

CẤU TRÚC THỐNG KÊ KHÔNG GIAN - THỜI GIAN VÀ ỨNG DỤNG TRONG KIỂM TRA CHẤT LƯỢNG SỐ LIỆU

TS. Nguyễn Đăng Quế

Trung tâm Tư liệu khí tượng thủy văn

Trong phạm vi bài này tác giả nêu lên khả năng ứng dụng các kết quả nghiên cứu cấu trúc thống kê trường các đại lượng khí tượng trong công tác kiểm tra, chỉnh lý số liệu thời gian thực và phi thực cũng như trong các bài toán khí tượng ứng dụng khác.

1. Mở đầu

Trong [1] chúng tôi đã có dịp đề cập đến một số vấn đề trong công tác chỉnh lý số liệu thời gian phi thực. Trong công trình đó chúng tôi đã nêu lên một nguyên tắc chung xây dựng hệ chỉnh lý số liệu tại một Trung tâm khí tượng, theo đó các phương pháp kiểm tra, chỉnh lý được thiết kế và xây dựng sao cho phù hợp với đặc thù của từng loại số liệu, từ đơn giản đến phức tạp, từ kiểm tra tại một điểm trong không gian và thời gian, đến kiểm tra trong không gian 2, 3, 4 chiều. Do tính chất đa dạng của các loại sai số có thể có trong số liệu quan trắc (đa dạng về tính chất sai số, đa dạng về nguồn gốc sai số, về mức độ và qui mô sai số...) ta cần lựa chọn các phương pháp thích hợp để có thể tận dụng tối đa thế mạnh của các phương pháp này cho mục đích phát hiện và sửa chữa sai số. Cuối cùng, với mục đích tổ hợp thế mạnh của các phép kiểm tra khác nhau vào trong cùng một phương pháp duy nhất người ta xây dựng phương pháp kiểm tra hỗn hợp. Đây là một phương pháp khá phức tạp nhưng cũng rất hiệu quả trong việc kiểm tra phát hiện và xử lý các loại sai số thường gặp trong số liệu quan trắc khí tượng.

Đĩ nhiên trên thực tế, do yêu cầu của công việc và nhiều điều kiện cụ thể (khả năng, trình độ, cơ sở vật chất kỹ thuật, sự bức bách cần thiết) người ta phải tạm chấp nhận sử dụng một Hệ kiểm tra, chỉnh lý với một mức độ hạn chế nào đó về số bước và độ phức tạp của bản thân các phương pháp.

Qua sơ đồ tổng hợp đã nêu trong [1] có thể thấy rằng các phương pháp kiểm tra, chỉnh lý số liệu có thể được xếp vào hai loại:

- 1) Kiểm tra, chỉnh lý chỉ dựa trên số liệu tại cùng một điểm trong không gian và thời gian;
- 2) Kiểm tra, chỉnh lý dựa trên số liệu tại các điểm khác nhau trong không gian và thời gian.

Đồng thời ta cũng thấy ngoài một số ít phương pháp có sử dụng đến các mối quan hệ nhiệt động lực (chủ yếu là các phương pháp kiểm tra theo chiều thẳng đứng đối với số liệu thám không vô tuyến), còn lại đa phần đều sử dụng các qui luật thống kê thu được từ các công trình nghiên cứu, tính toán trên chuỗi số liệu quan trắc của các đại lượng khí tượng. Dưới đây ta sẽ xem xét cụ thể khả năng sử dụng các kết quả nghiên cứu cấu trúc thống kê vào nhiệm vụ thiết kế xây dựng các phương pháp kiểm tra, chỉnh lý thời gian phi thực đối với số liệu điều tra cơ bản về khí tượng.

