

MỘT SỐ VẤN ĐỀ TRONG CÔNG TÁC CHÍNH LÝ SỐ LIỆU THỜI GIAN PHI THỰC

PTS. Nguyễn Đăng Quế
Trung tâm Tư liệu KTTV

I - Mở đầu

Nghiệp vụ lưu trữ tư liệu của Ngành ta đã hình thành từ lâu, song mọi việc được tiến hành theo phương thức thô sơ, kho lưu trữ duy nhất là kho giấy; phương thức thu thập chủ yếu là sổ sách báo biểu và được gửi về theo chế độ phi trực tuyến; công tác kiểm tra, chỉnh lý được thực hiện bằng tay.

Trên thế giới việc hiện đại hoá công tác tư liệu đã được tiến hành cách đây ba bốn thập kỷ. Cùng với sự phát triển như vũ bão của khoa học và công nghệ, các khâu cơ bản của dây chuyền công tác tư liệu đều được hiện đại hoá, tự động hoá và tin học hoá với mức độ ngày càng cao.

Trong thời gian vừa qua, công tác tổ chức của Ngành có một số cải tiến. Trung tâm tư liệu và các Đài khu vực lần lượt được hình thành và đi vào hoạt động. Với cách tổ chức này chúng ta có điều kiện tập trung đầu tư về mọi mặt nhằm hiện đại hoá từng bước các công tác của Ngành nói chung và công tác tư liệu nói riêng.

Để thực hiện nhiệm vụ trên đây, trong phạm vi bài này, chúng tôi mạnh dạn nêu ra một vài suy nghĩ về việc từng bước hiện đại hoá công tác tư liệu của Ngành.

1 - Chức năng chuyên môn cơ bản của các Trung tâm tư liệu

- Lưu trữ và phục vụ số liệu quan trắc cơ bản và các thông tin đã được xử lý cho công tác nghiên cứu khoa học và triển khai ứng dụng theo yêu cầu của các đối tượng thuộc Trung tâm chịu trách nhiệm.

- Triển khai nghiên cứu thử nghiệm và áp dụng kỹ thuật công nghệ mới đối với công tác tư liệu.

- Thường xuyên trao đổi với các Trung tâm khác về thông tin và số liệu trên cơ sở kỹ thuật và công nghệ thông dụng.

- Đào tạo và huấn luyện về công tác tư liệu cho các đối tượng thuộc Trung tâm quản lý.

2 - Một số khái niệm

a) Chế độ thời gian thực (real - time): Là chế độ trong đó thông tin được thu thập, xử lý và sử dụng chỉ một số giờ sau thời điểm quan trắc.

Phân biệt các mức thời gian thực sau đây:

- Early cut - off time : xử lý số liệu trong phạm vi 2 giờ sau thời điểm quan trắc

- Intermediate cut - off time : - 4 giờ

- Main operational cut-off time : - 6 giờ

- Final cut - off time : - 10 giờ hoặc 12 giờ

Chế độ thời gian thực chủ yếu được sử dụng trong công tác dự báo phục vụ.

b) Chế độ thời gian phi thực (non - real time) : là các công việc được tiến hành sau kỳ quan trắc một khoảng thời gian dài hơn.

Chế độ thời gian phi thực được dùng chủ yếu trong các Trung tâm tư liệu để xử lý số liệu phục vụ cho công tác lưu trữ và phục vụ tư liệu.

c) Khái niệm trực tuyến (on-line): Dùng để chỉ chế độ thu thập truyền phát thông tin trên sơ sở kênh liên lạc điểm nối điểm (số liệu được mã hoá và truyền trực tiếp).

d) Khái niệm phi trực tuyến (off-line): Là chế độ thu thập truyền phát thông tin bằng các phương tiện khác không trực tiếp và có độ trễ về thời gian.

II - Công tác kiểm tra, chỉnh lý số liệu thời gian phi thực

1 - Cách đặt vấn đề

Với mục đích hiện đại hoá công tác tư liệu của Ngành chúng ta còn có rất nhiều công việc phải làm từ khâu thu thập, xử lý, lưu trữ, quản lý và phục vụ số liệu. Công tác này phụ thuộc vào nhiều vấn đề như kinh phí, máy móc trang thiết bị, trình độ cán bộ, phương thức quản lý. Việc hiện đại hoá sẽ tiến hành từng bước một và như vậy, quy trình công nghệ chung về công tác tư liệu cũng cần phải từng bước được cải tiến cho phù hợp.

