

"APLICACION DE TECNICAS HEURISTICAS Y ESTADISTICAS A LA CARACTERIZACION Y PRONOSTICO DE LAS RAFAGAS SOLARES".

Isabel Ferro Ramos

Instituto de Geofísica y Astronomía
Academia de Ciencias de Cuba

RESUMEN

Se exponen los principales resultados obtenidos en el Departamento de Astronomía del IGA, Academia de Ciencias de Cuba sobre las ráfagas solares, luego de aplicar técnicas masivas de análisis de datos tales como análisis de arrumazón, reconocimiento de patrones en sus vertientes estadística y heurística, análisis de componentes principales, técnicas no paramétricas, generación automatizada de hipótesis y técnicas de inteligencia artificial, entre otras. Las mismas permitieron, por ejemplo, establecer diversas clasificaciones de las ráfagas, caracterizar aquellas que invaden la umbra de las manchas, definir el nuevo concepto de destello típico, poner de manifiesto la importancia de las variables en las bandas X y radio del espectro, y crear dos sistemas expertos para el pronóstico respectivo de la situación destellante y de la ocurrencia o no de ráfagas en las próximas 24 horas de la observación del grupo de manchas asociado. En la implementación práctica a través de un gran sistema experto que tenga como base de conocimiento estos y otros resultados de la literatura internacional, se comienza a trabajar actualmente con el objetivo de lograr un servicio de pronóstico de las ráfagas solares.

INTRODUCCION

El avance obtenido en los últimos años en el conocimiento de las ráfagas solares (Neiding 1986, Stepanov y Obridko 1987, Haish y Rodono 1989), unido a las posibilidades que la implementación computacional ha abierto en el empleo de técnicas heurísticas y estadísticas, ha hecho factible el pronóstico de las mismas con una efectividad entre 80 - 90 %. En el Departamento de Astronomía del IGA, Academia de Ciencias de Cuba, se vienen empleando desde hace unos 10 años, métodos heurísticos y estadísticos para el diagnóstico y pronóstico de esos fenómenos.

MATERIALES Y METODOS

Los datos sobre ráfagas solares fueron tomados de las publicaciones internacionales "Quarterly Bulletin on Solar Activity" y "Solar Geophysical Data", y los datos sobre manchas solares, utilizados muchas veces como índices pronosticadores, se tomaron del Servicio de Manchas Solares del IGA, ACC. También fueron utilizados los datos de ráfagas

recopilados por Kaastra 1985.

Los principales métodos utilizados fueron:

1.- Taxonomía numérica, a través del método de agrupamiento por pares ponderados (Davis 1973) mediante la coeficiente de similaridad de Gower (Gower 1971) y a través del método de Ward mediante la distancia euclidiana.

Se utilizaron además los métodos jerárquicos aglomerativos (Anderber 1973): conexión simple, conexión completa, centroide, mediana, promediación entre los grupos y promediación dentro del nuevo grupo.

2.- Métodos de la estadística no paramétrica como los ya tradicionales de Kolmogorov-Smirnov, criterio de los signos, prueba de las rachas (Gluber 1973) y métodos de correlación semiheurística para variables cualitativas (Ferro 1987)

3.- Reconocimiento de patrones a) en su vertiente heurística a través de la teoría de test (Zhurabliov 1978, Bravo 1983) y b) en su vertiente estadística a través del análisis discriminante (Davis 1973).

4.- Análisis de componente principales (Davis 1973).

5.- Generación automatizada de hipótesis de (método GUHA) mediante el cuantificador de asociación (Hajek y Valdés y de la Cruz 1984).

6.- Métodos de la inteligencia artificial mediante un algoritmo basado en la entropía y otros basados en los coeficientes λ asimétrico, λ simétrico y de reducción de error (Gil Chaviano 1989).

DESARROLLO Y RESULTADOS

Los métodos enumerados se aplicaron para tratar de investigar diversos aspectos de los destellos y lograr su pronóstico. Esos diversos temas, y los resultados principales obtenidos en cada uno de ellos, son:

1.-Clasificación de los destellos con uno y dos máximos de intensidad.

Se obtuvo una clasificación para los destellos que presentan un solo máximo de intensidad a través de las cinco variables ópticas más utilizadas, a saber: tiempo de subida al máximo, tiempo total de vida, tiempo relativo de subida al máximo, área en la escala ordinal acostumbrada y brillantez cualitativa. Las clases formadas fueron: destellos de larga duración, destellos brillantes de mediana duración, destellos brillantes de corta duración, destellos normales impulsivos, destellos normales de impulsividad media y destellos no brillantes lentos.