2. Về cấu trúc thống kê và khả năng ứng dụng

Một trong những đặc điểm cơ bản của khí quyển là tồn tại thường xuyên các chuyển động loạn lưu với cường độ và quy mô khác nhau. Trường các đại lượng khí

tượng luôn có độ biến động lớn. Sự biến đổi trong không gian và thời gian của các giá trị tức thời của các đại lượng khí tượng đều kèm theo những đặc tính rất phức tạp. Việc đo đạc các giá trị tức thời trong những điều kiện tương tự nhau đem lại kết quả có phần khác nhau. Nói cách khác, các đại lượng khí tượng quan trắc được đều mang đặc tính của một đại lượng ngẫu nhiên, tuy mức độ có khác nhau ở các đại lượng khác nhau. Do vậy, từ lâu các nhà khí tượng học đã xem xét các chuỗi số liệu khí tượng như chuỗi các đại lượng ngẫu nhiên và áp dụng cơ sở lý thuyết hàm ngẫu nhiên để nghiên cứu các quy luật thống kê và mô tả chúng [6]. Thực chất của phương pháp này là thay vì xem xét các đặc trưng của các giá trị tức thời của từng số liệu riêng biệt người ta nghiên cứu các quy luật chung nhất của một tập hợp thống kê trên cơ sở thoả mãn một số điều kiện nhất định.

Để thể hiện các quy luật đó, các nhà khí tượng học thường sử dụng các đặc trưng thống kê sau đây:

- Trung bình nhiều năm tại các thời điểm khác nhau trong không gian;
- Độ tản mạn, độ lệch quân phương trung bình;
- Hệ số tương quan theo chiều thẳng đứng (giữa các mức theo độ cao);
- Hàm tương quan theo chiều nằm ngang trên các mặt đẳng áp chuẩn;
- Xác định sai số quan trắc bằng phương pháp gián tiếp;
- Các giá trị cực trị và biên độ dao động;
- Dạng phân bố thống kê;
- Đánh giá tính đồng nhất và đẳng hướng của từng trường, từng chuỗi số liệu.

Bài toán nghiên cứu cấu trúc trường các đại lượng khí tượng đã được các nhà khí tượng học trên thế giới quan tâm từ rất lâu. Khối lượng các công trình đã công bố về vấn đề này rất lớn, đa phần tập trung giải quyết cho vùng vĩ độ cao và vĩ độ trung bình.

Ở vùng nhiệt đới, vấn đề này được quan tâm nghiên cứu muộn hơn vì nhiều lí do. Nhưng do mức độ quan trọng của nó không chỉ đối với vùng này mà còn đối với toàn cầu nên đã thu hút được sự quan tâm của nhiều nhà khoa học trên toàn thế giới.

Đối với vùng Đông Nam Á và trên lãnh thổ Việt Nam cũng đã có nhiều công trình nghiên cứu được thực hiện [2], [3], [4], [5].

Dưới đây ta chỉ xem xét khả năng và sự cần thiết áp dụng các kết quả nghiên cứu nêu trên phục vụ cho việc kiểm tra, chỉnh lý các loại số liệu điều tra cơ bản về khí tượng tại Việt Nam (kể cả thời gian thực và phi thực).

a. Về khả năng ứng dụng cho mục đích kiểm tra tại một điểm

Trước tiên, ta cần kiểm tra tính đồng nhất của chuỗi số liệu. Biết rằng chuỗi số liệu đồng nhất của một đại lượng ngẫu nhiên là chuỗi có tổng độ lệch khỏi giá trị trung bình nhiều năm bằng "0". Vì vậy, khi ta đã xác định được giá trị trung bình cho một chuỗi số liệu quan trắc nào đó, nhưng sau đó phát hiện thấy khi tăng độ dài của chuỗi sẽ xuất hiện một cách hệ thống độ lệch có cùng một dấu. Đó là trường hợp chuỗi số liệu bất đồng nhất. Đối với các chuỗi này việc đầu tiên cần làm là khử tính bất đồng nhất.

