THE APPLICATION OF DIGITAL SIGNAL PROCESSING METHODS TO COSMIC RAY ANISOTROPY DATA

H.E. Bergeson, D.J. Cutler, J. F. Davis,* and D.E. Groom

Physics Department, University of Utah
Salt Lake City, UT 84112 (USA)

In analyzing data from the Utah Anisotropy Detector we have abandoned the traditional "harmonic dial" method in favor of discrete Fourier transform (DFT) and related techniques. An important advantage of the approach is that one can tap the vast literature and experience of other fields, notably electrical engineering and radio astronomy. White noise and signal size distributions are predictable. Sidelobe interference can easily be studied. Trend removal techniques have been applied to weather modulation and to sideband suppression in the case of periodic data windows. Special window functions and linear prediction methods ("maximum entropy") have also been studied, with the conclusion that direct DFT methods are preferable.

1. Introduction. Digital signal processing has become a mature and sophisticated specialty, and it has found recent application to areas as diverse as radio astronomy and infrared spectroscopy. When faced with analyzing data from the Utah Anisotropy Detector, we therefore chose not to re-invent the wheel. In this paper we discuss the application of standard digital signal processing techniques to a specific data set, our underground muon counting rate data for the past 1.3 years. While the methods are far from novel, we believe their application to a cosmic ray problem has enough novel aspects to merit discussion. The bag of available tricks is large and the clarity gained in analyzing the data this way is enormous, but at the same time there are many pitfalls and the Fourier transform does not contain the whole story.

2. The Data Sample. The detector has been in nearly continuous operation since 1978 Jan. 5. In this paper we consider the interval 1978 Jan. 1 through 1979 May 1, arbitrarily starting our clocks at the beginning of 1978. The detector ran 81% of the time, with several 7 to 14 day "notches" in the "data window" due to power and tape drive failures in the summer of 1978. The average "on" counting rate is 4.47 muons/sec, for a total of 1.52×10^8 counts in the sample. Data summaries are recorded every half hour, and may be summed into wider bins whenever convenient by the off-line programs. Most of the analysis reported here used 2-hr bins, since wider bins smeared peaks in the frequency domain and narrower bins necessitated more computer time. A typical data segment is shown in Fig. 1, where we have plotted fractional deviations from the average for 360 2-hr intervals. Points have a statistical uncertainty of 0.56%. The long-term trends are due to atmospheric effects, which at our energy are dominated by temperature changes in the upper atmosphere.

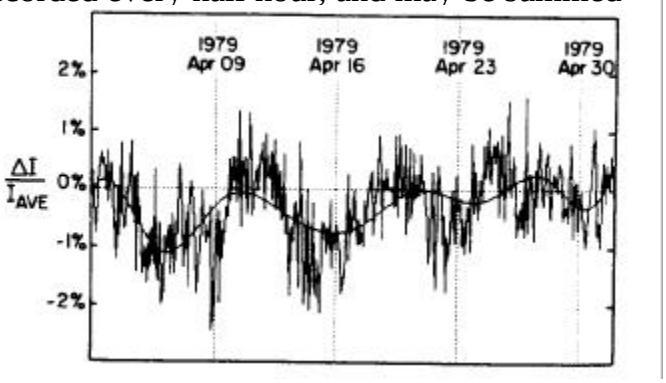

Fig.1. Typical data segment showing about 6% of the total data, binned into 2-hr intervals of about 32000 counts each.

3. DFT's and FFT's.[1] We define the discrete Fourier transform (DFT) of the data at frequency f_k as

$$X(f_k) = \frac{2}{aM} \sum_{m=0}^{M-1} x_m \exp(-2\pi i\ f_k m\Delta t) \qquad (1)$$

where $x_m = W_m\ \ddot{A}I_m/l_{ave}$. Here a is the "duty cycle" (81% in this case), M is the total number of data points, W_m is a "window function" (0 or 1, depending on whether the apparatus is working in the interval), $\ddot{A}I_m$ the deviation of the intensity from the overall average I_{ave}, and Ät the bin width. The normalizing factor in front of the summation is not conventional, but has been chosen so that $\hat{i}(f_k)\ \ |X(f_k)|$ is the desired anisotropy, i.e., the fractional amplitude of a harmonic with frequency f_k. (We have found it convenient to use the *amplitude* of the FT, rather than the usual *power*, in order to obtain vertical scales linear in the desired anisotropies.) The highest frequency which can be seen without aliasing is the Nyquist frequency 1/2Ät.

The fast Fourier transform (FFT) is similarly defined, except that f_k is chosen to be k/NÄt. The number of data points N must be chosen equal to a power of 2 in most top end to achieve this or to increase the frequency-domain point density.

