

NGHIÊN CỨU PHÁT TRIỂN HỆ THỐNG DIỄN GIẢI SAU MÔ HÌNH CÓ KHẢ NĂNG TỰ CẬP NHẬT DỰA TRÊN PHƯƠNG PHÁP UMOS VÀ LỌC KALMAN

Phần I: Phương pháp luận

NCS. Võ Văn Hòa, TS. Lê Đức, ThS. Đỗ Lệ Thùy

ThS. Dư Đức Tiến, CN, Nguyễn Mạnh Linh, CN. Nguyễn Thanh Tùng

Trung tâm Dự báo KTTV Trung ương

Bài báo này trình bày một số kết quả nghiên cứu ứng dụng các phương pháp thống kê có khả năng tự cập nhật sai số theo thời gian bao gồm UMOS và lọc Kalman (KF) để nâng cao chất lượng dự báo các yếu tố khí tượng bề mặt tại điểm trạm từ các sản phẩm dự báo của mô hình HRM và GSM. Các mô hình thống kê UMOS và KF được xây dựng tách biệt cho từng tập nhân tố dự báo là các sản phẩm dự báo từ mô hình HRM và GSM, và áp dụng cho 130 trạm quan trắc khí tượng bề mặt dựa trên chuỗi số liệu từ năm 2003 đến năm 2009. Phần I của bài báo sẽ giới thiệu khái quát về phương pháp xây dựng các mô hình UMOS và KF nói trên. Các kết quả nghiên cứu và đánh giá sẽ được trình bày trong phần II của bài báo.

1. Mở đầu

Ngày nay, các hệ thống dự báo thời tiết số trị là một thành phần quan trọng trong việc trợ giúp các dự báo viên đưa ra bản tin dự báo thời tiết. Một hệ thống dự báo số gồm ba thành phần: hệ thống phân tích, mô hình dự báo (tất định hoặc tổ hợp) và hệ thống diễn giải dự báo. Trạng thái khí quyển hiện tại sau khi được tạo ra nhờ hệ thống phân tích sẽ được sử dụng làm điều kiện ban đầu cho mô hình dự báo. Với sự phát triển của tốc độ tính toán cũng như mức độ chính xác trong mô phỏng vật lý, các mô hình hiện tại có thể cung cấp dự báo với chất lượng tốt cho các trường khí quyển trên cao. Tuy nhiên, với các trường khí tượng bề mặt, tác động trực tiếp đến đời sống hàng ngày, kết quả dự báo từ mô hình thường có sai số lớn. Một ví dụ điển hình là dự báo lượng mưa hay nhiệt độ tại độ cao hai mét. Những sai số này có thể được hiệu chỉnh bằng các phương pháp thống kê (Wilks, 2006) thông qua thành phần thứ ba của hệ thống dự báo số, đó là hệ thống diễn giải dự báo.

Trong thời kỳ phát triển ban đầu của dự báo số, có hai hướng tiếp cận thống kê sau mô hình với sản

phẩm dự báo từ mô hình là Perfect Prog - PP (Klein và cộng sự, 1959) và MOS (Glahn và Lowry, 1972). Tư tưởng cơ bản của hai phương pháp này dựa trên thiết lập quan hệ thống kê giữa dự báo từ mô hình với quan trắc bằng phương pháp hồi quy tuyến tính. Chính xác hơn, PP xây dựng quan hệ thống kê giữa các quan trắc với một độ trễ thời gian nhất định. Sau đó, với giả định mô hình không có sai số, dự báo từ mô hình được sử dụng thay cho quan trắc trong phương trình hồi quy khi dự báo. Như thế so với MOS, PP có hai hạn chế: không tính đến sai số hệ thống của mô hình và không thể sử dụng các đại lượng dẫn xuất (không quan trắc được) làm nhân tố dự báo như tốc độ thẳng đứng hay phân kỳ.

Trong thập kỷ 80, một số các trung tâm dự báo lớn trên thế giới đã xây dựng hệ thống diễn giải dự báo dựa trên phương pháp MOS, cung cấp dự báo tại các điểm trạm quan trắc khí tượng như Mỹ (Carter và cộng sự, 1989), Canada (Brunet và cộng sự, 1988), Hà Lan (Lemcke và Kruizinga, 1988), Úc (Woodcock, 1984), Anh (Francis và cộng sự, 1982). Tuy nhiên, vào những năm 90 người ta bắt đầu nhận thấy những hạn chế của MOS, dẫn đến giảm dần sử dụng MOS trong diễn giải dự báo. Vấn đề nằm ở

