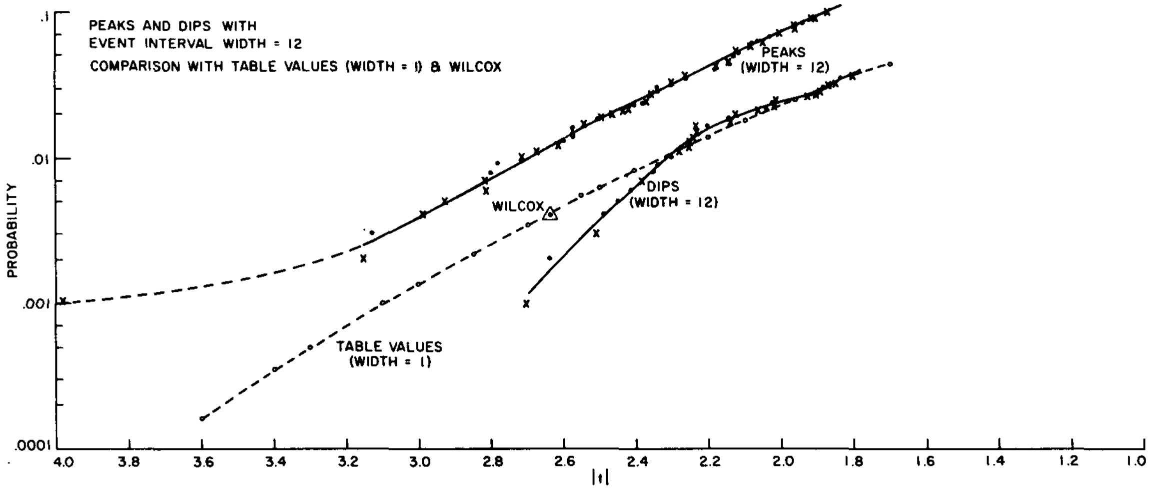

FIG. 3. As in Fig. 2, but with event interval width = 12 units for the solid (randomization runs) curves. The dashed curve (tabled values) is for implied event interval width = 1 unit, as in Fig. 2.

Unauthenticated | Downloaded 07/22/24 01:09 AM UTC

that this tabled value is based on the normality assumption, with event interval width = 1; i.e., the exact position of the response event is assumed to be defined.

For both real and randomized superposed epochs, the background or B-mean is obtained from six final column means, scattered over the superposed epoch, outside the maximum-width event interval. The scattering is chosen so that autocorrelation between the column values is negligible; thus, for the parametric significance test of t, a realistic number of degrees of freedom can be estimated. The t value calculated for all superposed epochs, real and randomized, is

$$t = [(E\text{-mean}) - (B\text{-mean})]/s_D,$$

where

$$s_D^2 \approx (SSD_E + SSD_B)/N_E N_B,$$

with

$N_E = 51$ = sample size of the response event, made up of a single "final column,"

$N_B = 6 \times 51$ = sample size of the background, made up of six scattered "final columns," and

SSD_E and SSD_B, the sums of the respective squared deviations. The number of degrees of freedom is $(N_E + N_B - 2)$.

For each randomized superposed epoch, a variety of event intervals, from 1 to 12 units wide, is used, and a t value is calculated for each final column in each selected event interval. Each such interval is then scanned for the largest and smallest t values therein, and these, then, represent the peak and dip for that interval for that superposed epoch.

For our real superposed epoch, that corresponding to the Wilcox *et al.* (1974) case, the large dip representing their SSB-VAI relation has a value $t = 2.645$, with a corresponding tabled (parametric test) probability value of 0.0043. Figs. 2, 3 and 4 show this value within a triangle labeled Wilcox; the dashed line, of which this point is part, shows the standard tabled t value versus probability relation, under the normality assumption, for our given degrees of freedom; the effective event interval width = 1, as shown.

The solid curves in these figures and in Fig. 5 are obtained from the randomized superposed epochs. Two computer runs enter the analysis, each involving 500 superposed epochs; each of these is constructed by superposing VAI values about 51 randomly selected key times, obtained from the 1964–70 interval. The crosses and dots defining these solid curves identify the two different randomization runs. Each probability value associated with a t value is calculated as the relative frequency with which t values as good as or better than the designated value occur among the 1000 random superposed epochs.

For the table values and dips curves in the figures, the t values along the abscissa are negative; t values corresponding to peaks are positive.

Figure 2 shows the results for the selected event interval width = 1. The dips curve in this figure should be compared with the table values curve and, in particular, with the Wilcox point. We see that the parametric test probabilities (the table values) are substantially more conservative (larger) than those obtained from the randomization runs. Thus, for VAI superposed epochs, any expectation that the parametric t-test, for t values as we have calculated them, would give overly optimistic results is not borne out.

The solid line randomization-run curves show that considerations of event interval width and distribution skewness enter strongly. Examination of Figs.

