Chương 6

QUAN HỆ THỐNG KÊ GIỮA CÁC BIẾN THUỶ VĂN

6.1. TỔNG QUAN

Hiện tượng thuỷ văn thường thường được hình thành bởi rất nhiều yếu tố, trong thực tế không thể xét đầy đủ được các yếu tố đó, trong nhiều trường hợp cũng không cần thiết phải xét như vậy. Vì thế khi xây dựng các mối quan hệ nhân quả chỉ cần phân tích những nhân tố về mặt định tính có thể xem như là chính đối với quá trình hình thành đặc trưng thuỷ văn nghiên cứu. Những nhân tố chính này quy định dạng cơ bản của mối quan hệ, còn những nhân tố khác không quan trọng bằng sẽ tạo nên môi trường phân tán đặc trưng cho mối quan hệ ngẫu nhiên.

Ngay cả trong các trường hợp khi mối quan hệ giữa các biến lượng nghiên cứu, thực chất là hàm số (điều này trong thực tế nghiên cứu thuỷ văn rất ít thấy), mối quan hệ được xây dựng theo tài liệu quan trắc sẽ không cho ta lời giải đơn trị, là do sai số đo đạc ngẫu nhiên được đưa vào mối quan hệ.

Vì lẽ đó, mà các nhà thuỷ văn thường không gặp quan hệ hàm số mà gặp những quan hệ thống kê, trong đó ứng với mỗi giá trị của đại lượng được lấy làm biến lượng độc lập sẽ có một tập hợp vô hạn những giá trị của đại lượng kia (hàm sẽ được mô tả bằng đường phân phối có điều kiện). Các đường phân phối có điều kiện sẽ thay đổi theo sự thay đổi của biến lượng độc lập. Mối quan hệ thống kê được ứng dụng rất rộng rãi trong mọi lĩnh vực thuỷ văn. Mối quan hệ này phải dựa vào các phương pháp đo đạc dòng chảy, và dựa vào các quan hệ này mà xây dựng các lược đồ tính toán, dự báo thuỷ văn. Mối quan hệ giữa dòng chảy sông ngòi và lượng mưa, mối quan hệ giữa lưu lượng hay mực nước ở các trạm quan trắc khác nhau trên một con sông đường lưu lượng, mối quan hệ dòng chảy của các sông nằm trong vùng đồng nhất về điều kiện địa vật lýv v ... đều là những thì dụ về việc sử dụng mối quan hệ thống kê trong thuỷ văn. Việc nâng cao mức độ phân tích khoa học các quá trình thuỷ văn, việc hoàn thiện các phương pháp toán học khái quát hoá các chuỗi thống kê vả việc sử dụng MTĐT đã tạo ra khả năng phát triển nhanh chóng sử dụng mối quan hệ thống kê vào nghiên cứu thuỷ văn.

Để khái quát hoá khái niệm quan hệ thống kê người ta sử dụng khái niệm quan hệ ngẫu nhiên, mà ứng với chúng không phải là chuỗi thống kê mẫu, mà là tập hợp đầy đủ các giá trị ngẫu nhiên nghiên cứu, khi dung lượng của chuỗi n tiến tới vô hạn hay đến một số hữu hạn N bao gồm toàn bộ khoảng biến thiên của biến lượng.

Như vậy, dung lượng mẫu càng lớn, mối quan hệ thống kê thực nghiệm càng tiến tới (xem như giới hạn của mình) quan hệ ngẫu nhiên. Việc khái quát hoá này cũng tương tự như khái quát hoá tần số thực nghiệm bằng khái niệm xác suất.

Khi giấu các biến lượng ngẫu nhiên x và y có mối quan hệ thống kê thì phân phối có điều kiện của biến ngẫu nhiên y thay đổi theo sự biến thiên của x. Ta nhớ rằng lượng phân phối có điều kiện của biến ngẫu nhiên y là luật phân phối của nó nhận được với điều kiện của biến ngẫu nhiên x nhận một giá trị xác định x_i. Rõ ràng là khái niệm luật phân phối có điều kiện có ý nghĩa nếu như xét đồng thời kê (x,y).

Khi giải quyết các bài toàn thực tế thường thì đề cập để những mẫu ngẫu nhiên có dung lượng nào đó được rút ra từ tổng thể. Điều đó có nghĩa là thường không xét mối quan hệ ngẫu nhiên mà là mối quan hệ thống kê. Ngoài ra khả năng sử dụng các quan hệ thống kê để dự báo và tính toán thuỷ văn phải được căn cứ vào giả thiết là ước lượng thích đáng của các mối quan hệ đó sẽ cho phép nhận được kết luận về quan hệ ngẫu nhiên có cơ sở chắc chắn. Một điều kiện quan trọng cho phép ta sử dụng các quan hệ thống kê để dự báo và tính toán các đặc trưng của chế độ thuỷ văn là sự cháp nhận giả thuyết cố định (hay giả thuyết dừng) của một loạt điều kiện hình thành các quan hệ này.

Khả năng ứng dụng của các mối quan hệ mà được làm sáng tỏ trên cơ sở tài liệu thực nghiệm, đối với tổng thể được đưa vào lý thuyết ước lượng các tham số mẫu của mối quan hệ chẳng hạn như ước lượng những dao động ngẫu nhiên của chúng.

Ước lượng này có giá trị đặc biệt khi chỉnh lý các đại lượng thuỷ văn tạo nên những chuỗi thường thường có dung lượng không lớn. Trong các trường hợp đó có thể các mối quan hệ thống kê rất phù hợp với tài liệu mẫu nhưng lại chệch so với quan hệ ngẫu nhiên.

Mối quan hệ ngẫu nhiên giữa hai biến lượng được mô tả đầy đủ nhất bằng hàm mật độ phân phối hai chiều. Còn mối quan hệ thống kê giữa hai biến lượng ấy được miên tả bằng biểu đồ lăng trụ tần suất. Những mô tả mối quan hệ thống kê và ngẫu nhiên giữa hai biến lượng như vậy sẽ được khái quát hoá đối với trường hợp mối quan hệ giữa n biến lượng. Các mối quan hệ này được mô tả bằng luật phân phối n chiều. Sự mô tả các mối quan hệ thống kê và ngẫu nhiên như vậy là đầy đủ nhất nhưng lại yếu cầu một lượng thông tin gốc rất lớn. Khi nghiên cứu các quá trình thuỷ văn những điều kiện này không thể thực hiện được. Vì vậy khi nghiên cứu các mối quan hệ thống kê nói chung và giữa hai biến lượng thuỷ văn nói riêng người ta thường sử dụng mối quan hệ gọi là tương quan, đây là mối quan hệ giữa giá trị được xác định của một đại lượng (đối số) và trị bình quân có điều kiện tương ứng của đại lượng kia (hàm số). Rõ ràng là mối quan hệ tương quan là dạng biểu diễn riêng của mối quan hệ thống kê.

Mối quan hệ tương quan được biểu diễn dưới dạng các phương trình tương quan hay phương trình hồi quy có thể là tuyến tính hoặc không tuyến tính. Sau đây chúng ta sẽ xét mối tương quan tuyến tính giữa các biến ngẫu nhiên. Trong trường hợp mối quan hệ không tuyến tính giữa các đặc trưng thuỷ văn cần nghiên cứu có thể biến đổi tài liệu gốc để cho mối quan hệ giữa các giá trị đã được biến đổi có dạng tuyến tính. Ta nhận thấy rằng phép biến đổi trên đây chính là chuyển luật phân phối gốc của đại lượng nghiên cứu sang dạng chuẩn.

Một số phương pháp biến đổi đó đã được nghiên cứu ở chương II. Cũng cần phải chú ý rằng phương pháp biến đổi đem dùng chỉ có ý nghĩa trong trường hợp khi yếu tố có trong mối quan hệ không tuyến tính giữa các đại lượng gốc được xác định là rất tin cậy.

Khi chuỗi tài liệu quan trắc ngắn thường có tình trạng nguy hiểm là lấy mối quan hệ tuyến tính để thay cho mối quan hệ không tuyến tính là do những phân tán không ngẫu nhiên của tài liệu tạo nên mẫu nhỏ.

6. 2. TƯƠNG QUAN TUYẾN TÍNH GIỮA HAI BIẾN

Trong thuỷ văn người ta rất hay sử dụng mối tương quan tuyến tính gữa hai biến lượng. Các mối quan hệ này được dùng kéo dài các chuỗi đặc trưng dòng chảy ra thời kỳ nhiều năm, để dự báo dòng chảy hay mực nước ở tuyến dưới theo tài liệu dòng chảy ở tuyến đo phía trên; tương tự như vậy đối với rất nhiều đặc trưng khác của chế độ thuỷ văn có thể xây dựng các quan hệ dự báo tính toán phụ thuộc vào các nhân tố xác định chung. Vì vậy chúng ta nghiên cứu mối quan hệ giữa hai đại lượng ngẫu nhiên không phải là trường hợp riêng của mối tương quan tuyến tính nhiều chiều, được trình bày ở mục sau, mà như là một bài toán độc lập.

Chúng ta sẽ xét các mối quan hệ cơ bản được mô tả bằng tương quan tuyến tính gữa hai biến lượng. Việc làm sáng tỏ mối quan hệ giữa các đặc trưng khí tượng thuỷ văn nghiên cứu sẽ được tiến hành trên cơ sở nghiên cứu các chuỗi của chúng. Khi đưa lên đồ thị các giá trị tương ứng x_i và y_i chúng ta nên nhóm ở mức độ nào đó phân bố theo quy tắc đường thẳng: y=ax + b phù hợp nhất với nhóm điểm đó. Điều đó đạt được trong trường hợp khi tổng bình phương khoảng lệch giữa tài liệu quan trắc được với giá trị tính toán được theo phương trình quy hồi là nhỏ nhất.

$$S = \sum_1^n [Yi - (ax_i + b)]^2 \qquad \min \qquad (6.1)$$

Các giá trị của tham số a và b thoả mãn với phương trình (6.1) ta tìm được khi cho đạo hàm của tổng đo theo các tham số trong không đạo hàm theo a.

$$\frac{dS}{da} = 2\sum_1^n (y_i - ax_i - b)(-x_i) = 0$$

Từ đó: $\sum_1^n x_i y_i + a\sum_1^n x^2{}_i + b\sum_1^n x_i = 0$ (6.2)

Đạo hàm theo b:

$$\frac{dS}{db} = 2\sum_1^n (y_i - ax_i - b) = 0$$

từ đó $\sum_1^n y_i - a\sum_1^n x_i - nb = 0$ (6.3)

Ta đặt trị bình quân của các biến lượng:

$$\frac{1}{n}\sum_1^n x_i = \overline{x}, \frac{1}{n}\sum_1^n y_i = \overline{y} \qquad (6.4)$$

Giải các phương trình (6.2) và (6.3) đối với các tham số a và b ta nhận được

$$a = \frac{\sum_1^n (x_i y_i - n\overline{xy})}{\sum_1^n (x^2{}_i - n\overline{x}^2)}, b = \overline{y} - a\overline{x}$$

Biểu thức của tham số a được gọi là hệ số hồi quy, có thể dẫn đấn dạng:

$$a = r\frac{\sigma_y}{\sigma_x} \qquad (6.5)$$

Hệ số tương quan r giữa các biến x và y có thể căn cứ vào các mẫu nghiên cứu tính theo công thức:

$$r = \frac{\sum_1^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_1^n (x_i - \overline{x})^2 \sum_1^n (y_i - \overline{y})^2}} \qquad (6.6)$$

Hệ số tương quan thường được sử dụng ở dạng sau:

$$r = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} \qquad (6.7)$$

Trong đó: cov(x,y) - hiệp biến (mônen hỗn hợp bậc hai) hay mômen quan hệ của các đại lượng x và y là kỳ vọng toán của tích các khoảng lệch x và y so với tần phân phối của chúng, nghĩa là:

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_1^n (x_i - \overline{x})(y_i - \overline{y})$$

hay $$r = \sqrt{a_{y/x} a_{x/y}} \qquad (6.8)$$

trong đó $a_{y/x}$ và $a_{x/y}$ hệ số hồi quy của y theo x và của x theo y.

Ta sẽ điểm lại những tính chất cơ bản của hệ số tương quan:

Nếu các biến x và y độc lập với nhau thì tổng của tích các khoảng lệch so với trị bình quân của chúng ở tử số các biểu thức (6.6) sẽ bằng 0 do đó hệ số tương quan cũng bằng 0. Trong trường hợp khi mối quan hệ giữa các biến lượng là hàm số (ngoài quan hệ tuyến tính ra) hệ số tương quan bằng cộng hay trừ 1 (±1). Khi đó mối tương quan, phụ thuộc vào mức độ chặt chẽ của nó, hệ số tương quan biến đổi trong khoảng ±1.

Hệ số tương quan tương ứng với trường hợp khi hàm số tăng theo sự tăng của đối số (mối quan hệ thuận), hàm số giảm khi đối số tăng sẽ được đặc trưng bằng hệ số tương quan âm (nội quan hệ nghịch).