El porcentaje de buena clasificación resultó ser de 88 % (Ferro y Alfonso 1989a).

Se obtuvo además una clasificación para los destellos con dos máximos de intensidad a través de las mismas cinco variables anteriores más el tiempo de subida al segundo máximo y el tiempo relativo de subida a este máximo. Las clases formadas se expresan por combinaciones complejas de los estados de las variables, lo cual es consecuencia del carácter más complejo de estos destellos o de la insuficiencia de esas siete variables para describirlos (Ferro y Alfonso 1989b). El porcentaje de buena clasificación resultó ser de 73 %.

2.-Clasificación de los destellos de pequeña área.

Se consideró que para cada fenómeno las 29 variables más utilizadas por el Servicio Internacional del sol y reportadas en QBSA y SGD. Resultó que la principal variable diferenciante es la brillantez y no el área. Los destellos se dividen ante todo en normales (según la brillantez) y brillantes. La segunda variable diferenciante resultó ser la forma de surgimiento de los destellos: D, como un punto brillante, E, como varios puntos brillantes, y F, como varios centros eruptivos (Ferro y Alfonso 1986a).

3.-Clasificación de los destellos caracterizados por las variables de la gama óptica, X y de radio del espectro. Se obtuvo que aparecen dos grandes grupos: uno integrado por destellos con brillantez normal y moderadamente energéticos y otro por destellos brillantes y muy energéticos donde están representados el 69 % de los destellos de dos bandas de la muestra. Las variables temporales y el índice espectral energético no muestran diferencias para estos dos grupos (Ferro 1989e).

4. Establecimiento de relaciones entre variables cuantitativas y cualitativas de los destellos. Se obtuvo que los destellos D muestran raramente dos máximos, son de corta duración (5 - 25 min) y son lentos, o sea, el tiempo relativo de la fase relámpago es mayor que el 20 % del tiempo total de vida (Ferro y Alfonso 1986a). Se observó que los destellos E y F no difieren entre sí de acuerdo a los anteriores parámetros. Se obtuvieron diversas relaciones (Ferro 1989e) entre las variables ópticas, X y radio del espectro.

Las más notables son

a) Para los destellos grandes, valores altos de la temperatura en el X blando se relacionan con valores bajos de los tiempos de subida en H alfa.

b) Para los destellos grandes brillantes, valores altos de la temperatura en el X blando se relacionan con valores altos de los tiempos de subida en H alfa.

c) se probó una fórmula que relaciona la energía en el X con la frecuencia máxima centimétrica f_{max} , el flujo máximo centimétrico F_{max} , el índice espectral λ_d y el área en H alfa A. El coeficiente de correlación es de 90 %.

$$E = A^{0.99} f_{max}^{0.31} F_{max}^{0.24} 10^{(0.12-0.1d)}$$

5.-Estudios estructurales de la actividad destellante de los centros activos. Se obtuvo (Ferro y Alfonso 1986b, Ferro 1986) que la cantidad de tipos de en un centro activo es proporcional a la cantidad de destellos que produce. Se definió la similaridad destellante S de un centro activo como el mínimo nivel de similaridad al cual todos los destellos integran un mismo grupo, o sea S expresa el nivel de parecido de todos los destellos del centro activo y resultó inversamente proporcional a la cantidad de destellos que produce el centro activo y no está relacionado con la edad del mismo.

Se halló que los centros activos con polaridad magnética irregular (PMI) presentan un valor bajo de S, o sea, producen destellos muy disímiles y son así también irregulares en sentido estructural.

Se obtuvo que un centro activo puede ser el asiento de destellos muy parecidos como muy diferentes, o sea porque los destellos surjan en una región contigua en el Sol, ellos necesariamente no tienen por qué parecerse: pueden ser incluso tan diferentes como los que produce todo el Sol en un intervalo largo de tiempo (1 año).

Se definió el destello típico de un centro activo mediante la introducción de una nueva aproximación algorítmica como aquel más parecido a todos los restantes tomados a la vez, cumple que en el caso de los grupos bipolares simples, ese destello definido como típico, se sitúa a lo largo del eje que conecta las dos manchas y tiende a ocupar una posición central. Temporalmente no se le halló regularidad alguna.

6.-Estudio de los destellos que cubren o invaden la umbra de las manchas (destellos Z).

