Geofísica Internacional (1992), Vol. 31, Num. 2, pp. 155-158

Two practical rules for forecasting the number of flares in an active region

I. Ferro Ramos and N. López Reyes
Institute of Geophysics and Astronomy, Institute of Mathematics, Cuban Academy of Sciences, Havana, Cuba.

RESUMEN

Mediante el empleo de técnicas de reconocimiento de patrones y sobre la base de relaciones previas, se formulan dos reglas prácticas y un criterio suficiente para pronosticar el número de destellos en una región activa. Se demuestra que la verosimilitud del método es superior a la del mejor pronóstico trivial, y para cada caso, se evalúa el porciento de pronósticos correctos o su efectividad. Se muestra que estas reglas, basadas en la informatividad de las variables, son independientes de la fase del ciclo.

PALABRAS CLAVE: Destellos solares, ciclo solar.

ABSTRACT

Employing pattern recognition methods and on the basis of previous relations, two practical rules and a suffient criterion are formulated for forecasting the number of flares in an active region. It is demonstrated that the likelihood of the method is superior to the best trivial forecast and for each case the percentage of correct forecasts or its effectivity is evaluated. It is shown that these rules, based upon the informativity of the variables, are independent of the phase of the cycle.

KEY WORDS: Solar flares, solar cycle.

INTRODUCTION

The total number of flares that will produce an active region at its passage through the disk, is a value for which there are no methods for forecasting.

The present paper continues a previous one (López and Ferro, 1987) in which we established three empirical relations obtained by a pattern recognition technique. Two of these rules are expressed by regression equations and the third one by a sufficient criterion.

The objective of this paper based upon an independent sample, is to demonstrate that the López and Ferro criteria are practical and effective.

DATA AND METHODS

In this paper we use three subsamples from López and Ferro 1987 and also an independent sample formed by 56 active centers belonging to years 1980-81. This sample was divided into two subsamples: a first one is described as "training sample" and the second one as "testing sample".

The method employed was the so-called test theory that is presented in López and Ferro, 1987 and is explained for the so-called CT algorithm in Bravo, 1983.

DISCUSSION AND RESULTS

1. Finding a possible relation between the forecasting criteria and the cycle's phase.

Some authors write (Burov, 1976; Bernshtein and Burov, 1976) that the solution of a pattern recognition can depend on the phase of the solar cycle. We suspected that the informativity of the variables does so also. To clarify this possibility, we calculated for the training sample the regression equations on the variables N (total number of flares in the passage of the active region through the disk) and M (total number of informative variables) as well as N and IB (number of different flares prior to the instant when the variable brightness becomes informative with weight equal to one).

It is found that the coefficients of the regression equations are very similar:

$N=4{,}50\ M + 4.43$ $r=0{,}88$ López and Ferro, 1987   (1)

$N=4{,}53\ M + 5{,}21$ $r=0{,}74$ Training sample   (2)

and

$N=1{,}70\ IB + 9{,}03$ $r=0{,}93$ López and Ferro, 1987   (3)

$N=1{,}43\ IB + 12{,}40$ $r=0{,}61$ Training sample   (4)

As an indirect test that the corresponding regressions do not differ we have plotted the straight lines (2) and (4) on the straight lines (1) and (3) respectively and show that they do not deviate from the standard error of regression (see fig. 1 and fig. 2). This plus the fact that the correlation coefficients do not differ significantly at the 5% level, show that the considered variables do not depend on the cycle phase.

