입력 강우 형태에 따른 수위예측 회귀모형의 성능 비교분석

Comparison and Analysis for Performance of Flood Stage Prediction Regression Model according to Type of Input Rainfall

최승용* · 한건연**

Choi, Seung Yong · Han, Kun Yeun

Abstract

Recently to overcome limitations of hydrological and physics based models for flood stage forecasting, the regression model as one of data-derived models has been widely adopted for predicting flood stage. One of the advantages of regression model is able to use data such as point rainfall without handling of data. But most of the rainfall-runoff models need converting point rainfall to basin mean rainfall. The objectives of this study are to compare performance of different flood stage predicting regression models according to point rainfall and basin mean rainfall and determine the most effective predicting flood stage regression model. To do this, the time scale was determined through the autocorrelation analysis of input data and different flood stage forecasting models were developed using point rainfall and basin mean rainfall. To evaluate the performance of established models, four statistical indices were used, namely; Root mean square error(RMSE), Nash-Sutcliffe efficiency coefficient(NSEC), mean absolute error(MAE), adjusted coefficient of determination(R^{*2}). The results show that the flood stage predicting regression model using point rainfall can carry out the river flood stage prediction better than the flood stage predicting regression model using basin mean rainfall.

Key words : flood stage predicting, regression model, data-derived model, point rainfall, basin mean rainfall

요 지

최근 수위 예측을 위한 수문학적, 물리적 기반 모형 등의 단점을 극복하고자 홍수예측을 위해 자료지향형 모형 중의 하나인 회귀모형이 널리 도입되고 있다. 이러한 모형의 장점 중 하나는 지점 강우와 같은 관측 자료를 아무런 처리 없이 그대로 사용할 수 있다는 것이다. 즉, 양질의 관측 데이터만 확보하면 성능 높은 수위 예측 모형을 구축할 수 있다. 그러나 대부분의 강우-유출 모형의 경우에는 지점 강우를 유역 평균 강우의 형태로 변환해야 하는 작업을 필요로 한다. 본 연구의 목적은 이러한 수위 예측 회귀모형 구축시 사용하는 지점 강우와 유역 평균 강우에 따른 홍수예측 성능을 비교·검토하고 이를 통해 적절한 수위 예측 회귀모형을 구축하는 것이다. 이를 위해 입력자료의 자기상관분석을 통해 독립변수의 시간 규모를 결정한 지점 강우와 유역 평균 강우를 이용한 홍수예측 모형을 구축하고 중랑천 유역의 다양한 홍수사상에 대해 적용하였다. 구축된 모형들의 성능을 평가하기 위해 평균제곱근오차, Nash-Sutcliffe 효율계수, 평균절대오차, 수정 결정계수와 같이 4개의 통계지표들을 사용하였다. 모의결과 지점 강우를 이용한 홍수예측 회귀모형이 유역 평균 강우를 이용한 홍수예측 회귀모형에 비해 좀 더 나은 예측 결과를 나타냈다.

핵심용어 : 홍수위 예측, 회귀모형, 자료 지향형 모형, 지점 강우, 유역 평균 강우

1. 서 론

수많은 비구조적 홍수방어 대책들 중의 하나로서 예보지점에서의 정확한 수위예측은 수자원 시스템의 계획 및 운영을 위해 필요하며, 특히 잠재적 홍수 피해를 저감하기 위해 가장 최선의 대책은 무엇인지 등을 결정하는데 있어 매우 중요한 역할을 한다.

최근의 수위 예측에 관한 방법은 기존의 수문학적, 물리적

기반의 강우-유출 모형이 요구하는 많은 양의 자료 및 모형의 검보정을 위한 시간과 노력(VanderKwaak and Loague, 2001)과 더불어 모형화에 대한 물리적 과정에 대한 이해 등의 여러 제한 사항들 때문에(Kokkonen and Jakeman, 2001), 최근 자료 지향형 기반의 모형이 홍수예측에 널리 도입되고 있다. 최근 널리 사용되고 있는 자료 지향형 모형의 기법으로는 회귀분석, 인공신경망, 퍼지추론 시스템 등이 있다. 이러한 모형들은 기존의 모형과 비교해서 물리적인 성질이 완

*정회원 · 국립방재연구소 책임연구원 (E-mail : ecofriend@korea.kr)

**정회원 · 경북대학교 공과대학 건축·토목공학부 교수 (교신저자)

전히 밝혀지지 않은 비선형(nonlinear)이거나 동적(dynamic)인 모형의 잡음자료(noisy data)를 처리하는데 용이하며, 모형을 구축하는데 있어서 복잡한 수학적 지식이 필요하지 않기 때문에 상대적으로 모형개발을 위한 절차가 간단하고, 결과 값을 얻기까지의 계산시간 역시 많이 소요되지 않는 장점을 가지고 있다. 또한 자료 지향형 기반의 모형들은 앞선 언급한 모형에 비해 적은 자료를 필요로 함에도 불구하고 홍수에 측에 있어서 정확도 높은 예측과 적은 계산소요 시간으로 인해 실시간 예보를 위한 적용성이 우수하다(Govindaraju, 2000). 이러한 자료 지향형 기반의 모형 중 회귀모형은 모형의 구축과정이 단순하고 기본 이론이 간단하여 홍수예측에 자주 사용되고 있다.

회귀모형을 이용한 국내 연구로는 윤용남과 원석연(1991)이 미계측 중소유역의 월유출량 산정을 위해 다중회귀모형을 사용하였고, 윤강훈과 신현민(1994)은 한강인도교 수위와 영향 인자간의 다중회귀분석을 통해 홍수위 예측을 수행한 바 있다. 윤강훈과 김태균(2003)이 레이더 강우를 이용한 다중회귀모형을 이용하여 평창강, 보청천, 위천 유역에 대해 모형의 적용성을 검토하였고, 성지연과 허준행(2009)은 다중선형 회귀모형의 다중공선성 문제를 해결하기 위해 독립변수를 조정하고 회귀계수를 재산정한 통계적 모형을 이용하여 수위를 예측하였다. 정동국과 이범희(2010)는 대전광역시 도심하천 구간에서 현재까지의 강우, 수위 및 상류지역의 수위를 기반으로 수위를 예측할 수 있는 회귀모형을 구성하여 2시간 후의 수위를 예측하였으며, 이범희(2010)는 회귀분석에 의한 도시홍수 예보시스템을 개발하여 수위를 예측하였다.

회귀모형을 이용한 국외 연구로는 수문학적 추정을 위한 적용성(Chaloulakou et al. (1999), McKerchar and Delleur (1974) and Yurekli et al. (2005)), 홍수예측(Tangborn and Rasmussen (1976), Phien et al. (1990) and Schilling and Wolter (2005)) 등과 같은 분야에 적용되었다. Huynh et al. (2003)은 다중선형회귀모형 구성에서 주요지점의 수위를 종속변수로 하여 주요지점과 상류지점의 이전시간 수위와 이전시간 강우량 자료를 독립변수로 하여 베트남의 홍강(Red river)에 적용결과 정확도 높은 예측 값을 얻어 적용성을 입증한 바 있다. Wu et. al. (2008) 자료분류 기법 중의 하나인 support vector machine(SVM)과 회귀식을 결합한 수위예측 모형을 개발하여 수위 예측에 적용하였고, Kisi는 영상 자료의 이산화 기법 중의 하나인 wavelet transform(WT)을 이용하여 회귀모형의 적용성을 검토하였다.