Khoảng lệch trung bình bình phương của các biến lượng so với bình quân số học của chúng được xác định theo các biểu thức:

$$\sigma_x\sqrt{\frac{\sum_1^n (xi-\overline{x})^2}{n}}; \quad \sigma_y=\sqrt{\frac{\sum_1^n (y_i-\overline{y})^2}{n}} \qquad (6.9)$$

- Tham số b có thể được viết dưới dạng:

$$b=\overline{y}-r\frac{\sigma_y}{\sigma_x}\overline{x} \qquad (6.10)$$

Với các đẳng thức (5.6) và (5.9) phương trình tương quan có thể biểu diễn dưới dạng:

$$\begin{aligned} &y=ax+b=ax+\overline{y}-ax \\ &y-\overline{y}=a(x-\overline{x})=r\frac{\sigma_y}{\sigma_x}(x-\overline{x}) \end{aligned} \qquad (6.11)$$

Đẳng thức vừa nhận được này là phương trình hồi quy của y theo x.

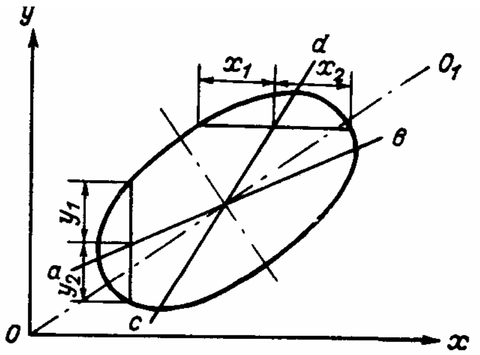
Tương tự ta nhận được phương trình quan hệ tuyến tính của x theo y có dạng:

$$x-\overline{x}=r\frac{\sigma_x}{\sigma x}(y-\overline{y}) \qquad (6.12)$$

Các phương trình (6.11) và (6.12) là những quan hệ độc lập khác nhau không thể nhận phương trình này từ phương trình kia được. Ta chú ý rằng dạng phương trình tương quan khác nhau của y theo x và của x theo y là do sự khác biệt của đặc tính thống kê trong quan hệ và không có liên quan gì với độ dài hữu hạn của tài liệu mẫu.

Các quan hệ trên nói chung đều đúng với các mẫu lấy từ bất kỳ luật phân phối nào của biến lượng ngẫu nhiên x và y. Nếu các biến lượng x và y phân phối theo luật chuẩn thì mỗi điểm của phương trình hồi quy là tâm của đường phân phối có điều kiện của biến ngẫu nhiên phụ thuộc (y), các giá trị y được lập nhóm quanh nó, các giá trị y này xuất hiện đồng thời (trong các lần thử khác nhau) với cùng một giá trị x nghiên cứu. Lúc này trong trường hợp riêng các đường phân phối có điều kiện cũng ứng với luật phân phối chuẩn có trị bình quân được tính bằng đẳng thức (6.4) và có phương sai xác định theo đẳng thức (6.9)

Dưới dạng tổng quát sự phân tán của những đại lượng có quan hệ tương đương với nhau, tuân theo luật phân phối chuẩn được biểu diễn theo phạm vi của elip phân tán (elíp xác suất như nhau) (hình 6.1). Đối với các chuỗi thống kê chuẩn độc lập với nhau elíp sẽ trở thành hình tròn, còn đối với mối quan hệ hàm số thì nó trở thành mối quan hệ tuyến tính đơn trị.



Hình 6.1 Sơ đồ quan hệ y = ax+b; y=a'x+b' đối với phân bố chuẩn biến x và y

Đường thẳng ab là đường hồi quy của y theo x nó chia các tuyến thẳng đứng của elíp ra làm 2 phần bằng nhau, và nó biểu diễn sự phân tán của giá trị y ứng với mỗi giá trị x. Giá trị phân tán lý luận được mô tả bằng quan hệ (6.9). Giá trị phương sai đặc trưng của hàm y là một số không đổi, không phụ thuộc vào x_i, vì vậy biểu thức (6.9) sẽ cho ta ước lượng sự phân tán của y. Đối với thiết diện ứng với giá trị x_i cố định cũng như đối với toàn bộ phương trình hồi quy nói chung đường chia đôi các cát tuyến nằm ngang song song với trục x.

Các đường ab và cd ứng với các phương trình (6.11) và (6.12) như trên đã rõ chúng chỉ trùng nhau khi các đại lương x và y có quan hệ hàm số với nhau:

Để kết luận về vấn đề này, vì các phương trình tương quan nhận được trên cơ sở các mẫu phải phù hợp với các quan hệ ngẫu nhiên, nên phải đánh giá độ chính xác và phương trình hồi quy và tham số của phương trình này. Để làm chỉ tiêu độ chính xác của phương trình hồi quy, người ta sử dụng khoảng lệch trung bình bình phương có điều kiện (sai số tiêu chuẩn) là khoảng lệch trung bình bình phương giữa các giá trị quan trắc và giá trị tính toán được theo phương trình hồi quy.

$$\sigma_y(x)\sqrt{\frac{\sum_{i=1}^{n}(y_{i,n}-y_p)^2}{n}} \qquad (6.13)$$

trong đó: $y_{i,q,tr}$ - Giá trị quan trắc được; y_{tt} - Gía trị tính toán được theo đường hồi quy.

Sử dụng hệ số tương quan thì biểu thức (6.13) sẽ có dạng:

$$\sigma_{y(x)} = \sigma y_0 \sqrt{1-r^2} \qquad (6.14)$$

Ở đây σ_{y0} - là khoảng lệch trung bình bình phương của chuỗi giá trị gốc y(hàm số); r - hệ số tương quan của phương trình đường hồi quy.

Biểu thức (6.14) cho thấy rằng chẳng hạn khi r=0,95 khoảng lệch trung bình bình phương của các giá trị nhận được theo phương trình hồi quy bằng 0,32, nghĩa là độ phân tán của các giá trị đó nhỏ gấp 3 lần so với độ phan tán của chuỗi biến lượng phụ thuộc gốc.

Khi nghiên cứu các tham số a và b được xem như là đại lượng biến đổi, chúng biến thiên từ mẫu này sang mẫu khác, độ chính xác của ước lượng có thể đặc trưng bằng các giá trị của sai số tiêu chuẩn tương ứng.

$$\sigma_b = \frac{\sigma_{y(x)}}{\sqrt{n}} - \frac{\sigma_y \sqrt{1-r^2}}{\sqrt{n}} \qquad (6.15)$$

Khi sử dụng để xây dựng đường hồi quy các tham số a và b có sai số ngẫu nhiên, chúng ta cho phép có sai số trong khi ước lượng tung độ của đường hồi quy. Sai số này có thể được đặc trưng bằng giá trị của phương sai tương ứng (bình phương sai số tiêu chuẩn).

$$\sigma_y^2(x_i) = \sigma_y^2(x)\left[\frac{1}{n-2} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] \qquad (6.17)$$

Phương sai $\sigma^2_{y(xi)}$ đặc trưng cho độ phân tán của tung độ đường hồi quy mẫu so với đường hồi quy của tổng thể.

Sau đây chúng ta sẽ nghiên cứu vấn đề độ chính xác về ước lượng hệ số tương quan mẫu.

Trong công trình của V.I Rômanôvski [111.tr 391] đã chứng minh được công thức sai số trung bình bình phương của hệ số tương quan.

$$\sigma_r = \frac{1-r^2}{\sqrt{n-1}}\sqrt{1+\frac{11r^2}{2n}+\frac{75r^2-13}{2n^2}}$$

mà khi n khá lớn (n > 25) được viết dưới dạng thường hay gặp:

$$\sigma_r \approx \frac{1-r^2}{\sqrt{n-1}} \qquad (6.18)$$

Khi n rất lớn và r gần bằng 1, phân phối của hệ số tương quan mẫu sẽ tiếp tới luật phân phối chuẩn có tần phân phối bằng r và khoảng lệch trung bình $=1-r^2$. Khi n là hằng số và r = 1 luật phân phối của hệ số tương quan càng lệch so với luật chuẩn.

Hệ số tương quan được tính theo mẫu có dung lượng hữu hạn n thường thường là nhỏ hơn hệ số tương quan của tổng thể, nghĩa là hệ số tương quan mẫu có chệch âm. Độ chệch này giảm khi n tăng.

Phân phối chuẩn của hệ số tương quan mẫu gần như được bảo tồn khi n không nhỏ lắm và r không lớn lắm. Trong các trường hợp khác (khi n nhỏ và r lớn) phân phối của r mẫu là không đối xứng.

Đối với hệ số tương quan tính theo các mẫu từ trong phân phối khác với luật chuẩn, luật phân phối của r mẫu nói chung là chưa biết vì thế việc ứng dụng hệ số tương quan thực nghiệm là khó khăn. Khi dung lượng của mẫu nhỏ (n < 50) và đặc biệt khi r lớn độ đánh giá mức độ phân tán ngẫu nhiên của hệ số tương quan mẫu người ta thường sử dụng phép biến đổi Fisher biến đổi này được dựa vào việc sử dụng biến lượng đặc biệt z có quan hệ hàm số với r bằng biểu thức.

$$z = \frac{1}{2}\ln\frac{1+r}{1-r} \quad (6.19)$$

Để xác định các giá trị z=f(r) nên sử dụng tại liệu của bảng 6.1

Phân phối z ngay cả đối với các mẫu không lớn rất gần với phân phối chuẩn trong thực tế không phụ thuộc vào n và giá trị thực r.

Sai số tiêu chuẩn z bằng:

$$\sigma = \frac{1}{\sqrt{n-3}} \quad (6.20)$$

Theo các giá trị δ_Z, và sử dụng số liệu bảng 6.1 ta có thể tìm được và cần đưa vào luật phân phối chuẩn sẽ xác định ở giới hạn nào đó những giá trị hệ số tương quan mẫu ứng với các mẫu khác nhau của xác suất tin cậy.

Trường hợp riêng sử dụng phép biến đổi Fisher là đồ thị hình 6.2

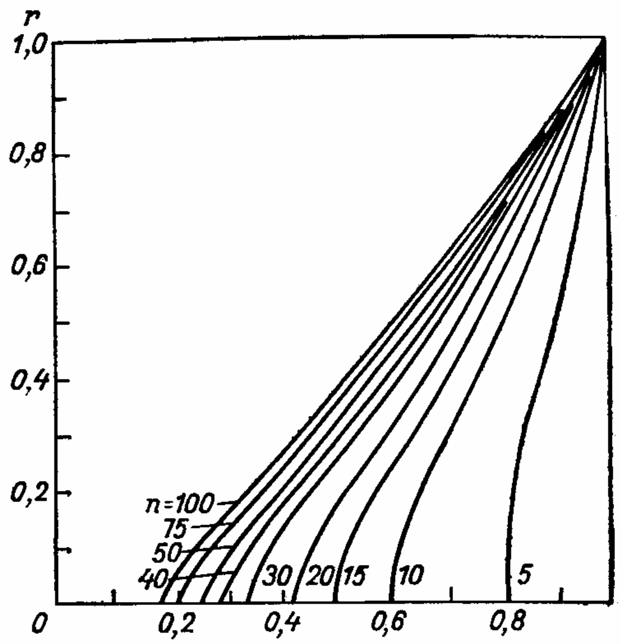
Bảng 6.1 Giá tri z = f(r)

r	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,1	0,10	0,11	0,12	0,13	0,14	0,15	0,16	0,17	0,18	0,19
0,2	0,20	0,21	0,22	0,23	0,24	0,26	0,27	0,28	0,29	0,30
0,3	0,31	0,32	0,33	0,34	0,35	0,37	0,38	0,39	0,40	0,41
0,4	0,42	0,44	0,45	0,46	0,47	0,48	0,50	0,51	0,52	0,54
0,5	0,55	0,56	0,58	0,59	0,60	0,62	0,63	0,65	0,66	0,68
0,6	0,69	0,71	0,72	0,74	0,06	0,78	0,79	0,81	0,83	0,85
0,7	0,87	0,89	0,91	0,93	0,95	0,97	1,00	1,02	1,05	1,07
0,8	0,10	1,13	1,16	1,19	1,22	1,26	1,29	1,39	1,38	1,42
0,9	1,47	1,53	1,59	1,66	1,74	1,83	1,95	2,09	2,30	2,65
0,99	2,65	2,70	2,76	2,83	2,90	2,99	3,11	3,25	3,45	3,80

Hệ số tương quan nhỏ nhất ứng với mức ử dụng 5% trong tổng thể, với các giá trị của hệ số đệ tính theo các mẫu có dung lượng khác nhau.

Một trong những bài toán tính toán thuỷ văn được giải quyết có sử dụng đến tương quan tuyến tính là việc chuyển các tham số của chuỗi đại lượng thuỷ văn được xác định theo mẫu ngắn sang giai đoạn dài. Cơ sở vật lý của lời giải đó là tính đồng bộ có trong dao động của các chuỗi thuỷ văn được nghiên cứu và của đặc trưng khí tương thuỷ văn nào đó có tương quan với đại lượng này. Lúc này đáng chú ý là đặc trưng (đối số) được có định trong suốt một thời kỳ dài là thời kỳ mà đại lượng (hàm số) thuỷ văn ta quan tâm.