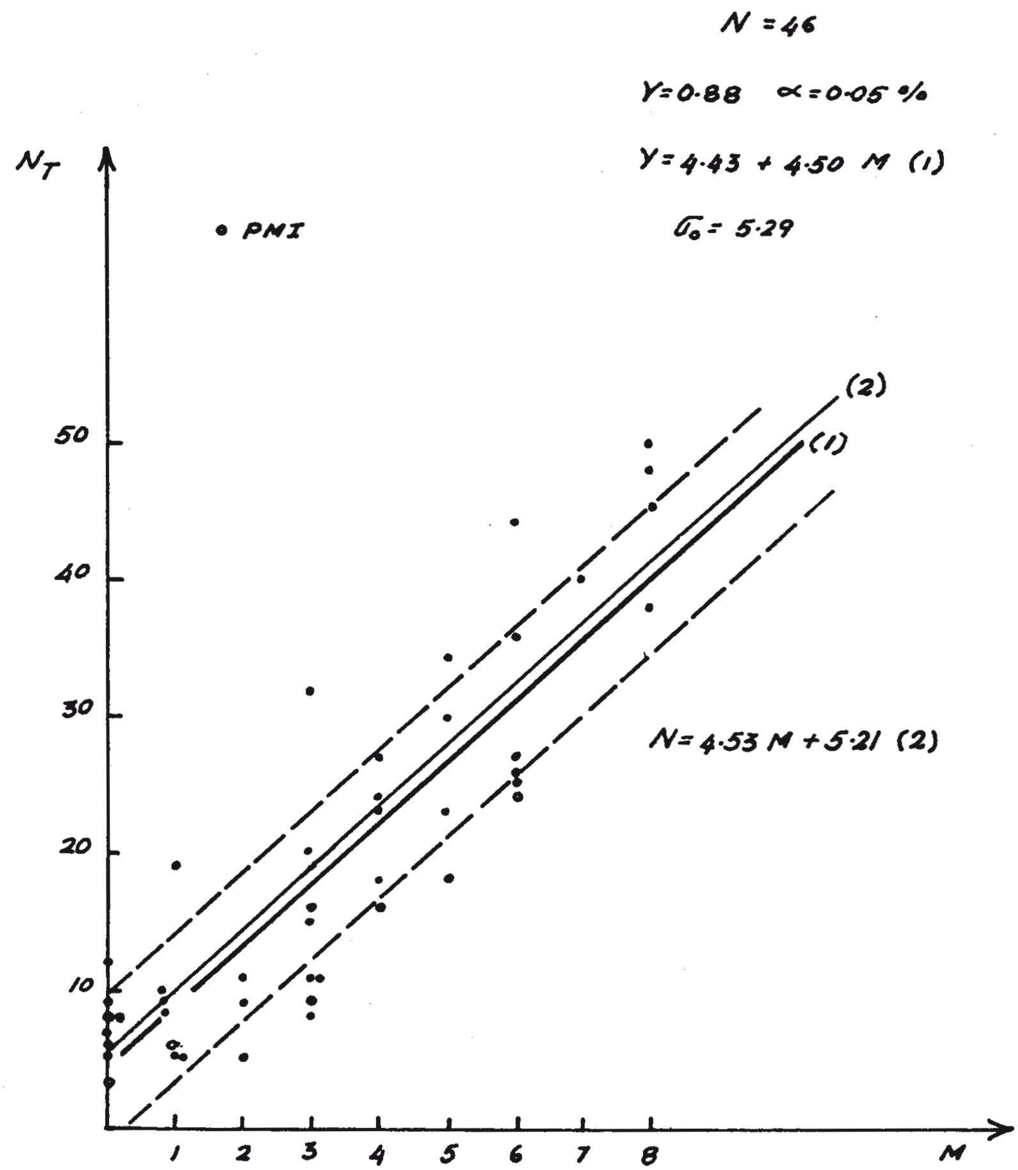

Fig. 1.

This was proved in a more explicite way by plotting M and IB separately versus the time of passage of the active center through the central meridian. No significant correlation was obtained.

2. Establishment of two practical rules for forecasting.

The following data elaboration was carried out for these two approaches: (a) taking as rules the relations (2) and (4) and evaluating them to obtain a forecast for the testing subsample (28 active centers) and (b) taking as rules the relations (1) and (3) and evaluating them for the training and testing subsamples (56 active centers).

For each case we constructed a contingency table between the values obtained and the forecast for intervals of 5 flares, because this is the standard error of the regression. Using Heidke's skill coefficient (Jakimec, 1980) and likelihood coefficient (Burov, 1981) we selected the best intervals in order to achieve the best forecast.

The following rules (see tables 1, 2 and 3) were achieved. Note that for each criterion the two approaches lead to the same results with respect to efficiency and number of intervals. For example, the criterion M permits the forecast for two and three intervals, the criterion IB only for two intervals. In the case of the M criterion for three intervals, the best results are obtained for a small number of flares and the worse one for a medium number of flares.