이러한 회귀모형의 경우 앞서 기술한 바와 같이 모형 구축을 위해 관측된 강우와 수위만을 이용하게 된다. 이때 예측 모형의 성능은 모형 구축을 위해 사용한 입력자료(관측 강우와 수위)의 시간 규모 및 입력자료의 형태에 따라 결정된다. 그러나 기존의 회귀모형을 이용한 연구에서는 입력자료의 시간 규모 및 형태에 따른 예측 성능을 평가한 연구는 미흡한 실정이다. 이에 본 연구에서는 입력자료의 자기상관분석을 통해 입력자료의 시간적 규모를 결정된 후 지점 강우와 유역평균강우를 이용한 수위예측 회귀모형을 구축하고 홍수예측 결과를 비교·검토하였다. 이때 독립변수 선정 방법으로는 시

간적 규모로 선정된 독립변수 중 종속변수와 밀접한 연관성이 있는 몇 개의 변수만을 선택하여 이를 회귀모형에 포함시키는 방법인 단계별 선택법을 이용하였다.

2. 회귀모형 이론

회귀분석이란 객관적으로 나타난 자료를 바탕으로 독립변수와 종속변수간의 상호관계(선형, 비선형 관계 등)를 분석하는 기법이다. 이러한 회귀분석은 변수들간의 관계를 정형적으로 나타낼 수 있으므로 독립변수와 종속변수간의 미래 추이를 예측하고 할 경우 유용하다. 회귀분석에 사용되는 변수는 크게 독립변수(independent variable)와 종속변수(dependent variable)로 구분되는데 하나의 독립변수와 하나의 종속변수를 포함하는 가장 단순한 회귀분석 형태를 단순회귀분석이라고 하고 여러 개의 독립변수가 있는 회귀분석을 다중회귀분석이라고 한다.

종속변수 Y 를 설명하는데 k 개의 독립변수인 $X_1, X_2, \dots, X_k$ 를 도입할 때 다중회귀모형은 다음과 같이 정의된다.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, \quad (1)$$

여기서, $\beta_0, \beta_1, \dots, \beta_k$ 는 회귀계수 또는 매개변수를 의미하고, ε_i 는 서로 독립이고 동일한 분포 $N(0, \sigma^2)$ 을 따르는 오차항이다. $\beta_j(j=1, \dots, k)$ 는 j 번째 독립변수 X_j 의 회귀계수(기울기)를 의미하는데, X_j 를 제외한 다른 모든 독립변수의 값이 고정된 상태에서 X_j 의 값이 한 단위 증가할 때의 Y 값의 평균변화량을 나타낸다. 그리고 X_{ji} 는 독립변수 X_j 의 i 번째 관측치를 의미한다. 다중선형회귀모형은 행렬로 표현함으로써 간단한 형태로 나타낼 수 있다. 식 (1)을 행렬형태로 표현하기 위해 β_0 에 형식적으로 '1'을 넣고 관측번호 i 를 1에서 n 까지 풀어 쓰면 다음과 같다.

$$\begin{aligned} Y_1 &= \beta_0 \cdot 1 + \beta_1 X_{11} + \beta_2 X_{21} + \dots + \beta_k X_{k1} + \varepsilon_1 \\ Y_2 &= \beta_0 \cdot 1 + \beta_1 X_{12} + \beta_2 X_{22} + \dots + \beta_k X_{k2} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 \cdot 1 + \beta_1 X_{1n} + \beta_2 X_{2n} + \dots + \beta_k X_{kn} + \varepsilon_n \end{aligned} \quad (2)$$

Y 행, X 행, 모수, 오차를 각각 묶어 벡터와 행렬을 이용하여 다음과 같이 나타낼 수 있다.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix}, \quad B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (3)$$

여기서 벡터는 모두 열벡터(column vector)인데 벡터 Y 와 ε 는 크기 $(k+1)$ 의 열벡터이며, X 는 크기 $n \times (k+1)$ 의 열벡터이다. 그리고 B 는 행렬이 된다. 이때 다중선형회귀모형은 다

음과 같은 간단한 형태로 나타낼 수 있다.

$$Y = X\beta + \varepsilon \quad (4)$$

여기서 오차항 벡터 ε 는 정규분포를 따르는 n 개의 확률변수 ε_i 들로 이루어졌기 때문에 다변량 정규분포를 따른다.

2.1 매개변수 추정

다중선형회귀모형의 회귀계수들의 추정 방법 중 가장 기본적인 방법인 최소자승법(least squares)이다. 최소자승법은 실제값과 추정치간 차이 즉, 회귀오차의 자승을 합한 값이 최소가 되는 표본 회귀식을 찾는 방법이다. 이 방법은 자료들을 표시하는 각 좌표들과 이 사이를 지나는 추정 회귀식과의 수직적 길이(차이)를 가장 짧게 하는 것으로, 각 수직적 길이를 제공하여 합한 값을 최소화 하는 것이다. 단순히 수직적 길이를 합하게 되면 양의 값과 음의 값의 합이 되기 때문에 적합한 결과를 얻지 못하게 되고 절대값을 사용할 경우에는 추후 미분계수 계산시 문제가 발생할 수 있다.

또한 회귀오차의 절대값을 합한 값을 최소화하는 방법보다 회귀오차의 자승을 합한 값을 최소화하는 방법을 선택한 이유는 큰 폭의 오차에 대해 보다 더 큰(승수배의) 가중치(별점) 부여함으로써, 독립변수 값이 동일한 평균치를 갖는 경우 가능한 한 변동폭이 적은 표본회귀선을 도출하기 위한 것이다.

오차를 최소화하기 위한 회귀계수 $\beta_0, \beta_1, \dots, \beta_k$ 를 구하기 위해서는 $\beta_0, \beta_1, \dots, \beta_k$ 에 대해 각각 편미분한 값이 0이 되면 된다. 오차항의 제곱합 Q 는 다음과 같이 표현할 수 있다.

$$Q = [Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})]^2 \quad (5)$$

최소자승법에 의한 회귀계수의 추정은 제곱합 Q 를 각 β_j 에 대하여 편미분하고 이를 0으로 하는 다음과 같은 연립방정식을 풀어 $\hat{\beta}_j$ 들을 구하는 것이다.

$$\begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2 \sum [Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})] = 0 \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum X_{1i} [Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})] = 0 \\ &\vdots \\ \frac{\partial Q}{\partial \beta_k} &= -2 \sum X_{ki} [Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})] = 0 \end{aligned} \quad (6)$$

위의 연립방정식을 만족하는 를 만족하는 라 하면 다음 관계가 성립하는데 이를 정규방정식이라 한다.