chỗ: khi xây dựng các phương trình hồi quy, tập số liệu mô hình dùng làm nhân tố dự báo phải đồng nhất. Theo đánh giá của Jacks và cộng sự (1990) để có được một quan hệ thống kê ổn định, cần ít nhất hai năm số liệu dự báo từ mô hình và thám sát. Sau đó khi sử dụng MOS, các đặc trưng của mô hình dự báo như các sơ đồ tham số hóa, độ phân giải,... cần được giữ nguyên như khi sử dụng mô hình để thiết lập quan hệ thống kê. Điều này rõ ràng là một hạn chế lớn của MOS, bởi ta biết rằng từ thập kỷ 90 với sự phát triển nhanh của tốc độ tính toán cũng như các hệ thống quan trắc, các mô hình thường xuyên được cập nhật với độ phân giải ngày càng cao hơn, các sơ đồ tham số hóa tinh tế hơn, trường phân tích chính xác hơn. Erikson và cộng sự (2002) cho ta thấy sai số hệ thống sẽ xuất hiện khi sử dụng MOS với mô hình được cải tiến. Do đó, khi mô hình có sự thay đổi ta phải đợi ít nhất hai năm mới có thể bắt đầu sử dụng MOS trong dự báo hoặc dự báo lại với mô hình đã thay đổi cho hai năm trước đó nhằm xác định lại các hệ số hồi quy. Với số phương trình hồi quy rất lớn (tại mỗi trạm, cho mỗi biến và mỗi hạn dự báo có một phương trình hồi quy riêng biệt), chi phí cho tái xây dựng hệ thống MOS khi mô hình thay đổi căn bản là rất lớn.

Để giải quyết hạn chế này của phương pháp MOS truyền thống khi mô hình liên tục có sự thay đổi người ta sử dụng các phương pháp thống kê có khả năng tự cập nhật. Thông tin sẽ được truyền vào phương trình dự báo ngay khi có những thay đổi trong mô hình dự báo. Hiện tại có hai phương pháp cho phép hệ phương trình dự báo tự cập nhật gồm: 1) Phương pháp lọc Kalman - KF (Simonsen 1991) và 2) Phương pháp UMOS. KF là một phương pháp đánh giá tối ưu trạng thái của một hệ thống thể hiện qua các biến trạng thái từ các quan trắc gián tiếp. Trong khi đó, UMOS về cơ bản vẫn sử dụng phương pháp MOS truyền thống nhưng đưa thêm khả năng tự cập nhật cho MOS thông qua một phương pháp lấy trọng số giữa hai tập dữ liệu cũ và mới khi có thay đổi trong mô hình (Wilson và Vallée, 2002). Nếu mô hình không có cải tiến nào đáng kể, UMOS sẽ trở thành phương pháp MOS thông thường. Phần I của bài báo này sẽ chỉ tập trung trình bày về cơ sở lý thuyết và phương pháp xây dựng

các mô hình thống kê có khả năng tự cập nhật sai số dựa trên phương pháp UMOS và KF. Các kết quả nghiên cứu và đánh giá sẽ được trình bày trong phần II của bài báo.

2. Phương pháp luận

a. Phương pháp UMOS

Theo như cách tiếp cận truyền thống của MOS, không dễ để có thể đưa ý tưởng cập nhật vào phương pháp, nếu chỉ nhìn MOS ở khía cạnh toán học quen thuộc của nó. Bài toán hồi quy tuyến tính đa biến MOS sử dụng có dạng như sau: xác định tập $p+1$ các hệ số tự do ai của phương trình dự báo yếu tố Y theo tập p các nhân tố dự báo Xi:

$$Y = a_0 + a_1X_1 + \dots + a_pX_p \quad (1)$$

Từ tập số liệu quá khứ và áp dụng dự báo tương lai. Các hệ số ai sẽ được xác định theo phương pháp bình phương tối thiểu dựa trên tập số liệu phụ thuộc dung lượng mẫu n. Viết dưới dạng ma trận, công thức tính vector hệ số ai có dạng sau:

$$a = (X^T X)^{-1} X^T y \quad (2)$$

trong đó y là vector yếu tố dự báo kích thước n, X là ma trận nhân tố dự báo kích thước $(p+1) \times n$, với mỗi vector cột, tương ứng vector nhân tố dự báo kích thước n. MOS không cố định trước tập p các nhân tố mà thực hiện hồi quy từng bước với p tăng dần từ 1 cho đến một giá trị nào đó thỏa mãn tiêu chuẩn dừng. Tại mỗi bước, bài toán hồi quy mà MOS thực hiện hoàn toàn đồng nhất với mô tả toán học ở trên.

Nếu thực hiện như đã mô tả ở trên, khi muốn cập nhật phương trình dự báo (1) với tập số liệu phụ thuộc đã được mở rộng thêm từ n lên $n+m$ (m là số ngày mới tích lũy thêm dữ liệu), toàn bộ quá trình phải được thực hiện lại từ đầu với hồi quy từng bước và tại mỗi bước phải giải các hệ phương trình tuyến tính (2). Với tập các nhân tố có thể vào khoảng 200, tại mỗi bước ta phải giải 200 hệ phương trình tuyến tính (1). Thông thường, mỗi phương trình dự báo sau khi thỏa mãn điều kiện dừng có khoảng 10 nhân tố, số lần giải phương trình sẽ là 2000 lần. Nhân lên với số trạm cần cập nhật, chi phí tính toán cho mỗi lần cập nhật sẽ rất lớn. Với tốc độ máy tính hiện nay, khối lượng tính toán như