Mối quan hệ thống kê của tài liệu quan trắc đồng có thể được sử dụng dưới 2 dạng sau đây. Một dạng sử dụng mối quan hệ thống kê là để khôi phục đại lượng thuỷ văn ta quan tâm cho toàn bộ thời kỳ tài liệu của đối số có được. Hướng thứ hai căn cứ vào việc sử dụng các phương trình ta quan sát, xác lập giữa các giá trị của tham số thống kê (trị bình quân và khoảng lệch tiêu chuẩn) ở đối tượng mà đối với n phải tiến hành kéo dài và ở đối tượng tương tự.



Hình 6.2 Giá trị cực tiểu của hệ số tương quan với mức sử dụng 5% trong tổng thể với các giá trị khác nhau của hệ số này trong các mẫu có dung lượng khác nhau.

Sử dụng cách thứ nhất để ta kéo dài chuỗi ngắn là giá trị trong đó được tính theo phương trình hồi quy. Chúng được khôi phục như vậy cho phép ta xác định các tham số nó ứng với thời kỳ quan trắc dài ở đối tương tự (trị bình quân để biến đổi), ngoài ra còn chứa một lượng thông tin bổ sung về sự lần lượt của các pha nước khác nhau trong thời kỳ nhiều năm.

Song cần phải chú ý là chuỗi khôi phục đó không được khôi phục chính xác như tài liệu quan tóc trước đây đặc trưng nghiên cứu. Vấn đề là ở chỗ các giá trị của hàm (trong trờng hợp này là của đại lượng phục hồi y) nhân theo phương trình hồi quy là trị bình quân của tập hợp sự thể hiện khi cố định giá trị của đối số x_i. Nhưng giá trị thực tế riêng biệt của hàm Y lệch tương đối nhiều so với đường hồi quy. Việc thay thế những giá trị phân tán xung đường hồi quy đó bằng kỳ vọng toán của chúng sẽ dẫn đến được khôi phục khác với chuỗi thực tế được san bằng những động X.X.Kirski và M.F.Menkel [82] đã chứng minh rằng biến đổi thực tế của đại lượng thuỷ văn được nghiên ứu bằng Cv/r, trong đó là hệ số biến đổi nhận được theo đã được khôi phục bằng phương trình hồi quy, còn r là hệ số tương quan của phương trình hồi quy. Tương tự như vậy để bảo đảm tính chất dao động chung của biến lượng y được biểu diễn bằng hệ số biến đổi thực Cv ta cần phải tăng khoảng lệch y-y tính theo phương trình hồi quy 1/r lần.

Thuật toán đó được dùng vào việc tính toán theo phương trình hồi quy gốc với trường hợp khi r=1 cũng giống như sử dụng lời giải, "duy nhất" ứng với phương trình.

$$y(x_i) = \bar{y} + \frac{\sigma_y}{\sigma_x}(x_i - \bar{x})$$

Khi các đường hồi quy của y theo x và của x theo y trùng nhau. Khi sử dụng phương pháp của G.P.Ivanôv cũng nhận được kết quả tương tự. Thực chất của nó là đối với đối tượng nghiên cứu không phải chuyển giá trị $\frac{y}{x}(xi - x)$ từ đối tượng tương tự, mà chuyển tần số P_{xi} đặc trưng cho giá trị $\frac{\delta y}{\delta x}(x_i - \bar{x})$

Đối với những chuỗi tuân theo luật phân phối chuẩn, việc chuyển từ tương tự sang các giá trị $\frac{\delta y}{\delta x}(x_i - \bar{x})$ hay P_x có nghĩa là thực hiện một phép toán tương đương chỉ khác nhau bằng hình dạng bên ngoài.

Đối với những luật phân xác suất không đối xứng, phương pháp Ivanôv là đúng hơn cả. Vì rằng giá trị P_{xi} nhận được theo đường tần suất thực nghiệm sẽ phản ánh được tính không đối xứng của nó.

Cần phải chú ý rằng khi hệ số tương quan nhỏ các chuỗi được khô phục có xét đến số hiệu chỉnh sẽ không phản ánh được dao động của đại lượng nghiên cứu trong một khoảng thời gian cụ thể, đối với khoảng thời gian này trong khi tính toán được sử dụng tài liệu quan trắc dao động của biến số x.

Phần khôi phục của chuỗi là một thí dụ điển hình chuyển đặc tính những dao động của chuỗi nghiên cứu mà không khôi phục chúng trong trình tự thời gian cụ thể.

Để chuyển các tham số thông kê sang thời kỳ nhiều n (không kéo dài chuỗi theo những giá trị tương ứng của đối tượng tương tự) người ta đã sử dụng các phương trình sau đây của Kriski và Menkel [82].

$$\bar{y}_N = \bar{y}_n + r\frac{\sigma_{yN}}{\sigma_{xN}}(\bar{x}_N - \bar{x}_n) \qquad (6.21)$$

$$\sigma^2_{yN} = \sigma^2_{2yN} + r^2 \frac{\sigma^2_{yN}}{\sigma^2_{xN}}(\sigma^2_{xN} - \sigma^2_{xn}) \qquad (6.22)$$

Trong đó, $\overline{Y}_N, \overline{x}_n$ trị bình quân của các đại lượng tương ứng trong thòi kỳ N (thời kỳ quan trắc nhiều năm ở đối tượng - tương tự) $\overline{x}_n, \overline{y}_n$ trị bình quân ứng với thời kỳ quan trắc ngắn n của đối tượng cần được kéo dài ước lượng khoảng lệch trung bình bình phương của y và x trong các thời kỳ đó; r - hệ số tương quan giữa các giá trị y và x xảy ra đồng thời.

Hệ số tương quan giữa các ước lượng của phương sai được lấy bằng r^2 trên cơ sở quan hệ gần đúng đã biết trong thống kê toán. Giá trị phương sai σ_{yN} được tính khi giải phương trình (6.22)

$$\sigma^2_{yN} = \frac{\sigma^2_{yn}}{1 - r^2(1 - \frac{\sigma^2_{xn}}{\sigma^2_{xN}})} \qquad (6.23)$$

Để đánh giá độ chính xác của các tham số nhận được cho thời kỳ nhiều năm người ta sử dụng các công tức sai số tiêu chuẩn.

$$\sigma_{\overline{y}N} = \pm \frac{\sigma_{yN}}{\sqrt{n}} \sqrt{1 - \frac{N-n}{N} r^2} \qquad (6.24)$$

$$\sigma_{\sigma yN} = \pm \frac{\sigma_{yN}}{\sqrt{2n}} \sqrt{1 - \frac{N-n}{N} r^4} \qquad (6.25)$$

Sau đây chúng ta sẽ xét một số ví dụ minh hoạ trình tự những đề xuất phân tích ở trên. Theo tài liệu quan trắc dòng chảy năm sông Dnepr - trạm Xmolensk (x) trong 82 năm (từ 1882) và sông ViaZma - trạm Xtarafa (Y) trong 18 năm (từ 1954) ta xây dựng phương trình hồi quy và sử dụng nó để kéo dài chuỗi tài liệu của trạm Xtaraia và để chuyển các tham số của chuỗi 18 năm sang thời kỳ nhiều năm. Đồ thị quan hệ của tài liệu trong thời kỳ quan trắc đồng bộ được biểu diễn trên hình 6.3.

Đối với thời kỳ quan trắc đồng bộ ta có $\sum xi = 134$ $\sum yi = 124$ $\overline{x} = 7{,}4 l/skm^2$. $\overline{y} = 6{,}9 l/skm^2$ $\sum(xi - x);(yi - y) = 73;$ $\sum(xi - \overline{x})^2 = 58;)$ $\sum(yi - \overline{y})^2 117;$ $\sigma_{xn} = 1{,}79$ $\sigma y = 2{,}55$. Ngoài ra ta đã biết $\sigma xN = 1{,}28, \overline{x}_N = 6{,}9$

Căn cứ vào chuỗi số liệu này ta nhận được:

$$a_{y/x} = \frac{\sum (x_i - \overline{x})(yi - \overline{y})}{\sum (xi - \overline{x})^2} = \frac{73}{58} = 1{,}27$$

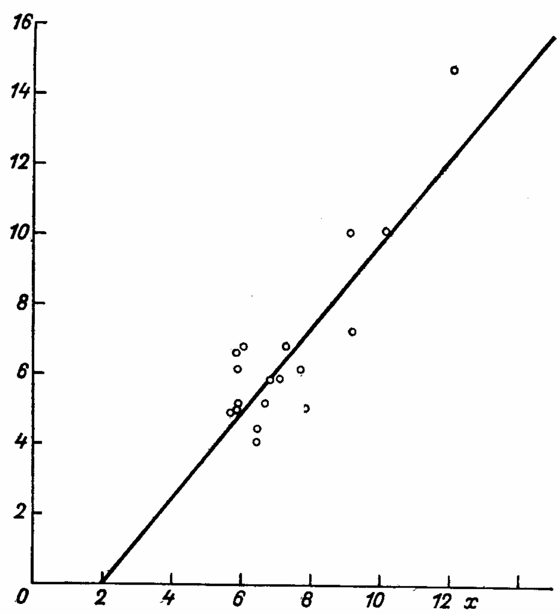
$$b = \overline{y} - a_{y/x}\overline{x} = 6{,}9 - 1{,}27.7{,}4 = -2{,}51$$

$$r = \frac{\sum (xi - \overline{x})(yi - \overline{y})}{\sum (xi - \overline{x})^2 \sum (yi - \overline{y})^2} = \frac{73}{\sqrt{58.117}} = 0{,}89$$

Trong trường hợp này phương trình hồi quy của y theo số có dạng:

$$Y(x) = 6{,}9 + 1{,}27\ (x - 7{,}4)\ \text{l/sKm}^2$$

hay là



Hình 6.3 Đồ thị quan hệ dòng chảy năm S. Viazmư - tr. Xtaraia (y) và s. Dnhepr ở tp. Smolensk (x)

$$Y(x) = 1{,}27x - 2{,}51\ \text{l/s Km}^2$$

Những tính toán tương tự đối với đường hồi quy của x theo y sẽ dẫn đến phương trình:

$$X(Y) = 0{,}62 + 3{,}11\ \text{l/s Km}^2$$

Khoảng lệch trung bình bình phương có điều kiện của biến lượng y đối với đường hồi quy ứng với đối số cho trước bằng:

$$\sigma_{y(x)1} = \sqrt{\frac{\sum(yi - \bar{y})^2 - a\sum(x_i - \bar{x})(y_i - \bar{y})}{n}} = \sqrt{\frac{117 - 1,27.73}{15}} = 1,24$$

Còn theo công thức (6.14)

$$\sigma_{y(x)2} = \sigma_y\sqrt{1 - r^2} = 2,55\sqrt{1 - 0,89^2} = 1,20$$

Sai số tiêu chuẩn của số hạng tự do b của phương trình hồi quy được xác định theo công thức:

$$\sigma_b = \frac{\sigma_{y(x)1}}{\sqrt{n}} = \frac{1,24}{\sqrt{18}} = 0,29$$

hoặc là

$$\sigma_b = \frac{\sigma_{y(x)2}}{\sqrt{n}} = \frac{1,20}{\sqrt{18}} = 0,28$$

Sai số tiêu chuẩn của hệ số hồi quy bằng

$$\sigma_a = \frac{\sigma_{y(x)1}}{\sqrt{\sum(xi - \bar{x})^2}} = \frac{1,24}{\sqrt{58}} = 0,16$$

hay là

$$\sigma_a = \frac{\sigma_{y(x)2}}{\sqrt{\sum(xi - \bar{x})^2}} = \frac{1,20}{\sqrt{58}} = 0,16$$

Trong trường hợp này các kết quả tính toán tham số của phương trình hồi quy có thể được biểu diễn dưới dạng

$$a \pm \sigma_a 1,27 \pm 0,16$$

$$b \pm \sigma_b - 2,51 \pm 0,29$$

Sai số tiêu chuẩn của tung độ phương trình hồi quy theo công tức (6.17) được biểu diễn bằng:

$$\sigma_{\overline{y}(xi)} = \sigma_{y(x)}\sqrt{\frac{1}{n-2}+\frac{(xi-\overline{x})^2}{\sum(x_i-\overline{x})^2}} = 1,24\sqrt{\frac{1}{16}+\frac{(xi-7,4)^2}{58}}.$$

Thí dụ: khi x_i =5$\delta_{y(x)}$=0,50l/s.km^2, khi x_i = 10 $\delta_{y(x)}$ $^{= 0,52}$l/skm^2 và khi == $\delta_{y(x)\ =}$ $_{0,29\ =}$ δb điều đó có thể có.

Sai số tiêu chuẩn của hệ số tương quan trong trường hợp này bằng:

$$\sigma = \frac{1-r^2}{\sqrt{n-1}} = \frac{1-0,89^2}{\sqrt{17}} = 0,05$$

Sai số xác suất bằng

ε = 0,67 = 0,03

Chúng ta sẽ tiến hành đánh giá sai số tiêu chuẩn hệ số tương quan bằng phép biến đổi Fisher. Để làm việc theo giá trị của hệ số tương quan r=0,89 và theo bảng ta sẽ xác định được giá trị của hàm z bằng z= 1,42. Sai số trung bình bình phương của z theo biểu thức (6.20) bằng

$$\sigma_z = \frac{1}{\sqrt{18-3}} 0,26$$

Như vậy trong phạm vi $\pm\sigma_z$ giá trị z bằng

$z_{tr} = z+\sigma_z = 1,42+0,28 = 1,68$

$z_d = z-\sigma_z = 1,42 - 0,26 = 1,16$

Tiếp theo với z_{tr} và z_d theo bảng 6.1 ta xác định được giới hạn trên và giới hạn dưới của r.

