Standard Errors at Harmonic Analysis on Cosmic Ray Data

By ERIC DYRING AND BENGT ROSÉN, University of Uppsala

(Manuscript received June 20, 1960)

Abstract

Standard errors of harmonic analysis have been calculated by three different methods. These methods are discussed and numerical calculations have been carried out from the results of harmonic analyses made on data from the neutron monitors in Uppsala and Murchison Bay. Periods covering from single days up to the mean of three years have been used for the comparison between the results. Although the standard error calculated according to the three methods include different types of variations, the results show good similarity. One of the methods also gives indications that only in some cases the addition of a second harmonic will be an improvement of the fit.

Introduction

Studies of cosmic radiation consist to a great extent of time series analysis. One important problem is the study of periodic intensity variations, such as diurnal, semidiurnal and 27-day. Very often harmonic analysis is used to approach these problems. Naturally it is important to make correct estimates of the standard errors in such analyses in order to interpret particular results.

Mathematical representation of intensity variations

Let u be the true intensity of the cosmic radiation and X_k the intensity value from a particular recorder, where k indicates the time. X_k will then be distributed around the mean u and with the standard deviation σ. The distribution will be theoretically equal to a Poisson distribution and we have

$$\sigma = \sqrt{u}$$

This is generally used in calculating standard errors. However, as our instruments will certainly have some sources of error, data from the various detectors will not give an exact Poisson distribution. Calculations have shown that for the nucleonic component measured with an international standard neutron monitor the distribution is approximately normal with the standard error $1.3\sqrt{\bar{u}}$ (Mc Cracken 1958, Dyring 1960). We may write:

$$X_k = u + e_k \qquad (1)$$

where e_k is a random variable with known distribution and with expectation

$$E(e_k) = 0$$

If we assume that the true variation during one 24-hour period consists only of first and second harmonics, we have for the true variation

$$u_k = u_d + R_1 \sin(\Theta + \alpha_1) + \\ + R_2 \sin(2\Theta + \alpha_2) = u_d + f(\Theta) \qquad (2)$$

where

$$\Theta = \frac{2\pi k}{n}, \quad k = 0, 1, 2, 3 \ldots\ldots n-1$$

and u_d is the mean of the true intensity during 24 hours. From a single recorder we would actually obtain:

$$X_k = u_d + f(\Theta) + e_k \tag{3}$$

If the assumption cf. eq. 2 is correct in all respects the random variable e_k will have a distribution identical with that of e_k in eq. 1. This cannot be expected to be exactly true. We make the new approach:

$$X_k = u_d + f(\Theta) + \varepsilon_k \tag{4}$$

ε_k is a residual, about which we do not make any other assumptions than that it does not behave too irregularly (see method B below).

Very often we want to estimate the average daily variation for a long period. This is usually done by making a harmonic analysis on averaged data. When taking a mean over N days according to eq. 4 we get

$$\frac{1}{N}\sum X_k = \frac{1}{N}\sum u_d + \frac{1}{N}\sum f(\Theta) + \frac{1}{N}\sum \varepsilon_k$$

If $f(\Theta)$ is constant from day to day we get:

$$\bar{X}_k = \bar{u}_d + f(\Theta) + \bar{\varepsilon}_k \tag{3$^+$}$$

where $\sigma^2_{\bar{\varepsilon}} = \frac{1}{N}\sigma^2_{\varepsilon}$.

If $f(\Theta)$ is not constant we may assume the same model as in eq. 4 but for averaged data:

$$\bar{X}_k = \bar{u}_d + f(\Theta) + \bar{\varepsilon}^+_k \tag{4$^+$}$$

Harmonic analysis procedure. We have data consisting of n equally-spaced observations X_0, X_1, $X_2 \ldots X_{n-1}$. By the common method of least squares we get estimates of u_d and the harmonic coefficients a and b in a diurnal component when making Q a minimum in:

$$Q = \sum_{k=0}^{n-1}\left(X_k - u_d - a\cos\frac{2\pi k}{n} - b\sin\frac{2\pi k}{n}\right)^2 \tag{5}$$

The amplitude R and the phase angle α are then estimated by

$$\hat{R} = \sqrt{\hat{a}^2 + \hat{b}^2} \tag{6a}$$

$$\hat{\alpha} = \operatorname{arctang}\frac{\hat{a}}{\hat{b}} \tag{6b}$$

where ^ indicates estimate.

The approaches (1) and (4) give different estimates of σ_a and σ_b.

