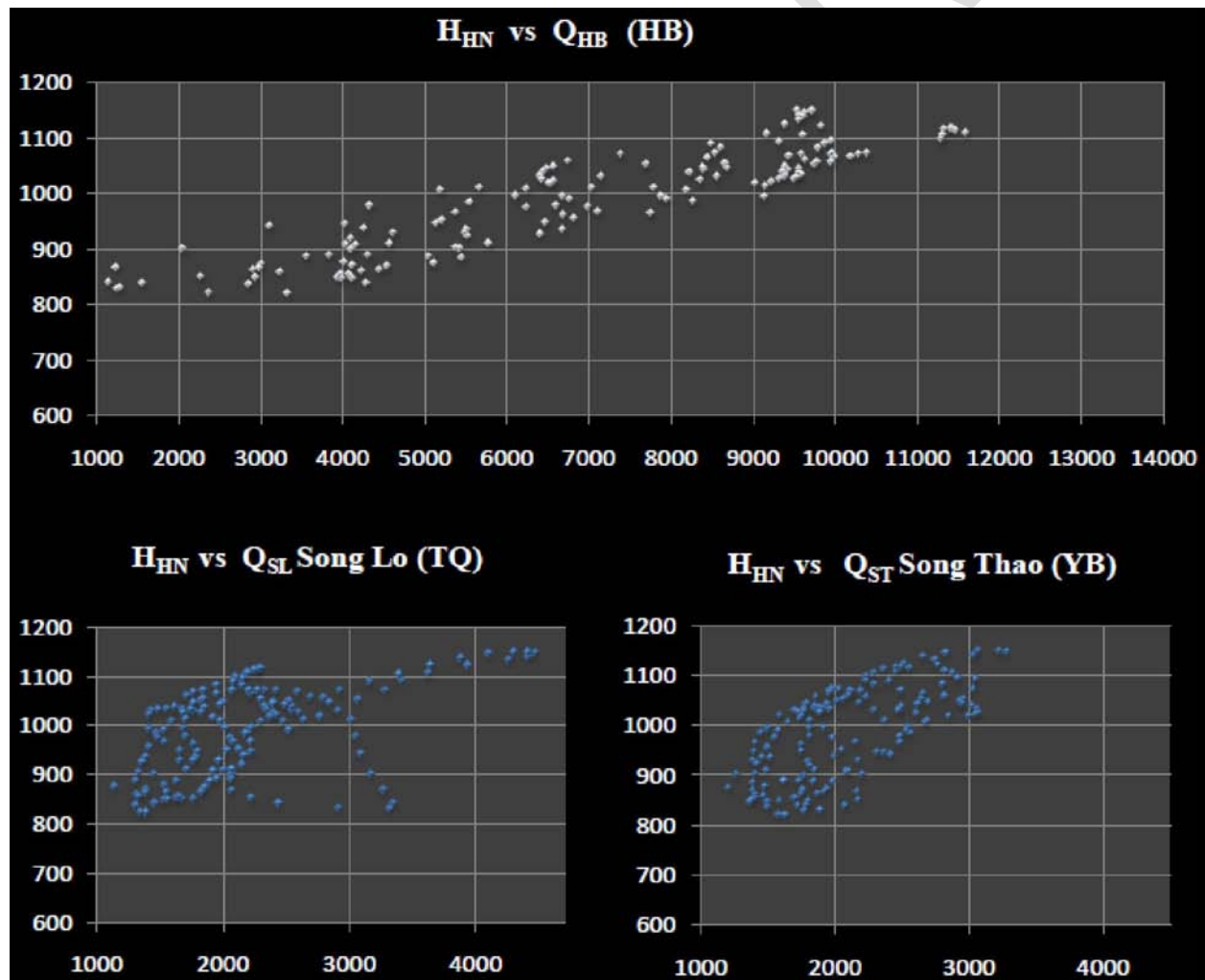


Mô hình hồi quy báo mức nước sông Hồng tại Hà Nội

Trần Trí Dũng – tridung@hn.vnn.vn

Phân tích sơ bộ: Mức nước sông Hồng đo ở Hà Nội là hàm của lưu lượng dòng chảy (m^3/sec) tại điểm đo. Đến lượt mình, lưu lượng dòng chảy sông Hồng tại Hà Nội là hàm của ba lưu lượng (i) Q_{HB} tại hạ lưu nhà máy thủy điện Hoà Bình; (ii) Q sông Lô, Q_{SL} , tại Tuyên Quang và (iii) Q sông Thao Q_{ST} tại Yên Bái. Lưu lượng dòng chảy của ba sông về đến Hà Nội cần thời gian giả thiết là 24 giờ¹. Như vậy, nếu biết Q_{HB} , Q_{SL} , Q_{ST} tại chu kỳ t , thì có thể dự báo mức nước sông Hồng tại Hà Nội tại chu kỳ $(t+4)$.

