**Journal of the Korean Society of Civil Engineers**
Vol. 33, No. 5: 1809-1827/ September, 2013
DOI: http://dx.doi.org/10.12652/Ksce.2013.33.5.1809

ISSN 1015-6348 (Print)
ISSN 2287-934X (Online)
www.kscejournal.or.kr

**Water Engineering** **수공학**

# 자료기반 실시간 홍수예측 모형의 비교·검토

최현구* · 한건연** · 노홍식*** · 박세진****
Choi, Hyun Gu*, Han, Kun Yeun**, Roh, Hong Sik***, Park, Se Jin****

# Comparison of Data-based Real-Time Flood Forecasting Model

**ABSTRACT**

Recently we need to take various measures to prepare for extreme flood that occur due to climate change. It is important that establish flood forecasting system to prepare flood over non-structure measures. The objective of this study is to develop superior real-time flood forecasting model by comparing the Neuro-fuzzy model and the multiple linear regression model. The Neuro-fuzzy model and the multiple linear regression model are established using same input data and applied for various flood events in Nakdong basin. The results show that the Neuro-fuzzy model can carry out flood forecasting results more accurately than the multiple linear regression model. This study can contribute to the establishment of a high accuracy flood information system that secure lead time in Nakdong basin.

**Key words :** Flood forecasting system, Real-time flood forecasting, Neuro-fuzzy model, Multiple linear regression model

**초 록**

기후변화로 인해 발생하는 이상홍수에 대비하기 위해서는 다양한 대책을 강구할 필요가 있다. 그 중 비구조적 대책으로 홍수예경보시스템을 구축하여 홍수에 대비할 수 있도록 하는 것이 중요하다. 본 연구의 목적은 실시간 홍수예측 시스템을 구축하기 위해 뉴로-퍼지 모형과 다중선형회귀 모형을 비교하여 우수한 실시간 홍수예측 모형을 개발하는데 있다. 이를 위해 같은 입력자료를 사용하여 뉴로-퍼지 모형과 다중선형회귀 모형을 구축하고 낙동강 유역의 다양한 홍수사상에 대해 적용하였다. 모의결과 뉴로-퍼지 모형이 다중선형회귀 모형보다 좀 더 나은 예측 결과를 나타내는 것을 확인할 수 있었다. 본 연구는 향후 낙동강 유역의 충분한 선행시간을 확보한 정확도 높은 홍수정보시스템의 구축에 활용할 수 있을 것으로 판단된다.

**검색어 :** 홍수예경보시스템, 실시간 홍수예측, 뉴로-퍼지 모형, 다중선형회귀 모형

## 1. 서 론

지구온난화를 유발하는 대기 중 온실가스는 산업혁명 이래 인간의 여러 활동에 기인하여 크게 증가하였다. 이러한 온실효과로 인한 기후변화의 영향으로 국지성 집중호우와 돌발성 호우가 국내 뿐 아니라 전 세계적으로도 많이 나타나고 있고, 그로 인한 이상홍수의 발생이 인명 및 재산피해를 날로 증가시키고 있는 추세이다. 그럼에도 불구하고 기존의 홍수대책의 경우 기존의 하천에 적용하는 홍수 예측 방법으로는 선행시간의 확보 등 여러 문제점으로 인해 한계가 있다. 홍수예측의 일반적인 목적은 홍수재해로 인한 인명 및 재산의 손실을 제거하거나 감소시키는 데 있다. 따라서 이러한 목적을 달성하기 위해서는 홍수의 예측의 결과를 신뢰성 높은 홍수

* 정회원 · 경북대학교 방재연구소 박사후연구원 (honorguard99@msn.com)
** 정회원 · 경북대학교 건축·토목공학부 교수 (kshanj@knu.ac.kr)
*** 경북대학교 건축·토목공학부 석사 (rhongs@naver.com)
**** 정회원 · 교신저자 · 경북대학교 건축·토목공학부 박사과정 (**Corresponding Author** · Kyungpook National University · sjin3615@naver.com)

**Received** February 15, 2013/ **revised** May 28, 2013/ **accepted** July 3, 2013

Copyright ⓒ 2013 by the Korean Society of Civil Engineers
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

예보 및 경보에서 정확하게 반영할 수 있어야하며 또한 충분한 선행시간의 확보가 필요하다. 홍수예측에서 예측결과는 정확해야 하며 사람들이 다가올 홍수에 준비할 수 있도록 충분한 예측 선행시간이 확보되어야 한다. 그러나 일반적으로 긴 예보선행시간을 확보하고자 할 경우 예측의 정확도가 떨어지게 되므로 유역의 특성 및 예보 목적을 고려하여 적용할 방법을 결정해야 한다.

현재 우리나라 홍수 예·경보 시스템은 홍수 예·경보를 위해서 강우 분석 및 예측의 과정과 홍수예측의 과정을 거치게 된다. 경우에 따라서는 강우 분석 및 예측의 과정 없이 실시간으로 전송되는 강우자료만을 이용하여 홍수예측을 수행하기도 한다. 이러한 예·경보 과정에서 몇 가지 문제점이 발생하는데 그 중의 하나가 강우분석 및 예측의 부분에서 발생한다. 현재 대부분의 경우 강우예측을 기상청에서 제공하는 수치예보자료(Regional Data Assimilation and Prediction System)에 의존하고 있다. 하류지역 주요 예보지점의 정확한 수위를 예측하기 위해서 상류유역의 정확한 강우예측자료는 반드시 필요하다. 하지만, 현재의 기술로는 정확한 강우예측이 어려운 것이 사실이며, 홍수 예측의 정확도는 강우-유출 모형의 정확도 못지 않게 입력된 예측 강우량의 정확도에 의존하게 된다.

홍수 예측과정에서 행해지는 기존의 수문학적 모형인 강우-유출 모형으로는 입력자료의 처리 등으로 인해 충분한 선행시간을 확보할 수 없다는 문제점이 있다. 또한 기존의 홍수예측 시스템에서는 지점 예보 위주의 점 단위 홍수 예·경보가 이루어지고 있는데 활용성이 점점 높아지는 하천 홍수터에서의 홍수 피해에 효율적으로 대비하고, 신속한 주민대피 계획 수립을 통한 인명 및 재산 피해를 최소화하는데 한계가 있다. 따라서 본 연구에서는 충분한 선행시간을 확보하기 위해 낙동강 유역에서의 실시간 홍수예측 시스템을 구축하는데 있어 자료기반 모형들을 이용해 접근해 보았다.

본 연구에서 사용한 뉴로-퍼지 기법을 이용한 국외 연구는 강우-유출모의(Smith and Eli, 1995; Jain et al., 2004), 하천유량예측(Imrie et al., 2000; Yarar et al., 2009), 강우예측(Luk et al., 2001; Ramirez et al., 2005) 등과 같이 다양한 분야에 적용되었다. Gautam and Holz(2001)는 적응형 뉴로-퍼지 기법을 강우-유출해석에 적용하였고, Nayak et al.(2005)은 3년간의 호우자료를 이용하여 뉴로-퍼지, 인공신경망, 퍼지 추론시스템을 이용하여 유량을 예측하였다. Kurtulus and Razack(2009)는 유출 응답에 대한 신경망과 뉴로-퍼지 모형간의 비교 연구를 수행하였으며 Talei et al.(2010)은 뉴로-퍼지 추론 시스템을 사용해 강우-유출 모델링에 적용하였고 Lohani et al.(2012)은 수문 시계열 모델링에 있어 뉴로-퍼지와 자기회귀를 비교 하였다.