$r_{tr} = 0,93$

$r_d = 0,82$

Khi sử dụng công thức (6.18), ta có:

$r_{tr} = 0,89 + 0,05 = 0,94$

$r_d = 0,89 - 0,05 = 0,84$

Sự khác nhau trong ước lượng r là do khi r lớn và dung lượng mẫu nhỏ luật phân phối của ước lượng mẫu sẽ lệch đi ít nhiều so với luật phân phối chuẩn.

Sử dụng các phương trình (6.21) và (6.23), ta tiến hành chuyển các tham số y và y sang thời kỳ nhiều năm.

$$\bar{y}_N = \bar{y}_n + r\frac{\sigma_{yn}}{\sigma_{xN}}(\bar{x}_N - \bar{x}_n) = 6{,}9 + 0{,}89\frac{2{,}55}{1{,}82}(6{,}9 - 7{,}4) = 6{,}26 l/skm^2$$

$$\sigma_{yN} = \frac{\sigma_{yn}}{\sqrt{1 - r^2\left(1 - \frac{\sigma^2_{xn}}{\sigma^2_{xN}}\right)}} = \frac{2{,}55}{\sqrt{1 - 0{,}89^2(1 - \frac{1{,}79^2}{1{,}82^2})}} = 2{,}58 l/skm^2$$

Trong trường hợp này $Cv_{yN} = \dfrac{\sigma_{yN}}{y_N} = 0{,}41$

Sử dụng phương trình tương quan y(x) = 1,27 x chuỗi dòng chảy sông Viazma có thể khối phục được cho t bộ thời kỳ mà có tài liệu quan trắc dòng chảy sông Dnepr trạm Xmôlensk.

Sử dụng chuỗi các tài liệu quan trắc được và tài liệu khôi phục theo phương trình tương quan ta nhận được $\bar{y}_N = 6{,}21 l/skm^2; \sigma_{yN} = 2{,}42 l/skm^2$ Trong trường hợp này sự khác nhau giữa các tham số nhận được theo chuỗi được khôi phục và theo tính toán bằng các phương trình (6.24) và (6.23) không lớn lắm điều đó là do mối quan hệ giữa dòng chảy trong thời kỳ quan trắc đồng bộ rất chặt chẽ (r = 0,89).

6 .3. TƯƠNG QUAN TÍNH TOÁN NHIỀU CHIỀU.

Trong nghiên cứu các quá trình thuỷ văn do nhiều tổ tạo nên, thí dụ như: khi xây dựng các lược đồ tính toán và dự báo đôi khi cần phải xác lập quan hệ tuyến tính giữa một số biến lượng với nhau. Để giải bài toán này người ta chú ý đến phép toán tương đương tuyến tính nhiều chiều. Thực chất của phương pháp này là sử dụng các lập luận cơ bản của phương pháp tương quan tuyến tính giữa hai biến lượng đối với trường hợp biến lượng mà ta quan tâm phụ thuộc vào số lượng tuỳ ý đối số x.

Cơ sở để tìm mối quan hệ là sử dụng tài liệu quan trắc đại lượng y và các đại lượng quy định nó là x_1, x_2 và x_n. Kết quả đo đạc đồng thời các đại lượng đó có thể biểu diễn dưới dạng

$Y_1, x_{11}, x_{21}, x_{31}, ... x_{j1}, ... x_{n1}$

$Y_2, x_{12}, x_{22}, x_{32}, ... x_{j2}, ... x_{n2}$

$Y_i, x_{1i}, x_{2i}, x_{3i}, ... x_{ji}, ... x_{ni}$

$y_m, x_{1m}, x_{2m}, x_{3m}, ... x_{jm}, ... x_{nm}$

$y, x_1, x_2, x_3, ... x_j, ... x_n$

Dựa vào những tài liệu đo đạc trên đây ta phải tìm một quan hệ tuyến tính giữa y và x1, x2, ... xn, theo nguyên tắc bình phương nhỏ nhất phù hợp nhất với tài liệu thực nghiệm. Lời giải nhận được rất đơn giản nếu như không nghiên cứu chính các giá trị gốc y và x1, x2, ... xn là các khoảng lệch của chúng so với trị bình quân.

$$y_1^0 = yi - \bar{y}; x^0 = x_{ij} - \bar{x}_j;$$
$$j = 1,2,...n$$
$$i = 1,2,...m$$

Trong trường hợp này phương trình hồi quy của tương quan tuyến tính nhiều chiều

$$y = \bar{y} + k_1(x_{ij} - x_i) + k_2(x_{2i} - \bar{x}_2) + ... + k_j(x_{ij} - \bar{x}_j) + ... + k_n(x_{ni} - \bar{x}_{ni})...$$

(6.26)

của giá trị y đối với các biến số x_1, x_2, ... x_n của nó được viết như là phương trình hồi quy của giá trị $y^0{}_i - y_{i0} - \bar{y}_0$ đối với khoảng lệch của đối số x_1, x_2, ... x_n so với trị bình quân của chúng.

$$y_1^0 = k_1 x_1{}^0 + k_2 x_2{}^0 + ... + k_n x_n{}^0 \qquad (6.27)$$

Rõ ràng là có m phương trình như vậy đối với số lượng quan trắc của giá trị y, x_1, x_2, ... x_n. Lúc này m == . Khi m >>n bài toán xác định các tham số sẽ không giải được khi m = n lời giải sẽ nhận được với độ chính xác thoả mãn với tài liệu gốc, nhưng lời giải này chỉ có ý nghĩa đối với các quan hệ hoàn toàn là hàm số.

Trường hợp hệ có số lượng phương trình lớn hơn số lượng tham số chưa biết sẽ là trường hợp cơ bản của việc xây dựng các phương trình hồi quy. Lời giải tốt nhất của hệ có phương trình thừa là tìm những giá trị của đại lượng chưa biết đó mà được liên kết với nhanh bằng các phương trình liên hệ nghiên cứu, khi xác định chúng các phương trình này khoảng lệch nhỏ nhất gữa các giá trị tính toán và quan trắc. nếu dùng tổng các khoảng lệch đó của chúng để đánh giá sẽ gặp phải trường hợp là các khoảng lệch lớn nhưng có dấu ngược nhau có thể bù trù lẫn nhau, trong khi đó của giá trị tuyệt đối của các khoảng lệch riêng biệt có thể là rất lớn.

Do đó lời giải tốt nhất của hệ phương trình được chấp nhận là lời giải mà trong đó tổng bình phương tất cả các khoảng lệch (hay sai số tính toán sử dụng phương trình hồi quy) có giá trị nhỏ nhất, vì vậy phương pháp giải này mang tên là phương pháp bình phương nhỏ nhất.

Như ta đã thấy ở bài 2 của chương này hệ số hồi quy trong phương trình tương quan liên kết hai biến lượng bằng y/x r_x giữa x và y được thay thế bằng tổ hợp các hệ số đó tính từng cặp đại lượng trong phương trình hồi quy. Chẳng hạn, trong trường hợp có 3 biến lượng y, x_1, x_2 phương trình hồi quy có dạng:

$$y^0 = \frac{\sigma_y}{\sigma x_1} . \frac{r_{yx1} - r_{yx2} r_{x1x2}}{1 - r^2_{x1x1}} \overset{0}{x}_1 + \frac{\sigma_y}{\sigma x_2} \frac{r_{yx2} - r_{yx1} r_{x1x2}}{1 - r^2_{x1x1}} \overset{0}{x}_2$$

Trong trường hợp có 4 biến lượng ta có:

$$\begin{aligned}
\overset{0}{y} = \frac{\sigma_y}{\sigma_{x1}} & \left[\frac{r_{x1y}(1 - r_{x2x3}2) - r_{x2v} r_{x2x3} - r_{x1x3}}{1 - r^2_{x2x3} - r^2_{x1x2} + 2 r_{x1x2} r_{x1x3} r_{x2x3}} + \right. \\
& \left. + \frac{r_{x2x3}(r_{ü2} r_{x1x3} + r_{yx3} r_{x1x2})}{1 - r^2_{x2x3} - r^2_{x1x2} + 2 r_{x1x2} r_{x1x3} r_{x2x3}} \right] \overset{0}{x}_1 + \\
& + \frac{\sigma_y}{\sigma_{x2}} \left[\frac{r_{yx2}(1 - r^2_{x1x3}) - r_{yx1} r_{x1x2} - r_{yx3} r_{x2x3}}{1 - r^2_{x2x3} - r^2_{x1x2} + 2 r_{x1x2} r_{x1x3} r_{x2x3}} + \right. \\
& \left. + \frac{r_{x1x3}(r_{2ü1} r_{x2x3} + r_{yx3} r_{x2x1})}{1 - r^2_{x2x3} - r^2_{x1x2} - r^2_{x1x3} + 2 r_{x1x2} r_{x1x3} r_{x2x3}} \right] \overset{0}{x}_2 + \qquad (6.29) \\
& + \frac{\sigma_y}{\sigma_{x3}} \left[\frac{r_{yx2}(1 - r^2_{x1x2}) - r_{yx1} r_{x3x1} - r_{yx2} r_{x3x2}}{1 - r^2_{x2x3} - r^2_{x1x2} - r_{x1x3} + 2 r_{x1x2} r_{x1x3} r_{x2x3}} + \right. \\
& \left. + \frac{r_{x1x2}(r_{ü1} r_{x2x3} + r_{yx2} r_{x1x3})}{1 - r^2_{x2x3} - r^2_{x1x2} - r^2_{x1x3} + 2 r_{x1x2} r_{x1x3} r_{x2x3}} \right] \overset{0}{x}_3
\end{aligned}$$

Các hệ phương trình viết dưới dạng (6.28) và (6.29) có thể được dùng đơn giản hoá và khái quát hoá cho trường hợp chung của các biến lượng bằng cách sử dụng định thức . Trong trường hợp đó các biểu thức tổng quát của hệ số hồi quy (kj) có thể viết dưới dạng:

$$kj = \frac{\sigma_y}{\sigma_{xj}} \cdot \frac{D_{yxj}}{D_{yy}} \qquad (6.30)$$

trong đó, σ_y khoảng lệch trung bình bình phương của biến lượng phụ thuộc (hàm số) σ_{xj} - khoảng lệch trung bình bình phương của biến lượng độc lập D_{yj} hay D_{yy} là những định thức con của định thức:

$$\begin{vmatrix} 1 & r_{yx1} & r_{yx2}\ldots. & r_{yx}\ j\ldots. & r_{yxn} \\ r_{x1y} & 1 & r_{x1x2}\ldots. & r_{x1xj}\ldots. & r_{xixn} \\ r_{x2y} & r_{x2x1} & 1 & r_{x2xj} & r_{x2xn} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ r_{xjy} & r_{xjx1} & r_{xjx2} & 1 & r_{xjxn} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ r_{xny} & r_{xnx1} & r_{xnx2}\ldots & r_{xnxj} & 1 \end{vmatrix} \qquad (6.31)$$

Định thức con là một phân của định thức gốc (D) . Trong trường hợp này, dòng đầu tiên và cột dọc ứng với biến lượng có trong ký hiệu của định thức con được xoá đi.Thí dụ định thức con thứ nhất D_{yy} nghĩa là định tức gốc D được xoá đi dòng thứ nhất và cột thứ nhất:

$$D_{yy} = \begin{vmatrix} 1 & r_{x1x2}\ldots & r_{xixj}\ldots & r_{x1xn} \\ r_{x2x1} & 1 & \ldots r_{x2xj}\ldots & r_{x2xn} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ r_{xjx1} & r_{xjx2} & 1\ldots & r_{xjxn} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ r_{xnx1} & r_{xnx2}\ldots & r_{xnxj}\ldots & 1 \end{vmatrix}$$

Định thức con $D_y x_1$ là định thức D được xoá đi dòng thứ nhất của cột thứ hai. Đó là định tức con thứ hai. Định tức con thứ ba (D_{yx2}) nhận được bằng cách xoá trong định thức D dòng thứ nhất và cột thứ ba của nó. Định tức con thứ tư D_{yx3} là xoá dòng thứ nhất và cột thứ tư trong định thức D, vv ... Những giải thích trên đều thuộc về trường hợp xác định hệ số hồi quy. Dưới dạng tổng quát định thức con D_{ij} nhận được bằng cách xoá đi dòng thứ i và cột thứ j của định thức D.