We are usually interested in computing WO in only a small fraction of the frequency range, e.g., 100 values between 0.95 and 1.05 solar days^{-1}. About 2.5 min of DEC System 20 time is required. Nearly the same amount of time is required for a 2^{14} point FFT, which yields about the same point density. Accordingly, we usually calculate the less cumbersome DFT except when a large fraction of the region below the Nyquist frequency is required.

A value $X(f_k)$ represents a best fit of a single harmonic with frequency f_k to the data, and in general is different then the simultaneous fit of several waves at *a priori* interesting frequencies. The non-independence is a result of the sidelobe interference discussed below.

4. Window Transforms. Suppose an ideal detector is turned on between t_0 and $t_0 + T$, and that a signal $\cos(2\eth f_0 t + \phi_0)$ exists. The FT is proportional to

$$X(f) = \frac{\sin\{2\pi(f-f_0)(T/2)\}}{\{2\pi(f-f_0)(t/2)\}} \exp i\{\phi_0 - 2\pi(f-f_0)(T/2-t_0)\} \qquad (2)$$

We recognize the first term as the Fraunhofer amplitude, with a first zero at $f - f_0 = 1/T$. A real (or noise) peak will thus have a characteristic width equal to the reciprocal of the sampling time; this scale determines the "jaggedness" of the frequency-domain data.

The FT of a periodic feature is thus the FT of the data window (in this case rectangular) shifted in phase and centered at the frequency of interest. The FT's of our window functions for the first half, second half, and total data segments are shown in Fig. 2. Data notches in the first half greatly enhance the sidelobes, while in the second case, where there are fewer notches, the window transform is very Fraunhofer-like. Since the width is proportional to 1/T, the total window transform is half as wide as the half-period windows.

DFT's in the vicinity of an inverse solar day for the data in the same time segments are shown in Fig. 3. Since the running period is about 1.3 years, peak widths for the full period are essentially the same as the separation of solar and sidereal frequencies, while in the, half-period segments they are

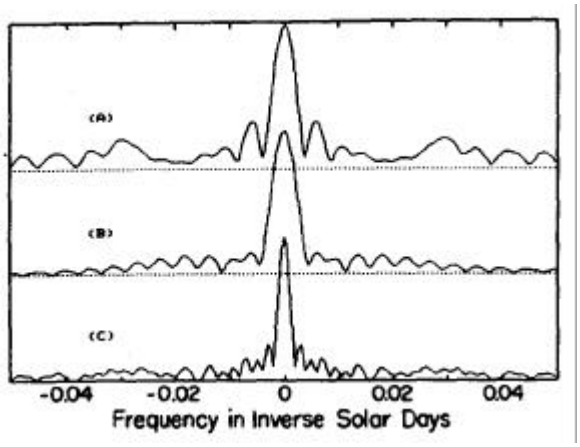

Fig. 2. Data window transform amplitudes for (a) the first half, (b) the second half, (c) total data interval.

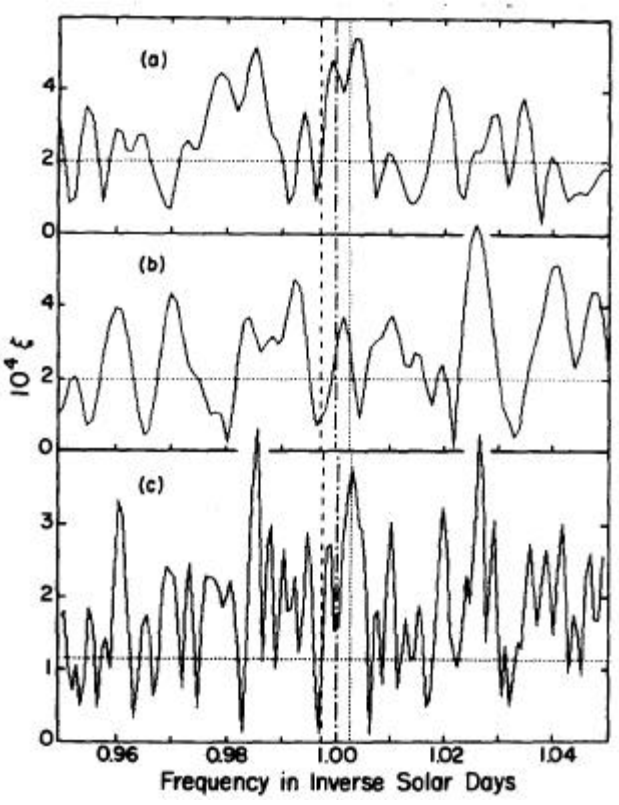

Fig. 3. DFT amplitudes in the vicinity of 1 day-I for (a) the first half, (b) the second half, and (c) total data interval. The dotted vertical line indicates the sidereal frequency, and the dashed line the anti-sidereal frequency.

twice as broad. We emphasize the sidelobe interference between solar and sidereal peaks, particularly in the first half. The notable differences in the spectra for two equivalent time intervals have also been characteristic of extensive Monte Carlo experiments.