국내에서도 최근 뉴로-퍼지를 수자원분야에 적용하기 위한 연구가 활발히 진행되고 있는데, 고영준(2001)은 하천의 시유출량 예측을 위해 퍼지-신경망을 적용하여 적용성을 검증하였고, 신성일(2002)은 1997~2002년 발생한 홍수사상에 대해서 금호강 중류부에 위치한 금호수위표지점에서 관측된 자료를 사용하여 홍수량 예측을 실시하였다. 또한, 최승용(2011)은 뉴로-퍼지를 이용해 중소하천에서의 수위예측과 나아가 하천에서의 홍수와 범람해석을 연계한 통합 시스템을 구축하였다.

또한 수자원 분야에 있어 회귀모형을 이용한 국외 연구로는 수문학적 추정을 위한 적용성(Chaloulakou et al., 1999; Yurekli et al., 2005), 홍수예측(Tangborn and Rasmussen, 1976; Schilling and Wolter, 2005) 등과 같은 분야에 적용되었다. Wu et al.(2008)은 자료분류 기법 중의 하나인 support vector machine(SVM)과 회귀식을 결합한 수위예측 모형을 개발하여 수위 예측에 적용하였다. 국내에서의 회귀모형에 관한 연구로는 윤용남과 원석연(1991)이 미계측 중소유역의 월유출량 산정을 위해 다중회귀모형을 사용하였고, 성지연과 허준행(2009)은 다중선형 회귀모형의 다중공선성 문제를 해결하기 위해 독립변수를 조정하고 회귀계수를 재산정한 통계적 모형을 이용하여 수위를 예측하였다.

실시간 홍수 예측에 관한 국외 연구로 근래에는 강우 자료를 분포형 모형에 적용하는 연구가 활발하게 수행 중이며(Carpenter et al., 2001), 일반적으로 확정론적인 모형이 강우-유출 관계에 사용되어 왔으나, 최근 들어 추계학적인 모형 및 신경망 모형이 자주 사용되기도 하며, 이들의 예측성과가 확정론적 모형의 결과와 비교되기도 하였다. 실시간 홍수예측에 관한 국내 연구로는 김경탁 등(2006)이 통계학적 모형을 이용하여 한강수계의 주요 지천에 대해 홍수예측을 실시한바 있다. 성지연 등(2009)은 통계적 모형을 이용하여 지천의 홍수예측을 실시하였으며, 정동국 등(2009)은 대전광역시의 도심하천 구간에서 현재까지의 강우, 수위 및 상류지역의 수위를 기반으로 수위를 예측할 수 있는 회귀모형을 구성하여 2시간 후의 수위를 예측하였다.

현재까지 뉴로-퍼지를 이용하여 주요 지점별 홍수위를 예측한 연구는 있으나 뉴로-퍼지 모형과 다중선형회귀 모형의 결과를 비교하고 그에 따른 하천해석의 비교를 통해 홍수방재에 직접적으로 적용한 사례는 없다. 따라서 본 연구는 뉴로-퍼지와 다중선형회귀 모형을 이용한 낙동강에서의 지점 홍수 예측에서부터 하도에서의 구간 홍수위 예측을 통해 실시간 홍수해석을 통합적으로 수행할 수 있는 시스템을 구축하고자 한다.

## 2. 실시간 홍수예측 모형의 구축

본 연구에서는 기존 물리적 모형을 기반으로 한 홍수예측이 가지는 한계점을 극복하고 실시간 홍수예측을 위한 홍수예측 모형을 구축하였다. 실시간 강우 및 수위, 방류량 자료만을 홍수예측을 위한 입력 자료로 사용하였으며 이를 통해 입력 자료 처리 등의

시간을 대폭 줄임으로써 충분한 선행시간을 확보하고자 하였다. 또한 실시간 홍수예측의 과정에서도 기존의 수문학적 기반의 강우-유출 모형의 경우 여러 단계의 추적과정에서 오차가 발생하고 또 계산 소요시간이 오래 걸렸으나 본 연구에서는 MATLAB 기반의 뉴로-퍼지 모형과 다중선형회귀 모형을 이용한 홍수예측을 수행함으로써 예측소요 시간이 짧고 정확한 실시간 홍수예측 모형을 구축하였다.

## 2.1 뉴로-퍼지 모형

퍼지 이론과 신경망을 결합한 뉴로-퍼지 모형의 기본 이론은 Jang(1992)에 의해 최초로 제안되었으며 그 이후 Brown and Harris(1994)에 의해 오늘날까지 연구 발전돼오고 있다. 신경망 이론과 퍼지 이론은 모두 특정 분야에 대해서 인간처럼 일을 할 수 있는 시스템을 구축하는 것에 관심이 있으나, 두 이론이 강점을 갖는 분야는 조금 다르다. 퍼지 이론은 논리적인 기반 하에서 애매한 정보를 처리, 추론하는데 적합하고 자연 언어적인 표현을 통해 고차원적인 처리에 유용하다. 반면 신경망 이론은 학습 능력을 가지고 있어서 시스템의 구성에 유연성이 크며, 데이터 기반의 처리에 우수하다. 이러한 장점들을 바탕으로 제시된 모형이 뉴로-퍼지 모형으로 신경망의 구조와 학습능력을 이용하여 제어환경으로부터 얻은 입출력 정보로부터 소속 함수와 규칙을 제어 대상에 맞게 자동으로 조정함으로써 수행된다.

### 2.1.1 모형의 구조

본 연구에서 사용된 뉴로-퍼지 모형의 구조를 설명하기 위해 퍼지 추론 시스템이 2개의 입력 값과 한 개의 출력 값을 갖는다고 가정하면 1차 Takagi-Sugeno 퍼지 모형은 다음과 같은 2개의 규칙을 갖는다.

Rule 1 : If $x$ is $A_1$ and $y$ is $B_1$, then $f_1 = p_1x + q_1y + r_1$
Rule 2 : If $x$ is $A_2$ and $y$ is $B_2$, then $f_2 = p_2x + q_2y + r_2$

Fig. 1은 1차 Takagi-Sugeno 추론 시스템을 나타낸 것이고 Fig. 2는 신경망을 도입한 뉴로-퍼지 모형의 구조이다.

뉴로-퍼지의 구조는 $O_{l,i}$로 표현되는 각 층(layer) 내에서 노드 출력을 갖는 5개의 층으로 구성되어 있다. 여기서 $O_{l,i}$는 층 $l$의 $i$번째 노드의 출력 값을 나타낸다. Fig. 2의 뉴로-퍼지 구조의 각 층에 대해 자세히 살펴보면 다음과 같다.

Layer 1 : 이 층의 모든 노드는 입력 노드로 단지 다음 층으로 외부적 신호만 보낸다.

Layer 2 : 이층의 모든 노드 $i$는 적응형 노드로 소속 함수로 작용한다.

$$O_{2,i} = \mu_{A_i}(x), i = 1,2$$
$$O_{2,i} = \mu_{B_{i-2}}(x), i = 3,4 \quad (1)$$

여기서 $x$ 또는 $y$는 입력 노드 $i$이고, $A_i$ 또는 $B_{i-2}$는 이 노드에 관련된 퍼지 집합(언어적 표시)이다. $O_{2,i}$는 퍼지 집합($A_1$, $A_2$, $B_1$, $B_2$)의 소속 정도를 나타내고 이의 전형적인 소속 함수는 Eq. (2)와 같다.