Thông qua việc so sánh các giá trị tính toán được trên chuỗi số liệu về hệ số đối xứng (A), hệ số lồi lõm (E) với các giá trị tương ứng của phân bố chuẩn ta có thể rút ra một số ứng dụng định hướng cho việc xây dựng phương pháp kiểm tra. Chẳng hạn, dựa trên giá trị độ lệch quân phương trung bình (σ) có thể sơ bộ đã biết được trong phạm vi $\pm 2\sigma$ có không dưới 75% và trong phạm vi $\pm 3\sigma$ - không dưới 89% giá trị

quan trắc. Trong trường hợp các giá trị A và E xấp xỉ bằng "0" - tức là chuỗi có phân bố gần chuẩn- trong phạm vi $\pm \sigma$ có khoảng 68%, $\pm 2\sigma$ có khoảng 95% và $\pm 3\sigma$ có khoảng 99,7% số liệu quan trắc.

Nếu $A \approx 0$ và $E > 0$ - tức là phân bố đối xứng thì xung quanh giá trị trung bình tập trung nhiều giá trị quan trắc hơn so với trường hợp phân bố chuẩn, nhưng tỷ lệ số liệu nằm ngoài phạm vi $\pm 3\sigma$ còn khá cao. Ngược lại nếu $A \approx 0$ nhưng $E < 0$ - xung quanh giá trị trung bình tập trung ít số liệu hơn song trong phạm vi $\pm 3\sigma$ tỷ lệ số liệu lại cao hơn ở phân bố chuẩn.

Những tính đặc thù trên đây cần được lưu ý khi xây dựng các phương pháp kiểm tra theo ngưỡng giá trị đối với từng đại lượng tại từng vùng và từng thời kỳ khác nhau.

Theo kết quả tính toán của chúng tôi [5], phân bố thống kê của chuỗi số liệu nhiệt độ và độ cao địa thế vị trên các mực từ 700 hPa trở lên có thể xem là phân bố chuẩn và có thể áp dụng được tiêu chuẩn trên đây để xây dựng phương pháp kiểm tra ngưỡng giá trị. Tính chất phân bố chuẩn tăng dần theo độ cao. Kết quả tính toán này đã được sử dụng để xây dựng các phương pháp kiểm tra, chỉnh lý cho số liệu của Đài Khí tượng Cao không, Viện Khí tượng Thủy văn và một phần số liệu hiện có trên máy tính tại Trung tâm Tư liệu khí tượng thủy văn. Bên cạnh đó, nhiều đại lượng khí tượng như gió, áp suất mặt đất, độ ẩm... có đường cong phân bố khác xa so với phân bố chuẩn. Trên đường cong phân bố của chuỗi số liệu một số đại lượng khí tượng hoặc tồn tại nhiều đỉnh hoặc là phân bố lệch trái hay lệch phải. Đối với những chuỗi số liệu này cần áp dụng các biện pháp khác mang tính cá biệt phù hợp với tính chất của từng dạng phân bố.

Vấn đề này chúng tôi sẽ đề cập đến trong một dịp khác.

b. Về khả năng ứng dụng các qui luật thống kê trong các phương pháp kiểm tra số liệu theo không gian và thời gian

Các phương pháp kiểm tra, chỉnh lý số liệu theo không gian đều dựa trên nguyên lý so sánh giá trị quan trắc với giá trị nội suy được từ số liệu quan trắc độc lập tại các điểm trong phạm vi không gian giới hạn cho phép. Cụ thể:

Nếu F_i^{OBS} - là giá trị quan trắc được tại trạm (i), ε - là giá trị độ lệch cho phép,

F_i^{NS} - là giá trị nội suy được từ số liệu quan trắc độc lập tại trạm (i). Sẽ xảy ra các khả năng:

Nếu $|F_i^{NS} - F_i^{OBS}| \leq \varepsilon$ - số liệu quan trắc được xem là đúng

và $|F_i^{NS} - F_i^{OBS}| > \varepsilon$ - nghi ngờ có sai số.

Để tính được giá trị F_i^{NS} người ta sử dụng các phương pháp nội suy đang dùng để phân tích trường. Ta sẽ xét cụ thể trường hợp sử dụng phương pháp nội suy tối ưu.