Se obtuvo (Ferro 1985) que los destellos Z son estadísticamente más brillantes y grandes que los no Z. Presentan con más frecuencia dos máximos y los centros activos donde surgen duran más y son más productivos en destellos. Los centros activos Z producen por lo general un solo destello Z con igual probabilidad de surgimiento antes o después del máximo de

desarrollo del centro activo. Y si el centro activo produjo varios destellos Z, ellos se distribuyen aleatoriamente entre el número total de los restantes destellos.

Se obtuvo además (Ferro 1987) que las variables más relacionadas con los destellos Z son: F, presencia de varios centros eruptivos; I, región activa muy extendida, H, erupción acompañada de un filamento oscuro.

7.-Estudio de las curvas de luz de los destellos.

Se obtuvo una clasificación de las curvas de luz de los destellos con un máximo de brillantez en cinco tipos: subdestellos normales, destellos brillantes y destellos Z y/o U y/o S brillantes.

Los parámetros temporales más importantes en las curvas de luz para el caso de los destellos con un máximo fueron el tiempo total de vida y el tiempo relativo de subida al máximo, independientemente del área y de la característica Z del destello. Para el caso de los destellos dobles, la importancia de las variables es diferente y depende del área y de la característica Z del destello.

De nuevo la brillantez resultó una variable diferenciante pues para los destellos dobles que presentan un tiempo de subida al máximo menor que 18 min, resulta que éste es proporcional al tiempo relativo de subida al segundo máximo, mientras que para aquellos con $t_1 > 18$ min, la proporcionalidad es inversa y en lo fundamental se crea debido a los destellos de brillantez B.

El único parámetro de las curvas de luz que es sensible a la fase del ciclo es la duración relativa de la fase relámpago: los destellos en el mínimo de actividad son más lentos (Ferro 1987).

8.-Informatividad o importancia de las variables de manchas relacionadas con los destellos.

Por algoritmos de inteligencia artificial se obtuvo para cada clase de Zurich, las variables que más y que menos condicionaban la presencia de destello. Las que más relacionadas están con los destellos son: para la clase A: la variación del número de poros y el área del grupo; para la clase B, la variación del área del grupo y el área de la mancha mayor; para la clase C, la fase del ciclo, la población y el área del grupo; para la clase D, el área del grupo, el tipo de penumbra, la población y la fase del ciclo; para la clase E, el tipo de penumbra y para la clase H, la fase del ciclo, el tipo de penumbra y el área del grupo.

El análisis también indica que ni el número de poros, ni la variación del área de la mancha mayor, están entre las variables con más participación en la producción de destellos.

9.-Informatividad o importancia de las variables propias de los destellos.

Se obtuvo (López y Ferro 1987) que de entre 14 variables propias de los destellos, reportadas por el Servicio Internacional del Sol, la F (presencia de varios centros eruptivos), el tiempo relativo de la fase relámpago y la brillantez, son las más importantes para diferenciar los destellos tomados por todo el Sol. El peso informacional de la variable área es mucho menor (0.14) que el de la variable brillantez (0.71), lo que ya de forma explícita indica su relativa importancia.

El estudio evolutivo de la informatividad de las variables a medida que el centro activo produce 1, 2, ..., N destellos, mostró que cuando cualquier variable se hace informativa con peso igual 1, sigue con ese peso durante todo el desarrollo del centro activo. Esto dio lugar para encontrar relaciones regresivas que se presentan en el siguiente epígrafe.

10.-Métodos de pronósticos

a) Mediante análisis discriminantes

Se obtuvieron las ecuaciones discriminantes lineales que permiten pronosticar si un grupo de manchas producirá destellos o no en las próximas 24 horas. La clase más fácil de pronosticar resultó ser la E, con una efectividad del 80 % pero por debajo del pronóstico ciego y la más difícil de pronosticar es la A con una efectividad de sólo 69 % pero significativa.

b) Mediante teoría de test.

Se obtuvieron ecuaciones regresivas entre la cantidad de destellos esperada en un momento dado que va produciendo una región activa (N) y la cantidad total de variables informativas con peso igual a uno hasta un momento dado (M). Y además, entre la cantidad de destellos que producirá la región activa en su paso por el disco y IB, el número de orden del destello diferente entre los producidos para el cual la variable brillantez comienza a hacerse informativa con peso igual uno. Se obtuvieron las siguientes reglas:

Si M es 0 ó 1, Nmin es 2-10; si M es 2-4, Nmin es 11-25 y si M es 4, Nmin es mayor que 25. Este criterio llamado M opera con una efectividad de 71 %.

