HE 2.2.30

Analytical Procedure for the Test of Nonlinearity in the Arrival Times of Air Showers

N. Ochi, and Large Area Air Shower (LAAS) group
Department of Physics, Okayama University, Okayama 700-8530, Japan

Abstract

An analytical procedure for the test of sporadic nonlinearity in a time series is proposed here. This is improved for unconfirmed chaotic feature in the arrival times of air showers advocated by Kitamura et al. in a recent paper. To handle the enormous quantity of data acquired by the LAAS group and to avoid ambiguity in human judgement, it is fully computerized and the significance level of an event can be quantified. The well-known Grassberger-Procaccia Algorithm for fractality and the surrogate data test for nonlinearity are the fundamentals of this procedure. Preliminary results are also reported.

1 Introduction:

Chaos is the irregular behavior of simple equations, and irregular behavior is ubiquitous in nature. When we observe an irregular time series, we can ask “Is this chaos or is this noise?” The confirmation of chaos in a time series provides some insight into modeling of the system, because the number of degrees of freedom can be estimated through the fractal dimension of the strange attractor. This is a primary motivation for studying chaos.

In recent years we have been investigating chaotic feature of cosmic-rays. Historically it was first reported by the Kinki University and Osaka City University groups that the time series of air shower arrival time intervals (ASATIs) show chaotic feature sporadically (Katayose et al., 1995; Ohara et al., 1995; Kitamura et al., 1997). However, their results have been questioned by Aglietta et al. (1996), who analyzed their own air shower data and concluded that ASATIs are consistent with the conventional view of white noise cosmic-rays.

Such a disagreement occurs frequently in the field of the chaotic time series analysis. This is due to the fact that distinguishing chaotic dynamics from stochasticity is often not an easy task, especially under the presence of experimental noises and the lack of sufficient data. As for ASATIs, the contamination by a substantial amount of noise is deduced during primaries’ journey from their source to earth-based detectors, thus the strange attractor in the phase space should be smeared. On the other hand, the sporadic behavior reported needs us to segment a time series into shorter pieces. Both of these fact leads to ambiguous results of the chaotic analysis, thus a statistical approach is needed. Fractal dimensions, which quantify the degree of complexity in a nonlinear time series, are also under influence of this ambiguity. As a result, we can no longer calculate them accurately, while the confirmation of chaos remains meaningful.

Under these circumstances, we have decided to take a more conservative approach to the analysis of ASATIs. We avoid accepting or excluding the presence of chaotic feature, but we calculate the significance level of the “chaos-likeness” in a time series statistically instead. To eliminate ambiguity in human judgement, the whole procedure is computerized. The fundamentals of our procedure are the Grassberger-Procaccia Algorithm (GPA) and the surrogate data test. As shown by Provenzale et al. (1992), the GPA can not distinguish chaotic feature from some kinds of nonlinear stochastic processes. Rather than try to exclude the latter case, we set a more modest goal of the detection of nonlinearity in ASATIs, because it is a sufficiently fruitful discovery for the cosmic-ray physics. This is an improved procedure from our previous one (Wada et al., 1997).

2 Analytical Procedure:

In this section we describe our analytical procedure for the test of sporadic nonlinearity in a time series. First, we outline the GPA, which is a standard technique to detect fractality in time series. Next, we explain the fitting and error calculation procedure; this is characteristic of our work. Finally, we mention the concepts and methods of the surrogate data test.

2.1 Grassberger-Procaccia Algorithm: The GPA, roughly speaking, quantifies the number of effective dimensions in a time series by the correlation dimension (Grassberger & Procaccia, 1983). The correlation dimension is an approximate value of the fractal dimension. From an observed scalar time series $\{x_i;\ i=1,..,N\}$, the higher dimensional vectors are created by the method of delay coordinates,

$$X_i = (x_i, x_{i+1}, \ldots, x_{i+(m-1)}) \qquad (i=1,2,\ldots,N-m+1)$$

where m is the embedding dimension and N is the number of data in a data set. The next step is to calculate the correlation integral,

$$C(r) = N'^{-1} \sum_{i=1}^{N-W} \sum_{j=i+W}^{N} \theta(r-|X_i-X_j|) \qquad N' = (N-W)(N-W+1)/2$$

where $\theta(x)$ is the Heaviside step function [$\theta(x)=0$ for $x\leq 0$, $\theta(x)=1$ for $x>0$] and W is the cut-off parameter (in this paper, W=4). C(r) is equivalent to the number of pairs of vectors with distance less than r. If we can find a wide "scaling region" in the log r vs. log C(r) plot, and if the slope (ν) of the scaling region is sufficiently smaller than m, the time series is (at least) nonlinear and ν is the correlation dimension. Here the first problem arises in the selection of the scaling region.

