PHÂN TÍCH PHỔ : MỘT CÔNG CỤ NGHIÊN CỨU TINH TẾ TRONG ĐIỀU TRA CƠ BẢN

Nguyễn Thuyết - Phòng Khoa học kỹ thuật

I - Bản chất và ý nghĩa của phương pháp phân tích phổ

Từ trước đến nay, người ta thường sử dụng toán thống kê để nghiên cứu và khai thác các chuỗi số liệu (coi như một hàm rời rạc theo thời gian hoặc theo không gian của các yếu tố đang được khảo sát). Vì vậy, ta thường rất quen thuộc với các đặc trưng thống kê như kỳ vọng toán, phương sai, hệ số tứ phân vị v.v... trong việc nghiên cứu các đặc tính của chuỗi số liệu hoặc như tần suất, hàm hiệp biến và tự hiệp biến v.v... trong việc khai thác chúng. Tóm lại, trong việc nghiên cứu và khai thác các chuỗi đo đạc, ta thường dùng chúng dưới dạng thể hiện là một hàm thời gian hoặc không gian. Thuật ngữ tần số thường ít gặp và được thay thế bằng khái niệm chu kỳ.

Khoảng vài chục năm gần đây, một đại lượng thống kê mới đang xuất hiện ngày càng phổ biến trong công tác nghiên cứu và khai thác các chuỗi quan trắc. Đó là phổ của các quá trình tự nhiên. Và phương pháp phân tích phổ đã trở thành một công cụ tinh tế trong điều tra cơ bản.

Thực ra, phổ của một đại lượng vật lý không phải là một khái niệm mới mẻ, mà chúng ta đã quá quen thuộc trong quang học và trong nhiều lãnh vực vật lý khác. Phổ của một đại lượng vật lý chính là cấu trúc sóng của đại lượng ấy. Trong điều tra cơ bản, người ta quan niệm một quá trình tự nhiên không diễn biến theo qui luật tuyến tính mà diễn ra từng đợt như một quá trình sóng. Vì vậy ứng dụng phântích phổ trong điều tra cơ bản có nghĩa là nghiên cứu cấu trúc sóng của một quátrình tự nhiên (nếu khảo sát phổ của một hàm thời gian hoặc hàm không gian), hay cấu trúc sóng của một trường yếu tố tự nhiên (nếu khảo sát phổ hai thứ nguyên "không - thời gian" của yếu tố ấy).

Phổ của một quá trình của một yếu tố tự nhiên, là một đại lượng thống kê suy ra từ một hàm thời gian hoặc không gian của yếu tố ấy, bằng một phép biến đổi toán học : phép biến đổi Fourier. Biến đổi một quá trình theo thời gian hoặc không gian thành một quá trình sóng không phải là cộng thêm vào một điều gì mới mẻ, mà chỉ là sắp xếp lại chuỗi quan trắc theo thứ tự sóng, thay cho theo thứ tự thời gian hoặc không gian. Dạng thể hiện mới này rõ ràng sáng sủa hơn, vì nó phù hợp với bản chất sóng của các quá trình tự nhiên.

Như vậy, trên một đồ thị phổ, những chỗ nhô cao ứng với một sóng thành phần của quá trình. Trong các kỹ thuật xử lý chuỗi số liệu, đã có phương pháp phân tích điều hòa ; tuy nhiên công cụ này thích hợp với các quá trình điều hòa,nghĩalà những quá trình có các chu kỳ rõ rệt. Phân tích phổ cho phép nghiên cứu kỹ hơn, chi tiết hơn, phát hiện được cả những chuẩn chu kỳ ứng với các bước sóng có đóng góp vào sự tạo thành cấu trúc sóng của quá trình dù là nhỏ.

Với cách tiếp cận như vậy, phổ là một hàm của biến tần số, nó có thể là một đại lượng vật lý bất kỳ. Trong công tác điều tra cơ bản nói chung, trong khí tượng, thủy văn và hải văn nói riêng, người ta hay sử dụng phổ của công suất, gọi là phổ công suất hay phổ năng lượng (công suất là một đại lượng tỉ lệ với bình phương biên độ, được định nghĩa bởi biểu thức :

$$\lim \frac{1}{T} \int_{-T/2}^{T/2} | f(t) |^2 dt.$$

trong đó f(t) là một hàm thời gian, $|f(t)|^2$ là công suất tức thời của f (t) và $|f(t)|^2$ dt là năng lượng toàn phần của f(t)).