El segundo criterio llamado IB estipula que si IB es de 0 a 5, NT es menor que 20; si IB es igual a 6, NT es aproximadamente 20 y si IB es mayor o igual a 7, NT es mayor que 20. Este criterio posee una eficiencia de 73 %.

c) Mediante los sistemas expertos FUL y FULGOR.

El primero de ellos (Ferro, Pérez Foval y Cepero 1991) pronostica si existe o no una situación destellante en el centro activo y el segundo, si ocurrirán destellos en las próximas 24 horas. Ambos operan sobre nueve variables de manchas y la fase del ciclo. Se logra una efectividad de 80 % para las clases de Zurich C y D. Para las restantes clases los porcentajes son

aún más altos pero no superiores al mejor de los pronósticos triviales, el pronóstico ciego.

CONCLUSIONES

1. Un resultado al parecer de interés es la importancia de la variable brillantez por sobre la variable área del destello. Es conocido desde tiempo atrás (Smith y Smith 1966) que estas variables correlacionan bien una con otra, no obstante se ha considerado como primaria y más significativa a la variable área (De Jager 1975). Esto se trasluce en los supuestos del programa CINOF y en que hace unos años los boletines de

actividad solar dejaron de publicar datos sobre el valor cuantitativo de la brillantez.

2. El método para pronosticar la cantidad total de destellos que producirá una región activa en su paso por el disco, es aún burdo, pero no conocemos que antes se haya intentado pronosticar esta magnitud.

3. El método de pronóstico basado en sistemas expertos es el más perspectivo de todos y actualmente se trabaja para lograr una base de conocimiento creada por algoritmos de aprendizaje automático y que incluya a la vez resultados de la literatura internacional.

BIBLIOGRAFIA

- Anderberg M. R., 1973, "Cluster analysis for applications", N.Y. Acad. Press, 372 pp.
- Bravo, A. 1983, Rev. Ciencias Matemáticas, vol. 4, No. 2, pp. 123 - 144.
- De Jager, C., 1975, Solar Physics, vol. 40, No. 1, pp 133 - 140.
- Ferro Ramos, I., 1985, Datos Solares No. 6, pp. 87 - 92.
- Ferro Ramos, I., 1986, Datos Solares No. 9, pp. 37 - 50.
- Ferro Ramos, I., 1987a, Datos Solares No. 5, pp. 81 - 91.
- Ferro Ramos, I., 1987b, Datos Solares No. 7, pp. 69 - 74.
- Ferro Ramos, I., 1989, Comunicaciones Cient. sobre Geofísica y Astronomía, No. 20, 15 pp
- Ferro Ramos, I., Alfonso Sosa J., 1986a, Datos Solares, No. 2, pp. 73 - 80
- Ferro Ramos, I., Alfonso Sosa J., 1989b, Datos Solares, No. 6, pp. 47 - 52
- Ferro Ramos, I., Alfonso Sosa J., 1989a, Comunicaciones Cient. sobre Geofísica y Astronomía, No. 6, 17 pp.
- Ferro Ramos, I., Alfonso Sosa J., 1989b, Comunicaciones Cient. sobre Geofísica y Astronomía No. 7, 13 pp.
- Gil Chaviano M., 1989, Tesis de grado, Escuela de Matemáticas, Universidad de la Habana.
- Gubler, E. V., Genkin, A. A. 1973, "Utilización de criterios no paramétricos en las investigaciones médico-biológicas. Leningrado, Medicina, 141 pp.
- Haish B., Rodono M., 1989, "Solar and Stellar Flares", Kluwer Ac. Pb., 511 pp
- Hajek P., Valdés J. J., de la Cruz A. V., 1984, Matematicky Ustav, Cekoslovenska Akademie Ved., No.9, 27 pp.
- Kaastra J. S., 1985, "Solar Flares: an electrodynamic model", Utrecht, 184 pp
- López Reyes N., Ferro Ramos, I., 1987, Datos Solares No. 3, pp. 71 - 79
- Neiding D. F. ed., 1986 "The lower atmosphere of solar flares". Proceed. NSO, Solar Maximum Mission, 503 pp.
- Smith G. J., Smith E.P. 1966, "Destellos solares", Mir, Moscú, 426 pp.
- Stepanov V. E., Obridko V.N., 1987, "Solar Maximum Analysis, VNU Science Press, 457 pp.
- Zhuravliov Y. U., 1978, Problemas de cibernética, No. 33