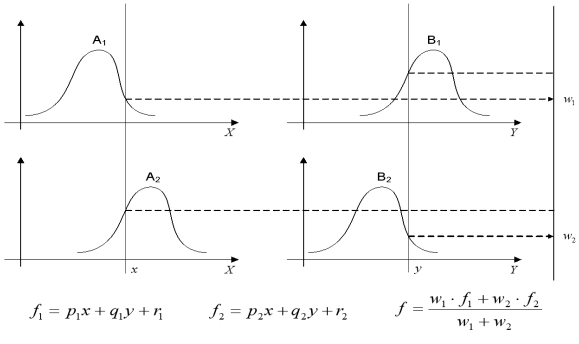
$$\mu_{A(x)} = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}} \quad (2)$$

여기서 $\{a_i, b_i, c_i\}$를 전건부 매개변수라 한다. 이 함수에서 매개변수 값이 변화함에 따라 퍼지 집합 $A$의 소속 함수는 다양하게 표현될 수 있다. 즉, 위의 소속 함수에서 $a_i$, $b_i$, $c_i$의 값을 변화시킴으로써 소속 함수를 사용자가 원하는 다양한 형태로 나타낼 수 있다는 것이다.

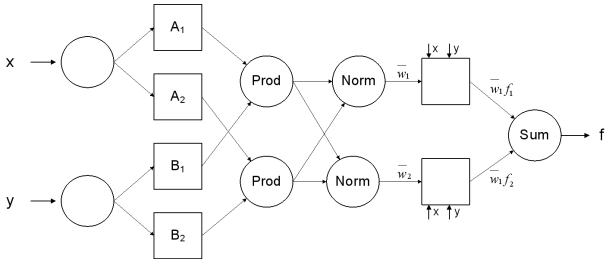
Layer 3 : 이층의 모든 노드는 Prod로 표시된 고정노드이며 출력은 모든 입력 신호의 곱으로 Eq. (3)과 같다.

$$O_{3,i} = w_i = \mu_{A_i(x)} \cdot \mu_{B_i(y)}, \quad i = 1,2 \quad (3)$$

각 노드의 출력 값은 규칙의 연결강도를 나타낸다.




Flg. 1. First-order Takagi-Sugeno Fuzzy Inference Scheme




Flg. 2. Structure of Neuro-Fuzzy Model

Layer 4 : 이층의 모든 노드는 Norm이라 표시된 고정 노드이며 $i$번째 노드와 관련된 규칙의 정규화된 연결강도를 나타낸다. 즉, $i$번째 노드는 전체 연결강도에 대한 $i$번째 노드의 연결강도의 비로 Eq. (4)와 같다.

$$O_{4,i} = \overline{w_i} = \frac{w_i}{\sum w_i} \quad i = 1, 2 \tag{4}$$

Layer 5 : 이층의 모든 노드는 노드 함수를 갖는 적응형 노드로 Eq. (5)와 같이 가중된 후건부 값을 계산한다.

$$O_{5,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{5}$$

여기서 $\overline{w_i}$는 층 4에서의 출력 값이고, $\{p_i, q_i, r_i\}$는 이 노드의 후건부 매개변수이다.

Layer 6 : 이 층의 노드는 오직 하나로 합으로 표시되며, 전체 시스템에 대한 추론 결과를 계산하기 위해 Eq. (6)과 같이 모든 입력 신호의 합으로 표현된다.

$$O_{6,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{6}$$

뉴로-퍼지의 최종 출력 값을 나타내는 Eq. (7)은 다음과 같이 후건부 파라미터의 선형조합으로도 나타낼 수 있다.

$$\begin{aligned} f &= \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \\ &= \overline{w_1}(p_1 x + q_1 y + r_1) + \overline{w_2}(p_2 x + q_2 y + r_2) \\ &= (\overline{w_1} p_1 + \overline{w_2} p_2) x + (\overline{w_1} q_1 + \overline{w_2} q_2) y + (\overline{w_1} r_1 + \overline{w_2} r_2) \end{aligned} \tag{7}$$

## 2.2 다중선형회귀 모형

회귀분석이란 둘 또는 그 이상의 변수들 간의 관계를 파악함으로써 어떤 특정한 변수(종속변수)의 값을 다른 한 개 또는 그 이상의 변수(독립변수)들로부터 설명하고 예측하는 통계적 기법이다. 이 분석방법에서는 먼저 변수들 간의 관계를 나타내는 타당한 수학적 모형을 이론적 근거나 경험에 바탕하여 설정하고, 변수들의 관측된 값을 이용하여 그 모형을 추정한 다음, 추정한 모형에 의해 변수들 간의 관계를 설명하든지 또는 예측 등의 분석에 응용하게 된다.

회귀분석은 상관관계의 연관성과 인과모형의 인과성을 종합한 개념으로 정리할 수 있다. 또한 계량적 종속변수와 하나 혹은 그이상의 독립변수들 간의 관련성을 분석하는데 있어 매우 강력한 분석력을 갖고 있으며, 적응성이 뛰어난 특성을 가지고 있다. 분석에 사용되는 변수는 크게 독립변수와 종속변수로 구분되는데 하나의 독립변수와 하나의 종속변수를 포함하는 가장 단순한 회귀분석 형태를 단순회귀분석이라고 하고 여러 개의 독립변수가 있는 회귀분석을 다중회귀분석이라고 한다. 회귀분석의 일반적인 형태는 1차 방정식의 함수관계로 나타난다.

### 2.2.1 회귀계수 추정기법

다중회귀모형의 기본이론은 다음과 같다. 반응변수 $Y$와 $p-1$개의 설명변수 $X_1, X_2, \ldots, X_{p-1}$가 다음과 같은 관계식을 가진다고 가정한다.

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} \tag{8}$$

즉, 반응변수 $Y$의 기댓값은 각 설명변수들의 일차함수로 나타내어진다. 위의 식에 개개의 관측점 $i = 1, 2, \cdots, n$에서 발생하는 오차를 나타내는 확률변수 $\varepsilon$을 더하면 다음의 모집단 다중회귀모형을 얻는다.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i \,, i = 1, 2, \cdots, n \tag{9}$$

확률변수 $\varepsilon$에 대한 가정은 단순회귀에서의 가정들과 같다. 즉, $\varepsilon_i$는 평균 0, 분산 $\sigma^2$을 가지며, 서로 독립이라 가정한다. 그리고 위의 식에서 $X_{ij}$는 $j$번째 설명변수 $X_j$의 $i$번째 관측치를 의미한다. 설명변수 $X_j$들은 측정오차가 없거나 또는 무시할 수 있을 정도로 작은 상수들이다. 회귀계수들에 대한 해석도 단순회귀모형에서와 비슷한데, 이를테면 $\beta_0$는 $Y$축의 절편, $\beta_j$는 $Y$와 $X_j$간의 기울기로 다른 설명변수들의 값들이 고정되었을 때 $X_j$가 한 단위 증가하면 $Y$의 평균값은 $\beta_j$만큼 변하게 된다. 다중선형회귀모형은 행렬로 표현함으로써 다음과 같은 간단한 형태로 나타낼 수 있다.

$$Y = X\beta + \varepsilon \tag{10}$$

### 2.2.2 단계별 선택법(Stepwise)

다중회귀모형의 분석에서 적절한 독립변수의 선택은 변수의 절약 측면에서 매우 유용한 방법이다. 이와 더불어 불필요한 변수를 모형에 포함시킴으로서 다중공선성과 같은 부수적인 문제점이 발생할 여지가 있으므로 종속변수와 가장 밀접한 연관성이 있는 몇 개의 변수만을 선택하여 이를 회귀모형에 포함시키는 방법이 변수 선택법이다. 본 연구에서는 다중선형회귀모형의 회귀계수들

의 추정 방법 중 단계별 선택법을 적용하였다. 변수를 선택하는 방법에는 다음과 같은 방법들이 있다.