II - Vài nét về kỹ thuật phân tích phổ

1. Hàm thời gian f(t) và phổ F (ω) của nó là một cặp biến đổi Fourier.

$$f(t) \longleftrightarrow F(\omega)$$

- Biến đổi Fourier thuận là :

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) \, \mathrm{Exp.}(-i\omega t) \, dt \qquad (1)$$

- Biến đổi Fourier ngược là :

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) \, \mathrm{Exp.}(i\omega t) \, d\omega \qquad (2)$$

Phương trình (1) biểu thị sự phân tích Fourier của f (t) và ngược lại phương trình (2) biểu thị sự tổng hợp Fourier của f (t), nghĩa là tổng hợp các thành phần phổ khác nhau thành hàm gốc f(t). Trong thực tế, chúng ta thường có sẵn hàm thời gian f (t) (tức chuỗi đo đạc), và tính toán phổ của nó chính là tính các thành phần của sự phân tích Fourier của nó. Biến đổi Fourier là một loại đặc biệt của biến đổi tích phân. Phương trình (1) là phương trình tích phân Fredhalm loại 1, với nhân Exp. ($-i\omega t$), còn phương trình (2) là nghiệm của phương trình (1), và ngược lại.

2. Từ định lý Parseval, người ta suy ra phổ công suất E (ω) là biến đổi Fourier thuận của hàm tự hiệp biến C_{11} (τ). (τ là số bước dịch chuyển).

$$C_{11}(\tau) \longleftrightarrow E(\omega)$$

Từ đó suy ra một cách tính ước lượng phổ công suất, gọi là phương pháp gián tiếp như sau :

a/- Tính hàm tự hiệp biến C_{11} (τ) của f(t), với số bước dịch chuyển tối đa là m (τ_{max} = m).

b/- Tính biến đổi Fourier của chuỗi m giá trị của hàm tự hiệp biến $C_{11}(\tau)$.

3. Với một chuỗi quan trắc có N + 1 số liệu, bước thời gian (khoảng cách thời gian giữa 2 quan trắc liên tiếp) là Δt, nếu ta tính phổ công suất từ chuỗi m giá trị của hàm tự hiệp biến, thì sẽ có các đặc trưng sau :

a/- Độ dài của trường số liệu : $T = N \cdot \Delta t$ (3)

b/- Tần số cao nhất của phổ là $\nu_m = \frac{1}{2 \cdot \Delta t} = \frac{N}{2T}$ (4)
Tần số này còn có tên riêng là tần số Nyquist.

c/- Tần số ứng với sóng thứ i là : $\nu_i = \frac{i\,\nu_m}{m} = \frac{i}{2m \cdot \Delta t}$ (5)

d/- Khả năng phân giải của phép phân tích phổ là bước tần số Δf, hay hiệu tần số của 2 sóng liên tiếp là :

$$\Delta f = \frac{1}{2m \cdot \Delta t} \tag{6}$$

e/- Sai số quân phương chuẩn của phép tính phổ là :

$$\varepsilon = \left(\frac{m}{N}\right)^{\frac{1}{2}} \tag{7}$$

Từ các đặc trưng trên suy ra những điều cần lưu ý về việc xác định các giá trị của m, Δt và T như sau :

3.1 - Xác định giá trị của m :

Từ các công thức (6) và (7) ta thấy rõ ràng giá trị của m có liên quan đến khả năng phân giải của phép phân tích phổ và đến sai số quân phương chuẩn của phép tính phổ nghĩa là đến độ tin cậy của ước lượng phổ tính được.

- Theo (6), hiển nhiên m càng lớn, Δf càng nhỏ, khả năng phân giải càng cao.
- Ngược lại, theo (7), m càng lớn, ước lượng phổ tính được càng ít tin cậy hơn.