trên có thể giảm thiểu nếu thực hiện tính toán song song cho nhiều trạm tại cùng một thời điểm. Khó khăn thực sự nằm ở phương thức thiết kế, lưu trữ và sử dụng số liệu cho bài toán cập nhật. Với những vấn đề cố hữu như trên, các trung tâm thường không thực hiện cập nhật phương trình dự báo theo ngày hay theo tháng mà thực hiện một lần duy nhất rồi áp dụng cho các năm kế tiếp hoặc cập nhật theo từng năm. Với dung lượng tập phụ thuộc đủ dài, cách thực hiện như vậy vẫn đảm bảo một kết quả tốt do phương trình dự báo sẽ ổn định trên một tập dữ liệu đủ dài. Tất nhiên, nếu tận dụng được tập số liệu mới để tăng dung lượng tập phụ thuộc, phương trình dự báo thu được sẽ có chất lượng tốt hơn.

Với các nghiên cứu của mình, Ross (1987, 1989, 1992) và sau đó là Wilson và Vallee (2002) đã chỉ ra rằng vẫn có thể thực hiện cập nhật cho MOS mà chi phí tính toán không lớn nếu thực hiện theo cách thông thường. Ý tưởng cập nhật được hình thành dựa trên biểu diễn lại dạng toán học của phương pháp hồi quy tuyến tính đa biến. Từ công thức (1.2) dễ thấy vector a được xác định thông qua ma trận XTX kích thước $(p+1) \times (p+1)$ và vector XTy kích thước $p+1$. Mỗi phần tử M_{ij} của ma trận XTX chỉ đơn giản là tích vô hướng giữa hai vector nhân tố dự báo kích thước n thứ i và j , trong khi mỗi phần tử V_i của vector XTy là tích vô hướng giữa vector nhân tố dự báo kích thước n thứ i với vector yếu tố dự báo kích thước n .

$$M_{ij} = \sum_{k=1}^n X_{ik} X_{kj} \quad (3)$$

$$V_i = \sum_{k=1}^n X_{ik} Y_k \quad (4)$$

Các phần tử này có đặc tính cộng, nghĩa là khi dung lượng mẫu tăng lên $n+1$, các phần tử M_{ij} hay V_i có thể xác định rất đơn giản bằng cách cộng thêm số hạng $X(i, n+1)X(j, n+1)$ và $X(i, n+1)Y(n+1)$ vào giá trị trước đó mà không cần phải tính lại toàn bộ cho $n+1$ số hạng. Khả năng cập nhật rõ ràng đã thể hiện ngay trong phân tích trên. Nếu muốn cập nhật, thay vì phải lưu chuỗi dữ liệu X_i hay Y với kích thước n tăng lên từng ngày, bài toán sẽ đơn giản hơn, nếu ta chỉ cần lưu các phần tử M_{ij} và V_i và thay đổi giá trị

của chúng từng ngày bằng cách cộng thêm một số hạng mới vào như trên. Cách thực hiện này sẽ tiết kiệm dung lượng lưu trữ với một ma trận và một vector kích thước không thay đổi $(p+1) \times (p+1)$ và $p+1$ so với ma trận nhân tố dự báo và vector yếu tố dự báo có dung lượng n tăng dần. Các hệ số ai cuối cùng vẫn hoàn toàn được xác định theo công thức (2) của bài toán hồi quy tuyến tính đa biến.

Tuy nhiên, bằng cách thay đổi điểm nhìn như vậy, phương pháp mới đòi hỏi tập các nhân tố p phải được xác định trước. Đây là chi phí phải trả cho việc đưa ý tưởng cập nhật vào MOS. Điều này cũng đồng nghĩa quá trình hồi quy từng bước sẽ không được thực hiện và các nhân tố dự báo phải cố định giống như trong phương pháp lọc Kalman. Một quá trình sơ tuyển cần được thực hiện tuyển chọn trước các nhân tố này. Thông thường p vào khoảng 10 và dung lượng lưu trữ được sử dụng cập nhật phương trình dự báo cho một trạm sẽ rất nhỏ với 132 số thực $(11 \times 11 + 11)$. Cách thức thực hiện này chính là UMOS dưới dạng đơn giản nhất và nếu so với MOS truyền thống thì khối lượng tính toán của nó là không đáng kể.

Trong thực tế, khối lượng tính toán còn giảm hơn nữa do không cần thiết phải giải hệ phương trình (1.2) từng ngày, do các phương trình dự báo MOS có tính ổn định cao. Một thay đổi nhỏ trên thang thời gian ngày của ma trận XTX sẽ không có tác động tới giá trị của tập hệ số a . Phương trình (2) chỉ cần giải trên quy mô tuần hoặc quy mô tháng và công việc hàng ngày chỉ liên quan đến cập nhật ma trận XTX và vector XTY bằng cách cộng thêm một số hạng thích hợp vào mỗi phần tử. Khối lượng tính toán chỉ tăng lên khi ta giải hệ phương (2) vào thời gian đã định trước theo tháng hay tuần.