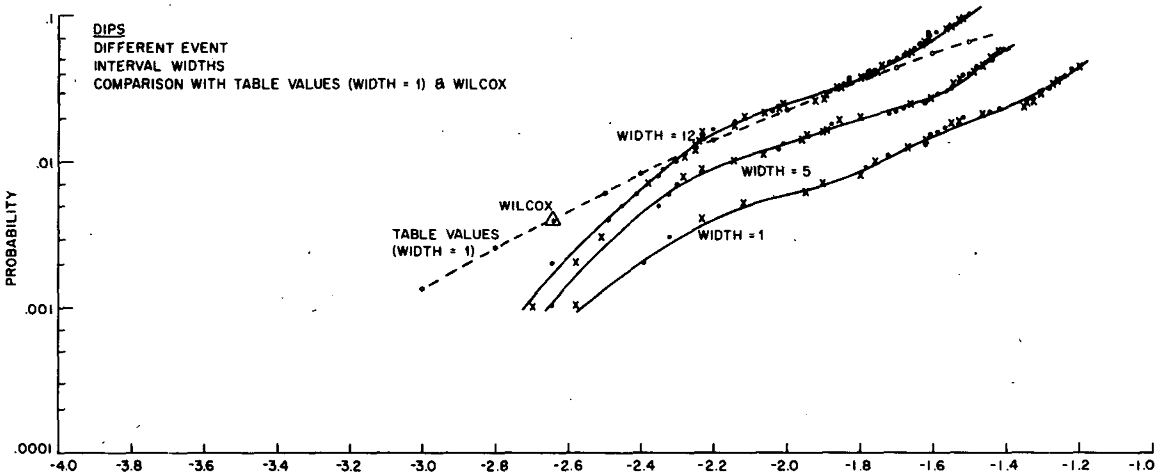

FIG. 4. The dashed (tabled values) curve is the same as for Figs. 2 and 3. The solid (randomization runs) curves are for superposed epoch dips with different event interval widths.

Unauthenticated | Downloaded 07/22/24 01:09 AM UTC

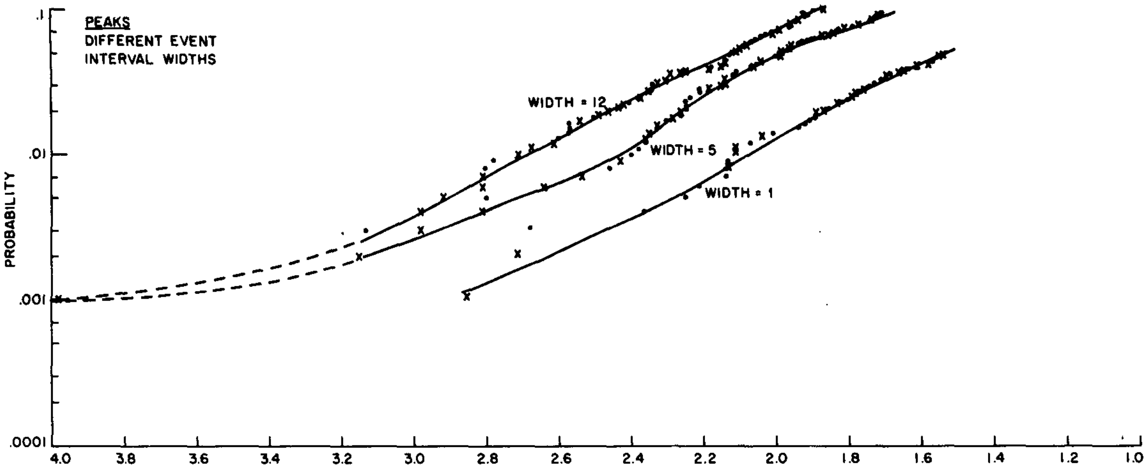

FIG. 5. As in Fig. 4, but with solid (randomization runs) curves for superposed epoch peaks.

2–5 indicates that, for a given event interval width, the peaks curve lies well above the dips curve, at least over the range of t values considered. Qualitatively, the randomization study indicates that it is easier to get a peak of a given t value than to get a dip of its negative value. This result is attributed to the skewness of the VAI distribution toward a high-value tail. Thus it is essential in assigning a correct significance level to a result that the appropriate type of randomization curve be referred to.

Examination of these figures also indicates the strong effect that event interval width has upon a probability value, or significance level. This effect is seen particularly clearly in Figs. 4 and 5, in which different widths are shown on the same plot. We see that by not taking into account the event interval (in effect, by using width = 1) we dramatically overestimate the strength of any apparent relation; this is particularly true in the case of peaks.

7. Conclusions

Parametric tests of Student's t values are compared with randomization tests of the same statistic. We find that, for VAI superposed epochs, parametric significance levels of t values are conservative relative to those obtained by Monte Carlo methods.

The skewness of the VAI distribution, although seemingly quite modest, leads to a substantially larger probability for a given t when associated with a peak than for the same amplitude t value associated with a dip.

The use of an implied event interval width = 1, as occurs in parametric testing, fails to take into account the fact that more realistic interval widths, typically of five units or more, yield substantially poorer significance levels.

The above points emphasize the importance of using randomization testing procedures.

A long introductory discussion considers the impact that post-selectivity has in falsely enhancing the apparent strength of a superposed epoch relation.

Acknowledgments. We thank R. H. Olson for his generosity in providing the tape of VAI values, and for his discussions concerning the characteristics of these data.

This research was supported in part by National Science Foundation Grant ATM 7911465.

REFERENCES

Hines, C. O., and I. Halevy, 1977: On the reality and nature of a certain sun-weather correlation. *J. Atmos. Sci.*, **34,** 382–404.

Shapiro, R., 1979: An examination of certain proposed sun-weather connections. *J. Atmos. Sci.*, **36,** 1105–1116.

Wilcox, J. N., P. H. Scherrer, L. Svalgaard, W. O. Roberts, R. H. Olson, and R. L. Jenne, 1974: Influence of solar magnetic sector structure on terrestrial atmospheric vorticity. *J. Atmos. Sci.*, **31,** 581–588.

——, L. Svalgaard, and P. H. Scherrer, 1976: On the reality of a sun-weather effect. *J. Atmos. Sci.*, **33,** 1113–1116.

Williams, R. G., and E. J. Gerety, 1978: Does the troposphere respond to day-to-day changes in solar magnetic field? *Nature*, **275,** 200–201.

Unauthenticated | Downloaded 07/22/24 01:09 AM UTC