$$\begin{aligned} \hat{\beta}_0 n + \hat{\beta}_1 \sum X_{1i} + \dots + \hat{\beta}_k \sum X_{ki} &= \sum Y_i \\ \hat{\beta}_0 \sum X_{1i} + \hat{\beta}_1 \sum X_{1i}^2 + \dots + \hat{\beta}_k \sum X_{1i} X_{ki} &= \sum X_{1i} Y_i \\ &\vdots \\ \hat{\beta}_0 \sum X_{ki} + \hat{\beta}_1 \sum X_{ki} X_{1i} + \dots + \hat{\beta}_k \sum X_{ki}^2 &= \sum X_{ki} Y_i \end{aligned} \quad (7)$$

식 (7)을 벡터-행렬식으로 표현하기 위해서 행렬 를 다음과

입력 강우 형태에 따른 수위예측 회귀모형의 성능 비교분석

같이 나타낼 수 있다.

$$X = \begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \dots & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} X_{2i} & \dots & \dots & \sum X_{1i} X_{ki} \\ \vdots & \vdots & \vdots & & & \vdots \\ \sum X_{ki} & \sum X_{ki} X_{1i} & \sum X_{ki} X_{2i} & & & \sum X_{ki}^2 \end{bmatrix} \quad (8)$$

식 (8)을 이용하면 식 (7)의 정규방정식이 다음과 같이 표현된다.

$$(X^T X) \hat{\beta} = X^T Y \quad (9)$$

따라서 β 의 최소자승추정량은 다음과 같다.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (10)$$

2.2 독립변수 선정 방법

다중회귀모형의 분석에서 적절한 독립변수의 선택은 변수의 절약 측면에서 매우 중요하다. 이와 더불어 불필요한 변수를 모형에 포함시킴으로서 다중공선성과 같은 부수적인 문제점이 발생할 여지가 있으므로 종속변수와 가장 밀접한 관련성이 있는 몇 개의 변수만을 선택하여 이를 회귀모형에 포함시키는 방법이 단계별 선택법(Stepwise)이다. 변수를 선택하는 방법에는 다음과 같은 방법들이 있다.

- ① 선정된 모든 독립변수를 사용하여 모형을 선택하는 방법
- ② 전진 선택법(Forward selection)으로 절편만 있는 최초 모형에서 설명력이 있다고 판단되는 독립변수를 하나씩 추가하면서 더 이상 추가될 독립변수가 없을 때까지 회귀분석을 수행하여 가장 적합한 모형 및 매개변수를 추정
- ③ 후진 소거법(Backward elimination)으로 모든 변수가 포함된 모형으로부터 출발하여 변수를 하나씩 제거하면서 최선의 모형을 구하는 방법. 즉, 모델에 나열된 모든 회귀 변수들을 포함하는 회귀모형을 구성한 한 후, 가장 기여도가 떨어지는 회귀 변수부터 차례로 소거해 나가는 방법
- ④ 단계별 선택법(Stepwise)은 전진 선택법과 후진소거법을 병행하여 최적의 모형을 선택하는 방법

전진선택법은 일단 모형 내에 들어 온 변수는 최종모형에 그 변수가 반드시 포함되어 있다. 그러나 새로운 변수가 모형에 추가되었을 때 독립변수간의 관계에 의해 기존의 독립변수의 설명력이 떨어지는 경우도 있을 수 있어, 한번 모형 내에 들어 온 변수라고 할지라도 새로운 완전모형에서 변수의 설명력을 재평가하여야 할 필요가 있다. 단계별 선택법은 이러한 개념을 도입한 모형 이다. 즉, 전진선택법에 의해 새로운 변수를 추가하고 추가된 변수에 의해 만들어진 새로운 완전모형에서 기존의 독립변수를 후진소거법에 의해 제거할 수 있는지를 판단하는 과정을 거치게 된다. 단계별 선택법에

의해 모형 선택을 할 때에는 두 가지 종류의 검정을 실시하여야 한다. 즉, 전진선택법에서 사용하였던 변수를 추가할 것인지를 대한 검정과 후진소거법에서의 검정인 변수를 제거할 것인지에 대한 검정을 필요로 한다. 단계별 선택법을 사용할 때 변수선택의 기준은 다음과 같다.

(1) 수정 결정계수(R^{*2})

$$R^{*2} = 1 - \frac{n-1}{n-k-1} \frac{SSE_k}{SST} \quad (11)$$

여기서 n 은 자료의 수, K 는 독립변수의 수, SSE_k 는 독립변수가 k 인 축소모형의 SSE 이다. 독립변수가 추가되면 결정계수인 R^2 값은 언제나 증가하게 된다. 따라서 독립변수의 개수가 다른 두 모형에 대해서는 결정계수 R^2 값의 비교만으로는 불충분하게 된다. 이를 보완하는 것이 수정결정계수이며, 이는 MSE 가 가장 작은 모형을 가장 좋은 모형으로 판정하는 기준이다.

(2) Mallows의 C_p

$$C_p = \frac{SSE_k}{MSE} + 2(k+1) - n \quad (12)$$

여기서 MSE 는 잔차제곱합인 SSE 의 평균이고, Mallows C_p 판정 기준은 선정된 모형의 예측 능력을 정량화하여 최적의 모형을 선택하는 방법으로 많이 사용되고 있다. 추정된 C_p 값은 $k+1$ 이며, 이 값에 가장 근접한 모형을 최적의 모형으로 선택한다. $C_p \approx k+1$ 인 k 중 최소인 것에 해당하는 모형을 최선의 모형으로 판정하는 기준이다.

(3) Akaike Information Criterion

$$AIC = n \log \frac{SSE_k}{n} + 2(k+1) \quad (13)$$

여기서 n 은 자료의 수, k 는 독립변수의 수이며, SSE_k 는 독립변수가 k 인 축소모형의 SSE 이다. 최적의 모형을 위한 판단을 위해서는 AIC 값이 가장 작은 값을 갖는 모형을 최적의 모형으로 선택한다(n 이 클 때는 C_p 와 동치).

3. 대상유역 및 입력자료 시간규모의 결정

3.1 대상유역 및 자료

개발 모형의 수위 예측에 대한 성능을 파악하기 위해 한강 유역의 중랑천에 적용하였다. 중랑천 유역은 동경 $127^{\circ}03' \sim 127^{\circ}08'$, 북위 $37^{\circ}40' \sim 37^{\circ}48'$ 사이에 위치하는데 본류인 중랑천은 경기도 양주군 주내면 산북리 불국산에서 발원하여 의정부시를 지나 남류하고, 서울시 성동구 사근동에 이르러 유역의 최대지류인 청계천이 합류하고, 성수대교 우안쪽으로 한강 본류에 유입한다.

중랑천 유역의 상류부 동쪽에 의정부시와 양주군과의 경계를 이루는 천령산맥이 위치하고, 유역의 동쪽에 용암산(EL. 476.9 m), 수락산(EL. 637.7 m), 불암산(EL. 508 m) 등이 왕숙천과 분수령을 이루고 있으며, 서쪽으로는 도봉산(EL. 730.5 m), 북한산(EL. 799.5 m) 등이 곡릉천, 창릉천과 분수령을 이룬다.