Biểu đồ ‘Scatter’ quan hệ mức nước Hà Nội $H_{HN}(t+4)$ với $Q_{HB}(t)$, $Q_{SL}(t)$ và $Q_{ST}(t)$ cho trong hình 1



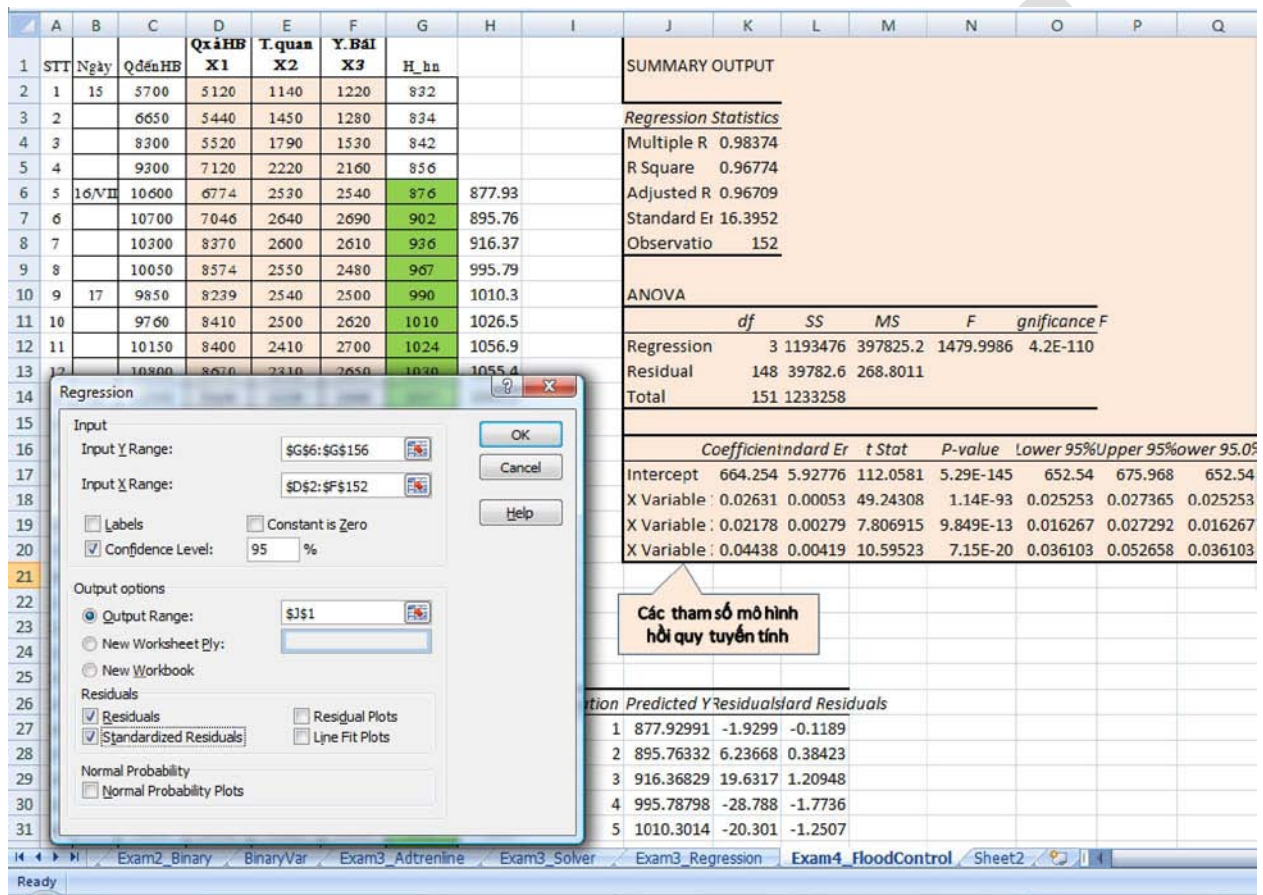
Hình 1- Quan hệ mức nước Hà Nội $H_{HN}(t+4)$ vs $Q_{HB}(t)$; $H_{HN}(t+4)$ vs $Q_{SL}(t)$ và $H_{HN}(t+4)$ vs $Q_{ST}(t)$.