① 모든 가능한 조합의 독립변수들에 대한 회귀분석을 수행하여 가장 적합한 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 회귀 모형을 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.

#### 2.2.2.1 수정 결정계수($R^{*2}$)

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

여기서 $n$은 자료의 수, $k$는 독립변수의 수, $SSE_k$ 는 독립변수가 $k$ 인 축소모형의 $SSE$ 이다. 독립변수가 추가되면 $R^2$ 값은 언제나 증가하게 된다. 따라서 독립변수의 개수가 다른 두 모형에 대해서는 $R^2$ 값의 비교만으로는 불충분하게 된다. 이를 보완한 것이 수정결정계수이다.

#### 2.2.2.2 Mallow의 $C_p$

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

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

#### 2.2.2.3 Akaike Information Criterion

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

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. 실시간 홍수예측 모형의 비교

### 3.1 대상유역

실시간 홍수예측 모형인 뉴로-퍼지 모형과 다중선형회귀 모형의 비교를 위해 낙동강 주요 8개 지류 중 밀양강과 금호강 유역에 대해 각각 모형을 구축하고 적용하였다. Fig. 3은 대상유역의 유역도를 나타내고 있다.

밀양강은 낙동강 제1지류로 낙동강수계 동남부에 위치하며, 유역면적은 1,421.26㎢로서 낙동강 유역의 약 6%이고, 유로연장은 101.5㎞이다. 유역형상계수는 0.138이며 평균경사는 20.36%로 유역이 대체적으로 완만한 경사를 보이고 있다. 본 연구에서는 시험유역 내 우량관측소를 내진, 밀양2, 산내(밀양), 청도, 용성, 산내(경주), 수월, 선리, 밀양댐, 운문, 운문댐, 삼랑진관측소의 12개소를 티센을 이용한 평균 강우량을 산정하기 위해 고려하였다. 수위관측소는 예측지점인 밀양2 수위관측소를 고려하였다.

금호강은 경산시, 대구광역시를 지나 낙동강 중류부 좌안(낙동강 하구로부터 168.49㎞ 지점)으로 합류되는 유로연장은 116.0㎞의 낙동강 제1지류 하천이다. 금호강의 유역면적은 2,092.42㎢로서 낙동강 유역면적 약 9%이며, 유역형상계수가 0.16으로 직사각형 형태의 유역형상을 지니고 있고, 평균경사는 27.39%로 상류지역을 제외하고는 유역이 대체적으로 완만한 경사를 보이고 있다. 유역의 토양형은 대체적으로 배수가 양호한 토양으로 이루어져 있고 유역

내에는 농업지역과 산림지역이 84%이상을 차지하고 있다. 시험유역 내 우량관측소는 죽장, 신령, 영천, 대구, 영대, 자인, 현서, 장천, 왜관, 기계1, 석산, 화북2, 가창, 용성, 청도, 대율관측소의 16개소가 운영되고 있으나 본 연구에서는 현서, 장천, 기계1, 왜관관측소는 배재하였고 티센을 이용한 평균 강우량을 산정하기 위해 동명, 서부, 고경, 와촌, 동촌, 산내(경주), 성주관측소를 고려하였다. 수위관측소는 성서 수위관측소를 고려하였다.

### 3.2 실시간 홍수예측 모형 구축

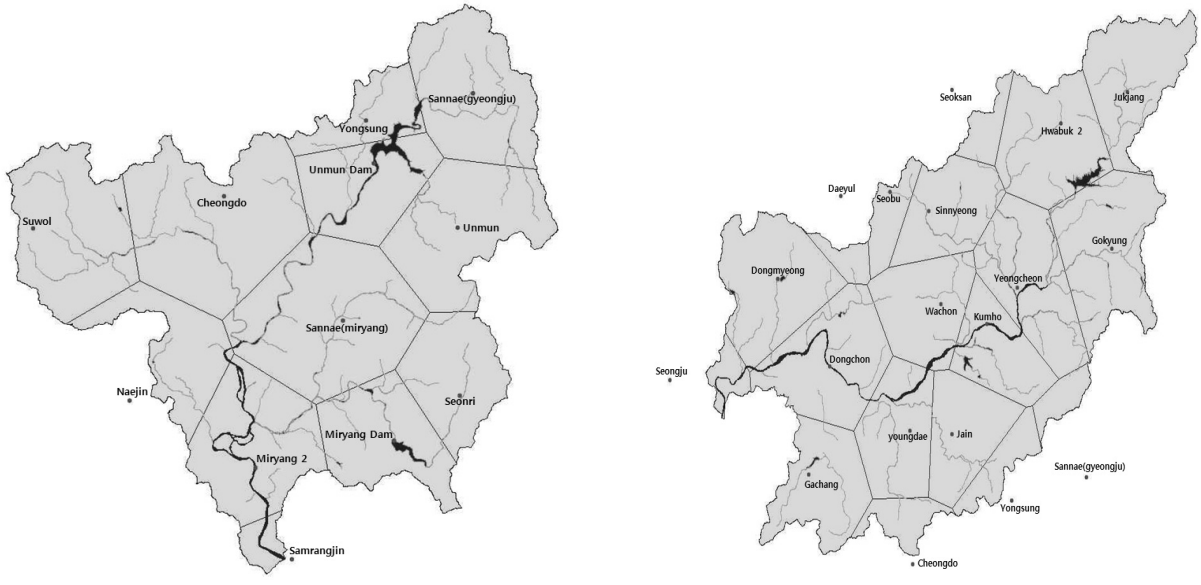
데이터 기반의 모형을 이용한 홍수예측에 있어서 중요한 사항은 수위에 영향을 미치는 강우, 수위, 유량, 방류량 자료의 시간적 분포를 입력자료 조합에 어떻게 구성할 것인지 또 얼마만큼의 자료를 사용할 것인지를 결정하는 것이다. 이에 본 연구에서는 입력자료 조합을 구성하여 최적 입력자료 조합을 선정하였으며 이를 이용하여 금호강, 밀양강 유역에서의 실시간 홍수예측 시스템을 구축해 보았다.

#### 3.2.1 최적 입력자료 조합의 선정

본 연구에서는 데이터들을 어떻게 구성하는 것이 수위(유량) 예측의 정확성이 가장 우수한지에 대한 검토를 통하여 데이터기반 모형을 이용한 홍수예측의 최적 입력자료의 조합을 선정하고자 하였다.

최적 입력자료 조합의 선정과 예측을 위해 사용한 호우사상의 특징은 Table 1에 나타내었으며 추후 모의에 대한 설명에 이해를 돕고 표기를 간단하게 하기 위해 각 호우사상에 대해 이름을 임의로 정하였다.