중랑천 유역의 유역면적은 296.04 km^2 , 유역연장 34.80 km , 유역평균 폭이 8.1 km , 유역의 평균고도는 107.2 EL.m 이며, 서울특별시와 의정부시가 $3/4$ 이상을 점하고 있는 고밀도 도시지역이고, 농경지는 상류부의 제 1, 2 지류 연변에 다소 분포하고 있으며, 하상의 구성 재료는 대부분이 가는 모래와 굵은 모래로 구성되어 있다. 하폭은 하구부에서 330 m , 서울 시계 부근에서 110 m 정도이고, 하상의 평균경사는 $1/1,150$ 이다(건설교통부, 2005). 그림 1은 대상유역의 수문관측소 현황을 나타내고 있다.

수위예측을 위한 지점으로 중랑천 하류 지점인 중랑교 수위관측소 지점을 선정하였다. 어떤 관측소에서 강우기록을 얻었다면 강우의 시·공간적 분포특성 분석을 위하여 단일 호우사상을 어떻게 선정해야 할 것인가는 매우 중요하다. 기존 강우의 시간분포에 대한 연구자들은 여러 가지 방법으로 다양한 단일 호우사상의 선정기준을 제시하였다. 이를 살펴보면 Huff방법은 연속자료에서 무강우시간 6시간 이상이면 단일 호우사상으로 선정하였고, Yen과 Chow 방법은 무강우가 없는 연속강우, Pilgrim과 Cordery 방법, Keifer와 Chu 방법은 지속시간별 최대우량을 단일 호우사상으로 선정하였다. 본 논문에서는 이중 Huff방법에 따라 무강우 시간이 6시간 이상인 것을 단일 호우사상으로 하여 모형을 구축하기 위한 입력 자료로 2002년부터 2005년까지의 17개 호우사상을 선정하고 선정된 기간의 호우 사상의 대상 유역에 대해 그림 1에서 보는 바와 같이 5개의 강우관측소의 강우자료와 중랑교 수위관측소의 수위를 사용하였다. 모형의 구축을 위해 강우자료는 대상 유역의 5개의 강우관측소의 각각의 지점 강우자료와 각각의 지점 강우량으로부터 유역의 평균 강우량 산정 기법 중 하나인 티센법을 이용하여 산정된 유역 평균 강우를 사용하였다.

티센법의 경우 산지유역에 사용할 경우 제한적이기는 하나 대상 유역인 중랑천 유역의 경우 도심유역으로 티센법을 사용하여도 큰 문제가 없을 것으로 판단하였다. 수위예측 회귀

그림 1. 대상유역 수문관측소 현황 및 티센망

표 1. 호우사상의 특성(유역 평균 강우)

호우사상 이름	강우시작	강우종료	총강우량 (mm)	강우지속기간 (hr)	호우원인	비고
J-1	02/04/29 08:00	02/04/30 16:00	81.1	32	기압골	Training
J-2	02/07/05 09:30	02/07/07 00:00	94.7	38.5	태풍	Training
J-3	02/08/04 07:00	02/08/07 23:00	361.3	88	기압골	Training
J-4	03/05/06 14:00	03/05/07 22:00	96.4	32	기압골	Training
J-5	03/07/21 19:00	03/07/23 08:00	192.1	37	장마전선	Training
J-6	03/08/06 11:00	03/08/07 07:30	65.9	20.5	기압골	Training
J-7	03/08/19 14:00	03/08/20 13:30	149.8	23	기압골	Training
J-8	03/08/23 04:30	03/08/28 07:00	319.8	122.5	기압골	Training
J-9	03/09/18 04:00	03/09/19 00:00	142.4	20	기압골	Training
J-10	04/07/03 18:30	04/07/05 10:30	83.7	40	기압골	Testing
J-11	04/07/11 18:30	04/07/13 07:00	138.9	36.5	기압골	Validation
J-12	05/06/26 17:00	05/06/27 18:30	136.3	25.5	기압골	Testing
J-13	05/07/01 00:30	05/07/03 15:00	118.3	62.5	장마전선	Testing
J-14	05/07/28 00:30	05/07/28 18:00	122.9	17.5	기압골	Testing
J-15	05/08/10 09:00	05/08/12 10:00	112.9	49.0	기압골	Validation
J-16	05/08/24 21:00	05/08/25 20:00	83.0	23.0	기압골	Validation
J-17	05/09/13 05:30	05/09/13 23:30	95.6	18.0	기압골	Validation

모형의 개발을 위해 training 자료로 2002년~2003년의 9개 호우사상을 사용하였고 2004년과 2005년의 8개 호우사상 중 4개를 test를 위한 자료로 나머지 4개의 호우사상을 모형의 검증에 위해 사용하였다. 표 1은 선정된 강우사상의 특성을 나타내고 있다.

3.2 독립변수 시간 규모의 결정

본 연구의 수위예측 회귀모형에서 독립변수로 사용되는 수위와 강우의 시간 규모를 결정하기 위해 자기상관분석(Autocorrelation analysis)을 수행하였다. 자기상관이란 시계열 자료(time-series data)가 일정 시간 단위에 따라서 시차상관(lag correlation)을 의미하는 것으로 주로 시계열 자료에서 나타난다. 즉, 어떠한 시계열자료가 주어질 때 현재 시간 t의 관측치가 이전 시간들의 관측치들에 대하여 상관이 있을 때 자기상관이 있다고 표현하며, 이러한 자기상관을 통해 과거 또는 현재 값이 현재 또는 미래 값을 예측하는데 쓰일 수 있는지 없는지를 판단할 수 있다. 자기상관계수는 식 (14)와 같이 계산된다.

$$\frac{\frac{1}{T} \sum_{t=\tau+1}^T (y_t - \bar{y})(y_{t-\tau} - \bar{y})}{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2} \quad (14)$$

그림 2는 30분 간격의 수위 시계열의 12차 자기상관분석을 통해 계산된 자기상관계수를 자기상관함수(Auto-correlation function)로 표현한 것이다. 그림에서 보는 바와 같이 차수(lag)가 커질수록 자기상관계수가 작아지는 것을 볼 수 있으나 12차까지도 0.6이상의 자기상관계수 값을 나타내고 있으며 95%의 신뢰도 구간에 포함되고 있음을 확인 할 수 있다.

입력 강우 형태에 따른 수위예측 회귀모형의 성능 비교분석

그림 2. 수위 시계열 자료의 자기상관함수

본 연구에서는 선행 예측 시간을 3시간, 즉 t+6으로 결정하고 자기상관계수가 그림에서 보는 바와 같이 0.75이상의 값을 갖도록 독립변수의 시간 규모를 결정하였다. 이를 통해 독립변수 수위자료의 시간적 규모를 t-4까지로 결정하였다.