Hai đòi hỏi này trái ngược nhau, vì vậy khi xác định m, trong thực hành phải chọn biện pháp thỏa hiệp hai đòi hỏi trên. Hơn nữa, theo (5), tần số thấp nhất ứng với bước sóng thứ 1 là $\nu_1 = \frac{1}{2m \cdot \Delta t} = \frac{1}{2T}$. Chu kỳ T_1 ứng với bước sóng thứ 1 là :

$$T_1 = \frac{1}{\nu_1} = 2m \cdot \Delta t \tag{8}$$

Nếu ta chọn $m = \frac{N}{2}$ thì chu kỳ dài nhất T_1 trong phổ tính được vừa bằng độ dài trường số liệu T, điều đó không đủ tin cậy, nên trong thực hành không được phép lấy $m = \frac{N}{2}$. Trong thực hành, theo kinh nghiệm của các nhà nghiên cứu về phổ thì tùy thuộc vào giá trị của N, khoảng giá trị của m tốt nhất là :

$$\frac{N}{10} \leqslant m \leqslant \frac{N}{5} \tag{9}$$

3.2 - Xác định giá trị của Δt.

Từ (4) và (6), ta thấy việc chọn Δt phải tùy thuộc vào mục đích nghiên cứu vì nó ảnh hưởng đến tần số cao nhất của phổ tính được, và đến khả năng phân giải của phép phân tích.

– Theo (4) thì chu kỳ ngắn nhất (ứng với sóng có tần số cao nhất) của phổ tính được dài bằng 2 lần bước thời gian Δt. Như vậy, với mục đích khảo sát các dao động chu kỳ ngắn ta phải chọn bước thời gian Δt nhỏ, thí dụ khi phân tích phổ để khảo sát các sóng thành phần của thủy triều, ta phải chọn $\Delta t = 1$ giờ và nhỏ hơn nữa ; còn khi nghiên cứu các qui luật biến đổi dài năm, ta phải chọn Δt lớn.

– Theo (6) thì nếu Δt càng lớn, Δf càng nhỏ và khảnăng phân giải càng cao. Tuy nhiên không phải ta được quyền tự do chọn Δt cho thích hợp với khả năng phân giải mong muốn, mà nó bị ràng buộc bởi những chu kỳ dao động mà ta cần khảo sát như ở trên đã nói.

3.3 – Xác định giá trị của T.

Việc chọn giá trị của T phụ thuộc nhiều yếu tố. Trước hết như đã biết, với giá trị Δt đã xác định, T càng ngắn, nghĩa là N càng nhỏ thì m càng nhỏ (theo bất đẳng thức 9), khả năng phân giải sẽ thấp. Tuy nhiên, khi đã đạt được những yêu cầu về khả năng phân giải và độ tin cậy của ước lượng phổ tính được rồi, nghĩa là đã xác định được giá trị của m và Δt rồi thì T cũng chỉ cần có độ dài thích hợp (Theo (9) sẽ là không vượt quá $10\, m . \Delta t$) ; vì có dài hơn cũng không đóng góp gì thêm cho khả năng phân tích. Trong thực tế, T thường bị giới hạn bởi chuỗi đo đạc ngắn ; và theo lý thuyết thì phổ đúng phải là phổ của một chuỗi thời gian vô hạn, với T hữu hạn nhất định sẽ có ảnh hưởng đến phổ tính được, thường gây hiện tượng thẩm thấu năng lượng, mà kỹ thuật phân tích phổ phải giải quyết hiện tượng này bằng những hàm cửa sổ (hay hàm làm trơn).

4. Hàm cửa sổ :

Trong thực tế, chúng ta dùng dạng thể hiện rời rạc của quá trình tự nhiên là chuỗi đo đạc để tính phổ. Các chuỗi này dù dài hay ngắn đều bị giới hạn, và kết quả là gây hiện tượng thẩm thấu năng lượng, thể hiện ở chỗ cho những giátrị âm của ước lượng phổ. Kỹ thuật dùng hàm cửa sổ là nhằm mục đích khử bỏ hiệntượng thẩm thấu năng lượng này, tức là loại bỏ những giá trị âm của ước lượng phổ.

Có nhiều loại hàm cửa sổ, nhưng trong thực hành thường dùng hai loại hàm cửa sổ Hanning – Tukey và hàm cửa sổ Hamming.