Sidelobes result from the sudden turn-on and turn-off of the data, and can be reduced if W_k in Eq. 1 is multiplied by a smoothly varying function, such as a Gaussian or cosine. Some statistical significance is lost, since data near the beginning and end of the interval are essentially thrown away. We have made such experiments, with the conclusion that the gains are not worth the price.

5. Phase. For the rectangular data window discussed above, the derivative of phase in Eq. 2 with respect to frequency is $-2\pi(T/2 + t_0)$. If t_0 is chosen so that the data window transform is real at f = 0 ($t_0 = T/2$ in the example), then the phase should be constant near the maximum of any persistent signal. We illustrate this in Fig. 4, which shows the phase corresponding to the DFT amplitude shown in Fig. 3(c). The sudden $\sim 12^h$ shifts are artifacts arising from the fact that principal values are plotted; it

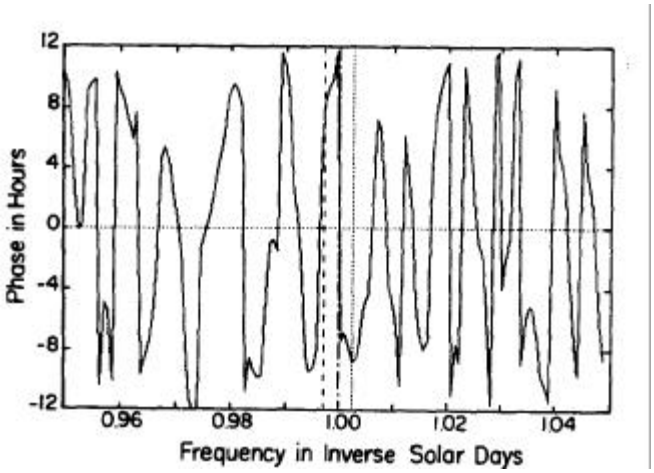

Fig. 4. The phase of the DFT for the total running period 01/01/78 to 05/01/79, with to chosen at 243 days to obtain a real data window DFT near f = 0.

is evident, for example, that the peak at 0.957 day^{-1} belongs at 19^h. The phase at the sidereal frequency is quite stable, while the phase of the noise peak at 1.027 day^{-1} shows a large slope.

6. Noise and Error Analysis. If at some f_0 there exists a genuine signal with amplitude $\hat{\imath}_0$ and phase ϕ_0 , then the real and imaginary arts of the FT independently distribute normally about this point with ó = 2/N, where N is the total number of events. Equal probability contours are thus circles in the complex plane, as is assumed in the usual harmonic dial analysis. If a result ($\hat{\imath}$, $\phi = \phi_0 - \theta$) is obtained in an experiment, then the probability that the correct answer lies in the area element $\hat{\imath}_0 d\hat{\imath}_0 \phi\theta$ at ($\hat{\imath}_0$, ϕ_0) is simply a Gaussian in the displacement from ($\hat{\imath}$, ϕ) times the area element. Frequency functions in $\hat{\imath}_0$, θ and a may be found by integrating over the unwanted variables. We obtain

$$P(\xi_0|\xi) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\{-(\xi - \xi_0)^2/2\sigma^2\} \sqrt{\xi_0/\xi}\, [\sqrt{2\pi x}\, e^{-x}\, I_0(x)] \qquad (3)$$

and

$$P(\theta|\xi) = \frac{1}{2\pi} \exp(-\xi^2/2\sigma^2) + \frac{y}{2\sqrt{\pi}} \exp(-\xi^2 \sin^2\theta/2\sigma^2)\{1 + \mathrm{sgn}(y)\,\mathrm{erf}(y)\} \qquad (4)$$

where $x = \hat{\imath}\hat{\imath}_0 / \sigma^2$ and $y = \hat{\imath}\cos\theta / 2\ \sigma$ [2] Both distributions become normal for $\hat{\imath} >> \sigma$, with $\langle \theta^2 \rangle = \sin^{-1}(\sigma / \hat{\imath}_0)$ as expected from the harmonic dial picture. In Eq. 3 the quantity in square brackets is within 20% of unity for x > 0.2, so we may understand the function as a Gaussian skewed to the high- $\hat{\imath}_0$ side by the factor $\hat{\imath}_0 / \hat{\imath}$. Given experimental values $\hat{\imath}$ and ó, one must integrate Eq. 3 to obtain confidence levels on $\hat{\imath}_0$, The sidereal frequency peak in Fig. 3(c) has an amplitude of 3.78 X 10^{-4}, and a = 1.15 X 10^{-4}. With 68% confidence, we find that $\hat{\imath}_0$, lies between 2.9 X 10^{-4} and 5.2 X 10^{-4}.