4. 수위예측 회귀모형의 구축

4.1 회귀식 선택의 검토

회귀식의 형태를 선정하는 것은 회귀모형 구축시 중요한 사항이다. 이에 본 연구에서는 잔차와 독립변수간의 산포도를 통하여 회귀식 선정을 검토하였다. 본래 잔차와 독립변수의 산포도에서는 어떠한 경향도 보이지 않는 것이 바람직하다(노형진과 정한열, 2001). 만약 독립변수와 잔차의 산포도에서 2차적인 경향 등이 나타난다면 회귀식의 구성에 해당 독립변수에 대한 2차항을 추가함으로써 비선형 회귀식을 구성하여야 한다. 그림 4는 모형 구축을 위해 사용된 독립변수와 잔차의 산포도 중 현재시간(t)에서의 강우와 잔차의 산포도를

각 예측 시간별로 나타내고 있다. 그림에서 보는 바와 같이 독립변수와 잔차 사이의 어떠한 경향도 나타내고 있지 않으므로 선형회귀식을 사용하여 모형을 구축하는 것으로 하였다.

4.2 모형 구축

자기상관분석을 통해 결정된 독립변수의 시간규모를 근거로 통계모형 구축을 위해 $R(t)$, $R(t-1)$, $R(t-2)$, $R(t-3)$, $R(t-4)$, $H(t)$, $H(t-1)$, $H(t-2)$, $H(t-3)$, $H(t-4)$ 의 강우와 수위관측 자료를 이용하였다. H 는 수위자료, R 은 강우자료를 나타내며 H 와 R 의 숫자들은 입력자료의 시간분포를 의미한다. 0은 현재 시간을 의미하고 -1은 현재시간으로부터 한 간격 앞선 시간의 자료를 의미한다. 본 연구의 경우 관측시간 간격이 30분으로 -1의 경우는 현재 시간으로부터 30분 이전 시간의 자료를 의미한다. 이때 모형 구축을 위해 사용된 강우의 경우 지점 강우자료와 유역 평균 강우의 두 가지 자료 형태를 구분하기 위해 지점 강우의 경우 각 강우관측소 별 번호를 부여하였고, 유역 평균 강우의 경우 $R_M(t)$ 과 같이 아래첨자

$M(\text{mean})$ 을 사용하여 나타내었다. 각 강우관측소의 번호는 편의를 위해 유역의 상류에서 하류로 가며 차례로 번호를 부여하였다. 즉, 의정부 강우관측소는 1, 장암 강우관측소는 2, 상계 강우관측소는 3, 퇴계원 강우관측소는 4, 우이 강우관측소는 5로 나타내었다.

4.2.1 지점 강우를 이용한 수위예측 회귀모형의 구축

단계별 선택법을 통해 독립변수가 종속변수의 예측에 기여하는지를 판단하여 독립변수의 추가와 제거를 수행하였고 이를 통해 수위예측 회귀모형을 구축하였다. 이때 강우 독립변수로는 유역 내 5개 관측소의 지점 강우를 사용하였다. 표 2는 수위예측 모형의 독립변수에 대한 추가 및 제거 결과와 각 독립변수의 개별회귀계수에 대한 t -검정 값 중 $t+1(0.5\text{hr})$ 선행의 결과를 나타내고 있다. 각 독립변수의 종속변수 Y 에 대한 영향력의 대소를 판단하기 위한 개별회귀계수의 유의성 검토에는 표에서 보는 바와 같이 t -검정 값을 이용하였다. t -검정 값이 높은 변수일수록 종속변수 Y 를 예측하는데 공헌도

그림 3. 독립변수 $R(t)$ 와 잔차의 산포도

표 2. 독립변수의 추가 및 제거 결과(선행시간 t+1)

독립변수	t-검정	독립변수	t-검정	독립변수	t-검정
R ₁ (t)		R ₃ (t)	10.552	R ₅ (t)	16.120
R ₁ (t-1)		R ₃ (t-1)	1.836	R ₅ (t-1)	3.810
R ₁ (t-2)	4.974	R ₃ (t-2)	-5.216	R ₅ (t-2)	-11.135
R ₁ (t-3)	3.180	R ₃ (t-3)	-5.629	R ₅ (t-3)	-4.764
R ₁ (t-4)	5.749	R ₃ (t-4)	-3.096	R ₅ (t-4)	-3.167
R ₂ (t)		R ₄ (t)	1.877	H(t)	56.088
R ₂ (t-1)	8.117	R ₄ (t-1)		H(t-1)	-5.515
R ₂ (t-2)	4.070	R ₄ (t-2)		H(t-2)	
R ₂ (t-3)		R ₄ (t-3)		H(t-3)	-9.418
R ₂ (t-4)	-2.574	R ₄ (t-4)		H(t-4)	-1.466
제거 독립변수					
R ₁ (t)	-0.222	R ₁ (t-1)	0.074	R ₂ (t)	1.349
R ₂ (t-3)	-1.379	R ₄ (t-1)	-0.505	R ₄ (t-2)	-1.386
R ₄ (t-3)	0.847	R ₄ (t-4)	-1.056	H(t-2)	-0.759

가 높다고 판단하며 식 (15)와 같이 계산한다.

$$t\text{-검정값} = \frac{\text{개별회귀계수값}}{\text{계수의표준오차}} \quad (15)$$

표에서 보는 바와 같이 최종 모형 구축을 위해 선정된 독립변수의 개별회귀계수의 t-검정 값은 유의성 기준인 1.414보

다 큰 값을 나타내고 있음을 통해 독립변수의 추가 및 제거가 잘 이루어진 것을 확인할 수 있다.

표 3은 지점 강우를 이용한 회귀분석을 통해 구축된 각 선행 시간별 수위예측 회귀모형을 나타낸 것이다. 표에서 보는 바와 같이 모형 구축을 위해 사용된 강우 독립변수의 경우 예측 선행시간에 따라 모형 구축 시 사용된 관측소와 시간분포가 다르다는 것을 확인할 수 있다.

4.2.2 유역 평균 강우를 이용한 수위예측 회귀모형의 구축

지점 강우를 이용한 수위예측 회귀모형 구축과 같은 방법으로 유역 평균 강우를 이용한 수위예측 회귀모형을 구축하였다. 표 3은 각 선행 예측시간별 수위예측 회귀모형의 독립변수 추가 및 제거 결과와 회귀계수 산정결과를 나타낸 것이다. 또한 각 독립변수의 개별회귀계수 t-검정 값을 제시하고 있다. 표에서 보는 바와 같이 30분 예측 모형의 경우 독립변수의 추가 및 제거과정을 통해 최종적으로 H(t-2), H(t-3), H(t-4)의 독립변수들이 제거된 상태로 수위예측 모형이 구축되었고 60분 예측 모형의 경우에는 R(t-4)와 H(t-4)의 독립변수들이 제거된 상태로 수위예측 모형이 구축되었다. 표 3과 4를 비교해보면 같은 t+1 선행시간의 경우라도 유역 평균 강우를 이용한 모형의 경우 모든 시간 분포의 강우가 모형 구축에 사용되는데 반해, 지점 강우를 이용한 모형의 경우에는 퇴계원 관측소의 경우 t시간의 강우만을 이용하여 모형 구축이 이루어진 것을 알 수 있다.