실시간 홍수예보의 관점에서 입력 자료의 처리 시간을 줄이기 위해 입력자료의 수는 5개~6개로 제한하고 입력 자료 조합을 구성하였다. 최적의 입력자료 조합을 선정하기 위해 강우자료와 수위자료 그리고 댐 방류량 30분 자료를 이용하여 Table 2와 같이 6개의 입력자료를 구성하였다. 구축된 입력자료와 N-3 사상을 이용하여 180분의 예측선행시간을 이용하여 예측을 수행하였으며 결과를 Table 3에 정량적으로 제시하였다. 선행 시간이 짧을수록 예측의




(a) Miryang (b) Kumho

Fig. 3. Study Area

Table 1. Character of Events for Selection of Optimal Input Data Combination

| Naming | Period | Total Rainfall (mm) | Rainfall Intensity (mm/hr) | Rainfall Type | Purpose |
|---|---|---|---|---|---|
| N-1 | 02/8/30 14:30～9/1 05:00 | 202.3 | 33.5 | Typhoon | Training |
| N-2 | 03/9/11 08:30～9/13 01:00 | 252.1 | 39.6 | Typhoon | Checking |
| N-3 | 06/7/08 14:30～7/12 23:30 | 352.8 | 43.9 | Typhoon | Testing |
| N-4 | 07/9/14 14:00～9/18 06:00 | 219.3 | 52.0 | Typhoon | Testing |
| N-5 | 11/7/06 12:30～7/11 14:00 | 423.5 | 34.8 | Rainy Season | Testing |
| N-6 | 12/9/13 22:30～9/17 17:00 | 196.9 | 29.7 | Typhoon | Testing |

Table 2. Input Data Combination

| Case | Naming | Input Data | | |
|---|---|---|---|---|
| | | Rainfall | Discharge | Stage |
| Case 1 | R0_H0123 | R(t) | - | H(t), H(t-1), H(t-2), H(t-3) |
| Case 2 | R01_H012 | R(t), R(t-1) | - | H(t), H(t-1), H(t-2) |
| Case 3 | R01_Q0_H012 | R(t), R(t-1) | Q(t) | H(t), H(t-1), H(t-2) |
| Case 4 | R012_Q0_H01 | R(t), R(t-1), R(t-2) | Q(t) | H(t), H(t-1) |
| Case 5 | R012_H01 | R(t), R(t-1), R(t-2) | - | H(t), H(t-1) |
| Case 6 | R0123_H0 | R(t), R(t-1), R(t-2), R(t-3) | - | H(t) |

Table 3. Statistics Results for Selection of Optimal Input Data Combination(N-3)

| | RMSE | CC | NSEC | RPE | RAD | MAD |
|---|---|---|---|---|---|---|
| R0_H0123 | 0.1747 | 0.9975 | 0.9943 | 3.0064 | 0.0195 | 0.0930 |
| R01_H012 | 0.1304 | 0.9987 | 0.9968 | 3.1717 | 0.0164 | 0.0782 |
| R01_Q0_H012 | 0.2207 | 0.9968 | 0.9910 | 4.4732 | 0.0325 | 0.1549 |
| R012_Q0_H01 | 0.2308 | 0.9966 | 0.9901 | 4.8627 | 0.0346 | 0.1646 |
| R012_H01 | 0.2162 | 0.9962 | 0.9913 | 11.3079 | 0.0204 | 0.0971 |
| R0123_H0 | 0.4162 | 0.9847 | 0.9679 | 0.9850 | 0.0532 | 0.2536 |

정확도가 향상 될 것은 분명하나 입력자료 조합 구성의 목적이 어떤 조합이 가장 정확도가 높은가를 보기위한 것이므로 선행시간의 길이에 따른 비교는 무의미하기 때문이다.

정량적 평가를 통한 최적 자료 조합의 선정을 위해 본 연구에서는 각 입력 조합에 따른 관측수위와 예측수위의 평균제곱근오차(Root Mean Square Error), 상관계수(Correlation Coefficient), Nash-Sutcliffe 효율계수(Efficiency Coefficient), 상대첨두오차(Peak Relative Error), 평균절대편차(Mean Absolute Deviation), 평균절대편차와 평균값 사이의 비인 RAD를 통계지표로 사용하였다. 사용된 통계지표를 간단히 설명하면 다음과 같다.

#### 3.2.1.1 평균제곱근오차(RMSE)

$$\mathrm{SE} = \sqrt{\frac{\mathrm{SSE}}{\mathrm{N}}} \tag{14}$$

여기서 SSE는 오차제곱의 총합이고 N은 자료의 수이다.

$$SSE = \sum_{i=1}^{N}(y_i - \hat{y_i})^2 = \sum_{i=1}^{N} e_i^2 \tag{15}$$

여기서 $y_i$ 와 $\hat{y_i}$는 관측 값과 예측 값이고, $e_i$ 는 관측 값과 예측 값 사이의 오차이다.

#### 3.2.1.2 상관계수(CC)

2개의 변수 $x$ 와 $y$ 가 있을 때에 $x$ 의 변화에 따라서 $y$ 도 변화하는 관계를 상관관계라고 한다. $x$ 가 증가하면 $y$ 도 증가하는 관계를 양의 상관관계, $x$ 가 증가하면 $y$ 가 감소하는 관계를 음의 상관관계라고 한다. 어느 쪽의 관계도 보이지 않는 경우를 무상관이라 한다.

상관계수는 2개의 변수 사이에 상관관계가 여부를 수치적으로 판단하는 지표로 이용한다. 상관계수는 통상 $r$ 이라는 기호로 표시하며 -1.0에서 1.0까지의 값을 가지고 두 데이터 집합 사이의 상관성을 나타낸다. 상관관계의 강도는 상관계수의 절대값 $|r|$ 또는 자승값 $r^2$ 으로 평가한다. 어느 쪽도 1에 가까울수록 상관이 강하다는 것을 의미한다. 상관계수의 크기가 0에 가까우면 두 데이터 집단간의 관련성이 거의 없다는 것으로 해석한다. 상관계수 $r$ 은 다음과 같은 순서로 산출한다.

$$r = \frac{S(xy)}{\sqrt{S(xx)S(yy)}} \tag{16}$$

여기서, $S(xx)$ 는 $x$ 의 편차제곱의 합이고, $S(yy)$ 는 $y$ 의 편차제곱의 합이며, $S(xy)$ 는 $x$ 와 $y$ 의 편차제곱의 합이다.

#### 3.2.1.3 Nash-Sutcliffe 효율계수(Efficiency Coefficient)

$$NSEC = \frac{ST - SSE}{ST} \tag{17}$$

$$ST=\sum_{i=1}^{N}(y_i-\bar{y})^2 \tag{18}$$

$$\bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i \tag{19}$$

여기서 ST는 통편차이고 $\bar{y}$ 는 $y$ 값 N개의 평균값이다.

#### 3.2.1.4 상대첨두오차(Relative Peak Error)

$$RPE=\frac{|\widehat{Q_p}-Q_p|}{Q_p}\times 100(\%) \tag{20}$$

여기서 $\widehat{Q_p}$ 는 예측 첨두유출량이고, $Q_p$ 는 관측 첨두유출량이다.