2.2 Fitting Error Calculation: The scaling region is limited from below by the size of the noise and by the finite number of data and from above by the finite size of the strange attractor. Since from an experimental point of view we can not estimate a priori the level of noise and the attractor size, procedures for determining the optimal scaling region unambiguously have to be established. To do so, we created the program which pick the best scaling region, where the root-mean-square error of the deviation of the plot from the fitted straight line over a given width of r becomes minimal. (An example of this fitting procedure is shown in Figure 1). We call this value of the minimal error "FE" (Fitting Error) assigned to the data set.

Not only for the determination of the best scaling region, FE can be used as a statistic that characterizes the quality of the scaling region. Since a lack of data leads to smaller ν even for a stochastic time series, we can not distinguish between nonlinear and stochastic time series by the value of ν. Instead, we can use FE because the stochastic time series do not make the scaling region owing to the large random fluctuation, which results in a larger FE.

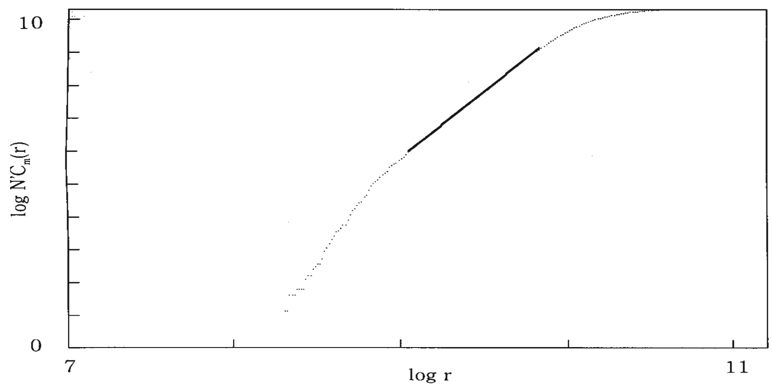

Figure 1: An example of the GPA and the determination of the best scaling region.

2.3 Surrogate Data Test: Skeptical researchers may say that the observed nonlinear feature can be explained as a result of the fluctuation or unknown effect of linear stochastic processes. We employ the surrogate data test to persuade them.

This test first specifies some linear process as a null hypothesis, then generates surrogate data sets which are consistent with this null hypothesis, and finally compares the values obtained from them with that from the original data. Since we can make many realizations of the null hypothesis, the distribution of the values is reliably established. If the value computed from the original data is significantly different than the distribution of values computed from the surrogate data then the null hypothesis is rejected and nonlinearity is detected. We can evaluate the significance level of the original value by a standard statistical procedure.

For a time series with no evidence of any dynamics, the null hypothesis is that the observed time series is fully described by independent random variables. Randomly shuffling the time-order of the original time series is a simple but satisfactory method to generate surrogate data sets. Of course, the value which will be evaluated here is FE.

3 Result: Simulation:

In this section we apply the preceding procedure to a computer-generated chaotic time series to check usefulness of it. The model is the Mackey-Glass equation, and parameters in the equation are chosen in accordance with Havstad et al. (1989).

We extracted 512 time series of N=256 from the equation, and applied the procedure to them. We used m=9, which is large enough for this model. We also analyzed 512 random time series of N=256 for comparison. The resulting FE distributions are shown in Figure 2. As expected, FE for the chaotic time series take smaller values than for the random ones. However, the difference is small and a considerable part of the distributions overlapped. This results in the fact that we can not say clearly which data sets are chaos or stochastic by the absolute value of FE.

A computer-generated time series is noise-free. For experimental time series, it is anticipated that distinguishing nonlinearity from stochasticity is a more difficult task.