In the white noise case $\hat{\imath}_0$, = 0 and Eq. 3 reduces to the usual Rayleigh distribution $P(\hat{\imath},0) = (d/d\hat{\imath})(\exp\{ -\hat{\imath}2 / 2\acute{o}2 \})$. Our noise spectrum is essentially flat above 0.7 day-1, with some evidence for a component proportional to 1/f at lower frequencies. We show the distribution of 7000 DFT points above this frequency in Fig. 5. The curve is the Rayleigh function with a adjusted for a best fit. We found ó = 1.11 X 10^{-4}, as compared with 2/N = 1.15 X 10^{-4}.

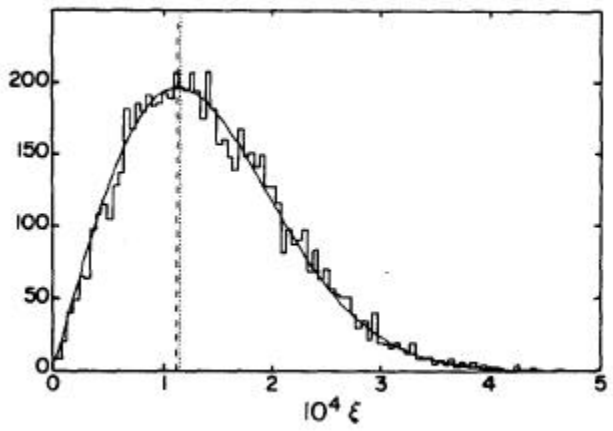

Fig. 5. The distribution of 7000 DFT points above 0.7 day^{-1}. The solid curve is a Rayleigh distribution with ó adjusted for a best fit. The vertical lines are the fitted ó (dashed) and 2/N (dotted).

7. Trend Removal. One may eliminate long-term trends such as weather effects by means of a digital version of an automatic volume control. The solid curve shown in Fig. 1 is a least-squares cubic spline fit to the data with 3.3 day knot spacing. The reduced X^2 is 1.1. Low frequency harmonics are suppressed, and the DFT in the vicinity of 1 day^{-1} is unaffected.

8 Linear Prediction Theory (Maximum Entropy).[3] Claims have been made that linear prediction methods yield better resolution than FT's with less computation. We have used the all-pole algorithm, also known as the "maximum entropy method," described by Andersen(4) to study simulated data with Gaussian noise. Any signals which we could detect with this method could also be seen using the DFT. For our data, where the noise is at least 10 times larger than the signal and where we have hundreds of cycles, the FT analysis is at least as good as linear prediction. It also gives the phase of the signal. which linear prediction does not, and produces a linear spectrum.

9. Solar Frequency Cuts. In an attempt to study possible solar modulation effects at the sidereal frequency, we performed a series of experiments in which 6 to 12 hour windows with 1 solar day spacing were imposed on the data. Before trend removal, an 0.5% peak at the annual frequency conspired with various harmonics of the window to produce a nearly bilaterally symmetric peak distribution about the solar frequency. Even with trend removal, the modulation of the sidereal peak by the second harmonic of the unsymmetric periodic data window produced a prominent peak at the antisidereal difference frequency. These experiments underscore the caution which must be exercised, especially if a continuous range of frequencies is not examined.

10. Acknowledgments. This work was supported by the National Science Foundation USA . We are also grateful for the Physics Department at the University of Utah for donating a large block of computer time which made the many computer experiments possible.

References

1. The literature on the subject is vast, but we have found A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing* (Prentice-Hall, 1975) to be our standard reference on the theory. E.O. Brigham, *The Fast Fourier Trans*form, (Prentice-Hall, 1974) has been a useful source of methods.

2. A similar situation is discussed by J. Linsley, Phy. Rev. Lett. _36, 1530 (1975). Tables and integral distributions of the "off-center Rayleigh distribution" are given in University of Utah internal report UUHEP-79/4 (unpublished).

3. A helpful review is given by J. Makhoul, Proc. of the IEEE 63, No. 4, 561 580 (Apr. 1975). The paper contains an extensive list of references.

4. N. Andersen, Geophysics 39, No. 1, 69-72 (Feb. 1974).

* Present Address: Los Alamos Scientific Laboratory, Los Alamos NM 87545 (USA)

Filename:	icrc794-194
Directory:	A:
Template:	C:\Program Files\Microsoft Office\Templates\Normal.dot
Title:	
Subject:	
Author:	
Keywords:	
Comments:	
Creation Date:	09/24/01 10:11 PM
Change Number:	1
Last Saved On:	09/24/01 10:11 PM
Last Saved By:	
Total Editing Time:	81 Minutes
Last Printed On:	10/25/01 1:18 PM

As of Last Complete Printing

Number of Pages:	5
Number of Words:	1,859 (approx.)
Number of Characters:	10,599 (approx.)