Hai ma trận XTX và vector XTY có tầm quan trọng đặc biệt trong phương pháp UMOS. Nếu coi vector yếu tố dự báo Y tương đương như các vector nhân tố dự báo khác thì rõ ràng với cùng bản chất tích vô hướng, phần tử V_i của XTY không có gì khác so với phần tử M_{ij} của XTX . Do đó, để đơn giản, thay vì phải tách biệt và lưu trữ hai ma trận và vector, người ta ghép vector XTY vào cột cuối ma trận XTX và gọi chung là ma trận SSCP (Sums of

Squares and Cross Products) với kích thước $(p+2) \times (p+1)$. Như tên gọi, SSCP thể hiện rõ bản chất tích vô hướng của nó. Tóm lại, thay vì cách tiếp cận hồi quy tuyến tính đa biến, UMOS đặt trọng tâm vào ma trận SSCP và thực hiện giải hệ phương trình tuyến tính trên ma trận này để xác định tập hệ số ai.

Cách tiếp cận này hoàn toàn trùng với phương pháp MOS truyền thống (tất nhiên trong trường hợp tập p nhân tố dự báo đã được cố định trước) và chỉ trở nên khác biệt nếu mô hình dự báo thay đổi. Nếu vẫn thực hiện theo MOS truyền thống, khi mô hình thay đổi ta phải đợi một khoảng thời gian đủ dài để có thể phát triển một phương trình dự báo mới trong khi không thể sử dụng phương trình dự báo cũ. Dự báo sau mô hình sẽ không thể thực hiện trong khoảng thời gian đợi này và cách giải quyết duy nhất có thể là chạy lại mô hình mới với thời gian trước đây nhằm xây dựng lại MOS. Tuy nhiên, khi thay đổi điểm nhìn sang ma trận SSCP, vấn đề sẽ được giải quyết đơn giản hơn nếu xác định ma trận SSCP trong thời gian chuyển đổi giả định một quan hệ tuyến tính chuyển dần từ ma trận SSCP cũ sang ma trận SSCP mới và kết thúc khi mô hình mới đã có được một chuỗi dự báo đủ dài. Giả thiết này không có cơ sở toán học (không tương đương với hồi quy tuyến tính đa biến) nhưng có thể chấp nhận về mặt ứng dụng và hiệu quả có thể có được trong giai đoạn chuyển đổi mô hình. Như vậy, khi mô hình thay đổi, ma trận SSCP sẽ được tính thông qua một phương pháp lấy trọng số giữa ma trận SSCP cũ và mới.

$$SSCP = w_0SSCP_o + w_nSSCP_n \quad (5)$$

Wilson và Vallee (2002) đưa ra một công thức thực nghiệm cho trọng số này dưới dạng:

$$w_n = w_{max} + (1 - w_{max}) \sqrt{1 - \frac{(N_n - N_{max})^2}{(N_{min} - N_{max})^2}}$$

$$w_0 = \frac{N_{max} - w_n N_n}{N_0} \quad \text{với } N_n \leq N_{max} \quad (6)$$

$$w_n=1, w_0=0 \quad \text{với } N_n > N_{max}$$

Trong đó w_n là trọng số cho ma trận SSCP mới,

w_0 : trọng số cho ma trận SSCP cũ, N_n : dung lượng mẫu từ tập số liệu mới, N_0 : dung lượng mẫu từ tập số liệu cũ, N_{min} : dung lượng mẫu tối thiểu mà khi vượt quá mô hình mới bắt đầu tác động tới ma trận SSCP, N_{max} : dung lượng mẫu tối đa mà khi vượt quá mô hình cũ được xem như không còn tác động tới ma trận SSCP, w_{max} : tham số điều khiển. Dạng hàm này được hai tác giả lựa chọn trên nguyên tắc:

- Kết quả từ mô hình cũ không ảnh hưởng đến phương trình dự báo mới sau một khoảng thời gian đủ dài

- Mô hình ngay sau khi thay đổi sẽ không tác động tới giá trị của ma trận SSCP cho tới khi tích lũy được một dung lượng tối thiểu

So với lọc Kalman, MOS truyền thống có điểm mạnh nằm ở khả năng phân mùa dự báo do chất lượng dự báo của mô hình cũng có đặc tính phân mùa. Phân mùa dự báo cũng có thể thực hiện dễ dàng cho UMOS thông qua hai tập p nhân tố dự báo khác nhau cho hai mùa. Để làm trơn quá trình chuyển mùa, UMOS giống như MOS có thể cập nhật thêm một số ngày của mùa còn lại trong thời đoạn chuyển mùa. Nếu tập nhân tố dự báo như nhau cho cả hai mùa, Wilson và Vallee (2002) sử dụng thêm một sơ đồ trọng số chuyển mùa cho ma trận SSCP giữa hai mùa đông và hè với trọng số có dạng hàm bậc thang $(2/3, 1/3)$, $(1/2, 1/2)$ cho tới $(1/3, 2/3)$. Các trọng số này còn được điều chỉnh bởi dung lượng mẫu ma trận SSCP mỗi mùa nhằm loại bỏ khả năng SSCP chịu tác động nhiều hơn bởi ma trận SSCP tương ứng với mùa có dung lượng lớn hơn.