– Hàm Hanning – Tukey có dạng :

$$W(t) = \begin{cases} \frac{1}{2}\left(1 + \cos \frac{\pi \tau}{m}\right) & \text{với } 0 \leq \tau \leq m \\ 0 & \tau > m \end{cases} \qquad (10)$$

– Hàm Hamming có dạng :

$$W(t) = \begin{cases} \left(0{,}54 + 0{,}46 \frac{\pi \tau}{m}\right) & \text{với } 0 \leq \tau \leq m \\ 0 & \tau > m \end{cases} \qquad (11)$$

Dùng hai hàm cửa sổ này dẫn đến việc làm trơn các ước phổ tính được bằng một phép tính lấy trung bình có trọng lượng.

- Theo Hanning - Tukey có các hệ số trọng lượng như sau :

$$\widetilde{E}_i = 0,25\ E_{i-1} + 0,5\ E_i + 0,25\ E_{i+1} \qquad (12)$$

- Theo Hamming có các hệ số trọng lương :

$$\widetilde{E}_i = 0,23\ E_{i-1} + 0,54\ E_i + 0,23\ E_{i+1} \qquad (13)$$

5. Hàm lọc :

Kỹ thuật tinh tế nhất của phép phân tích phổ là việc dùng hàm lọc. Mục đích chủ yếu của việc sử dụng hàm lọc là để đặc tả dải tần số mà mình quantâm nhất trong phép phân tích phổ, hoặc để loại bỏ những dao động do nhiễu, không phải là bản chất của quá trình tự nhiên. Cần phân biệt, tránh nhầm lẫn hàm lọc với hàm cửa sổ. Hàm lọc được sử dụng ngay trong bước chuẩn bị số liệu, tạo ra một tập số liệu mới trong đó đã loại bỏ hoặc giảm bớt những tính chất mà mình không cần quan tâm, nhằm mục đích làm nổi rõ những đỉnh phổ mà mình quan tâm. Còn hàm cửa sổ được sử dụng khi tính ước lượng phổ để làm trơn, loại bỏ giá trị âm của ướclượng phổ.Trong một phép phân tích phổ có thể cũng sử dụng cả hàm lọc và hàm cửa sổ, và bao giờ phép lọc cũng phải thực hiện trước.

Với bản chất như vậy, phép lọc có phân thành 3 loại :

- Lọc bỏ tần số cao, cho phép loại bỏ các dao động có chu kỳ ngắn,thường là những nhiễu xen lẫn vào các sóng thành phần của hiện tượng.

- Lọc bỏ tần số thấp, cho phép loại bỏ các dao động chu kỳ dài, điều đó cho phép loại bỏ hoặc làm giảm bớt những thành phần biến đổi dài năm để tập trung nghiên cứu những thành phần biến đổi ngắn ngày của hiện tượng.

- Lọc bỏ cả 2 loại tần số cao và thấp chỉ giữ lại các sóng trung gian.

Cũng như hàm cửa sổ, hàm lọc có nhiều loại. Ở đây chúngtôi chỉ giới thiệu một hàm lọc mà chúng tôi đã dùng để viết chương trình tính phổ bằng ngôn ngữ FORTRAN. Hàm lọc này là một phép tính trung bình trượt có trọng lượng mà hệ số trọng lượng là các hệ số khai triển của nhị thức Niutơn :

$$W_n = \frac{k\,!}{n\,!\ (k-n)\,!}$$

Trong đó, k - bậc của nhị thức ,

$n = 0, 1, \ldots\ldots k.$

Giả sử ta có tập số liệu đo đạc $\{X_i\}$; $i = 1, 2, \ldots \ldots N$, thì phép lọcbỏ tần số cao cho ta một tập mới $\{X_i'\}$ mà :

$$X_i' = \frac{\sum\limits_{n=0}^{k} W_n\ X_{i+n-\frac{k}{2}}}{\sum\limits_{n=0}^{k} W_n} \qquad (14)$$

Tập số liệu lọc bỏ tần số thấp $\{\widetilde{X}_i\}$ có được bằng cách tính :

$$\widetilde{X}_i = X_i - X_i' \qquad (15)$$

Với tập số liệu X_i nếu ta lọc nhẹ một lần nữa bằng tập hệ số $\{W_n\}$ sẽ được tập số liệu $\{X_i''\}$ chỉ còn các sóng trung gian.