Method A

With the assumption in (3) and (3$^+$) we have

$$X_k = u_k + e_k$$

If the sampling distribution of X_k is true Poissonian we have for the m:th harmonic (Kendall),

$$\hat{\sigma}_{a_m} = \hat{\sigma}_{b_m} = \sqrt{\frac{2}{n}\bar{X}_k}$$

and if we take care of the correction constant

$$\hat{\sigma}_{a_m} = \hat{\sigma}_{b_m} = 1.3\sqrt{\frac{2}{n}\bar{X}_k}$$

Method B (Compare BARTELS 1935)

$$X_k = u_k + \varepsilon_k \quad \text{(cf. eq. 4 and 4}^+\text{)}$$

We assume no *a priori* information other than

1. $E(\varepsilon_k) = 0$ and the variance σ^2 (independent of k)

2. ε_k is normally distributed. This assumption should not be taken to literally as long as the distribution of ε_k is not too irregular.

According to the theory of least squares we can estimate σ^2_ε as

$$\hat{\sigma}^2_\varepsilon = \frac{1}{n-3}\sum_{k=0}^{n-1}\left(X_k - \hat{u}_d - \hat{a}\cos\frac{2\pi k}{n} - \hat{b}\sin\frac{2\pi k}{n}\right)^2 \tag{7}$$

The function $f(\Theta)$ is assumed to consist of a 1st harmonic. We know that $\hat{u}_d$, $\hat{a}$, $\hat{b}$, are independent random variables, normally distributed with resp. means u_d, a, b, and that

$$\frac{(n-3)\hat{\sigma}^2_\varepsilon}{\sigma^2_\varepsilon}$$

is x^2-distributed with $n-3$ degrees of freedom. Eq. 7 can be shown to be equal to:

$$\sigma_\varepsilon^2 = \frac{1}{n-3}\left\{\sum_{k=0}^{n-1}(X_k - \hat{u}_d)^2 - \frac{n}{2}\hat{R}^2\right\} \quad (8)$$

If our function $f(\Theta)$ consists of first and second harmonics we get in the same way:

$$\sigma_\varepsilon^2 = \frac{1}{n-5}\left\{\sum_{k=0}^{n-1}(X_k - \hat{u}_d)^2 - \frac{n}{2}\hat{R}_2^2\right\} \quad (9)$$

and the estimates

$$\hat{\sigma}_a^2 = \hat{\sigma}_b^2 = \frac{2}{n}\hat{\sigma}_\varepsilon^2$$

As the methods A and B give the estimates of σ_a and σ_b, the next step is to get the estimates of σ_R and σ_α.

Let F be a function of a number of uncorrelated random variables α, β, γ – – – –. By approximating $F(\alpha, \beta \ldots)$ by the linear part of its Taylor development the following formula for σ_F^2 is given:

$$\sigma_F^2 = \left(\frac{\partial F}{\partial \alpha}\right)^2 \sigma_\alpha^2 + \left(\frac{\partial F}{\partial \beta}\right)^2 \sigma_\beta^2 + \ \ldots (10)$$

It can easily be proved that for the harmonic coefficient a and b

$$\text{cov}\ (a, b) = 0$$

Eq. 10 gives the estimates of σ_R and σ_α

From eq. 6a and 6b we get the approximations:

$$\hat{R} \approx \sqrt{a^2+b^2} + \frac{a}{\sqrt{a^2+b^2}}(\hat{a}-a) + \frac{b}{\sqrt{a^2+b^2}}(\hat{b}-b)$$

$$\hat{\alpha} \approx \text{arctag}\ \frac{a}{b} + \frac{b}{a^2+b^2}(\hat{a}-a) - \frac{a}{a^2+b^2}(\hat{b}-b)$$

which give

$$\hat{\sigma}_R = \sigma_a = \hat{\sigma}_b$$

$$\hat{\sigma}_\alpha = \frac{\hat{\sigma}_a}{R} = \frac{\hat{\sigma}_b}{R}$$

and

$$\text{cov}\ (\hat{R}, \hat{\alpha}) = 0$$

Method C

A third method, which differs from the methods A and B can be used. It is built on the problem of "cloud of points" (Bartels 1932, Dorman 1957) and can be used only for long periods. The harmonic coefficients a and b are calculated for each day during the period. In the general case, i.e. when a and b are correlated over the period error limits for the pair (a, b) are obtained by a probable error ellipse.

If we assume non-correlation between a and b the probable error ellipse will become a circle.

