Statistical Procedure to Test Significance in the Analysis of Cosmic Ray Data by Superposed Epoch Method–I

Badruddin and Y.P. Singh
Department of Physics, Aligarh Muslim University, Aligarh-202 002, India

Abstract

In this paper we describe a technique, based on t-test, with an application to test the level of significance of results obtained on the basis of superposed epoch (Chree) analysis. Since most of the data acquired in the heliosphere show a solar cycle variation, a suitable method is also described for the data transformation (removal of solar cycle variation) before it is subjected to test a 'genuine' effect. A comparison of the results of significance test, before and after the data transformation, is also presented.

1. Introduction

Superposed epoch (Chree) analysis is often used to demonstrate an effect or a periodicity. This method of analysis [1] was originally applied for studying the time variation of geophysical data. In its first application, Chree [1] reported a 27-day periodicity (recurrence tendency) in geomagnetic data. Since then this method of analysis is being used in several disciplines either for testing the relationship between two diverse phenomena or to search for periodicities in the data. In addition to cosmic ray physics, various fields of research in which this method of analysis is often used include solar, magnetospheric, heliospheric, ionospheric and atmospheric physics as well as astrophysics, space biology and meteorology/climatology etc. Although a powerful method, an appropriate procedure to test the level of significance (statistical reality) of the obtained results is still lacking. This is highly desirable as, in the absence of a suitable test, a spurious/undesirable signal may appear as a 'genuine' effect.

2. Method

Before, evaluating the statistical significance of the 'effect of interest'(e.g. Forbush decrease), the data is transferred (corrected) to remove the effects other than one under study. In our data (cosmic ray counts) the other effect, that has to be removed, is due to solar cycle variation. For this purpose, we have adopted a procedure for data transformation without affecting the sample average and the 'genuine' effect (signal). This has been done by transforming (shifting) the data

pp. 3639–3642

sets of each epoch to the level of sample mean. In this way we are able to remove the solar cycle effect without altering the sample mean and 'genuine' signal, in each epoch. Then a formal test procedure is applied [2-5].

If X_1, X_2, X_3,..., X_n be the random sample with parameter μ and σ^2, and a subsample of k values. Then for a sample of size n, whose mean value is $\bar{X}$ and for a subsample of k values, the following expressions give the upper and lower confidence limit for $\bar{X}$ and $\bar{X}_k$, both with confidence level 100 $(1-\alpha)$%, with $(n-2)$ df,

$$\bar{X} \pm t_{\alpha/2,n-2} S/\sqrt{n-1} \tag{1}$$

$$\bar{X}_k \pm t_{\alpha/2,n-2} S \left\{ \frac{k(n-2)}{n-k-k\left(\frac{\bar{X}_k-\bar{X}}{S}\right)^2} \right\}^{-1/2} \tag{2}$$

After superposed epoch analysis, the significance of deviation from mean is tested using (1) and (2). These expressions are different from that of [5] in the sense that in above equations S is sample standard deviation instead of sample variance.

3. Results

To illustrate the application of this procedure, we have considered, for the superposed epoch analysis, the daily average cosmic ray data before and after arrival of interplanetary shocks producing Forbush decreases. In order to demonstrate the effect of solar cycle variation on the statistical analysis, we have selected Forbush decreases occurring during different phases of solar cycle. Three typical events along with the average superposed epoch result are shown in Fig. 1 (upper left panel); different levels (before zero day) in three events are because of the solar cycle effect. These events, after the removal of solar cycle effects, are shown in lower left panel. Fig. 2 shows the mean superposed intensity for all the shocks considered without correction and after correction. It should be noted that the two profiles are exactly similar.

4. Discussion

In Fig. 2 (upper right panel) we have shown the statistical results (95% confidence interval of mean intensity and the decrease in intensity due to shocks) without correcting the data for solar cycle effect. Although key days in the superposed epoch analysis correspond to shocks, all producing clear Forbush decreases, the test results appear to show that the decrease observed on +1 day cannot be considered to be statistically significant.