#### 3.2.1.5 평균절대편차(MAD)

$$MAD=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y_i}| \tag{21}$$

#### 3.2.1.6 RAD

$$RAD=\frac{MAD}{\bar{y}} \tag{22}$$

Table 3에서 보는 바와 같이 입력자료로 강우자료가 많이 사용된 R0123_H0 입력 자료 조합을 제외한 나머지 입력 자료 조합들의 평균제곱근오차는 0.2m 내외의 값을 보였다. 상관계수에서도 R0123_H0 입력 자료 조합을 제외한 나머지 입력 자료 조합들은 0.99이상의 높은 상관관계를 보임을 확인할 수 있었다. 나머지 다른 통계 지표들의 경우에도 R0123_H0 입력 자료 조합을 제외한 나머지 입력 자료 조합들 간의 차이는 크게 나지 않음을 알 수 있었다. 한 가지 특이한 점은 R0123_H0 조합의 RPE 값이 가장 좋은 결과를 냈다는 것인데 피크구간을 가장 정확하게 예측한 것이 아니라, 전체 예측 값 중에서 가장 높은 값과 실측값에 피크값과의 차이가 가장 적다는 의미로 해석 할 수 있다.

R01_H012의 입력 자료 조합이 나머지 입력 자료 조합에 비해 평균제곱근오차의 경우에 약 31%~74% 정도로 작게 나타났고 다른 모든 통계지표에서 다른 조합들 보다 우수한 결과를 나타냈다. 이와 같은 결과를 통해 본 연구에서는 R01_H012의 입력 자료 조합이 가장 정확한 수위 예측을 하는 것으로 판단하였고 향후 실시간 홍수 예측을 위한 최적의 입력 자료로 선정하였다. 즉 t 시간의 강우, t-1 시간의 강우, t 시간의 수위, t-1 시간의 수위, t-2 시간의 수위를 최적 입력 자료로 구축하였으며, 강우와 수위는 대표강우와 수위관측소의 시계열자료를 입력자료로 사용하였다.

#### 3.2.2 뉴로-퍼지 모형 구축

실시간 홍수예측 시스템을 구축하기 위해서 Table 1에 명시한 최적 입력자료 조합에서 선정한 6개의 강우사상 기간을 대상유역에 적용하였고 유역별로 티센을 이용해 평균 강우량을 각각 산정하였다. 뉴로-퍼지 모형의 적용성을 검토하기 위해서는 훈련, 확인, 검증의 세 단계를 거치게 되는데, 이 중 훈련 과정과 확인 과정을 통해 소속 함수의 매개변수를 결정하게 된다. 이 때 훈련과 확인 과정에 사용하는 호우 사상에 따라 소속 함수의 매개변수 값이 달라지게 되며 뉴로-퍼지 모형의 결과 또한 차이가 나게 된다. 본 연구에서는 앞서 선정된 6개의 호우사상을 이용해 홍수예측 모형을 구축하는데 있어 과거 사상들을 이용해 미래 사상을 예측하는 관점에 비추어 N-1, N-2 사상을 훈련 및 확인을 위해 사용하였고, N-3사상은 최적입력자료 조합에 사용을 하였으며, N-4, N-5, N-6 사상을 검증에 사용하였다. 이렇게 함으로써 과거사상을 이용해 구축한 모형이 얼마나 미래에 발생할 홍수를 잘 예측할 수 있는지 검증 할 수 있었다.

모의 시간의 단축 및 최적의 입력 공간 분할을 위해서 subtractive clustering 기법을 적용하여 각 입력 변수당 2개의 소속 함수와 2개의 규칙을 이용한 뉴로-퍼지 모형을 구축하였다. 또한 입력자료(수위와 강우의 조합)와 출력 자료(수위)의 비선형성을 고려하여 소속 함수의 형태로는 가우스 함수를 사용하였으며 홍수예측의 선행시간은 30, 60, 90, 120, 150, 180분의 6가지 경우로 하여 각 선행시간별 소속 함수의 매개변수를 결정하였다.

#### 3.2.3 다중선형회귀 모형 구축

본 연구에서는 독립변수가 종속변수의 예측에 기여하는지를 판단하여 독립변수의 추가와 제거를 수행하였고 이를 통해 다중선형회귀 홍수예측 모형을 구축하였다. 독립변수의 시간규모는 뉴로-퍼지 모형과의 비교를 위해 같은 입력 자료를 사용하였다. 즉, t시간의 강우, t-1시간의 강우, t시간의 수위, t-1시간의 수위, t-2시간의 수위를 입력 자료로 구축하였다. 두 모형간의 결과를 비교하기 위해 독립변수 뿐 아니라 예측 지점, 강우사상 등 모든 조건을 동일하게 하였다. Table 4와 5는 다중선형회귀 모형에 적용한 밀양강과 금호강유역의 각 선행 예측시간별 홍수예측 모형의 회귀계수 산정결과와 각 독립변수의 개별회귀계수 t-검정 값을 나타내고 있다. 표에서 보는 바와 같이 30분 예측 모형의 경우 독립변수의 추가 및 제거과정을 통해 최종적으로 R(t)의 독립변수들이 제거된 상태로 홍수예측 모형이 구축되었고 60분 예측 모형의 경우에는 R(t-1)의 독립변수들이 제거된 상태로 홍수예측 모형이 구축되었다. Table 4에서보는 바와 같이 밀양강 유역에서의 30분 후의

**Table 4.** Established Result of Multiple Linear Regression model(miryang)

| | t=30min Model | | t=60min Model | | t=90min Model | |
|---|---|---|---|---|---|---|
| | Coefficient | t-testing | Coefficient | t-testing | Coefficient | t-testing |
| Intercept | -0.00847 | -5.0810 | -0.01947 | -7.8826 | -0.03735 | -2.6082 |
| R(t) | | | 0.03432 | 13.8766 | 0.03052 | 2.1307 |
| R(t-1) | 0.01541 | 9.2475 | | | 0.03640 | 2.3237 |
| H(t) | 1.32495 | 70.8781 | 1.61687 | 58.5314 | 1.91162 | 15.5500 |
| H(t-1) | | | | | -0.40036 | -1.7108 |
| H(t-2) | -0.32531 | -17.2870 | -0.61756 | -22.1925 | -0.51269 | -3.8798 |
| | t=120min Model | | t=150min Model | | t=180min Model | |
| | Coefficient | t-testing | Coefficient | t-testing | Coefficient | t-testing |
| Intercept | -0.05789 | -10.6848 | -0.08096 | -4.2167 | -0.10309 | -4.6022 |
| R(t) | | | 0.04944 | 2.5806 | 0.08921 | 3.9815 |
| R(t-1) | 0.10752 | 19.8460 | 0.09875 | 4.8460 | 0.10381 | 4.3560 |
| H(t) | 2.32582 | 19.8402 | 2.29747 | 15.9789 | 2.10364 | 12.5103 |
| H(t-1) | -1.32847 | -11.2981 | -1.30092 | -9.0249 | -1.10858 | -6.5760 |
| H(t-2) | | | | | | |

**Table 5.** Established Result of Multiple Linear Regression model(Kumho)

| | t=30min Model | | t=60min Model | | t=90min Model | |
|---|---|---|---|---|---|---|
| | Coefficient | t-testing | Coefficient | t-testing | Coefficient | t-testing |
| Intercept | 0.00159 | 1.9630 | 0.00616 | 3.6536 | 0.01239 | 4.6094 |
| R(t) | 0.00890 | 10.9782 | 0.02983 | 17.5834 | 0.05772 | 21.4751 |
| R(t-1) | | | | | | |
| H(t) | 2.03880 | 41.4062 | 2.57986 | 90.9792 | 3.14685 | 70.0504 |
| H(t-1) | -1.20800 | -12.8655 | -1.59416 | -56.3590 | -2.17488 | -48.5351 |
| H(t-2) | 0.16513 | 3.6193 | | | | |
| | t=120min Model | | t=150min Model | | t=180min Model | |
| | Coefficient | t-testing | Coefficient | t-testing | Coefficient | t-testing |
| Intercept | 0.02061 | 5.6793 | 0.03069 | 6.7540 | 0.04340 | 1.8468 |
| R(t) | 0.09318 | 25.6781 | 0.13525 | 29.7647 | 0.14304 | 6.0854 |
| R(t-1) | | | | | 0.04520 | 1.7392 |
| H(t) | 3.56524 | 58.7857 | 3.83127 | 50.4507 | 3.82686 | 33.6730 |
| H(t-1) | -2.61097 | -43.1588 | -2.89825 | -38.2600 | -2.92059 | -25.9301 |
| H(t-2) | | | | | | |