¹ Tác giả không phải là kỹ sư chuyên ngành thủy văn nên giả thiết đưa ra có thể chưa chính xác.

Hình 1 cho thấy mức nước Hà Nội $H_{HN}(t+4)$ vs $Q_{HB}(t)$ gần tuyến tính; trong khi $H_{HN}(t+4)$ vs $Q_{SL}(t)$ và $H_{HN}(t+4)$ vs $Q_{ST}(t)$ không rõ ràng. Vì thế, hai mô hình được trắc nghiệm:

a. Mô hình tuyến tính

Bảng tính Excel, hộp thoại 'Regression' và các tham số mô hình hồi quy tuyến tính xin xem hình 2.



Mô hình hồi quy tuyến tính quan hệ mức nước sông Hồng tại Hà Nội $H_{HN}(t+4)$ vs $Q_{HB}(t)$ & $Q_{SL}(t)$ & $Q_{ST}(t)$ theo biểu thức 1.31 có các đặc trưng thống kê $R^2 = 0.9677$; $ESS = 39783$; sai số chuẩn $S_e = 16.4$ (cm).

$$H_{HN}(t+4) = 0.0263 Q_{HB}(t) + 0.0218 Q_{SL}(t) + 0.0444 Q_{ST}(t) + 664.25 \quad (1)$$

Trong đó $Q_{HB}(t)$ lưu lượng dòng chảy qua hạ lưu nhà máy thủy điện Hoà Bình (m^3/sec) tại thời điểm t ; $Q_{SL}(t)$ lưu lượng dòng chảy sông Lô đo tại Tuyên Quang (m^3/sec) tại thời điểm t ; $Q_{ST}(t)$ lưu lượng dòng chảy sông Thao đo tại Yên Bái (m^3/sec) tại thời điểm t ; và

$H_{HN}(t+4)$ là mức nước sông Hồng tại Hà Nội (cm) tại thời điểm chậm sau bốn thời đoạn (t+4) hay chậm sau $6 \times 4 = 24$ giờ.

Giả sử dùng (1) dự báo mức nước sông Hồng tại Hà Nội (cm) khi biết: $Q_{HB}(t) = 9960 \text{ m}^3/\text{sec}$; $Q_{SL}(t) = 2080 \text{ m}^3/\text{sec}$; $Q_{ST}(t) = 1980 \text{ m}^3/\text{sec}$. Sau 24 giờ, mức nước sông Hồng tại Hà Nội (cm) với mức độ tin cậy 95% sẽ giao động trong phạm vi $\pm 2 S_e$ là: kỳ vọng $H_{HN}(t+4) = 0.0263 \cdot 9960 + 0.0218 \cdot 2080 + 0.0444 \cdot 1980 + 664.25 = 1075.8$ cm; thực đo = 1072 cm; max: $1075.8 + 2 \cdot 16.4 = 1108.4$ cm; min: $1075.8 - 2 \cdot 16.4 = 1043.0$ cm.

Hình 2. Bảng tính Excel, hộp thoại ‘Regression’ để xác định các tham số mô hình hồi quy tuyến tính quan hệ mức nước sông Hồng tại Hà Nội.

b. Mô hình phi tuyến tính

Dùng ‘Solver’ để tìm giá trị các tham số mô hình phi tuyến có dạng đa thức mũ tổng quát

$$H_{HN}(t+4) = A_1 * Q_{HB}^{m1}(t) + A_2 * Q_{SL}^{m2}(t) + A_3 * Q_{ST}^{m3}(t) + A_0 \quad (2)$$