Ở đây chúng tôi chỉ bàn đến công tác kiểm tra, chỉnh lý theo phương thức hiện đại. Nói đúng hơn là phương thức mà chúng ta cần phải tiệm cận dần dần trong một tương lai gần đây nhất trên cơ sở các khâu liên quan cũng song song được hiện đại hoá.

Mục đích của việc kiểm tra phi thời gian thực là để bảo đảm tiêu chuẩn cao nhất có thể được về chất lượng của số liệu điều tra cơ bản trước lúc lưu trữ và phục vụ chúng cho người tiêu dùng.

Công tác kiểm tra chất lượng phi thời gian thực phải được tiến hành cho toàn bộ số liệu điều tra cơ bản kể cả số liệu đã qua kiểm tra thời gian thực (real-time quality control) khi truyền qua hệ thống thông tin liên lạc quốc tế (GTS). Tuy vậy, nhiệm vụ cụ thể của việc kiểm tra chất lượng phi thời gian thực phụ thuộc vào chức năng của các Trung tâm lưu trữ. Công tác kiểm tra ở mọi Trung tâm đều phải được tiến hành trực tiếp trong dây chuyền nghiệp vụ càng sớm càng tốt ngay sau khi nhận được số liệu cần lưu trữ.

2 - Phân loại sai số, nguồn gốc phát sinh

Việc đo đạc các tham số vật lý, trong đó có quan trắc các yếu tố khí tượng thủy văn là một công việc xấp xỉ hóa tính chất tự nhiên của yếu tố đó. Các giá trị của yếu tố khí tượng được xem là tổng của hai thành phần:

$$F = f + \sum \delta_i,$$

trong đó : - f : giá trị thực của tham số vật lý. Trên thực tế giá trị này chúng ta cũng không được rõ mà chỉ chấp nhận một giá trị khả dĩ nhất do chúng ta đo đạc được.

- δ_i : các loại sai số. Trên thực tế các sai số luôn luôn tồn tại. Chúng ta chỉ có thể hạn chế chúng đến mức thấp nhất có thể được mà thôi.

Có thể xếp loại các nguồn gốc sai số sau đây:

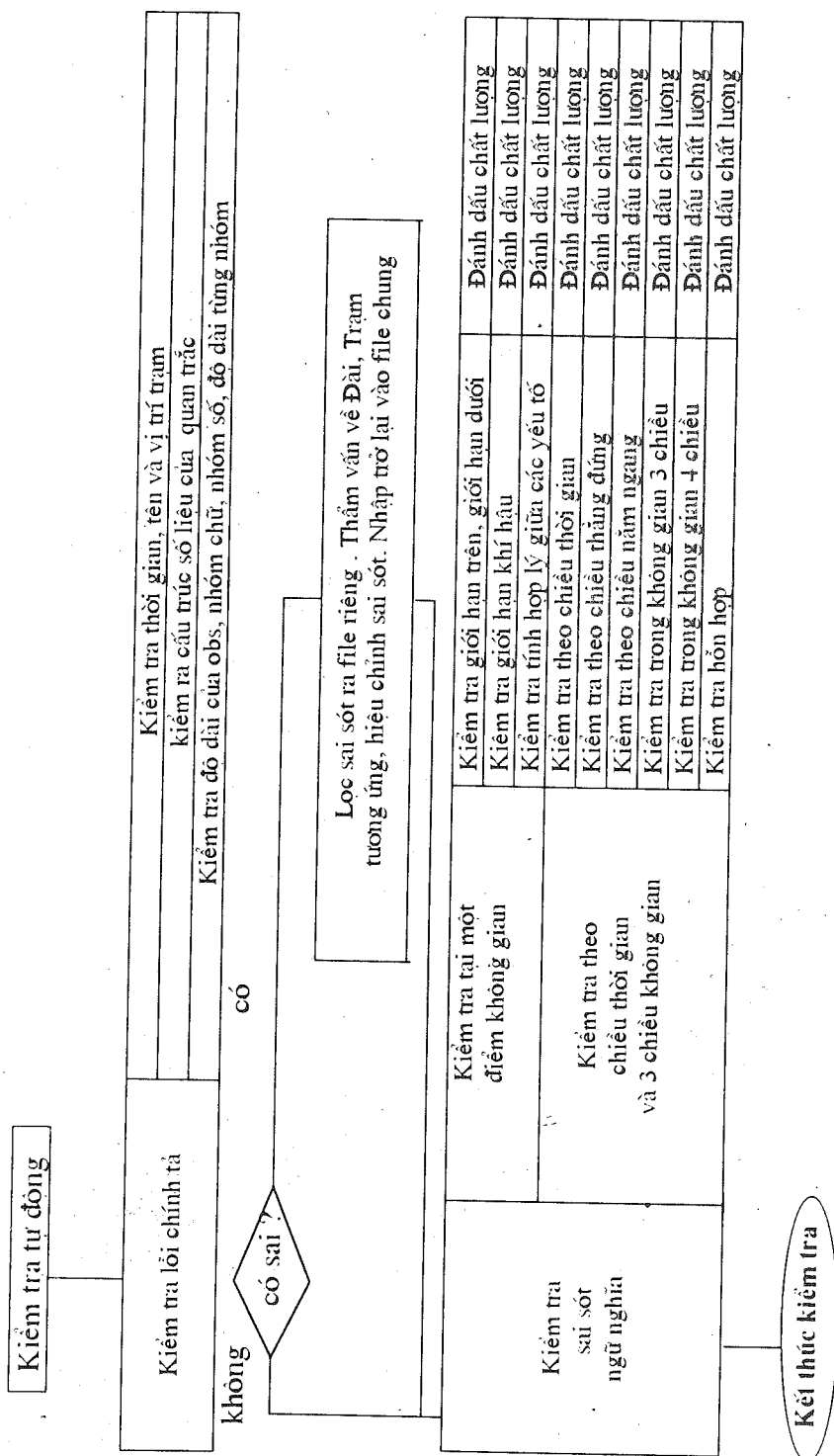
- Sai số do dụng cụ đo đạc (trong quá trình sản xuất, kiểm định, vận chuyển và lắp đặt tại điểm quan trắc...),
- Sai số do quá trình quan trắc trực quan tại trạm,
- Sai số do đọc trị số trên dụng cụ đo đạc,
- Sai số do quá trình xử lý sơ bộ tại trạm,
- Sai số do quá trình mã hoá và khai mã,
- Sai số do quá trình truyền tin từ trạm về các trung tâm thu thập,
- Sai số do quá trình chuẩn bị và nhập liệu lên máy tính.