표 3. 지점 강우를 이용한 수위예측 회귀모형 구축 결과

모형	절편	의정부 강우				
		R ₁ (t)	R ₁ (t-1)	R ₁ (t-2)	R ₁ (t-3)	R ₁ (t-4)
H(t+1) 예측모형	0.0165			0.0028	0.0017	0.0032
H(t+2) 예측모형	0.0378			0.0058	0.0054	0.0058
H(t+3) 예측모형	0.0632	0.0038	0.0047	0.0109	0.0085	0.0108
H(t+4) 예측모형	0.0927	0.0103	0.0115	0.0163	0.0136	0.0106
H(t+5) 예측모형	0.1232	0.0178	0.0199	0.0239	0.0158	0.0101
H(t+6) 예측모형	0.1515	0.0247	0.0254	0.0253	0.0172	0.0090
장압 강우						
R ₂ (t)	R ₂ (t-1)	R ₂ (t-2)	R ₂ (t-3)	R ₂ (t-4)		
	0.0057	0.0031		-0.0019		
0.0148	0.0119	0.0032	-0.0082	-0.0030		
0.0287	0.0144		-0.0106	-0.0087		
0.0384	0.0103	-0.0063	-0.0183			
0.0411		-0.0136	-0.0197			
0.0356		-0.0148	-0.0220			
상계 강우						
R ₃ (t)	R ₃ (t-1)	R ₃ (t-2)	R ₃ (t-3)	R ₃ (t-4)		
0.0070	0.0015	-0.0043	-0.0040	-0.0024		
0.0060		-0.0063				
	-0.0053	-0.0070				
-0.0079	-0.0085			-0.0120		
-0.0092	0.0000			-0.0194		
				-0.0204		

표 3. 지점 강우를 이용한 수위예측 회귀모형 구축 결과(계속)

퇴계원 강우				
$R_4(t)$	$R_4(t-1)$	$R_4(t-2)$	$R_4(t-3)$	$R_4(t-4)$
0.0012				
				-0.0025
				-0.0046
				-0.0074
우이 강우				
$R_5(t)$	$R_5(t-1)$	$R_5(t-2)$	$R_5(t-3)$	$R_5(t-4)$
0.0089	0.0023	-0.0067	-0.0028	-0.0018
0.0198		-0.0128	-0.0083	-0.0052
0.0264		-0.0207	-0.0127	-0.0061
0.0305		-0.0270	-0.0140	
0.0288		-0.0286	-0.0161	
0.0233	-0.0063	-0.0291	-0.0162	
중랑교 수위				
$H(t)$	$H(t-1)$	$H(t-2)$	$H(t-3)$	$H(t-4)$
1.2444	-0.1615	0.0000	-0.1048	0.0000
1.3547	-0.1531	-0.1728	0.0000	-0.0766
1.2348	0.0000	0.0000	-0.3113	0.0000
1.2609	0.0000	-0.2183	0.0000	-0.1492
1.2153	0.0000	0.0000	-0.3511	0.0000
1.1949	0.0000	0.0000	-0.3560	0.0000

표 4. 유역 평균 강우를 이용한 수위예측 회귀모형 구축 결과

	$H(t+1)$ 예측모형		$H(t+2)$ 예측모형		$H(t+3)$ 예측모형	
	계수	t-검정	계수	t-검정	계수	t-검정
절편	0.0199	33.454	0.0547	44.671	0.0933	50.323
$R(t)$	0.0158	26.481	0.0385	31.433	0.0565	30.460
$R(t-1)$	0.0067	9.365	0.0083	5.689	0.0129	5.774
$R(t-2)$	-0.0048	-6.730	-0.0082	-5.689	-0.0110	-4.973
$R(t-3)$	-0.0018	-2.639	-0.0031	-2.186	-0.0040	-1.905
$R(t-4)$	-0.0018	-2.984			-0.0034	-1.761
$H(t)$	1.4645	82.739	1.6260	38.733	1.5280	27.697
$H(t-1)$	-0.4895	-28.936	-0.6630	-8.689	-0.6369	-12.080
$H(t-2)$			-0.1495	-2.062		
$H(t-3)$			0.1196	3.207		
$H(t-4)$						
	$H(t+4)$ 예측모형		$H(t+5)$ 예측모형		$H(t+6)$ 예측모형	
	계수	t-검정	계수	t-검정	계수	t-검정
절편	0.1411	55.957	0.1824	77.165	0.2211	60.215
$R(t)$	0.0680	26.964	0.0743	31.433	0.0793	21.601
$R(t-1)$	0.0137	4.609	0.0174	5.689	0.0188	4.339
$R(t-2)$	-0.0097	-3.314	-0.0087	-5.689	-0.0073	-1.722
$R(t-3)$						
$R(t-4)$						
$H(t)$	1.4647	18.938	1.3489	38.733	1.2534	11.127
$H(t-1)$	-0.7070	-8.394	-0.6546	-8.689	-0.6276	-5.116
$H(t-2)$						
$H(t-3)$						
$H(t-4)$	0.0805	3.347	0.1022	-0.226	0.1330	3.800

5. 모의결과 및 비교검토

5.1 모의결과

구축된 각각의 회귀모형을 이용하여 수위예측을 수행하였다. 모형의 검증에 위해 사용한 4개의 호우 사상에 대한 결과 중 J-11과 J-16 호우사상의 선행시간별 예측 결과를 그림 4와 5에 나타내었다. 그림 4에서 보는 바와 같이 30분 선행기간의 경우에는 지점 강우와 유역 평균 강우를 이용한 수위 예측 회귀모형 모두 예측 수위와 관측 수위가 상당히 잘 일치하고 있음을 알 수 있으나, 유역 평균 강우를 이용한 회귀 모형의 수위 예측결과가 다소 크게 나타나고 있는 것을 알 수 있다. 이러한 경향은 예측 선행시간이 커질수록 뚜렷하게 나타나고 있는데, 특히 침투 예측에 있어서 유역 평균 강우를 이용한 회귀모형의 예측 결과가 지점 강우를 이용한 회귀 모형의 예측 결과에 비해 과대 예측되는 경향을 보이고 있음을 확인할 수 있다. 이러한 이유는 면적평균강우 산정시 지

점강우를 유역에 평균화하는 과정에 기인한 것으로 판단된다.

J-16 호우사상의 예측 결과인 그림 5를 보면 두 회귀모형 간의 예측 성능 차이를 좀 더 명확히 확인할 수 있다. 그림에서 보는 바와 같이 30분 선행시간의 경우에는 두 모형 모두 상당히 만족할만한 예측 결과를 보이고 있으나 선행시간이 길어질수록 유역 평균 강우를 이용한 수위예측 회귀모형의 경우에는 수위 상승부와 침투부근에서 상당히 과대 예측하는 경향을 보이고 있다. 이에 반해 지점 강우를 이용한 수위예측 회귀모형의 경우에는 120분 선행시간까지는 수위 상승부와 침투부근에서 수위예측 결과가 관측치와 비교적 잘 일치하는 것을 확인할 수 있었다. 물론 이후의 선행시간에서는 지점 강우를 이용한 수위예측 회귀모형의 경우에도 침투부근에서 다소 과대 예측하는 경향을 보이고는 있으나 이 경우에도 유역 평균 강우를 이용한 모형에 비해서는 우수한 예측 성능을 보이고 있음을 확인할 수 있었다.