Trong phép lọc này, k nhận các giá trị chẵn và thuật ngữ "lọc nhẹ" được hiểu là lọc với giá trị của k nhỏ hơn.

Cần lưu ý thuật ngữ tần số cao và thấp ở đây không phải với ýnghĩa tuyệtđối, mà là tương đối so với giá trị Δt. Những sóng có tần số cao là những sóng có tần số lân cận tần số Nyquist của phép phân tích phổ.

x

x x

Chúng tôi đã xây dựng được một chương trình tính phổ một thứ nguyên theo các kỹ thuật và thuật toán đã dẫn ở trên. Ngoài ra còn có các bộ lọc khác như các bộ lọc để thể hiện các sóng nằm trong dải tần 0 - 0,8 cpd,và thể hiện các sóng bội bậc chẵn hoặc bậc lẻ của các sóng nhật triều và bán nhật triều, thích hợp cho các phép phân tích phổ các quá trình thủy triều (thuật toán các l` lọc này được trình bày trong cuốn The analysis of tides của Babriel Godin).Một bộ lọc theo phương pháp lấy trung bình trượt có trọng lượng, nhưng hệ số trọng lượng W_n là các giá trị của hàm phân bố Gauss cũng đã được viết thành một chương trình con trong bộ chương trình chung. Về hàm cửa sổ, ngoài hàm cửa sổ Hanning - Tukey còn có các chương trình con viết cho các hàm cửa sổ của Parden và của Maruyama.

Chương trình không những dùng để tính phổ đơn của một quá trình, mà còn để tính phổ giao của hai quá trình X (t) và Y (t) nhằm mụcđích nghiên cứu mối liên kết giữa các sóng thành phần của 2 quá trình này. Việc điều khiển tính phổ đơn hay phổ giao được thực hiện bằng cách so sánh giá trị của biến hình thức KHOÁ. Chương trình còn được viết một cách mềm dẻo để có thể tùy ý chọn phương án tính phổ có lọc hay không lọc số liệu, hoặc lọc các loại khác nhau. Điều này được thực hiện bằng phép gán giá trị cho biến MR. Chương trình đã được dùng để tính phổ của nhiệtđộ và lượng mưa ở Hà nội, phổ của các hệ số hoàn lưu kinh hướng Katxơ.Kết quả được chấp nhận và cho phép thấy được những chuẩn chu kỳ phù hợp với công trình nghiên cứu của nhiều tác giả.

Phân tích phổ là một công cụ nghiên cứu tinh tế,nó tỏ ra "mạnh" hơn các phép phân tích được dùng trước nó. Tuy nhiên, nó vẫn chỉ là một công cụ toán học, cho phép phát hiện một cách khách quan các chuẩn chu kỳ thông qua việc xácđịnh các đỉnh trong đồ thị phổ. Song phần giải thích cơ chế vật lý của các chuẩn chu kỳ này lại là nhiệm vụ của các nhà chuyên môn, vì thế việc kết hợp giữa các cánbộ toán với cán bộ chuyên môn là rất cần thiết và "lý thú" trong công tác phân tích phổ. Mặt khác, cần tuân thủ nghiêm ngặt những qui định trong kỹ thuật phân tích để tránh các "đỉnh phổ giả" trong tính toán.

Trên đây giới thiệu một số kinh nghiệm trong quá trình làm phân tích phổ.Một số công thức quen thuộc dễ dàng tìm thấy trong các cuốn "Some applications of statistics to Meteorology" của Hans. A. Panofsky và Glenn W. Brier, và "Measurement and anelysis of random data" của Iulius S. Bendatt và Allan G. Piersol (2 cuốn này đều có bản dịch tiếng Nga ở Thư viện Tổng cục). Các bạn nào muốn tham khảo chi tiết về lý thuyết phân tích phổ có thể tìm đọc cuốn "Spectral analysis in Geophysics"của Markus Bath./.