$$SSCP = \sum_{s=a}^b w_s \frac{\min(N_a, N_b)}{N_s} (w_{ns}SSCP_{ns} + w_{os}SSCP_{os}) \quad (7)$$

Với chỉ số s chạy từ a đến b tương ứng với hai mùa, w_s là trọng số mùa cho mùa s, w_{os} và w_{ns} : trọng số như trong (1.6), N_s mà cụ thể hơn N_a và N_b là dung lượng mẫu của mùa tương ứng.

Như mô tả ở trên, nếu UMOS thực hiện phân mùa với hai tập nhân tố khác nhau, quá trình cập nhật của UMOS sẽ tách thành hai quá trình độc lập

riêng cho mùa đông và mùa hè. Phương pháp thực hiện như vậy so với nguyên tắc liên tục tự cập nhật sẽ bị gián đoạn tại hai thời điểm chuyển mùa. Nếu thực hiện nguyên tắc này, UMOS phải cố định tập nhân tố dự báo cho cả hai mùa và quá trình phân mùa theo nhân tố dự báo không được thực hiện theo phương thức phân mùa của MOS truyền thống. Ngoài ra, tập nhân tố cố định thích hợp cho mô hình cũ, nhưng có thể không thích hợp cho mô hình mới cũng có thể dẫn đến chất lượng dự báo giảm. Để giải quyết vấn đề, UMOS sẽ được thực hiện phức tạp hơn tương tự như quá trình hồi quy từng bước của MOS truyền thống.

Tư tưởng của phương pháp có thể hiểu được nếu quan sát công thức (2) khi ta có thể rút ra từ ma trận XTX một ma trận con tương ứng vector XTY một vector con chỉ chứa p' biến mà ta quan tâm từ p biến tổng thể, sau đó giải hệ phương trình có dạng tương tự như (2) nhưng cho ma trận con và vector con này để tìm các hệ số ai cho tập p' biến này. Có thể rút ra nhiều tập p' biến khác nhau từ tập p và đây là cơ sở để thực hiện hồi quy từng bước. Ta sẽ bắt đầu từ p' bằng 1 và giải hệ phương trình dạng (2) trên mỗi tập con với một phần tử có thể để lựa chọn nhân tố đầu tiên, sau đó tiếp tục tăng p' và dừng cho đến khi đạt được một tiêu chuẩn dừng nào đó. Theo phương pháp này, tập p các nhân tố dự báo ban đầu sẽ không cần cố định và sử dụng trong phương trình dự báo giống như lọc Kalman. Ta chỉ cần chọn ra một bộ p nhân tố đủ lớn các nhân tố có khả năng trở thành nhân tố dự báo nhưng cũng không quá lớn để duy trì một ma trận $SSCP$ có dung lượng nhỏ. Tập con các nhân tố được tuyển chọn cho phương trình dự báo sẽ thay đổi khi chuyển từ tuần sang tuần, tháng sang tháng, mùa sang mùa hoặc từ mô hình cũ sang mô hình mới. Phương trình dự báo do đó sẽ biến đổi liên tục, "trơn" khi chuyển mùa hay chuyển đổi mô hình. Tuy nhiên, do thực hiện tuyển chọn nhân tố, cách thực hiện này đòi hỏi phải có một tập dữ liệu phụ thuộc. Như vậy, để làm trơn quá trình chuyển mùa hay chuyển đổi mô hình, bên cạnh khối lượng tính toán tăng thêm, dung lượng lưu trữ cũng tăng thêm. Tất nhiên, nếu so với MOS truyền thống thì chi phí này vẫn nhỏ hơn. Với tập dữ liệu phụ thuộc cần chú ý trong thời gian chuyển đổi mô hình,

UMOS sẽ không thực hiện lưu tập phụ thuộc với độ dài n tăng dần từng ngày như MOS truyền thống mà cố định độ dài tập training bằng N_{max} trong công thức (6). Đây là dung lượng mẫu tối đa để một phương trình dự báo khi xây dựng theo MOS truyền thống ổn định.

b. Phương pháp KF

Phương pháp lọc Kalman (gọi tắt là KF - Kalman Filter), được đặt theo tên tác giả của bộ lọc, ra đời năm 1960 trong công trình của R. E. Kalman mô tả một bộ lọc đệ quy cho phép đánh giá trạng thái của một hệ động lực tuyến tính. Một cách khái quát, KF là một tập hợp các phương trình toán học mô tả một phương pháp đệ quy cho phép đánh giá trạng thái ẩn của một hệ động lực với sai số thấp nhất từ số liệu đo gián tiếp về hệ này, thừa nhận quan trắc và các quy luật mô tả hoạt động của hệ tồn tại một độ bất định nào đó. Từ hệ động lực ở đây có thể áp dụng rất đa dạng trong thực tế như quá trình chuyển động của một vệ tinh, sự phát triển của một nền kinh tế hay chuyển động của khí quyển. Hai ví dụ sau đặc trưng cho các hệ động lực phi tuyến và KF có những gần đúng để thực hiện đánh giá trạng thái của các hệ động lực này. Tuy nhiên, phần trình bày ở đây sẽ chỉ giới hạn ở các hệ động lực tuyến tính.