Sau đây chúng ta sẽ xem xét mối quan hệ của hệ các định thức với biểu thức dạng (6.28). căn cứ vào biểu thức (6.30) các hệ số hồi quy đối với phương trình ba biến lượng.

$$\overset{0}{y} = k_1 \overset{0}{x}_1 + k_2 \overset{0}{x}_2 \qquad (6.32)$$

là

$$k_1 = \frac{\sigma_y}{\sigma_{x1}} \cdot \frac{D_{yx1}}{D_{yy}}$$

$$k_1 = \frac{\sigma_y}{\sigma_{x2}} \cdot \frac{D_{yx2}}{D_{yy}}$$

Đối với trường hợp nghiên cứu định thức có dạng:

$$D = \begin{vmatrix} 1 & r_{yx1} & r_{yx2} \\ r_{yx1} & 1 & r_{x1x2} \\ r_{yx2} & r_{x1x2} & 1 \end{vmatrix} \qquad (6.33)$$

Các định thức con của định thức này trong biểu thức hệ số hồi quy bằng:

$$D_{yx1} = \begin{vmatrix} r_{yx1} & r_{x1x2} \\ r_{yx2} & 1 \end{vmatrix} = r_{yx1} - r_{yx2} r_{x1x1},$$

$$D_{yx2} = -\begin{vmatrix} r_{yx1} & 1 \\ r_{yx2} & r_{x1x2} \end{vmatrix} = -\left[r_{yx1} r_{x1x2} - r_{yx2}\right] = r_{yx2} - r_{yx1} r_{x1x2},$$

$$D_{yy} = \begin{vmatrix} 1 & r_{x1x2} \\ r_{x1x2} & 1 \end{vmatrix} = 1 - r_{x1x2}^2$$

phương trình hồ quy sẽ có dạng (6.28). Các nguyên tố của định thức (6.33) là những hệ số tương quan từng đôi một giữa các biến nghiên cứu, được xác định theo phương trình

$$r_{jk} = \frac{\sum_{1}^{m}(x_j - \bar{x}_j)(x_k - \bar{x}_k)}{\sqrt{\sum_{1}^{m}(x_j - \bar{x}_j)^2 \sum_{1}^{m}(x_k - \bar{x}_k)^2}} \qquad (6.34)$$

Hệ số tương quan toàn phần hay hệ số tương quan tổng giữa biến lượng phụ thuộc vào tất cả các biến lượng độc lập được xác định theo biểu thức:

$$R = \sqrt{1 - \frac{D}{D_{yy}}} \qquad (6.35)$$

Nếu các hệ số tương quan từng đôi một đặc trưng cho mức độ quan hệ giữa hai biến lượng biến thiên từ - đến 1, thì hệ số tương quan tổng hợp, có giới hạn biến thiên từ 0 đến 1.

Mối quan hệ tuyến tính (chứ không phải quan hệ bất kỳ) giữa hai biến lượng mà không có $r=0$ và $R=0$ giữa n biến lượng. Trong trường hợp có mối quan hệ hàm số giữa các biến lượng $r = \pm 1,0$; $R = 1,0$.

Hệ số tương quan tổng hợp R khi $r = 1$ luôn luôn lớn hơn bất kỳ hệ số tương quan từng đôi một nào ở trong định thức D, ngoài đường chéo, ở đây $r_{yy} = r_{x1x1} = r_{x2x2} = ... = r_{x2x2} = ... = r_{x1xn} = 1$.

Ta nhận thấy rằng hệ số tương quan tổng hợp có thể được tính như thế là hệ số tương quan từng đôi một giữa các giá trị quan trắc được của biến lượng phụ thuộc vào những giá trị tính toán theo đường hồi quy.

Khoảng lệch trung bình bình phương của các giá trị quan trắc (y_{qtr}) so với các giá trị tính toán được theo phương trình hồi quy (y_{tt}) đặc trưng cho độ chính xác của phương trình hồi quy mẫu đem dùng, có thể được tính theo công thức:

$$\sigma_y = \sqrt{\frac{\sum_{1}^{m}(y_H - y_p)^2}{m}} \qquad (6.36)$$

hay là

$$\sigma_y = \sigma_0 \sqrt{1 - R^2} \qquad (6.37)$$

R=1, $\sigma_y = 0$, chứng tỏ rằng giá trị quan trắc và tính toán được theo phương trình (6.30) hoàn toàn phù hợp với nhau.

Khi R=0; $\sigma_y = \sigma_0$ thì suy ra việc sử dụng phương trình hồi quy sẽ không có ý nghĩa.

Để đánh giá sai số trung bình bình phương của hệ số phương trình hồi quy (k_j) ta sử dụng công thức:

$$\sigma_{kj} = \sqrt{\frac{m\sigma_y}{(m-n)P_{kj}}}, \quad j = 1,2,...n, \qquad (6.38)$$

trong đó m - số số hạng của chuỗi được dùng để xây dựng phương trình hồi quy n số biến lượng độc lập.

$$P_{kj} = \frac{\Delta}{\Delta_{jj}}$$

Δ_{jj} định thức con của định thức, nhận được bằng cách xoá bỏ dòng thứ j và cột thứ j trong định thức.

Từ phương trình (6.38) rút ra sai số của hệ số hồi quy (trong các điều kiện như nhau) sẽ tăng lên khi số lượng biến lượng tăng. Với độ dài của chuỗi các giá trị thuỷ văn không lớn lắm việc sử dụng số biến lượng nhiều hơn bốn đê xây dựng phương trình hồi quy nhiều chiều sẽ nhận được những hệ số hồi quy kém chính xác và lời giải về mặt thống kê là kém ổn định.

Biểu thức tổng quát của sai số trung bình bình phương của hệ số hồi quy k_j đối với trường hợp ba biến lượng được viết dưới dạng:

$$\sigma_{ki} = \frac{\sigma y}{\sigma_{x1}\sqrt{(m-2)(1-r_{x1x2}^2)}} \qquad (6.39)$$

$$\sigma_{k2} = \frac{\sigma y}{\sigma_{x2}\sqrt{(m-2)(1-r_{x1x2}^2)}} \qquad (6.40)$$

Đối với trường hợp 4 biến lượng ta có:

$$\sigma_{k1} = \frac{\sigma_y}{\sigma_{x1}} \sqrt{\frac{1 - r_{x2x3}^2}{(m-3)(1 - r_{x1x2}^2 - r_{x1x3}^2 - r_{x2x3}^2 + 2r_{x1x2}r_{x1x3}r_{x2x3})}},$$

$$\sigma_{k2} = \frac{\sigma_y}{\sigma_{x2}} \sqrt{\frac{1 - r_{x2x3}^2}{(m-3)(1 - r_{x1x2}^2 - r_{x1x3}^2 - r_{x2x3}^2 + 2r_{x1x2}r_{x1x3}r_{x2x3})}},$$

$$\sigma_{k3} = \frac{\sigma_y}{\sigma_{x3}} \sqrt{\frac{1 - r_{x2x3}^2}{(m-3)(1 - r_{x1x2}^2 - r_{x1x3}^2 - r_{x2x3}^2 + 2r_{x1x2}r_{x1x3}r_{x2x3})}}.$$

Khả năng sử dụng công cụ tương quan tuyến tính nhiều chiều trong thực tế nghiên cứu được gắn liền với việc ứng dụng MTĐT. Thí dụ chương trình được lập dưới dạng ngôn ngữ ALGOL - 6.0 đã được dùng ở GGI để nghiên cứu quan hệ giữa nhiều biến lượng đối với bài toán kéo dài các chuỗi thuỷ văn sang thời kỳ nhiều năm. Để giải bài toàn này phải dựa vào các mối quan hệ thống kê của nhiều biến lượng.

Với các chuỗi tài liệu đặc trưng của chế độ thuỷ văn thường thường có độ dài khác nhau và những năm không quan trắc được, chương trình sẽ xét trước bài toán lượng thông tin gốc của tình hình bằng cách bổ sung vào chỗ không có tài liệu bằng các số không. Chương trình sẽ xét trước việc tìm tự động những biến lượng độc lập hiệu quả nhất và khôi phục theo các biến lượng đó những giá trị không quan tác được.

6.4. ỨNG DỤNG PHƯƠNG PHÁP TƯƠNG QUAN TUYẾN TÍNH NHIỀU CHIỀU ĐỂ KÉO DÀI CÁC CHUỖI THUỶ VĂN NGẮN SANG THỜI KỲ NHIỀU NĂM.

Chúng ta sẽ xét trình tự việc kéo dài các giá trị dòng chảy năm sang thời kỳ nhiều năm của sông Kivir - trạm Miatuxôvô bằng cách sử dụng sông tương tự là các sông Vuokxy - trạm nhà máy thuỷ điện X (x_1) và sông Nêva - trạm Pêtrôkrêpost (x_2)

Đối với sông Xvir - trạm Miatuxôvô có tài liệu quan trắc trong các thời đoạn 1881 - 1940 và 1945 - 1951. Để minh hoạ phương pháp xây dựng phương trình hồi quy ta giả thiết rằng ở điểm này thông tin về dòng chảy năm chỉ có trong 20 năm (1928 - 1940 và 1945 - 1951) được quan trắc đồng bộ trên ba sông nghiên cứu. Dựa vào tài liệu quan trắc đồng bộ tròn 20 năm đó ta tìm các hệ số tương quan của từng đôi một: r_{yx1} = 0,74; r_{yx2}=0,88; r_{x1x2} = 0,55. Trong trường hợp này định thức (D) và các định thức con của nó sẽ bằng:

$$D = \begin{vmatrix} 1 & 0,7 & 0,88 \\ 0,74 & 1 & 0,55 \\ 0,88 & 0,55 & 1 \end{vmatrix} = 1 - 0,55 - 0,77 - 0,30 + 2.0,74.0.88.0,55 = 0,1$$

$$D_{yx1} = \begin{vmatrix} 0,74 & 0,55 \\ 0,88 & 1 \end{vmatrix} = -(0,74 - 0,88.5,55) = -0,26,$$

$$D_{yx2} = \begin{vmatrix} 0,74 & 1 \\ 0,88 & 0,55 \end{vmatrix} = -(0,74.0,55 - 0,88) = -0,47,$$

$$D_{yy} = \begin{vmatrix} 1 & 0,55 \\ 0,55 & 1 \end{vmatrix} = 1 - 0,55^2 = -0,0,70.$$

Hệ số tương quan tổng hợp bằng:

$$R = \sqrt{1 - \frac{D}{D_{yy}}} = \sqrt{1 - \frac{0,10}{0,70}} = 0,93$$

Việc tính toán trị bình quân và khoảng lệch trung bình bình phương của chuỗi gốc sẽ nhận được kết quả sau:

$$\bar{y} = 0,876, \bar{x}_1 = 9,79, \bar{x}_2 = 8,75$$
$$\sigma_y = 1,71; \sigma_{x1} = 1,96; \sigma_{x2} = 1,58$$

Các hệ số hồi quy của phương trình bằng:

$$k_1 = \frac{1,71.0,26}{1,96.0,70} = 0,32$$
$$k_2 = \frac{1,71.0,47}{1,58.0,70} = 0,73$$

Với những tài liệu gốc trên ta sẽ xây dựng phương trình hồi quy dạng:

$$y - \bar{y} = k_1(x1 - \bar{x}_1) + k_2(x_2 - \bar{x}_2)$$

hay là f:

Y-8,76 = 0,32 (x_1 - 9,79) + 0,73 (x2 - 8,57) nghĩa là:

$$Y=0,32x_1 + 0,73x_2 - 0,63 \qquad (6.41)$$

Khoảng lệch trung bình bình phương của mối quan hệ vừa nhận được bằng:

$$\sigma_{\bar{y}} = \sigma_y \sqrt{1-R^2} = 1,71\sqrt{1-0,93^2} = 0,63$$

Khoảng lệch trung bình bình phương của các hệ số hồi quy k_1 và k_2 theo công thức (6.39) và (6.40) bằng:

$$\sigma_{k1} = \frac{0,63}{1,96\sqrt{(20-2)(1-0,55^2)}} = 0,09$$

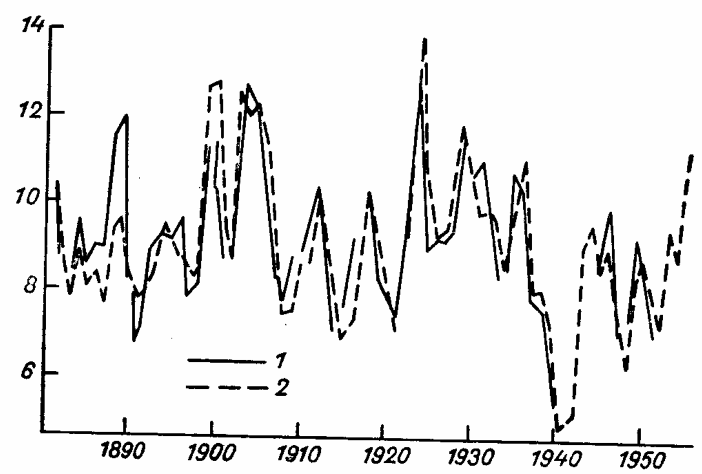
$$\sigma_{k2} = \frac{0,63}{1,58\sqrt{(20-2)(1-0,55^2)}} = 0,11$$

Những giá trị k_1 và k_2 vừa nhận được chứng tỏ rằng các hệ số hồi quy có độ tin cậy cao

Phương trình hồi quy (6.41) sẽ được dùng để tính các giá trị dòng chảy năm nước sông Xvir trạm Miatuxôvô trong thời kỳ 1859 - 1958 nghĩa là trong thời kỳ mà ở các sông tương tự có tài liệu quan trắc. Từ dưới dòng chảy năm đã được khôi phục trong thời kỳ đó ta tiến hành nghiên cứu các mẫu của các khoảng thời gian riêng biệt. Trị bình quân và hệ số biến đổi của các mẫu quan trắc và khôi phục đã được trình bày trong bảng 6.2. Trong bảng 6.2 đã tiến hành so sánh những giá trị tính toán với các giá trị quan trắc để đánh giá độ chính xác của phương trình dùng để kéo dài. Trong các bài toán thực tế kéo dài tài liệu rõ ràng là sự so sánh đó không thể tực hiện được. Sự so sánh dòng chảy năm tính toán và quan trắc được của sông Xvir - trạm Miatuxôvô (hình 6.4) cho thấy rằng sự phù hợp của chúng rất tốt.