Trong phương pháp này (hiện nay được sử dụng rộng rãi nhất trên thế giới), ta cần giải hệ phương trình đại số tuyến tính:

$$\sum_{j=1}^n \mu_{ij} P_j + \eta_i P_i = \mu_{oi} \quad (i=1, 2, \dots, n)$$

Trong đó P- là hệ số trọng lượng; μ_{ij} - là hàm tương quan giữa trạm (i) và trạm (j) của đại lượng cần kiểm tra; η_i - ước lượng sai số quan trắc, n - là số trạm ảnh hưởng dùng để nội suy.

Ở đây các kiến thức về quy luật thống kê được sử dụng thông qua hàm tương quan (μ) của các đại lượng khí tượng trong không gian. Cụ thể: Nếu kiểm tra theo chiều thẳng đứng μ_{ij} - là hệ số tương quan giữa số liệu của các mực i và j ; nếu kiểm tra trên từng mặt đẳng áp - hàm tương quan trong không gian 2 chiều; nếu kiểm tra trong không gian 3 chiều - hàm tương quan trong không gian 3 chiều; nếu kiểm tra theo thời gian là hàm tương quan theo thời gian.

Ngoài hàm tương quan (μ), trong lúc triển khai phương pháp này còn cần dùng đến một số tham số thống kê khác như tính đồng nhất và đẳng hướng của trường độ lệch, giá trị chuẩn khí hậu, độ tản mạn, độ lệch quân phương trung bình (σ)... Tất cả các tham số này đều biến thiên theo mùa và khu vực địa lý.

Một số kết quả tính toán nghiên cứu đặc trưng cấu trúc thống kê trường các đại lượng khí tượng cơ bản tại vùng Đông Nam Á và khu vực địa lý bao quanh lãnh thổ Việt Nam đã được nêu trong [2], [3], [4]. Các kết quả nghiên cứu này đã được dùng để xây dựng phương pháp kiểm tra số liệu khí tượng cao không trong vùng địa lý nêu trên. Phương pháp này đang được tiếp tục hoàn thiện và thử nghiệm để áp dụng cho các loại số liệu điều tra cơ bản khác của Ngành.

3. Kết luận

Trên đây là một số vấn đề về bài toán nghiên cứu cấu trúc thống kê trường các đại lượng khí tượng, bước đầu ứng dụng chúng trong việc xây dựng và sử dụng các phương pháp kiểm tra, chỉnh lý số liệu điều tra cơ bản của Ngành. Vì ý nghĩa khoa học và thực tiễn ứng dụng nên vấn đề này cần được các nhà khoa học quan tâm nghiên cứu nhiều hơn nữa.

Tài liệu tham khảo

1. Nguyễn Đăng Quế. Một số vấn đề trong công tác chỉnh lý số liệu thời gian phi thực.- *Tạp san Khí tượng Thủy văn*, số 1/1999.
2. Nguyễn Đăng Quế. Cấu trúc thống kê thẳng đứng của các đại lượng khí tượng cơ bản ở vùng Đông Nam Á.- *Tạp san Khí tượng Thủy văn*, số 2, 1996, tr. 17-31.
3. Nguyễn Đăng Quế. Cấu trúc thống kê trường các đại lượng khí tượng cơ bản tại vùng Đông Nam Á.- *Tuyển tập báo cáo tại Hội nghị cao không lần thứ IV*, tháng 1/1996.
4. Nguyễn Đăng Quế, Hoàng Phương Hồng. Đặc điểm cấu trúc thống kê trường gió trên cao trên lãnh thổ Việt Nam và khu vực địa lý lân cận.- *Tuyển tập công trình Hội thảo khoa học Cơ học thủy khí với Thiên nhiên kỷ mới*, Hội Cơ học Việt Nam-Hội Cơ học thủy khí, Hà Nội, 2001.
5. Nguyễn Đăng Quế. Phân bố thống kê số liệu cao không trên lãnh thổ Việt Nam. *Tuyển tập công trình Hội nghị khoa học*. Đài Khí tượng Cao không, Hà Nội, 2001.
6. Kazakavich D. I. Các cơ sở hàm ngẫu nhiên và việc ứng dụng chúng trong khí tượng thủy văn.- NXB Khí tượng Thủy văn Leningrát, 1971.