3 - Các phương pháp kiểm tra

Các phương pháp thường dùng để phát hiện sai sót đối với số liệu điều tra cơ bản trước khi đưa vào lưu trữ được phân loại thứ tự tăng dần mức độ phức tạp như sau:

- a) Kiểm tra lỗi chính tả
- b) Kiểm tra sự thoả mãn của các đại lượng khí tượng khác nhau đối với các giới hạn trên và giới hạn dưới.
- c) Kiểm tra sự thoả mãn của các đại lượng khí tượng khác nhau đối với các giới hạn được xác định bằng phương pháp tính toán thống kê. Những giới hạn này thường là các hàm số của vị trí địa lý, thời gian trong năm; đôi lúc còn phụ thuộc vào các yếu tố khác nữa.
- d) Kiểm tra tính hợp lý về mặt vật lý giữa các yếu tố khác nhau trong cùng một quan trắc.

4-Sổ đồ chung về công tác kiểm tra chỉnh lý số liệu tại các trung tâm tự liệu



e) Kiểm tra tính hợp lý về mặt vật lý theo phương nằm ngang bằng cách so sánh đồng thời với các giá trị quan trắc ở các trạm lân cận hoặc với trường phân tích.

f) Kiểm tra tính hợp lý về mặt vật lý theo chiều thẳng đứng (kiểm tra tĩnh học hay còn gọi là kiểm tra thủy tĩnh học) đối với số liệu cao không cũng như số liệu khí tượng thủy văn biển.

g) Kiểm tra tính hợp lý giữa số liệu quan trắc của các kỳ liên tiếp trên cùng một trạm.

Trong nhiều trường hợp ta có thể tiến hành đồng thời một số phương pháp kiểm tra khác nhau.

Cần phân biệt sự khác nhau về mặt logic giữa các loại sai số hiển nhiên (sai số thô) cần được loại bỏ hoặc được thay bằng các đại lượng tính toán nội ngoại suy và các giá trị bị nghi ngờ là sai. Trong cả hai trường hợp các đại lượng đó đều phải được đánh dấu lại.

Trong sai số quan trắc còn có phần đóng góp của các nhiễu động quy mô nhỏ. Đối với các loại số liệu khác nhau mức độ đóng góp có khác nhau (đối với số liệu khí tượng bề mặt ảnh hưởng của nhiễu động quy mô nhỏ lớn hơn nhiều so với ảnh hưởng đối với số liệu khí tượng trên các mực trên cao thuộc khí quyển tự do...). Việc phân biệt đâu là nhiễu động và đâu thực sự là sai số quan trắc là một việc làm không hiện thực ít nhất là trong thời gian hiện nay.