However, as shown in Fig. 2 (lower right panel), when the statistical test was done after correcting data as described above, the observed decrease is statistically significant. The 95% confidence interval for the minimum intensity is

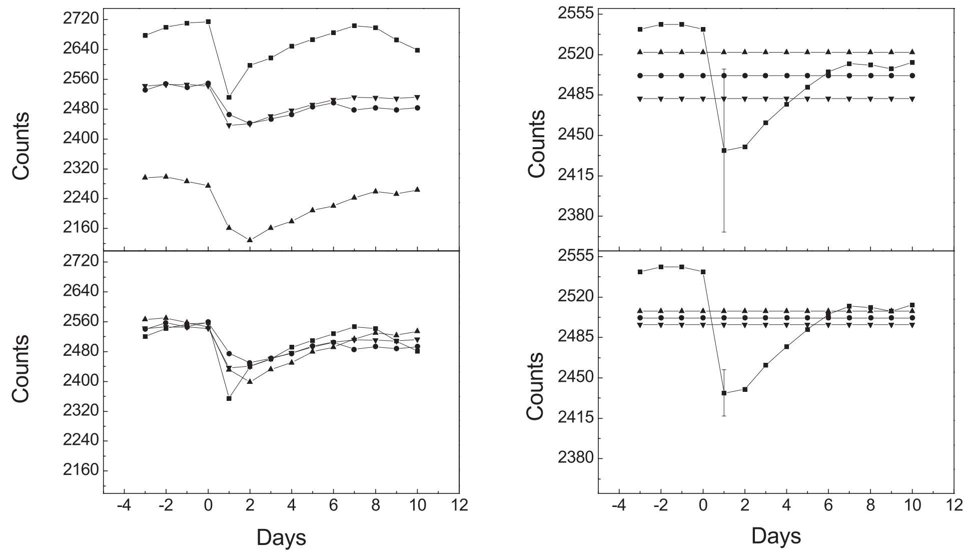

Fig. 1. Three typical FD events (square, circle, triangle) alongwith their average profile (inverted triangle) before (upper left panel) and after (lower left panel) the correction for solar cycle effect.

Fig. 2. Mean superposed intensity alongwith the 95% confidence interval for minimum intensity (vertical bar) and for mean intensity (upper and lower horizontal bars) alongwith the mean counts (middle horizontal bar) before (upper right panel) and after (lower right panel) the correction for solar cycle effect.

calculated by using equation(2). It is shown by vertical line. The mean intensity for all days is shown by a horizontal line (middle). The 95% confidence limits for mean intensity are calculated by using equation (1) and are shown by two horizontal lines (upper and lower), both in upper and lower figure in the right. If part of the vertical solid line representing the 95% confidence interval of decrease goes into the area bounded by the two horizontal lines representing 95% confidence interval of average intensity, the observed effect (decrease) could have come about simply through random fluctuations of the intensity [5]. If, however, the vertical solid line, does not go across the lower horizontal line, it is very unlikely that the observed value of decrease is due to random fluctuations and we can suggest a physical reason between the shock and subsequent decrease in cosmic ray intensity.

5. Conclusions

A statistical procedure, based on t-test, to assess the statistical reality of an average variation as obtained from superposed epoch analysis is described.

A procedure to transfer (correct) the data for other (superimposed) real effect is given. This procedure is useful for the space data, in general, and cosmic ray data, in particular, where other effect (e.g. solar cycle effect) of a large magnitude is present in the data sets. The importance of this correction has also been demonstrated, for cosmic ray data, as an example.

6. References

1. Chree C. 1932, Philos Trans R Soc London, Ser A 213, 245
2. Devore J.L. 2000, Probability and Statistics (Duxbury, California)
3. Forbush S. E. et al. 1982, Rev Geophys Space Phys 20, 971
4. Fruend J.E. 1981, Modern Elementary Statistics (Prentice-Hall, Calif.)
5. Haung Y., Lee Y. 1975, J Geophys Res 80, 2863