Như vậy bảy giá trị tham số cần tìm là $A_1, m1; A_2, m2; A_3, m3$ và A_0

Bảng tính Excel, hộp thoại ‘Solver’ và các tham số mô hình hồi quy phi tuyến xin xem hình 3. Trong bảng tính có đặt tên (define name) một số ô để dễ theo dõi. Có thể lập bảng tính theo các bước sau đây

- i. Đặt tên “m1”, “A₁”, “m2”, “A₂”, “m3”, “A₃”, và “A₀”, cho một số ô từ I2 đến O2 để giữ các giá trị tham số m1, A₁, m2, A₀.
- ii. Dùng ô L5 để thể hiện giá trị hàm mục tiêu ESS

Để tính ESS, ta cần tính mức nước theo hàm hồi quy (2) cho mỗi mẫu dữ liệu. Những giá trị mức nước ước tính \hat{Y}_i được tạo ra trong cột I:

Công thức cho ô I6: $=A1_*D2^m1_+A2_*E2^m2_+A3_*F2^m3_+Ao$
(Copy đến I7 ÷ I157)

Tổng bình phương sai số ESS được tính cho ô L5:

Công thức cho ô L5: $=SUMXMY2(G6:G157,I6:I157)$

Lưu ý rằng công thức trong ô L5 chính là phương trình hàm mục tiêu

Phần bảng tính mở rộng để tính R^2 , RSS và TSS như sau:

Công thức tính giá trị trung bình của mức nước Hà Nội cho ô G4 đặt tên H_hn_ave :
 AVERAGE (G6:G157)

Để tính $RSS = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ là tổng bình phương biến động của mỗi giá trị ước tính \hat{Y}_i xung quanh giá trị trung bình \bar{Y} thường gọi là tổng bình phương hồi quy (regression sum of squares-RSS) trong cột J:

Công thức cho ô J6: =(I6-H_hn_ave)^2
 (Copy đến I7 ÷ I157)

Tổng bình phương hồi quy RSS được tính cho ô J5:

Công thức cho ô J5: =SUM(J6:J157)

Để tính $TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2$ là tổng bình phương biến động của mỗi giá trị thực Y_i xung quanh giá trị trung bình \bar{Y} trong cột K:

Công thức cho ô K6: =(G6-H_hn_ave)^2
 (Copy đến K7 ÷ K157)

Tổng bình phương TSS được tính cho ô K5 đặt tên TSS:

Công thức cho ô K5: =SUM(K6:K157)

Công thức tính giá trị độ đo đánh giá sự phù hợp giữa mô hình hàm hồi quy và tập dữ liệu R^2 cho trong ô I3

Công thức cho ô I3: =1-L5/TSS

Mô hình hồi quy phi tuyến tính quan hệ mức nước sông Hồng tại Hà Nội $H_{HN}(t+4)$ vs $Q_{HB}(t)$ & $Q_{SL}(t)$ & $Q_{ST}(t)$ theo biểu thức 2 có các đặc trưng thống kê $R^2 = 0.9703$; $ESS = 36647$; sai số chuẩn $S_e = 15.736$ (cm).

$$H_{HN}(t+4) = 0.4673 * Q_{HB(t)}^{0.7093} + 0.8412 * Q_{SL(t)}^{0.6221} + 141.236 * Q_{ST(t)}^{0.1634} + 153.54 \quad (3)$$

Trong đó $Q_{HB}(t)$ lưu lượng dòng chảy qua hạ lưu nhà máy thủy điện Hoà Bình (m^3/sec) tại thời điểm t ; $Q_{SL}(t)$ lưu lượng dòng chảy sông Lô đo tại Tuyên Quang (m^3/sec) tại thời điểm t ; $Q_{ST}(t)$ lưu lượng dòng chảy sông Thao đo tại Yên Bái (m^3/sec) tại thời điểm t ; và $H_{HN}(t+4)$ là mức nước sông Hồng tại Hà Nội (cm) tại thời điểm chậm sau bốn thời đoạn (t+4) hay chậm sau $6 \times 4 = 24$ giờ.