4 Result: Air Shower Arrival Time Intervals:

In this section we report preliminary results on the nonlinear analysis of ASATIs $\{x_i;\ x_i=t_i-t_{i-1}$, where t_i is the arrival time of the i-th air shower$\}$. Data used here was taken by four stations of the Large Area Air Shower (LAAS) group (Ochi et al., 1999). They are Okayama University, Okayama University of Science, the No.31 building of Kinki University, and Nara Industrial University. In order to check sporadic behavior in ASATIs, an original time series was subdivided into data sets with N data points. The analysis was carried out separately for each data set. The total number of air showers used is about $7.5*10^5$, which make 11804, 5887, 2928 data sets of N=256, 512, 1024, respectively. We used again m=9.

Applying the surrogate data test, the most significant data set was 3.9 σ for a data set of N=256 of Kinki University, as shown in Figure 3. Considering the 20619 data sets inspected, the probability of getting this level of significance or greater is 0.66, which is not significant at all.

Next, we searched for simultaneous significant data sets among the four stations. From data sets on 1996 December 13, we found the coincident appearance of 3.1 σ, 1.9 σ, 2.0 σ, and 2.2 σ data sets overlapping 12.5 hours. The chance probability for this event was 0.18, which is still too weak to claim the detection.

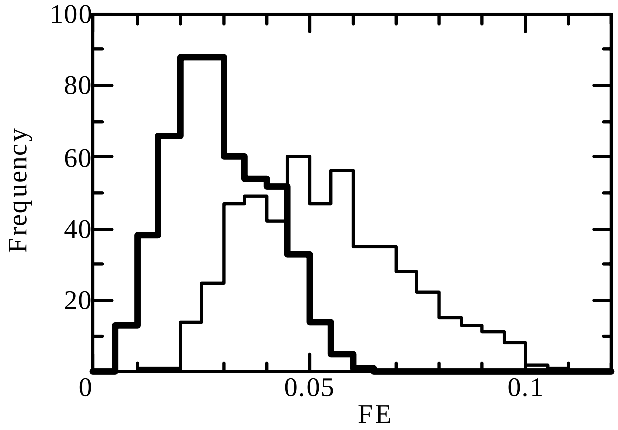

Figure 2: The FE distributions for the chaotic data sets(*thick line*) and for the random data sets(*thin line*).

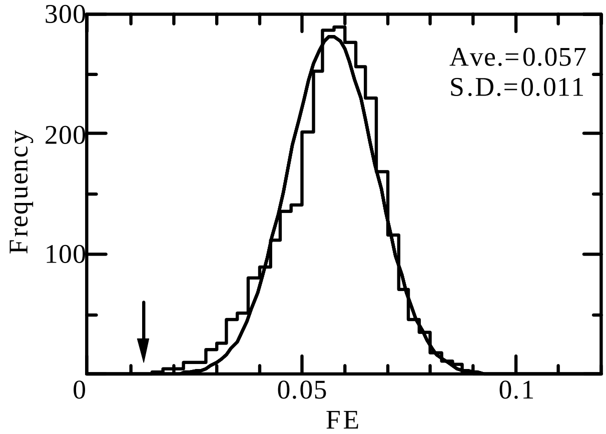

Figure 3: The result of the surrogate data test for the most significant data set. *Histogram:* the FE distribution for surrogate data sets. *Solid Curve:* a Gaussian fit to the histogram. *Arrow:* the original FE (3.9 σ).

5 Conclusion:

An analytical procedure for the test of sporadic nonlinearity in a time series is proposed here. This is a fully computerized algorithm to handle the enormous quantity of data acquired by the LAAS group and to avoid ambiguity in human judgement. This procedure can quantify the significance level of an event statistically. The usefulness of the procedure is ascertained by the computer simulation, while the preliminary results on ASATIs did not show nonlinearity.

We have not improved a procedure for distinguishing chaos from nonlinear stochasticity. This could be done by means of the space-time-separation plot or other methods. We will continue this work, at the same time we need to accumulate more air shower data.

References

Aglietta, M., et al. 1996, Europhys. Lett., 34, 231
Grassberger, P. & Procaccia, I. 1983, Physica D, 9, 189
Havstad, J. W., et al. 1989, Phys. Rev. A, 39, 845
Katayose, Y., et al. 1995, Proc. 24th ICRC (Rome), 1, 301
Kitamura, T., et al. 1997, Astrop. Phys., 6, 279
Ochi, N., et al. 1999, HE 6.1.20, this conference
Ohara, S., et al. 1995, Proc. 24th ICRC (Rome), 1, 289
Provenzale, et al. 1992, Physica D, 58, 31
Wada, T., et al. 1997, Proc. 25th ICRC (Durban), 7, 245