이러한 결과는 J-11과 J-16 호우사상에 대한 예측치와 관

그림 4. 수위예측 회귀모형의 선행시간별 수위예측 결과(J-11)

그림 5. 수위예측 회귀모형의 선행시간별 수위예측 결과(J-16)

그림 6. 수위예측 회귀모형간 오차 비교(J-11)

측치간의 오차를 나타낸 그림 6과 7을 살펴보면 더욱 명확하게 알 수 있다.

J-11 호우 사상의 1.5시간 후의 예측치와 관측치간의 오차를 도시한 그림 6의 좌측 그림을 살펴보면 지점 강우를 이용한 수위예측 회귀모형이 유역 평균 강우를 이용한 회귀모

형에 비해 오차가 작게 나타나고 있음을 알 수 있고 이러한 결과는 3.0시간 후를 예측한 그림 6의 우측 그림을 통해 재 확인할 수 있다. J-16 호우 사상의 예측결과에 대한 오차를 나타낸 그림 7의 경우에도 그림에서 보는 바와 같이 1.5시간과 3.0시간 후의 예측 결과에 대해 지점 강우를 이용한 수위

그림 7. 수위예측 회귀모형간 오차 비교(J-16)

표 5. 정량적 평가를 위한 통계지표

통계지표	식
평균제곱근오차(RMSE)	$\sqrt{\frac{\sum_{i=1}^N (v_i - \hat{y}_i)^2}{N}}$
Nash-Sutcliffe 효율계수 (NSEC)	$\frac{\sum_{i=1}^N (v_i - \bar{y})^2 - \sum_{i=1}^N (v_i - \hat{y}_i)^2}{\sum_{i=1}^N (v_i - \bar{y})^2}$
평균절대오차(MAE)	$\frac{1}{N} \sum_{i=1}^N v_i - \hat{y}_i $

여기서 v_i 와 $\hat{y}_i$ 는 각각 관측치와 예측치, $\bar{y}$ 와 $\hat{\bar{y}}$ 는 관측치의 예측치의 평균, N 은 자료의 개수이다.

예측 회귀모형이 유역 평균 강우를 이용한 회귀모형에 비해 오차가 작게 나타나고 있음을 알 수 있다, 이를 통해 지점 강우를 이용한 수위예측 회귀모형이 유역 평균 강우를 이용한 회귀모형에 비해 예측 능력을 보이고 있음을 확인할 수 있었다.

본 연구에서 구축된 수위예측 회귀모형을 이용한 4개의 검증 사상에 대한 예측의 정확도를 정량적으로 평가하기 위해 앞서 제시한 수정 결정계수(R^2 , adjusted coefficient of determination)와 평균제곱근오차(Root Mean Square Error), Nash-Sutcliffe 효율계수(Efficiency Coefficient), 평균절대오차 (Mean Absolute Error)와 같은 통계지표를 사용하였다. 평균 제곱근오차는 관측값과 예측값의 차이를 다룰 때 흔히 사용되는 통계지표로 정밀도(precision)를 표현하는데 적합하며 평균절대오차는 관측값과 예측값의 편차 절대값을 산술 평균한 것이다. 사용된 통계지표를 간단히 설명하면 다음과 같다.

표 6. 수위예측 회귀모형간 성능 평가를 위한 통계지표 값의 비교

호우사상	모형	RMSE						NSEC					
		t+1	t+2	t+3	t+4	t+5	t+6	t+1	t+2	t+3	t+4	t+5	t+6
J-11	지점 강우	0.024	0.051	0.084	0.114	0.137	0.163	0.999	0.991	0.964	0.933	0.902	0.862
	평균강우	0.030	0.062	0.092	0.118	0.143	0.167	0.996	0.987	0.956	0.928	0.894	0.856
J-15	지점 강우	0.026	0.065	0.089	0.119	0.143	0.157	0.998	0.989	0.958	0.918	0.878	0.841
	평균강우	0.035	0.070	0.102	0.133	0.156	0.171	0.995	0.975	0.929	0.863	0.821	0.783
J-16	지점 강우	0.008	0.023	0.032	0.045	0.061	0.077	0.999	0.995	0.975	0.968	0.918	0.887
	평균강우	0.011	0.028	0.040	0.055	0.070	0.086	0.998	0.990	0.972	0.958	0.912	0.872
J-17	지점 강우	0.044	0.083	0.113	0.147	0.165	0.189	0.997	0.963	0.908	0.866	0.810	0.796
	평균강우	0.049	0.091	0.120	0.151	0.177	0.200	0.995	0.959	0.897	0.847	0.796	0.743
호우사상	모형	MAE						R^2					
		t+1	t+2	t+3	t+4	t+5	t+6	t+1	t+2	t+3	t+4	t+5	t+6
J-11	지점 강우	0.008	0.018	0.029	0.027	0.045	0.053	0.998	0.988	0.963	0.933	0.902	0.861
	평균강우	0.010	0.021	0.031	0.041	0.052	0.065	0.996	0.986	0.956	0.928	0.894	0.855
J-15	지점 강우	0.009	0.020	0.030	0.037	0.051	0.061	0.995	0.979	0.950	0.906	0.866	0.837
	평균강우	0.012	0.025	0.038	0.052	0.067	0.077	0.991	0.960	0.909	0.853	0.801	0.762
J-16	지점 강우	0.004	0.007	0.013	0.018	0.027	0.045	0.998	0.994	0.982	0.978	0.927	0.895
	평균강우	0.005	0.009	0.014	0.023	0.035	0.047	0.998	0.993	0.974	0.954	0.915	0.874
J-17	지점 강우	0.012	0.023	0.041	0.067	0.087	0.108	0.987	0.943	0.901	0.856	0.797	0.749
	평균강우	0.016	0.032	0.050	0.071	0.094	0.115	0.984	0.939	0.889	0.835	0.779	0.721

검증사상 4개의 수위예측 결과에 대해 평균제곱근오차, 효율계수, 평균절대오차, 수정 결정계수를 각각 계산하여 표 6에 나타내었다.

표에서 보는 바와 같이 평균제곱근오차는 30분 선행시간의 경우 4개의 검증 사상 모두에 대해 지점 강우를 이용한 수위예측 회귀모형의 결과가 0.008 m~0.044 m로 유역 평균 강우를 이용한 회귀모형의 결과인 0.011 m~0.049 m에 비해 작게 나타나고 있음을 알 수 있다. 예측 선행시간이 길어질수록 평균제곱근오차가 두 모형 모두 조금씩 증가하는 것을 확인할 수 있었으나 이 경우에도 지점 강우를 이용한 수위예측 회귀모형의 결과가 유역 평균 강우를 이용한 회귀모형에 비해 우수한 결과를 보이고 있음을 확인 할 수 있다. Nash-Sutcliffe 효율계수(NSEC)의 경우에도 모든 선행시간에 대해 지점 강우를 이용한 수위예측 회귀모형이 유역 평균 강우를 이용한 회귀모형에 비해 최소 0.001에서 최대 0.058의 높은 값을 나타내고 있음을 확인할 수 있었다. 평균절대오차와 수정 결정계수에 대한 두 모형간 결과를 비교해보면 예측 선행시간별로 다소 차이는 있으나 지점 강우를 이용한 수위예측 회귀모형이 유역 평균 강우를 이용한 회귀모형에 비해 우수한 예측 성능을 보이고 있음을 확인 할 수 있다.