Tóm tắt các đặc trưng thống kê của (1) và (3) cho trong bảng 1

| Loại mô hình | Đặc trưng thống kê |
|---------------------------|--|
| a. Mô hình tuyến tính (1) | $R^2 = 0.9677$; $ESS = 39783$; sai số chuẩn $S_e = 16.4$ (cm). |
| b. Mô hình phi tuyến (3) | $R^2 = 0.9703$; $ESS = 36647$; sai số chuẩn $S_e = 15.7$ (cm). |

Bảng 1 Tóm tắt các đặc trưng thống kê của (1) và (3)

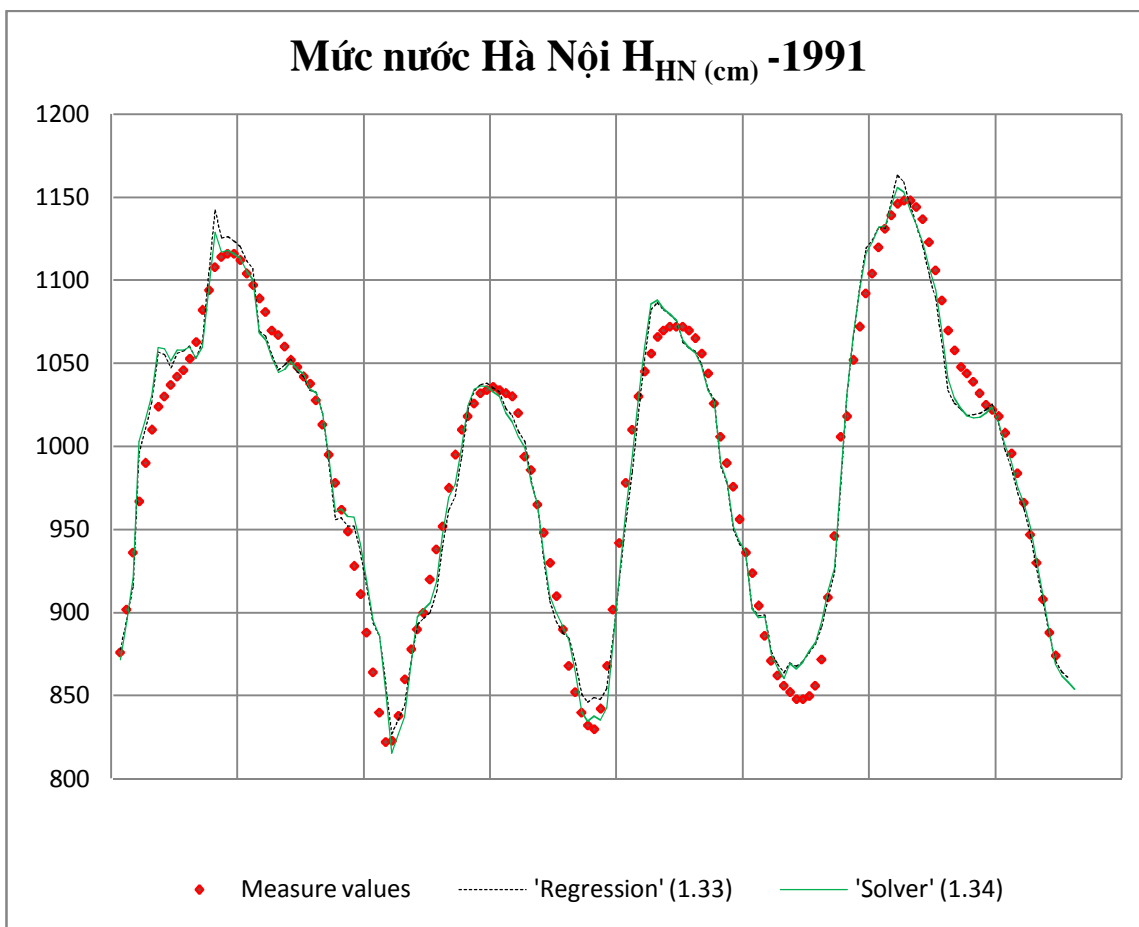
Nhận xét: với cùng một dạng mô hình hồi quy đa thức nhiều biến, mô hình phi tuyến nói chung có thể phù hợp hơn mô hình hồi quy tuyến tính (so sánh theo những đặc trưng thống kê Bảng 1). Điều này được thể hiện rõ khi so sánh mức nước sông Hồng thực đo tại Hà Nội và mức nước dự báo trước 24 giờ tính theo mô hình hồi quy (1) và (3) cho trong hình 4.