Chúng ta sẽ nghiên cứu dạng khá đầy đủ của việc sử dụng tương quan tuyến tính nhiều chiều ở thí dụ kéo dài tài liệu dòng chảy năm nước sông XOZ - trạm Xmavgorod bằng cách sử dụng 8 sông tương tự được trình bày ở bảng 6.3

Ta nhận thấy rằng việc quan trắc thực tế dòng chảy năm sông Xôz - trạm Clavgorod đã được thực hiện trong một thời kỳ khá dài (1897 - 1968), vì vậy lược đồ nghiên cứu sau đây đối với tuyến này chỉ có ý nghĩa về mặt phương pháp luận. Những tính toán trình bày sau đây, trong đó việc lựa chọn sông tơng tự đã đợc thực hiện trên máy tính điện tử M - 220. Khi tiênd hành nghiên cứu phương pháp được trình bày sau đây người ta lấy thời kỳ quan trắc đồng bộ trên sông Xôz và các sông tương tự là 12 năm (1957 - 1968).



Hình 6.4. Sự dao động nhiều năm của dòng chảy năm s.Xtr.r - tr Miatuxôvô .

1 - Theo tài liệu quan trắc được.

2. Theo tài liệu khôi phục được bằng phương trình hồi quy

Bảng 6.2 Trị bình quân và hệ số biến đổi của dòng chảy năm s.Xvir trạm Miatuxôvô quan trắc và tính toán được trong các thời kỳ khác nhau bằng cách sử dụng phương trình tương quan

Thời kỳ (năm)	Chuỗi quan trắc		Chuỗi đã khôi phục	
	Trị bình quân (l/skm^2)	Hệ số biến đổi Cv	Trị bình quân (l/skm^2)	Hệ số biến đổi Cv
1928 - 1940				
1945 - 1951	8,76	0,19	8,76	0,19
1881 - 1927	9,38	0,16	9,25	0,18
1881 - 1940				
1945 - 1951	9,31	0,16	9,11	0,18
1959 - 1958	-	-	9,16	0,18

Chương trình sẽ xét trước việc thiết lập phương trình hồi quy tuyến tính nhiều chiều của biến lượng phụ thuộc (trạm kéo dài) với tất cả các biến lượng độc lập (trạm tương tự). Lúc này phải tiến hành chọn tất cả tổ hợp có thể có của các trạm tương tự điều đó đã dẫn đến việc xây dựng C_n^k phương trình, trong đó n - tổng số điểm tương tự dùng để xây dựng các phương trình, k số trạm tương tự dùng để xây dựng các phương trình riêng. Trong trường hợp nghiên cứu này số trạm tương tự n = 8.

Vì thế, số phương trình có một trạm tương tự bằng $C^1_8 = 8$

có hai trạm tương tự - $C^2_8 = 28$

có ba trạm tương tự - $C^3_8 = 56$

có bốn trạm tương tự - $C^4_8 = 70$

có năm trạm tương tự - $C^5_8 = 56$

có sáu trạm tương tự - $C^6_8 = 28$

có bảy trạm tương tự - $C^7_8 = 8$

Như vậy từ các phương trình trên đối với việc sử dụng trực tiếp người ta chọn sao cho $^{kj}/_{ki} = 2$ nghĩa là phương trình hồi quy được công nhận là tin cậy, nếu như sai số trung bình bình phương của hệ số hồi quy nhỏ hơn hai lần giá trị tuyệt đối của hệ số hồi quy.

Bảng 6.3 Tài liệu về các sông tương tự dùng để kéo dài dòng chảy năm s.Xoztr Xlavgorod sang thời kỳ nhiều năm.

STT	Sông - trạm	Diện tích lưu vực km^2	Số năm quan trắc	Tham số		
				xl/skm^2	Cv	Cs
1.	1.	1.	1.	1.	1.	1.
0	Xoz - Xlavgôd	17.700	68	6,01	0,29	1,11
1	Dnepr - Xmôlensk	14.100	81	6,88	0,27	0,91
2	Xoz - Gômel	38.900	66	5,26	0,30	0,87
3	Bêrêzina - Bobruisk	20.200	79	5,89	0,20	0,99
4	Đexna - Briansk	12.400	70	5,47	0,30	0,94
5	Đexna - chernigôv	81.400	75	3,97	0,29	0,79
6	Dnepr - Rêchisa	58.200	68	6,34	0,23	0,52
7	Dnepr - Orsa	28.000	81	6,96	0,27	0,92
8	Priplat - Nôzưn	97.200	83	3,78	0,30	0,30

Nếu như trong một phương trình nào đó dù chỉ đối với một phương trình hồi quy có $-\frac{kj}{\sigma k_j} \geq 2$ thì phương trình hồi quy đó sẽ không được sử dụng trong tính toán. Xuất phát từ điều kiện đó trong thí dụ này việc kéo dài dòng chảy năm sang thời kỳ nhiều năm sẽ không phù hợp với phương trình nào đó có bốn và nhiều hơn sông tương tự. Trong số 56 phương trình có 3 sông tương tự. Trong số 56 phương trình có 3 sông tương tự (của 4 biến lượng), theo điều kiện đó chỉ đề cập có 3 phương trình ; trong số 28 phương trình có hai tương tự chỉ đề cập 8 phương trình, và cuối trong số 8 phương trình có một sông tương tự sẽ dề cập tất cả các phương trình đó.

Sau đây chúng ta sẽ trình bày các phương trình hồi quy thoả mãn với điều kiện $-\frac{k_j}{k_j} \geq 2$; trong phương trình này chỉ số ở các biến lượng ứng với số thứ tự của sông trong bảng 6.3

$Y = 1{,}36\ x_7 + 0{,}49\ x_5 + 0{,}69\ x_1 - 0{,}50$ (1)

$Y = 0{,}28\ x_7 + 0{,}88\ x_2 - 0{,}28\ x_8 - 0{,}51$ (2)

$Y = 0{,}71\ x_2 + 0{,}53\ x_6 - 0{,}38\ x_8 + 0{,}29$ (3)

$Y = 0{,}37\ x_1 + 0{,}61\ x_3 - 0{,}72$ (4)

$Y = 0{,}34\ x_7 + 0{,}62\ x_5 + 0{,}35$ (5)

$Y = 0{,}45\ x_7 + 0{,}50\ x_3 - 0{,}56$ (6)

$Y = 0{,}61\ x_7 + 0{,}58\ x_5 - 0{,}41$ (7)

$Y = 1{,}36\ x_2 + 0{,}45\ x_8 + 0{,}57$ (8)

$Y = 1{,}48\ x_4 + 1{,}61\ x_5 - 2{,}50$ (9)

$Y = 0{,}44\ x_4 + 0{,}82\ x_8 - 0{,}52$ (10)

$Y = 027 x_1 + 0{,}71\ x_2 - 0{,}38$ (11)

$Y = 0{,}56\ x_1 + 1{,}03$ (12)

$Y = 0{,}71\ x_7 + 1{,}05$ (13)

$Y = 1,08\ x_2 + 0,30$ (14)

$Y = 1,10x_3 - 1,35$ (15)

$Y = 0,58\ x_4 + 2,80$ (16)

$Y = 1,91\ x_5 - 1,25$ (17)

$Y = 0,91\ x_6 - 6,39$ (18)

$Y = 1,05\ x_8 + 1,185$ (19)

Việc đánh hệ số hồi quy ($\frac{kj}{\sigma k_j}$) của các phương trình vừa nhận được trình bày ở bảng 5.4. Trong cũng có trình bày các hệ số tương quan nhiều chiều(R), số năm quan trắc (n) được dùng để xây dựng các phương trình hồi quy và tổng độ dài của chuỗi (N) bao gồm các giá trị dòng chảy quan trắc và khôi phục được.

Căn cứ vào giá trị $\gamma = \frac{k_j}{\sigma k_j}$ có thể đánh giá được phương trình nào là thích hợp. Đối với các phương trình có một tương tự, tương ứng được xác định bằng giá trị $\frac{k_j}{\sigma k_j}$ duy nhất phương trình có giá trị lớn hơn là tin cậy hơn. Đối với các phương trình có 2 , 3 và nhiều biến lượng, phương trình tốt nhất là phương trình có giá trị $\min = \frac{k_j}{\sigma k_j}$ là nhỏ nhất. Chẳng hạn như đối với phương trình (3) có giá trị nhỏ nhất $\gamma_{\min} = \frac{k_j}{\sigma k_j}$, nhưng trong khi đó nó lại là lớn nhất trong số các giá trị nhỏ nhất khác (2,07 , 2,54). Như vậy phương trình (3) là tin cậy hơn cả.

Các hệ số tương quan nhiều chiều lớn là do hệ số tương quan từng đôi một của trạm được kéo dài với trạm tương tự mà có tài liệu quan trắc trong toàn bộ thời kỳ là lớn.

Ma trận hệ số tương quan tường đôi một là đối xứng , vì vậy nó chỉ được tính theo đường chéo chính. Việc đánh số các cột và các dòng được thực hiện theo số trong bảng 6.3.

Bảng 6.4. Ước lượng các tham số của đường hồi quy.

Phương trình	kj/kj			R	n năm	N năm
	j=1	j=2	j=3			
1.	2.	3.	4.	5.	6.	7.
(1)	2,13	3.91	2,07	0,98	11	58
(2)	3,89	5,97	2,42	0,99	12	63
(3)	4,00	4,08	3,49	0,99	12	62
(4)	5,38	6,23		0,99	11	79
(5)	4,20	5,03		0,99	12	63
(6)	7,67	5,27		o,99	11	77
(7)	8,40	2,23		0,97	12	69
(8)	10,7	2,66		0,98	12	66
(9)	4,13	4,49		0,91	12	70
(10)	3,12	3,28		0,86	12	70
(11)	3,24	5,34		0,98	11	64
(12)	8,18			0,93	11	82
(13)	10,6			0,95	12	81
(14)	12,1			0,96	12	66
(15)	8,00			0,93	11	80
(16)	3,09			0,68	12	70
(17)	3,54			0,72	12	75
(18)	15,4			0,98	12	68
(19)	3,26			0,70	12	83

Theo các phương trình hồi quy (1) - (19) ta đã khôi phục được tài liệu dòng chảy năm của sông Xoz-trạm Xlavgorod trong thời kỳ quan trắc của các trạm tương tự. Tiếp theo, theo chuỗi năm bao gồm dòng chảy năm quan trắc được trong thời kỳ từ 1957 đến 1968 và các giá trị được khôi phục, ta tính dược các tham số của chuỗi (Y,Cv,Cs) và trình bày trong bảng 6.6. từ tài liệu của bảng này sẽ rút ra được sự hội tụ tốt nhất của các tham số mẫu được khôi phục bằng tài liệu quan trắc trong toàn bộ thời kỳ (Y = 6,10 ; cv = 0,29 ; cs = 1,11) ứng với ba phương án đầu (tiến hành kéo dài theo ba trạm tương tự). Khi có 19 phương án kéo dài điều tất nhiên khi sinh ra là phải sử dụng một cách hợp lý nhất. Rõ ràng là ứng với các ước lượng của phương trình trình bày ở bàng 6.4 và 6.6, trước hết là nên sử dụng phương trình (3) có giá trị min lớn hơn cả $\gamma\min = \frac{k_j}{\sigma k_j}$. Song phương trình này cho phép ta kéo dài chuỗi dòng chảy sông Xoz trong trạm vi tài liệu quan trắc đồng bộ có ở các sông tương tự x_2,x_6,x_8 (sông Xoz = trạm Gomel, sông Đnepr - trạm Retrixa, sông Pripiat - trạm Mozur) nghĩa là trong các giai đoạn 1900 - 1930; 1935 - 1940; 1944 - 1956. Theo nức độ tin cậy phương trình (2) đứng ở vị trí thứ hai được dùng để khôi phục dòng chảy của những năm là có tài liệu quan trắc đồng bộ trên các sông tương tự x_7, x_2, x_8, không cùng với thời kỳ được dùng trong tính toán theo phương trình (1). Trong trường hợp này dường như là thời kỳ 1931 - 1934.