Một hệ động lực tuyến tính được đặc trưng bởi vector trạng thái x kích thước N , cung cấp thông tin cần thiết để có thể mô tả hệ. Vector này biến đổi theo thời gian theo quy luật tuyến tính và thể hiện mặt động lực của hệ thống:

$$x_k = A \times x_{k-1} + B \times u_{k-1} + w_{k-1} \quad (8)$$

trong đó x_k là vector trạng thái thời điểm hiện tại, x_{k-1} là vector trạng thái thời điểm trước đó, u_{k-1} là vector điều khiển hệ thống tại thời điểm trước đó, w_{k-1} là vector đặc trưng cho độ bất định của hệ thống do quy luật tuyến tính mô tả thông qua ma trận A chỉ là gần đúng. Ma trận A kích thước $N \times N$ có tên gọi ma trận chuyển dịch trạng thái mang thông tin về quy luật hoạt động của hệ thống. Ma trận B được đưa vào phương trình trên đặc trưng cho các quá trình bên ngoài điều khiển hệ thống đang xét tác động thông qua vector điều khiển u_{k-1} . Các bài toán

KF thông thường không sử dụng B và u.

Vấn đề đặt ra là ta biết quy luật hoạt động của hệ thống qua ma trận A, nhưng không thể xác định trực tiếp trạng thái x_k từ các phép đo thích hợp (trạng thái bị ẩn) mà chỉ có được các quan trắc gián tiếp z_k kích thước M có quan hệ tuyến tính với x_k :

$$z_k = H \times x_k + v_k \quad (9)$$

Với vector v_k đặc trưng cho sai số hay nhiễu khi thực hiện đo, H là ma trận kích thước $M \times N$ mô tả quan hệ được giả định tuyến tính giữa biến đo được với biến trạng thái cần xác định. H được gọi là ma trận quan trắc. Để bài toán xác định, ta cần giả định một số thông tin cho hai vector sai số w_k và v_k . Lọc Kalman xem đây là hai vector ngẫu nhiên, độc lập và tuân theo phân bố Gauss với trung bình bằng 0 và ma trận hiệp biến (covariance) lần lượt là Q và R:

$$\begin{aligned} w_k &\sim N(0, Q) \\ v_k &\sim N(0, R) \end{aligned} \quad (10)$$

Ngoài ra, để đơn giản hóa các ma trận Q, R, A, H đã được bỏ qua chỉ số dưới nhưng thực tế có thể biến đổi theo thời gian. Để đánh giá tối ưu trạng thái x_k với z_k đo được, lọc Kalman sẽ cực tiểu hóa hiệp phương sai của sai số đánh giá. Như vậy, ở bước thứ k, sử dụng phương trình (8) ta dễ dàng thu được đánh giá tiên nghiệm (đánh giá sơ bộ) $\hat{x}_k^-$ của vector trạng thái x từ các thông số của bước trước đó. Sau đó, khi có quan trắc z_k ta cần điều chỉnh lại đánh giá tiên nghiệm sao cho đánh giá hiệu chỉnh thu được phù hợp với quan trắc hiện có. Đánh giá này có tên gọi đánh giá hậu nghiệm và được ký hiệu bởi $\hat{x}_k$. Đánh giá tiên nghiệm và hậu nghiệm sẽ có sai số:

$$\begin{aligned} e_k^- &= x_k - \hat{x}_k^- \\ e_k &= x_k - \hat{x}_k \end{aligned} \quad (10)$$

Ma trận hiệp biến của các sai số trên có dạng:

$$\begin{aligned} P_k^- &= E(e_k^- e_k^{-T}) \\ P_k &= E(e_k e_k^T) \end{aligned} \quad (11)$$

với E là toán tử trung bình. Kalman giả định $\hat{x}_k$ là một hàm tuyến tính của $\hat{x}_k^-$ và z_k và thu được công thức sau dưới dạng ma trận:

$$\hat{x}_k = (I - KH) \times \hat{x}_k^- + K \times z_k \quad (12)$$

Ma trận I-KH xuất hiện bên cạnh $\hat{x}_k^-$ với dạng trên có được sau một số biến đổi toán học. Để dẫn đến dạng trên ma trận này, ban đầu được giả định độc lập với K nhưng để thỏa mãn một số ràng buộc nó phải có dạng I-HK. K có tên gọi độ lợi (gain) và được xác định thông qua cực tiểu hóa hiệp phương sai của sai số hậu nghiệm:

$$K = P_k^- H^T (H P_k^- H^T + R)^{-1} \quad (14)$$

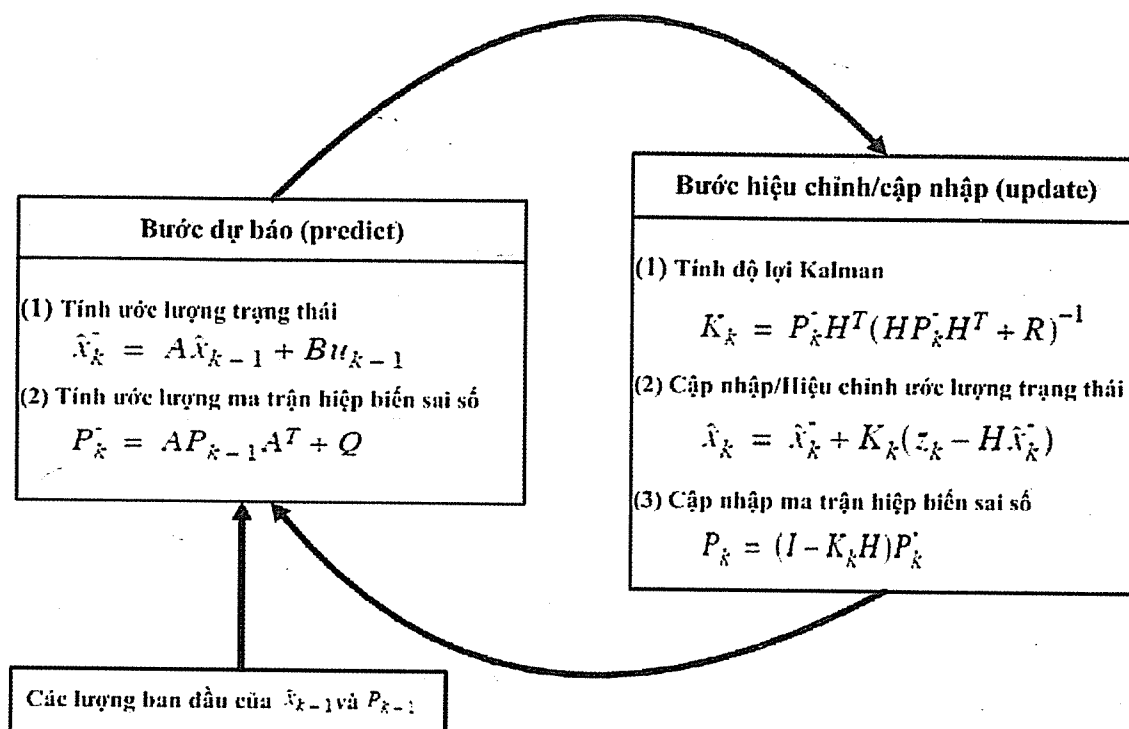
Cụ thể hơn về quá trình xử lý toán học dẫn đến hai công thức (13) và (14) có thể xem trong Grewal và Andrews (2001). Quan tâm chủ yếu đến mặt ứng dụng của phương pháp, đề tài sẽ không trình bày biến đổi chi tiết toán học ở đây bởi quá trình xử lý khá phức tạp và dài. Theo công thức trên, K biến đổi theo thời gian và cần được ký hiệu lại bởi K_k . Cuối cùng, để có thể thực hiện quá trình đệ quy của KF, ta cần thêm giá trị ban đầu của $\hat{x}_0$ và P_0 . Như vậy, khi thực hiện lọc Kalman sẽ bao gồm hai bước (xem hình 1) như sau:

1. Dự báo (Predict): bước này đòi hỏi phải biết giá trị tại bước trước đó của $\hat{x}_{k-1}$ và P_{k-1} . Dựa trên hai giá trị này, $\hat{x}_{k-1}^-$ và P_{k-1}^- sẽ được xác định.
2. Hiệu chỉnh: $\hat{x}_{k-1}^-$ và P_{k-1}^- sẽ được hiệu chỉnh dựa trên quan trắc z_k . Cụ thể, K_k sẽ được tính theo (14), dựa vào đó xác định $\hat{x}_k$ và P_k .