In this case σ_R and σ_α can be calculated from eq. 10 with:

$$\hat{\sigma}_a^2 = \frac{1}{N-1}\sum(a_i - \bar{a})^2, \quad \hat{\sigma}_b^2 = \frac{1}{N-1}\sum(b_i - \bar{b})^2$$

where N is the number of days.

Discussion

If we use method A, we get from eq. 3 the random variable e_k the standard deviation of which will be estimated by $\sqrt{u_k}$ (cf. eq. 1). σ_R and σ_α will then depend only on the uncertainty of the individual observations, i.e. the distribution from which they are drawn will be wholly decisive. The magnitudes of σ_R and σ_α will not be affected by the closeness of the fit. This means that if we have a number of observational series, where all observations are of the same order of magnitude X_k and drawn from a Poisson distribution, then σ_R and σ_α will be constant for the different observational series.

The residual ε_k in eq. 4 may be interpreted as

$\varepsilon_k = e_k + r_k$

e_k = random variable as above.

r_k = specification error.

Method B will give us σ_R and σ_α which will be dependent both on all sources of variation and on the closeness of fit. There may be some difficulties in interpreting σ_R and σ_α as standard errors, as they take into account all actual variations, for example the day-to-day variability in the true form of the daily variation, and there is a question whether this can be regarded as random. At least for long periods these variations may be assumed random and σ_R and σ_α will be standard errors. This is also supported by the numerical comparison between the methods below.

Method C will give us σ_R and σ_α. These estimates will be dependent on all the variations as in method B. The very tedious calculation will make this method unpractical.

Method A offers:

1. Standard errors of the fitted harmonic parameters R and α.
2. The statistical fluctuations of the individual observations alone are decisive.
3. σ_R and σ_α are independent of the closeness of fit.

Method B offers:

1. Standard errors of the fitted harmonic parameters R and α.
2. σ_R and σ_α depend on all variations during the period of observations.
3. σ_R and σ_α are dependent on the closeness of fit.

Method C offers:

1. Standard errors of the fitted harmonic parameters R and α.
2. σ_R and σ_α depend on all variations during the period of observations.
3. σ_R and σ_α will be independent of the closeness of fit.

Comparison with numerical results

To make a comparison between the methods, results from harmonic analysis of neutron monitor data from the two cosmic ray stations at Uppsala (60° N, 18° E) and Murchison Bay (80° N, 18° E) have been used. Different periods from single days up to the mean of 3 years have been examined.

Bihourly values have been used. The data are corrected for trend and atmospheric effects in the usual way. Each observation is expressed in per cent of the daily mean. Values are obtained for the amplitude of first and second harmonics in per cent of the daily mean and for the time of maximum in hours and minutes. The standard errors according to the three methods are shown in tables I and II. Some periods are chosen according to distinct phase shifts, which are shown by a vector sum diagram made over solar rotation periods (SANDSTRÖM, DYRING, LINDGREN 1960). The single days are randomly chosen during a cosmically calm period 20 Dec. 1958 to Jan. 1959. To get a rough information of method C when assuming a and b non-correlated, we have calcu-

Table I. Numerical results from the two methods A and B for the neutron monitor in Uppsala.

Type of period	Time	The standard error of the amplitude of 1:st harmonic in per cent		
		I	II	III
		Acc. to eq. 11	1:st harmonic fitted. Standard error acc. to eq. 8	1:st and 2:nd harmonics fitted. Standard error acc. to eq. 9
Single days	22/12 1958	0.240	0.315	0.211
	26/12 1958	0.240	0.217	0.172
	29/12 1958	0.240	0.264	0.258
	30/12 1958	0.240	0.231	0.214
	3/1 1959	0.240	0.163	0.115
	7/1 1959	0.240	0.140	0.093
Period s with typical phase shifts	31/8 56—7/3 57	0.017	0.027	0.026
	8/3 57—20/7 57	0.021	0.021	0.023
	21/7 57—3/3 58	0.016	0.021	0.020
	1/5 58—30/4 59	0.013	0.017	0.015
Periods of different length	35 days	0.041	0.063	0.072
	8/3 57—20/7 57	0.021	0.021	0.023
	12 months (1958)	0.013	0.010	0.011
	26/8 57—30/4 59	0.010	0.012	0.011
	1/9 56—31/8 59	0.007	0.011	0,011

Table II. Numerical results from the three methods A, B and C for the neutron monitor in Murchison Bay.