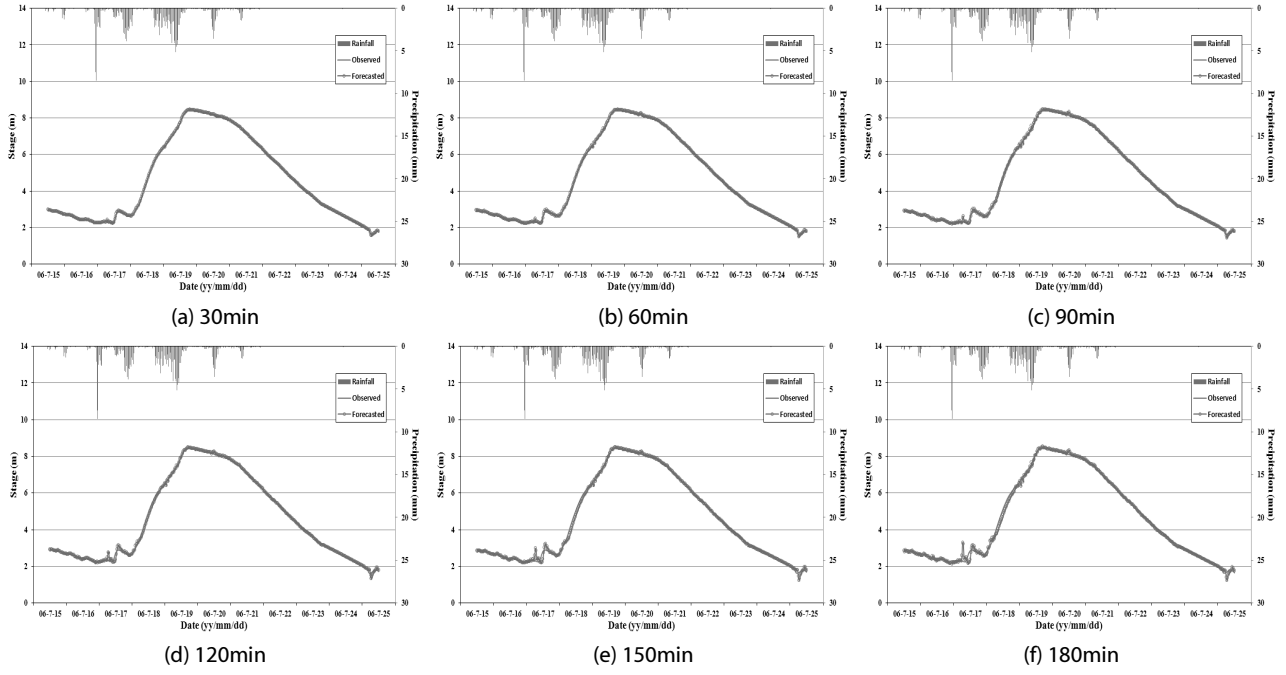
홍수예측 모형은 Eq. (23)과 같이 표현될 수 있으며, 다른 예측시간에 대해서는 표에 제시된 절편과 상수를 이용하여 식을 구성하여 예측을 수행하였다.

$$H(t+1) = -0.00847 + 0.001541R(t-1) + 1.32495H(t) - 0.32531H(t-2) \quad (23)$$

## 3.3 실시간 홍수예측 결과

### 3.3.1 뉴로-퍼지 모형 실시간 홍수예측 결과

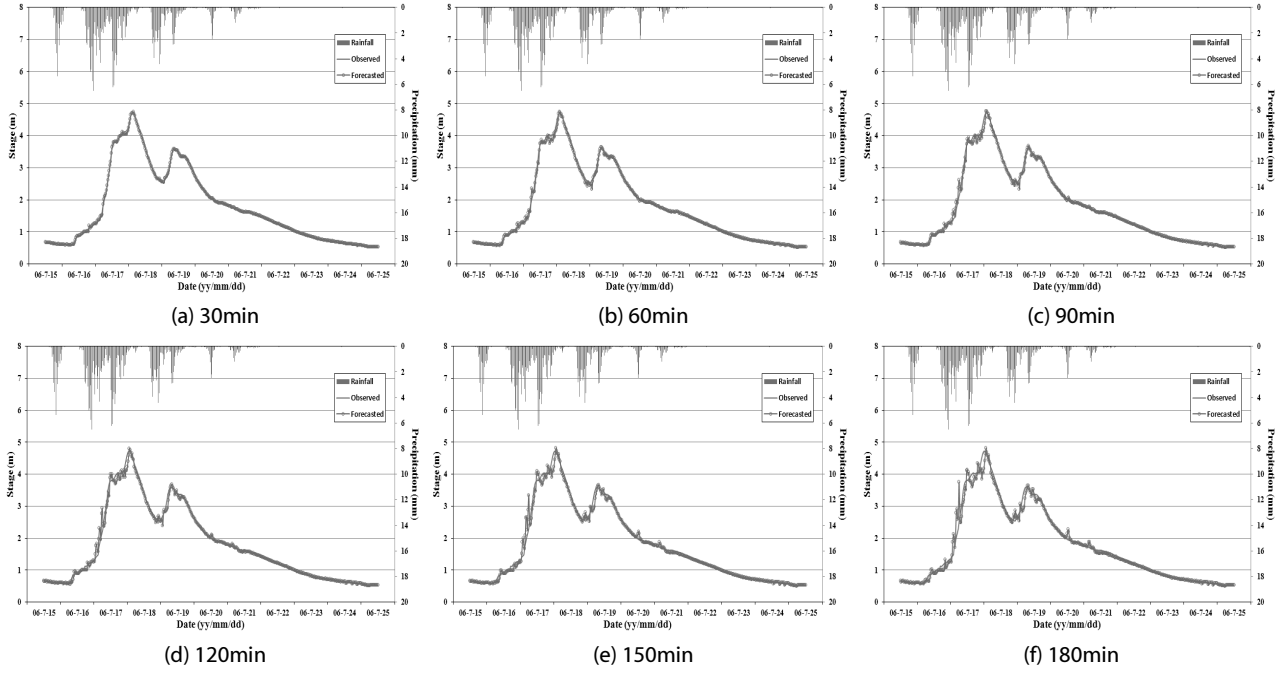
앞의 과정을 통해 유역별로 구축된 뉴로-퍼지 모형을 이용하여 6개의 호우 사상에 적용한 결과 중 밀양강 N-4 호우사상과 금호강 N-4 호우사상의 선행시간에 따른 결과를 Fig. 4와 5에 나타내었다. 특히 밀양강 유역에서는 Fig. 4와 같이 180분 선행시간까지 거의 일치하고 있음을 확인할 수 있으며, Fig. 5에서 보는 바와 같이