Lọc Kalman bắt đầu được áp dụng vào dự báo sau mô hình tại ECMWF từ các công trình nghiên cứu của Simonsen (1991) và Perisson (1991). Các nghiên cứu sau đó nói chung không có gì khác so với ý tưởng ban đầu của hai tác giả này. Hệ động lực tuyến tính trong trường hợp này được mô tả thông qua bộ các hệ số ai của phương trình dự báo (1). Đây chính là vector trạng thái của hệ thống. Theo cách nhìn này, hệ động lực mà ta quan tâm có thể hiểu như một dự báo viên với các tham số đặc trưng ai. Ma trận A của hệ này được xác định đơn giản bằng ma trận đơn vị I. Điều khiển bên ngoài bởi B không được xét đến ở đây. Vector quan trắc z_k do đó sẽ là quan trắc y của bài toán hồi quy (1). Theo ngôn ngữ của KF, ta cần xác định các hệ số ai trong khi chỉ có các quan trắc y, trong đó, quan trắc y quan hệ với các hệ số ai theo công thức (1). Điều này có nghĩa các nhân tố dự báo sẽ đóng vai trò các phần

tử của ma trận H. Bây giờ, vấn đề còn lại là xác định dạng của ma trận Q và R của hai sai số ngẫu nhiên w và v. Chất lượng dự báo của KF hoàn toàn được

quyết định bởi hai ma trận này. Hai ma trận này thường được giả định có dạng đường chéo với các phần tử trên đường chéo có giá trị như nhau.



Hình 1. Sơ đồ mô tả bước dự báo và hiệu chỉnh của lọc Kalman

3. Kết luận

Bài báo này đã giới thiệu khái quát về cơ sở lý thuyết của hai phương pháp thống kê hiện đại và có khả năng tự cập nhật sai số theo thời gian là UMOS và KF. Các phân tích và hướng tiếp cận xử lý để ứng dụng hai phương pháp này để cải tiến chất lượng dự báo của các hệ thống MOS truyền thống tại Việt Nam cũng đã được đề xuất. Việc triển khai thực hiện

các mô hình UMOS và KF trong điều kiện nước ta là hoàn toàn khả thi. Trong nghiên cứu này, chúng tôi đã phát triển bộ mô hình thống kê UMOS và KF dựa trên các thuật toán nói trên và thử nghiệm cho 130 trạm khí tượng quan trắc bề mặt của Việt Nam để dự báo nhiệt độ không khí, nhiệt độ tối cao và tối thấp, nhiệt độ điểm sương, tốc độ và hướng gió, và vận lượng mây. Các kết quả đánh giá sẽ được trình bày chi tiết trong phần II của bài báo.

Tài liệu tham khảo

1. Brunet N., Verret R. and Yacowar N., 1988: An objective comparison of model output statistics and "perfect prog" systems in producing numerical weather element forecasts. *Wea. Forecasting*, 3, 273-283.
2. Carter G. M., Dallavalle P. J., and Glahn H. R., 1989: Statistical forecasts based on the national meteorological centre's weather prediction system. *Wea. Forecasting*, 4, 401-412.
3. Francis P.E., A.P. Day and G.P.Davis, 1982: Automated temperature forecasting, an application of Model Output Statistics to the Meteorological Office numerical weather prediction model. *Meteorol. Mag.*, 111, 73-87
4. Glahn H. R. and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather

- forecasting. *J. App. Meteor.*, 11, 1203-1211.
5. Jacks E., J.B. Bower, V.J. Dagostaro, J.P. Dallavalle, M.C. Erikson and J.C. Su, 1990: New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, clout amount and sea surface wind. *Wea. Forecasting*, 5, 128-138.
 6. Klein W.H., B.M.Lewis and I. Enger, 1959: Objective prediction of five-day mean temperature during winter. *J. Meteor.*, 16, 672-682.
 7. Lemcke C. and Kruizinga S., 1988: Model output statistics forecasts: Three years of operational experience in the Netherlands. *Mon. Wea. Rev.*, 116, 1077-1090.
 8. Person A., 1991: Kalman filtering - A new approach to adaptive statistical interpretation of numerical meteorological forecasts. Lectures and papers presented at the WMO training on the interpretation of NWP products in terms of local weather phenomena and their verification, WMO, Wageningen, the Netherlands, XX-27-XX-32.
 9. Ross G. H., 1987: An updatable model output statistics scheme. Programme on Short and Medium range Weather Prediction. PSMP Rep. Series, No 25, World Meteorological Organization, 25-28.
 10. Ross G. H., 1989: Model output statistics using an updatable scheme. Preprints, 11th Conf. On Probability and Statistics in Atmosphere Sciences. Monterey, CA, Amer. Meteor. Soc., 93-97.
 11. Ross G. H., 1992: Probability of precipitation using an updatable MOS model. Preprints, 12th Conf. On Probability and Statistics in Atmosphere Sciences. Toronto, ON, Canada, Amer. Meteor. Soc., 190-194.
 12. Simonsen, C., 1991: Self adaptive model output statistics based on Kalman filtering. Lectures and papers presented at the WMO training on the interpretation of NWP products in terms of local weather phenomena and their verification, WMO, Wageningen, the Netherlands, XX-28-XX-33
 13. Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, Second Edition, 649 pp.
 14. Wilson L. J., and M. Vallée, 2002: The Canadian Updatable Model Output Statistics (UMOS) system: Design and development test. *Wea. Forecasting*, 17, 206-222.
 15. Woodcock F., 1984: Australian experimental model output statistics forecasts of daily maximum and minimum temperature. *Mon. Wea. Rev.*, 112, 2112-2121