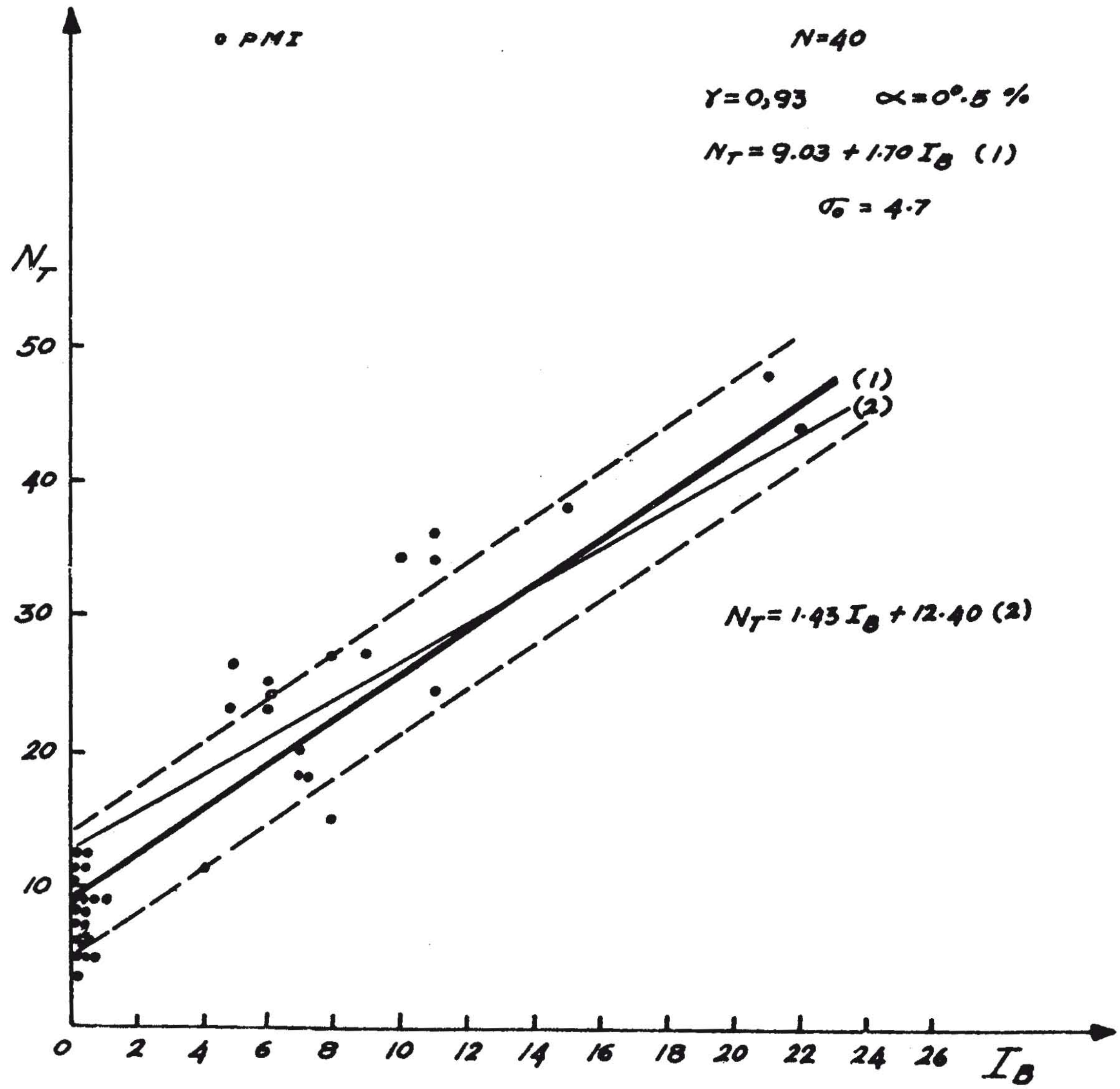

Fig. 2.

Table 1.

M Criterion for three intervals. Effectiveness 71%

If M = 0 or 1	N_{min} =	2 - 10
= 2,3 or 4	=	11 - 25
= 4	=	>25

Table 2

M. criterion for two intervals. Effectiveness 84%

If M = 0 to 4	N_{min} < 25
> 4	> 25

Table 3

IB criterion for two intervals. Effectiveness 73%

If IB = 0 to 5	NT< 20
6	~ 20
> 7	> 20

The comparison between M and IB criteria is possible for the case of two intervals. It turns out that the M criterion is better.

In all cases the obtained likelihood is higher than the blind forecast. It was necessary to prove this in order to show that our forecasting method is not trivial.

3. Testing a sufficient criterion for forecasting.

In López and Ferro, 1987 it was found that a sufficient condition for an active center to produce 20 or more flares is that the variable "area (A)" or "U (two-ribbon flare)" is informative with a weight equal to 1. When these variables are not informative with that weight the forecast is not possible.

These criteria were tested on the training and testing subsamples. They worked only for 26% of the cases, but when they did the effectiveness (number of correct forecasts) reached 80%.

4. Search for better forecasting variables.

On the basis of subsamples (López and Ferro, 1987) it was tested whether a better correlation exists between the variables "number of informative variables until a certain time" and "number of flares until that time". The correlation obtained was 89% and since the previous rule had shown a correlation of 88%, it is not practical to adopt this new forecasting criterion.

Also it was tested whether a better correlation exists between the total number of informative variables with weight one at the final stage of flare production by the active region, and the total number of different flares produced. For this case the correlation is 89% and again it is not practical to take this new regression instead of (1); i.e., it is better to forecast the total number of flares and not the total number of different flares.

ACKNOWLEDGMENTS

Our thanks go to Michele Monzon and Clara Serra of the Astronomical Department, for their work in data processing.

BIBLIOGRAPHY

BERNSHTEIN, P.B.and V.A. BUROV, 1976. *In:* Birth and evolution of active regions at the Sun, Nauka, Moscow, pp. 186-189.

BUROV, V.A. *Ibid,* 142-145.

BUROV, V.A. , 1981. *Phys. Solariterr.,* 16, 106-118.

JAKIMEC, W. *et al.*, 1980. *Solar Terrestrial Proc., 3,* p. C54.

LOPEZ, N. and I. FERRO, 1987. *Solar Data,* 3, 71-79.

BRAVO, A., 1983. *Revista Ciencias Matemáticas, IV,* 2, 123-144.

I. Ferro Ramos and N. López Reyes
Institute of Geophysics and Astronomy
Cuban Academy of Sciences
Av. 212 No. 2906
La Lisa, La Habana, Cuba.