Tương tự như vậy phương trình (3) được dùng để bổ sung chuỗi đã được khôi phục bằng lượng thông tin trong các thời kỳ 1866 - 1887 và 1895 - 1899.

Như vậy ta đã nhận được chuỗi bằng cách sử dụng ba trạm tương tự theo những phương trình hồi quy khác nhau.

Các tham số của chuỗi dòng chảy năm sông Xoz - trạm Xlavgorod đã được khôi phục như trên bằng : $Y = \frac{k_3}{\sigma k_3}$; Cv = 0,29 ;Cs = 0,76; nghĩa là rất phù hợp với tài liệu quan trắc được trong tất cả thời kỳ quan trắc.

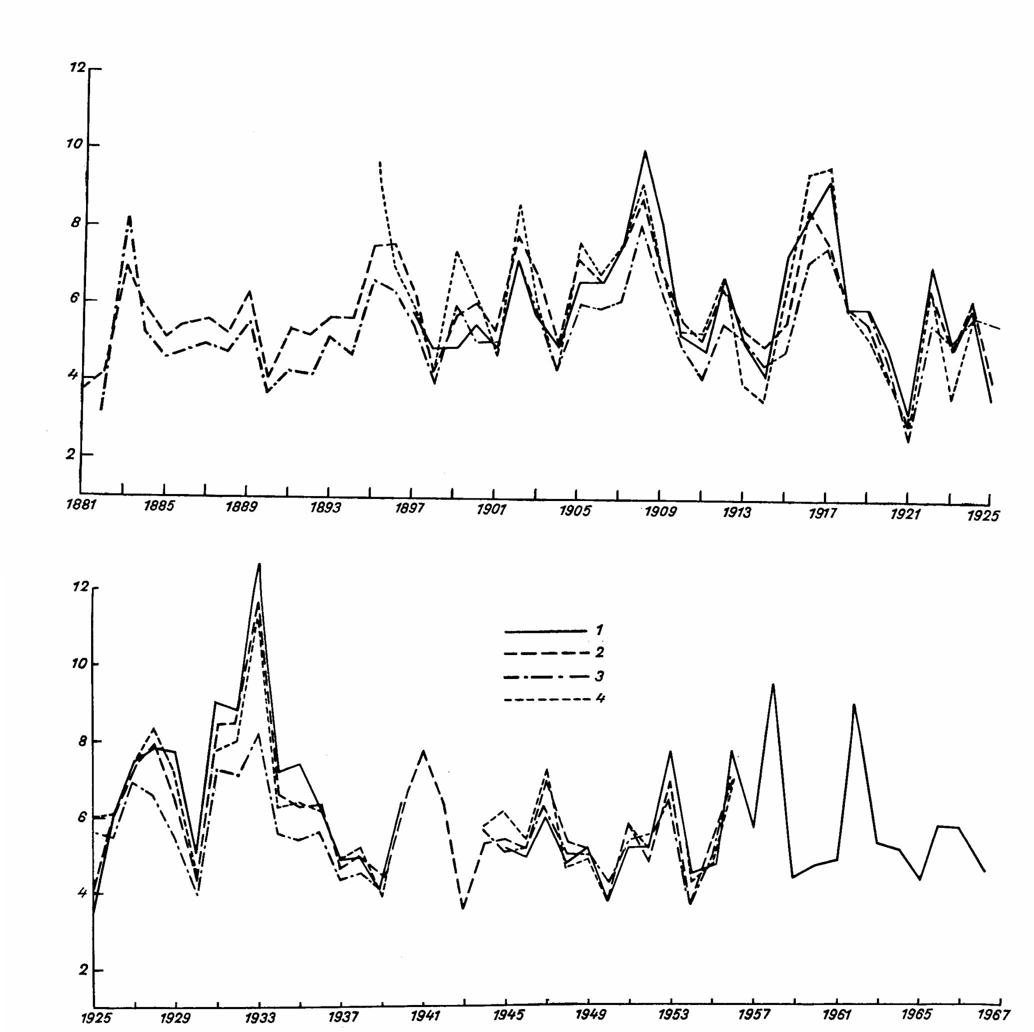
Bảng 6.5. Ma trận của hệ số tương quan từng đôi một.

r	0	1.	2.	3.	4.	5.	6.	7.	8.
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
0.	1	0,38	0,85	0,96	0,75	0,79	0,81	0,91	0.45
1.		1	0,98	0,79	0,60	0,73	0,68	0,85	0,23
2.			1	0,83	0,64	0,73	0,70	0,87	0,28
3.				1	0,79	0,75	0,85	0,90	0,48
4.					1	0,62	0,59	0,84	0,39
5.						1	0,64	0,73	0,31
6.							1	0,68	0,40
7.								1	0,38
8.									1

1) Số thứ tự đã cho theo bảng 6.3.

Tương tự như vậy ta đã nhận được chuỗi khôi phục bằng cách sử dụng hai trạm tương tự và một trạm tương tự. Trong các trường hợp đó hiệu quả của việc kéo dài là nhỏ hơn so với ba trạm tương tự. Thật vậy, các tham số phân phối đối với chuỗi vừa khôi phục theo hai trạm tương tự bằng cách sử dụng các phương trình (4), (5) tương tự (9) bằng : $\overline{Y} = 5{,}45$,Cv = 0,23 , Cs = 0,78. ĐIều đó phù hợp với tài liệu quan trắc kém hơn so với việc kéo dòi theo ba trạm tương tự. Cuối cùng, sử dụng các phương trình (12) - (14), (16) tương tự và (17) ta nhận được các tham số theo một trạm tương tự $\overline{Y} = 5{,}89$, Cv = 0,25, Cs = 0,92.

Sau đây ta sẽ trình bày bảng thông báo tổng quát về các thời kỳ khôi phục dòng chảy sông ngòi theo ba, hai, và một trạm tương tự .



Hình 6.5 - Sự dao động nhiều năm của dòng chảy năm s.Xoz-tr.Xlavgorod.

1 - tài liệu quan trắc được , 2 - dòng chảy đã được khôi phục tương đương tương tự, 3 - dòng chảy đã được khôi phục theo tương tự, 4 - dòng chảy đã được khôi phục theo 3 trạm tương tự.

+ 3 trạm tương tự : 1900 - 1930. 1935 - 1940. 1944 - 1956

Theo phương trình (3)

1931 - 1934 theo phương trình (2)

1886,1887, 1885 - 1889 theo phương trình (1).

+ 2 trạm tương tự , 1882 - 1917, 1921 - 1939, 1945 - 1920, 1940 theo phương trình , 1944 theo phương trình .

+ 1 trạm tương tự , 1895 - 1930, 1935 - 1940, 1942, 1944 - 1956. Theo phương trình

1931 - 1934 theo phương trình .

1882 - 1894 theo phương trình .

1881 theo phương trình .

1941 - 1943 theo phương trình .

Bên cạnh sự hội tụ khá tốt của các tham số phải nhận được theo các chuỗi quan trắc tương tự và khôi phục được. Còn thấy có sự phù hợp hoàn toàn thoả mãn được rút ra hình 6.5.

Những thí dụ sử dụng công cụ toán học của đường quy nhiều chiều để kéo dài các chuỗi thuỷ văn ngắn sang thời kỳ nhiều năm vẫn không cụ thể hết được các khả năng ứng dụng của nó trong tính toán thuỷ văn. Như vậy tương tuyến tính nhiều chiều được dụng vào mục đích dự báo là thiết lập được phương trình hồi quy tuyến tính nhiều chiều của biến phụ thuộc (preditant - đặc trưng được dự báo) các biến độc lập (prediktor - các yếu tố tạo nên hiện tượng phải dự báo). Trường hợp riêng của mỗi quan hệ tuyến tính tương tự là phương trình hồi quy của dòng chảy sông ngòi mặt cắt khống chế với tài liệu dòng chảy của các trạm trên các sông nhánh.

Ngoài ra, có rất nhiều thí dụ ta đã biết, về việc sử dụng tương quan tuyến tính nhiều chiều để xác lập mối quan hệ của đặc trưng thuỷ văn nào đó (dòng chảy năm của con sông, dòng chảy rắn, dòng chảy nhỏ nhất, hệ số điều tiết trong năm ...) với các yếu tố hình thành nên nó.

Công cụ toán học tương quan tuyến tính chỉ có quyền sử dụng khi giữa các biến lượng gốc có mối quan hệ tuyến tính hay gần như là tuyến tính. Việc đánh giá độ tuyến tính của mối quan hệ có thể được xác định từ nhận thứ bản chất này hay căn cứ vào phân tích sơ bộ tài liệu gốc.

Thông thường mức độ tuyến tính giữa các biến lượng thuỷ văn được xác định bằng đồ thị cùng với việc phân tích các điểm lệch xa nhất so với đường quan hệ giả thiết. Khi tiến hành phân tích các quan hệ đồ thị cần phải đặc biệt chú ý đến lời giải về mối quan hệ không tuyến tính, vì khi số điểm ít trên đường của đồ thị và sự phân tán của chúng lại lớn dường như mức độ không tuyến tính có thể là do sự phân bổ các đường điểm tạo nên. Sự phân tích như vậy của mối quan hệ đồ thị thường là rất nhiều. Lúc này có thể sử dụng các phương pháp định lượng đánh giá mức độ tuyến tính của mối quan hệ đã được nghiên cứu trong thống kê toán. Trong trường hợp các chuỗi gốc không tuyến tính phương pháp ưu thế nhất xác định phương trình hồi quy là sự chuẩn hoá sơ bộ và tuyến tính hoá các biến lượng gốc. Những phương pháp chuẩn hoá các biến lượng thuỷ văn đã được nghiên ưứu ở chương II. Vấn đề này được trình bày cụ thể nhất là trong cuốn khảo cứu của G,A. Alekxeev (9). Trong thực tế thuỷ văn thường chuyển về dạng tuyến tính bằng phép biến đổi Lôgarit các biến lượng.

6.5. ƯỚC LƯỢNG HÀM TƯƠNG QUAN KHÔNG GIAN CỦA CÁC ĐẶC TRƯNG THUỶ VĂN (DÒNG CHẢY SÔNG NGÒI).

Thời gian gần đây tương quan tuyến tính nhiều chiều bắt đầu được dùng để nội suy không gian các đặc trưng thuỷ văn và hợp lý hoá lưới trạm thuỷ văn. Giải các bài toán này phải dựa vào việc sử dụng hàm tương quan không gian của các yếu tố đặc trưng cho chế độ thuỷ văn, có đánh giá tĩnh đồng nhất về mặt thống kê của chúng. Ngoài ra hàm tương quan không gian có thể hữu ích khi kéo dài các chuỗi dòng chảy năm sang thời kỳ chuẩn nhiều năm trong khi phân tích tính đồng bộ và không đồng bộ của các dao động dòng chảy sông ngòi. Người đầu tiên dùng hàm tương quan không gian dòng chảy năm các sông phụ thuộc vào khoảng cách giữa các trọng tâm của lực vực đó là N.V.Xomôv (136). Trong công trình này hàm tương quan không gian được tính theo dòng chảy năm của các sông và trung bình ở Liên Xô được phân bố cách đều nhau đến 9000 km.

Các hàm tương quan không gian N.V Xômôv tính được đã cho phép ta đánh giá được sự đồng bộ hay không đồng bộ của dòng chảy trên các sông lớn mà được chú ý tới khi nghiên cứu lược đồ hệ thống năng lượng duy nhất, trong phạm vi đó có sự điều tiết năng lượng theo lượng nước chảy qua. Về sau G.P. Kalinin (58) đã tính được các hàm tương quan không gian của dòng chảy năm của các sông lớn và trung bình trên Trái Đất. Người ta đã xác định được là hệ thống tương quan từng đôi một (r) giảm đi một khoảng cách (L) tăng trong phạm vi khoảng cách 0 - 2500 km mối tương quan dòng chảy của các sông nghiên cứu là dương, khi tiếp tục tăng khoảng cách thường tưhờng thấy có sự tương quan âm yếu và nó bằng không khi L = 9000 kmu. Ta nhận thấy rằng các quan hệ thực nghiệm r = 1 (L) thường nhận được khi bình quân hoá theo các nấc khoảng cách của số hệ số tương quan từng đôi một, điều đó nói đúng ra là hợp lý đối với hàm tương quan không gian đồng nhất và đẳng hướng khi mà khoảng lệch của hệ số tương quan từng đôi một so với quan hệ bình quân r = f(L) là do những biến động ngẫu nhiên của hệ số tương quan mẫu. Trong trường hợp hàm (L) không đồng nhất, sự bình quân hoá nó có thể đem đến sai lầm của mối quan hệ có tược tự nhiên r = f(L). Việc đánh giá tính đồng nhất của các hàm tương quan không gian là đặc biệt quan trọng, nếu như xét trước việc sử dụng r = f(L) trong tính toán thống kê sau này, thí dụ như đánh giá độ chính xác của việc nội suy không gian và hợp lý hoá lưới trạm.

Ý đồ đánh giá tính đồng nhất thống kê của hàm tương quan và sử dụng nó để nội suy không gian và sử dụng nó để nội suy không gian đã được G.A.Alekxeev thực hiện trong công trình [9].