| 1 | Ngày | QđếnHB | QxảHB X1 | T. quan X2 | Y. Báo X3 | Các tham số mô hình hồi quy phi tuyến | | m1 | A ₁ | m2 | A ₂ |
|----|--------|--------|-------------|---------------|--------------|---------------------------------------|----------|---------|----------------|---------|----------------|
| 2 | 15 | 5700 | 5120 | 1140 | 1220 | | | 0.70931 | 0.46733 | 0.62214 | 0.8412 |
| 3 | | 6650 | 5440 | 1450 | 1280 | | R^2 | 0.97028 | | | |
| 4 | | 8300 | 5520 | 1790 | 1530 | 989 | S_e | 15.736 | RSS | TSS | ESS |
| 5 | | 9300 | 7120 | 2220 | 2160 | H_hn | Regress' | Solver' | 1196548 | 1233258 | 36,647 |
| 6 | 16/VII | 10600 | 6774 | 2530 | 2540 | 876 | 878 | 872 | 13804.7 | 12800.2 | |
| 7 | | 10700 | 7046 | 2640 | 2690 | 902 | 896 | 895 | 8897.76 | 7593.06 | |
| 8 | | 10300 | 8370 | 2600 | 2610 | 936 | 916 | 921 | 4595.46 | 2823.66 | |
| 9 | | 10050 | 8574 | 2550 | 2480 | 967 | 996 | 1003 | 190.217 | 490.098 | |
| 10 | 17 | 9850 | 8239 | 2540 | 2500 | 990 | 1010 | 1016 | 725.028 | 0.74277 | |
| 11 | | 9760 | 8410 | 2500 | 2620 | 1010 | 1027 | 1031 | 1728.57 | 435.216 | |
| 12 | | 10150 | 8400 | 2410 | 2700 | 1024 | 1057 | 1060 | 4975.38 | 1215.35 | |
| 13 | | 10800 | 8670 | 2310 | 2650 | 1030 | 1055 | 1059 | 4874.92 | 1669.69 | |
| 14 | 18 | 12500 | 8460 | 2220 | 2660 | 1037 | 1047 | 1051 | 3867.2 | 2290.76 | |
| 15 | | 13300 | 8620 | 2160 | 2810 | 1042 | 1056 | 1058 | 4776.68 | 2794.37 | |
| 16 | | 13600 | 9965 | 2160 | 2920 | 1046 | 1058 | 1058 | 4750.87 | 3233.27 | |
| 17 | | 13500 | 11600 | 2200 | 2820 | 1053 | 1060 | 1060 | 5043.9 | 4078.33 | |
| 18 | | | | | | | | | 58.83 | 5455.57 | |
| 19 | | | | | | | | | 0.69 | 8623.32 | |
| 20 | | | | | | | | | 41.4 | 10996 | |
| 21 | | | | | | | | | 25.3 | 14128.1 | |
| 22 | | | | | | | | | 61.5 | 15590.5 | |
| 23 | | | | | | | | | 6636 | 16093.9 | |
| 24 | | | | | | | | | 87.9 | 16093.9 | |
| 25 | | | | | | | | | 5434 | 15095 | |
| 26 | | | | | | | | | 06.9 | 13193.2 | |
| 27 | | | | | | | | | 23.4 | 11634.2 | |
| 28 | | | | | | | | | 62.9 | 9972.39 | |
| 29 | | | | | | | | | 1.41 | 8438.6 | |
| 30 | | | | | | | | | 6.34 | 6538.64 | |

Hình 3 Bảng tính Excel, hộp thoại ‘Solver Parameter’ để xác định các tham số mô hình hồi quy phi tuyến

So sánh mức nước sông Hồng thực đo tại Hà Nội và mức nước dự báo trước 24 giờ tính theo mô hình hồi quy tuyến tính ‘Regression’ (1) và phi tuyến ‘Solver’ (3). Những điểm đỉnh và đáy của

diễn biến mức nước, mô hình phi tuyến 'Solver' cho kết quả phù hợp hơn so với mô hình hồi quy tuyến tính 'Regression'. Xin xem hình 4



Hình 4. mức nước sông Hồng thực đo tại Hà Nội và mức nước dự báo trước 24 giờ .