III - Phương hướng cần tiệm cận trong công tác kiểm tra, chỉnh lý số liệu điều tra cơ bản

1 - Tại các Đài khu vực (hoặc các Trạm nếu có đủ điều kiện)

- Nhập các số đọc từ máy móc, thiết bị đo đạc về các yếu tố khí tượng thủy văn lên máy tính.

- Chạy bộ chương trình do Trung ương soạn thảo và cài đặt để xử lý sơ bộ (các loại hiệu chỉnh tại trạm).

- Kiểm tra số liệu dựa trên sự thỏa mãn giới hạn trên và giới hạn dưới.

- Kiểm tra số liệu dựa trên sự thỏa mãn các giới hạn thống kê khí hậu.

- Kiểm tra tính hợp lý giữa các yếu tố khác nhau trong cùng một quan trắc.

- Kiểm tra theo chiều thẳng đứng (đối với số liệu cao không và khí tượng thủy văn biển).

- Đối chiếu lại các sai sót đã được chương trình phát hiện và hiệu chỉnh nếu có thể được.

- Truyền số liệu về Trung ương bằng phương pháp trực tuyến hoặc phi trực tuyến theo chế độ thời gian thực cho Trung tâm Quốc gia Dự báo và thời gian phi thực cho Trung tâm Tư liệu .

So với công tác nghiệp vụ hiện nay, trong phương thức mới không đặt vấn đề làm báo biểu hàng tháng tại các trạm quan trắc. Số liệu từ đài trạm gửi về chỉ có số quan trắc và số liệu đã có trên máy tính.

2 - Tại Trung tâm tư liệu KTTV

- Thu thập đầy đủ các loại số liệu từ các Đài, Trạm gửi về.

- Phân loại số liệu theo các chỉ tiêu thống nhất (theo thời gian, theo trạm, theo khu vực, theo lưu vực, v.v...).

Sau đó sẽ tổ chức công tác kiểm tra, chỉnh lý theo mức độ tăng dần độ phức tạp và chặt chẽ về mặt vật lý của các phương pháp.

- Kiểm tra sai sót chính tả (sai về cấu trúc từng loại số liệu, kiểm tra về số chữ, mã số, số nhóm, độ dài mỗi nhóm, độ dài toàn bộ quan trắc, tên trạm, biểu số trạm, thời gian quan trắc v.v...).

- Kiểm tra giới hạn trên, giới hạn dưới.

- Kiểm tra giới hạn khí hậu đối với từng yếu tố, từng trạm hoặc từng vùng.

- Kiểm tra tính hợp lý giữa các loại số liệu trong cùng một quan trắc.

- Kiểm tra theo chiều thẳng đứng.

- Kiểm tra theo không gian hai, ba, bốn chiều.

- Kiểm tra theo chiều thời gian.

- Kiểm tra hỗn hợp.

Việc lựa chọn các tổ hợp phương pháp kiểm tra còn phụ thuộc vào điều kiện, yếu tố cụ thể.

Các loại số liệu có sai số do các phương pháp kiểm tra trên đây phát hiện ra, nếu có điều kiện sẽ thẩm vấn trở lại các Đài, Trạm để khẳng định lại và chỉnh lý nếu có thể được.

Ngược lại, chúng ta cần đánh dấu chất lượng số liệu dựa trên bảng phân loại sai số và mức độ sai số để cảnh tỉnh người sử dụng.

IV- Kết luận và kiến nghị

Để thực hiện thành công quy hoạch phát triển ngành Khí tượng Thủy văn từ nay đến năm 2000 và 2010, tùy khả năng và điều kiện cụ thể, chúng ta cần tiến hành hiện đại hoá từng bước nhưng phải thực hiện với phương châm năm bắt và, nếu có thể được, đón đầu những thành tựu mới về phát triển khoa học kỹ thuật nói chung và công nghệ tin học nói riêng. Có như vậy thì đồng vốn bỏ ra hiện nay mới có ý nghĩa trong thời gian tương lai.

Trên đây, với tư cách là một cán bộ khoa học, chúng tôi xin mạnh dạn nêu một số ý kiến nhằm cung cấp thông tin cho các nhà quản lý và hoạch định kế hoạch phát triển công tác tư liệu nói riêng và công tác của ngành Khí tượng Thủy văn nói chung với mong muốn thúc đẩy sự nghiệp chung.

Tài liệu tham khảo

1- Guide on the Global Data Processing system. WMO N^o305 . 1976.

2 - Nguyễn Đăng Quế - Các phương pháp xử lý số liệu khí tượng.
Bài giảng cao học (tài liệu đánh máy) Hà Nội, 1995.