Sau đây chúng ta sẽ xét trình tự tính toán hàm tương quan không gian và đánh giá tính đồng nhất của chúng với thí dụ dòng chảy năm của các sông thượng lưu khu vực sông Đnepr. Ta thấy rằng việc tính toán như vậy thực tế có thể thực hiện trên MTĐT. cụ thể là thí dụ nghiên cứu sau đây đã được tiến hành theo chương trình viết dưới dạng ngôn ngữ ALGOL của M.V.Zorin.

Thuật toán của chương trình sẽ xét trước những tính toán các tham số sau đây

a) Trị bình quân số học đối với tất cả ácc chuỗi quan trắc

$$\overline{x}j = \frac{\sum_{i=1}^{nj} x_j}{n_j}$$

b) Khoảng lệch trung bình bình phương và hệ số biến đổi đối với tất cả các chuỗi quan trắc.

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{nj} (x_i - x_j)}{n_j}}$$

$$C_{vj} = \frac{\sigma_j}{x_j}$$

c) Hệ số tương quan từng đôi một đối với thời kỳ quan trắc đồng bộ.

$$r_{jk} = \frac{\sum_{i=1}^{nxk}(Xik - xk - \overline{xk})(XÞ - \overline{X}j)}{\sigma_x \sigma_j n_{kj}}$$

trong đó n_{kj} - số năm quan trắc đồng bộ giữa các năm.

Căn cứ vào những tính toán vừa thực hiện được theo các công thức này có thể xây dựng được quan hệ giữa hệ số tương quan từng đôi một với khoảng lệch cách giữa các trọng tâm của lưu vực. đối với các đặc trưng dòng chảy sông ngòi; và với các khoảng giữa các điểm quan trắc đối với lượng mưa. Các điểm thực nghiệm trong trường toạ độ r và L thường được phân bố trên một băng rất rộng.

Sự phân tán của hệ số tương quan từng đôi một trong trường toạ độ r. L có thể có quan hệ với những biến động ngẫu nhiên của nó được tạo nên bởi các mẫu dùng trong tính toán hữu hạn.

Độ chính xác của hệ số tương quan từng đôi một đối với nấc khoảng cách (ΔL) cho trước sẽ được tăng lên khi số năm quan trắc đồng bộ dùng để tính toán tăng lên.

Để vẽ đường quan hệ r = f (L) người ta tính trị bình quân có tỷ trọng theo năm quan trắc đồng bộ của hệ số tương quan từng đôi đối với cách nêu khoảng cách (L = 50 km) bằng công thức:

$$r_{itr} = \frac{\sum_{i=1}^{N} rjnj}{\sum_{j=1}^{N} nj}$$

trong đó N - số điểm trong mỗi nấc khoảng cách.

Đường hồi quy r = f(L) trong trường điểm thực nghiệm được vẽ theo các điểm của giá trị bình quân có tỷ trọng của hệ số tương quan và trị bình quân số học tương ứng với khoảng cách đối với mỗi nấc.

Đường hồi quy vừa nhận được làm mối quan hệ thực r f(L) ứng với bản chất của mối tương quan không gian của yếu tố nghiên cứu. Người ta cũng cho rằng khoảng cách chênh lệch của các điểm thực nghiệm so với quan hệ này là do những biến động ngẫu nhiên của tài liệu mẫ. Giả thiết này hay chính là giả thiết không cần phải tiến hành kiểm tra thống kê. Sự kiểm tra này nên sử dụng phép biến đổi Fisher. Phép biến đổi này sẽ đem lại những kết quả khá tốt, ngay cả khi số năm quan trắc đồng bộ không không lớn và giá trị r lớn.

Phép biến đổi trên có xét đến số hiệu chỉnh độ chênh lệch $\frac{r}{r(n-1)}$ có dạng

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r} + \frac{r}{2(n-1)}$$

các giá trị mẫu Z được phân phối theo luật chuẩn có phương sai được xác định theo phương trình (6.20).

Tiếp theo phải tính các giá trị $z \pm \delta$, $z + 3\delta_z$, vì thế ta tính được các giới hạn trên và dưới tương ứng với từng hệ số tương quan r ik thực tế. Để làm hệ số tương quan thực (r thực) người ta lấy các giá trị của nó từ hệ bình quân r = f(L). So sánh hệ số tương quan thực và thực tế và biết được các giới hạn tin cậy phía trên và phía dưới đối với hệ số tương quan thực tế ta sẽ xác định được số điểm rơi vào trong cá khoảng $\pm 5\delta$, $\pm 2\delta r$ và $\pm 3\delta r$. Số hệ tương quan thực nghiệm rới vào mỗi vùng như trên được biểu diễn dưới dạng phần trăm so với tổng số trường hợp được so sánh với xác suất lý luận của luật phân phối chuẩn.

Nếu như xác suất thực nghiệm và lý luận gần bằng nhau thì hàm tương quan không gian được coi là đồng nhất, hay chính xác hơn là giả thiết không ban đầu được chấp nhận. Trong trường hợp ngược lại khi có sự khác nhau lớn giữa xác suất thực nghiệm với lý luận, giải thuyết không bị loại và chấp nhận giả thiết trên - giả thuyết hàm tương quan không gian không đồng nhất. Trong trường hợp như vậy trường nguyên tố nghiên cứu gốc cần phải phân chia ra làm các vùng đông nhất nhỏ hơn, đối với từng vùng cần phải xây dựng hàm tương quan không gian và lại đánh giá tính đồng nhất của nó.

Bên cạnh phương pháp cơ bản này đánh giá tính đồng nhất của các hàm tương quan không gian ta có thể sử dụng những phương pháp khác đơn giản hơn mà đôi khi còn hiểu hơn. Trong số các phương pháp đó có chỉ tiêu phù hợp giữa hàm phân phối lý luận với tài liệu tược nghiệm mà ta đã xét ở chương IV. Trong trường hợp sử dụng chi tiêu phù hợp ta sẽ xác định được sự phù hợp giữa các hàm phân phối lý luận và thực nghiệm của hệ số tương quan mẫu. Để làm hàm phân phối thực nghiệm của hệ số tương quan mẫu ta sử dụng các hệ số tương quan từng đooi mọt rơi vào trong phạm vi khoảng cách ($\Delta I = 50$ m). Để làm phân phối chuẩn có khoảng lệch trung bình bình phương của hệ số tương quan mẫu:

$$\sigma_r = \frac{1 - r_{tr}^2}{n_{tb} - 1} \quad (6.42)$$

và của các gí trị z $\quad \sigma_z = \frac{1}{\sqrt{n_{tb} - 3}}$

trong đó $\quad n_{tb} = \frac{\sum_1^N ni}{N}$ số điểm rơi vào khoảng ΔL.

Để ước lượng kỳ vọng toán nên lấy trị bình quân tỷ trọng theo sau năm quan trắc đồng bộ của hệ số tương quan từng đôi một (hay tham số Z) đối với khoảng cách (ΔL).

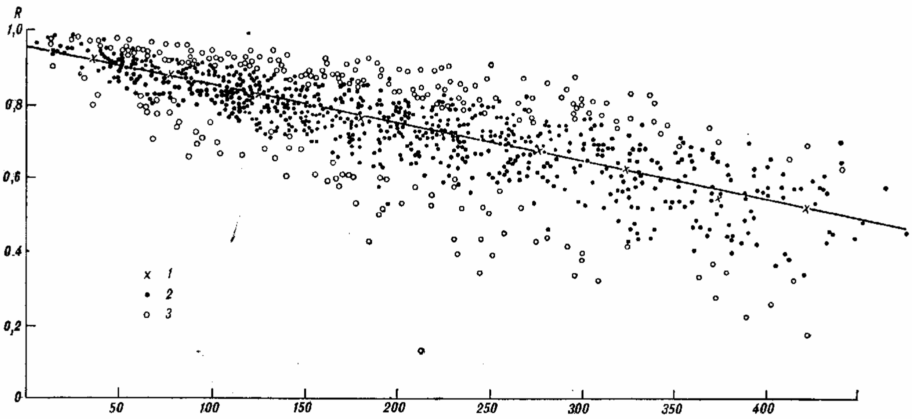
Sự phù hợp của hàm phân phối thực nghiệm với lý luận được xác lập bằng cách sử dụng chỉ tiêu Kôlmôgôrôv. Để giải bài toán này có thể sử dụng chỉ tiêu phù hợpkhác thí dụ như χ^2 hoặc ω^2, có hiệu lực hơn nhưng cần phải tính toán nhiều hơn. Việc xác định sự phù hợp của hàm phân phối thực nghiệm với hàm lý luận đối với nấc khoảng cách ΔL cũng như việc chấp nhận tính đồng nhất của hàm không không gian đối với nấc đó. Nếu như xác định được sự phù hợp của các phân phối lý luận và thực nghiệm đối với tất cả ΔL thì suy ra là mối quan hệ $r = f(L)$ đồng nhất trên toàn bộ khoảng cách.

Ngoài ra, đánh giá tính đồng nhất của hàm tương quan không gian theo các nấc khoảng cách có thể được thực hiện bauừg cách sử dụng phân phối F của Fisher để đánh giá tính đồng nhất của phương sai lý luận và thực nghiệm. Lúc này ta sẽ xác định được tính đồng nhất của phương sai thực nghiệm của các hệ số tương quan tương đôi một tính theo công thức bình thường:

$$\sigma_r^2 = \frac{\sum_1^n (r_i - \bar{r}_{tb})^2}{n}$$

hay là $\sigma_z^2 \frac{\sum_{1}^{n}(zi - \bar{z})}{n}$

và phương sai lý luận được xác định theo biểu thức (6.17) hay (6.18).



Hình 6.6 Hàm tương quan của các biến đặc trưng dòng chảy sông Dnhepr

Trên hình 6.6 trình bày hàm tương quan không gian của dòng chảy năm của các sông thượng lưu Đnhepr, kể cả sông Xoz và Bêrêzina. Phương trình hồi quy r = f(L) đã nhận được theo các giá trị bình quân của hệ số tương quan từng đôi một đối với mỗi nấc của khoảng cách (ΔL = 50 km), nhưng điều đó được ký hiệu bằng các chữ thập.

Hàm tương quan không gian tính được theo 991 hệ số tương quan từng đôi một kể từ 45 chuỗi tài liệu quan trắc. Số năm quan trắc đồng bộ trong tính toán hệ số tương quan từng đôi một trung bình chiếm khoảng 20,5 năm.

Việc đánh gía tính đồng nhất của hàm tương quan không gian của dòng chảy năm trên các sông thượng lưu sông Đnhepr được tiến hành bằng cách sử dụng tất cả các chỉ tiêu đồng nhất kể trên. Việc đánh giá này cho thấy rằng hàm tương quan không gian nghiên cứu là đồng nhất. Vì vậy khoảng lệch của hệ số tương quan bằng từng đôi một so với đường hồi quy r = f (L) là do sai số của hệ số tương quan mà sai số này là do các mẫu dùng trong tính toán có độ dài hữu hạn.

Hàm tương quan không gian, theo ý kiện của G.A.Alekxeev - có thể dùng để xác định sai số đo đạc của đặc trưng thuỷ văn nghiên cứu dưới dạng một phần trăm số khoảng lệch tiêu chuẩn tự nhiên (δ).

$$\delta = \sqrt{\frac{1}{r(0)} - 1}, \qquad (6.43)$$

trong đó r (0) - giá trị của hàm tương quan không gian khi L = 0.

Trong thí dụ đang xét r (0) = 0,95 vì vậy τ = 0,23 hay gần bằng một phần tư sự biến đổi tiêu chuẩn tự nhiên của dòng chảy năm.

Với sự đề nghị đó cần phải chú ý rằng việc xác định giá trị r(0) trên cơ sở ngoại suy hàm tương quan không gian là phép đoán hết sức điều kiện, điều đó có thể đi đến những sai số không thể lường trước được trong khi xác định giá trị δ. Việc sử dụng hàm tương quan không gian đồng nhất để nội suy các đặc trưng thuỷ văn theo không gian sẽ dẫn đến việc xác định ma trận của hệ số tương quan từng đôi một giữa cá giá trị của nguyên tố ở điểm cần nội suy với tài liệu quan trắc ở các hàm gần nhất, Khi đã biết tương quan không gian quan trắc bào gồm các điểm nội suy ta dễ dàng xác định được ma trận của hệ số tương quan (từng đôi một và sẽ đưa vào tính toán phương trình hồi quy tuyến tính nhiều chiều ...). Theo phương trình hồi quy ta tiến hành nội suy nguyên tố nghiên ở bất kỳ điểm nào. Trong trường hợp đó, sai số của phép nội suy.

Ta nhận thấy rằng lời giải tương tự dễ dàng thực hiện được trong trường hợp trường đồng nhất và dao động đẳng hướng. Khi không có sự đồng nhất theo lãnh thổ của trị bình quân thường thường người ta sử dụng phép biến đổi dạng xi - x hay xi/x. Khi không có sự đồng nhất của khoảng lệch trung bình bình phương cần phải dựa vào phương trình hồi quy tuyến tính nhiều chiều nhưng giá trị nội suy của thang số này, tất nhiên là nếu như phép nội suy có cơ sở thống kê và vật lý chắc chắn.