6. 결 론

본 연구에서는 최근 수위 예측을 위해 널리 사용되고 있는 자료지향형 모형 중의 하나인 회귀모형의 입력 강우 형태에 따른 수위예측 회귀모형을 구축하고 모의 결과를 비교·검토하였다. 본 연구의 결과를 요약하면 다음과 같다.

- 1) 수위예측 회귀모형의 독립변수에 대한 시간 규모를 결정하기 위해 입력자료에 대한 자기상관분석을 수행하였다. 자기상관분석 결과를 통해 선행 예측 시간을 3시간, 즉 $t+6$ 으로 결정하고 자기상관계수가 0.75이상의 값을 갖도록 독립변수의 시간 규모를 결정하였으며 그 결과 선행 유출(수위)시간을 $t-4$ 로 결정하였다.
- 2) 지점 강우와 유역 평균 강우를 독립변수로 단계별 선택법을 통한 변수의 추가 및 제거를 통해 수위예측 회귀모형을 구축하였다. 이때 회귀식의 형태를 선정하기 위해 잔차와 독립변수간 산포도를 이용하였고, 그 결과 선행회귀식을 모형구축을 위해 사용하였다. 또한 각 모형의 구축단계에서 모형의 유의성을 검토하기 위해 개별회귀계수에 대한 t -검정을 수행하였고, 검정 결과 구축된 모형이 유의함을 확인할 수 있었다.
- 3) 구축된 모형을 중량천 유역의 다양한 홍수 사상에 대해 적용하였다. 두 모형의 수위예측에 대한 성능을 비교하기 위해 수위예측결과와 오차 등을 나타낸 그림을 제시하였고, 평균제곱근오차, Nash-Sutcliffe 효율계수, 평균절대오차, 조정 결정계수와 같은 4개의 통계지표들을 사용하였다. 모의결과 지점 강우를 이용한 수위예측 회귀모형이 유역 평균 강우를 이용한 회귀모형에 비해 좀 더 나은 예측 결과를 보이고 있음을 확인 할 수 있었다.
- 4) 이러한 결과는 지점 강우를 이용한 경우가 수위예측에

기여하는 최적의 독립변수(강우)의 공간 및 시간 분포를 결정할 수 있기 때문인 것으로 판단된다. 이에 반해 유역 평균 강우를 이용한 경우 강우의 공간 분포를 일정하게 가져가게 되고 또 이 경우 시간분포의 의미가 일정부분 왜곡됨으로써 예측 성능이 감소한 것으로 판단된다. 향후 연구에서는 좀 더 다양한 유역 평균 강우 산정 기법에 따른 검토와 함께 강우관측소의 밀도에 따른 연구가 필요할 것으로 판단된다.

감사의 글

본 연구는 국토해양부가 출연하고 한국건설교통기술평가원에서 위탁 시행한 건설기술혁신사업(08기술혁신F01)에 의한 차세대홍수방어기술개발연구단의 연구비 지원에 의해 수행되었습니다.

참고문헌

- 건설교통부 (2005) 주요지천 홍수예보 프로그램 개선.
 성지연, 허준행 (2009) 통계적 모형을 이용한 지천 홍수예측, **한국수자원학회 학술발표회논문집**, pp. 1524-1527.
 윤강훈, 김태균 (2003) 레이더 예측 강우를 이용한 다중회귀 유출 모형의 개발, **대한토목학회 학술발표회 논문집**, pp. 2187-2198.
 윤강훈, 신현민 (1994) 한강인도교 수위와 영향인자간의 다중회귀 분석에 의한 홍수위 예측모형, **한국수자원학회지**, 제27권, 제3호, pp. 55-69.
 윤용남, 원석연 (1991) 미계측 중소유역의 월유출량 산정을 위한 다중회귀모형 연구, **한국수자원학회지**, 제24권, 제3호, pp. 71-82.
 이범희 (2010) 회귀분석에 의한 도시홍수 예보시스템의 개발, **대한토목학회논문집**, 제30권, 제4-B호, pp. 347-359.
 정동국, 이범희 (2009) 회귀기법을 이용한 도시홍수위 예측모형의 개발, **한국수자원학회논문집**, 제43권, 제2호, pp. 221-231.
 Chaoulakou, A., Assimacopoulos, D. and Lekkas, T. (1999) Forecasting daily maximum ozone concentration in the Athens basin. *Environmental Monitoring and Assessment*, Vol. 56, pp. 97-112.
 Govindaraju, R.S. (2000) Artificial neural networks in hydrology :hydrologic applications. *Journal of Hydrologic Engineering*, Vol. 5, No. 2, pp. 124-137.
 Huynh Ngoc Phien and Nguyen Duc Anh Kha. (2003) "Flood forecasting for the upper reach of the Red River Basin, North Vietnam". <http://www.wrc.org.za>
 Kisi, O. (2010) Wavelet regression model for short-term streamflow forecasting. *Journal of Hydrology*, Vol. 389, pp. 344-353
 Kokkonen, T.S. and Jakeman, A.J. (2001) A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. *Water Resources Research*, Vol. 37, No. 9, pp. 2345-2353.
 McKerchar, A.I. and Delleur, J.W. (1974) Applications of seasonal parametric linear stochastic models to monthly flow data. *Water Resources Research*, Vol. 10, pp. 246-255.
 Phien, H.N., Huong, B.K. and Loi, P.D. (1990) Daily forecasting with regression analysis. *Water SA*, Vol. 16, No. 3, pp. 179-184.
 Tangborn, W.V. and Rasmussen, L.A. (1976) Hydrology of North Cascades region, Washington-part 2: a proposed hydrometeorological streamflow prediction method. *Water Resources*

- Research*, Vol. 12, pp. 203-216.
- VanderKwaak, J.E. and Loague, K. (2001) Hydrologic-response simulations for the R-5 catchment with a comprehensive physics-based model. *Water Resources Research*, Vol. 37, No. 4, pp. 999.
- Wu, C.L., Chau, K.W. and Li, Y.S. (2008) River stage prediction based on a distributed support vector regression. *Journal of Hydrology*, Vol. 358, pp. 96-111.
- Yurekli, K., Kurung, A. and Ozturk, F. (2005) Testing residuals of an ARIMA model on the Cekerek stream watershed in Turkey. *Turkish Journal of Engineering and Environmental Sciences*, Vol 29, pp. 61-74.

© 논문접수일 : 11년 08월 23일
© 심사의뢰일 : 11년 08월 31일
© 심사완료일 : 11년 09월 21일