Type of period	Time	The standard error of the amplitude of 1:st harmonic in per cent			
		I	II	III	IV
		Acc.to eq. 11	1:st harmonic fitted. Standard error acc.to eq. 8	1:st and 2:nd harmonic fitted. Standard error acc.to eq. 9	Acc. to eq. 12
Single days	22/12 1958	0.234	0.214	0.212	
	26/12 1958	0.234	0.406	0.388	
	29/12 1958	0.234	0.171	0.164	
	30/12 1958	0.234	0.178	0.195	
	3/1 1959	0.234	0.196	0.211	
	7/1 1959	0.234	0.180	0.191	
Periods with typical phase shifts	10/9 57—2/8 58	0.014	0.018	0.017	0.020
	3/8 58—18/11 58	0.025	0.044	0.025	0.034
	19/11 58—29/5 59	0.019	0.026	0.018	0.023
Periods of different length	3/8 58—18/11 58	0.025	0.044	0.025	0.034
	12 months (1958)	0.013	0.008	0.006	0.018
	26/8 57—30/4 59	0.010	0.016	0.016	0.014

lated $\hat{\sigma}_{\bar{R}}$, as only one-day values of R_1 from Murchison Bay neutron monitor are available.

The calculation of the standard errors is made in four ways.

I. Method A

$$\sigma_R = 1{,}3\sqrt{\frac{2}{X}}\cdot 100 \qquad (11)$$

where X is the total of the observed values over the period, $\hat{\sigma}_R$ is expressed in per cent.

II. Method B. Fitting first harmonic.

III. Method B. Fitting first and second harmonics.

IV. Method C.

$$\hat{\sigma}^2_{\bar{R}} \approx \frac{1}{N(N-1)}\sum(R_i - \bar{R})^2 \qquad (12)$$

in per cent.

Conclusion

A comparison of the results shows that the magnitudes of the standard error obtained by the three methods are surprisingly similar. This seems to hold for both single days and long periods. From this we can conclude that our approach with harmonic analysis is good, i.e. the specification errors are small. As the neutron monitor will measure the cosmic ray intensity with a big opening angle, which will mean that a sharply defined anisotropy will be smoothed out during a 24-hour rotation of the earth, the result can be expected.

Method C seems to be more similar to method B than to method A. This can be expected when the closeness of fit is good, as they both take account of all variations occurring during the time from which the data of the analysis are taken and method A does not do this.

By comparing II and III in the tables we can have information about the existence of a second harmonic. In only 6 of the total 25 different cases the second harmonic causes a marked reduction of the standard error, indicating the existence of an important semidiurnal component. These 6 cases are distributed on both single days and long periods. For the two stations Uppsala and Murchison Bay the semidiurnal component in most cases produces very little improvement in the fit.

As only the single days and the calendar year 1958 give identical periods for the two stations the material is too small to make possible any conclusion concerning a correlation between the two stations.

The question of which method will be the best to estimate the standard errors of the harmonic coefficients is still open. It depends on the use of the results. Method A with a correction of the Poisson standard deviation is favourable according to practical calculating aspects as long as the fit is good. A further comparison between the methods will be carried out not only for neutron monitor data but also for meson telescope data.

Acknowledgement

The authors wish to express their thanks to doctor Arne E. Sandström for his great interest and encouragement in this paper.

They are also very grateful to doctor N. R. Parson for many valuable discussions.

The investigation is sponsored by the Air Research and Development Command, United States Air Force, through its European Office, which is gratefully acknowledged.

REFERENCES

BARTELS, J., 1932: Statistical Methods for Research on Diurnal Variations; *Terrestrial Magnetism and Atmospheric Electricity*, vol. **37**, p. 293.

BARTELS, J., 1935: Random Fluctuations, Persistence, and Quasi-Persistence in Geophysical and Cosmical Periodicities; *Terrestrial Magnetism and Atmospheric Electricity*, vol. **40**, p. 12.

DORMAN, L. I., 1957: *Cosmic Ray Variations*, Translation prepared by Technical Documents Liaison Office, Wright-Patterson, Air Force Base, Ohio, p. 95.

DYRING, E., 1960: *Statistical Studies of the Performance of Cosmic Ray Recording Instruments.* In preparation.

KENDALL, M. G., 1948: *The advanced Theory of Statistics, Part II* p. 433.

MC CRACKEN, K. G., 1958: *Thesis*, University of Tasmania.

SANDSTRÖM, A. E., DYRING, E., LINDGREN, S., 1960: The daily variation of the cosmic ray nucleonic component at Murchison Bay and Uppsala, *Tellus* **12**, pp. 332—347.