(a) 30min (b) 60min (c) 90min

(d) 120min (e) 150min (f) 180min

**Fig. 4.** Forecasting Results of Neuro-Fuzzy model(miryang, N-4)



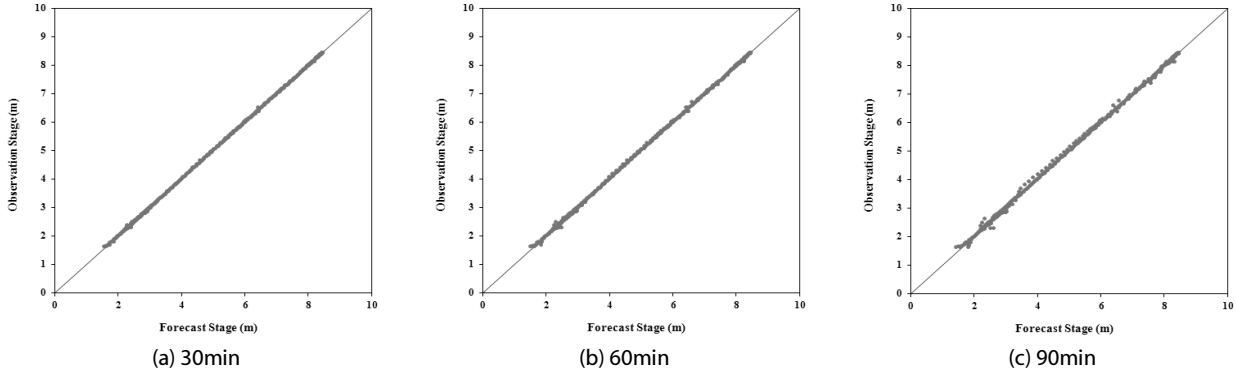

(a) 30min (b) 60min (c) 90min

(d) 120min (e) 150min (f) 180min

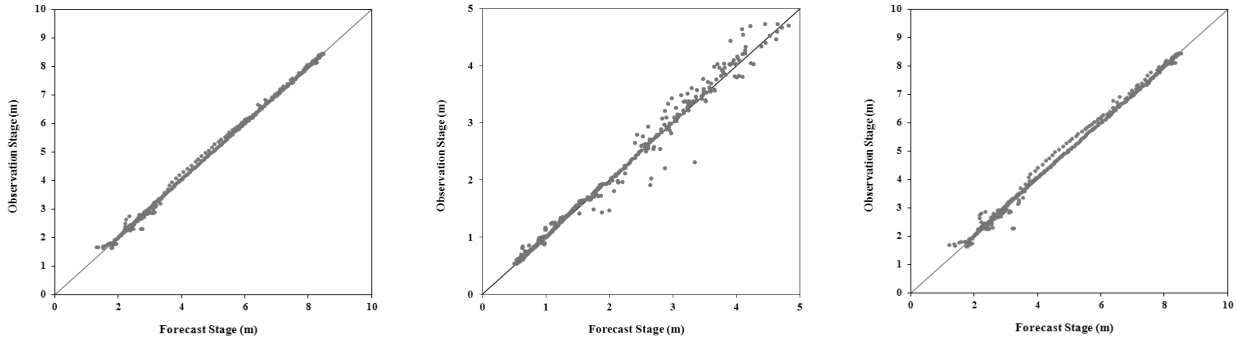
**Flg. 5.** Forecasting Results of Neuro-Fuzzy model(Kumho, N-4)

금호강 유역에서는 뉴로-퍼지를 이용한 예측 수위와 관측 수위가 90분 선행시간까지는 거의 일치하고 있음을 알 수 있고 나머지 선행시간에 대한 예측 수위도 대체적으로 관측 수위와 일치하고 있음을 확인할 수 있다. 물론 선행시간이 길어질수록 수위 상승부의 예측 정확도가 다소 감소하는 양상을 보이고 있으나 선행시간을 감안한다면 예측의 정확도가 우수한 것으로 판단된다.

각 선행시간별 예측의 정확도를 좀 더 자세히 알아보기 위해 밀양강 N-4 호우사상의 관측 결과에 대한 적합도 결과를 Fig. 6에

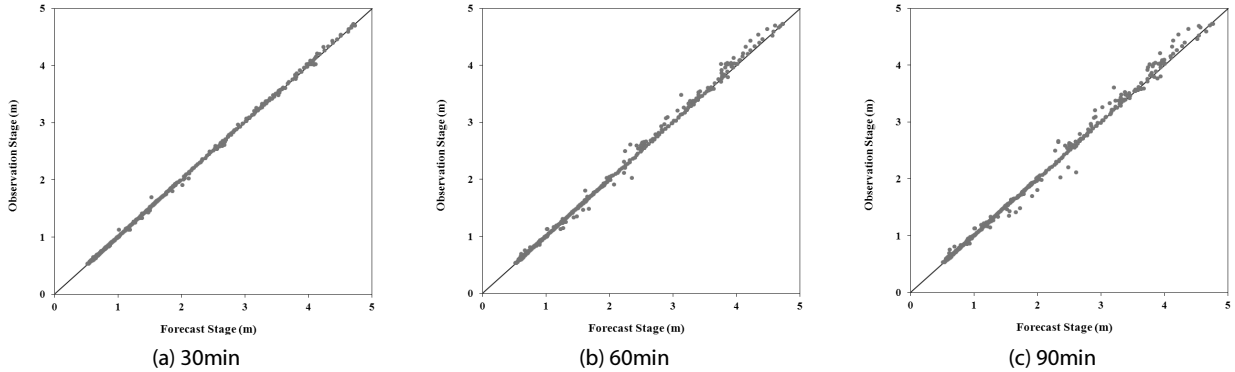



(a) 30min (b) 60min (c) 90min

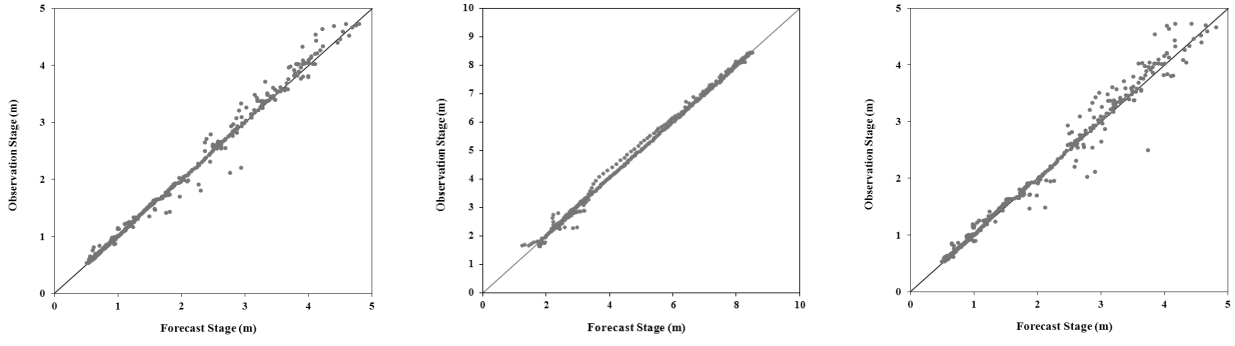



(d) 120min (e) 150min (f) 180min

FIg. 6. Suitability Results of Neuro-Fuzzy model(miryang, N-4)




(a) 30min (b) 60min (c) 90min




(d) 120min (e) 150min (f) 180min

FIg. 7. Suitability Results of Neuro-Fuzzy model(Kumho, N-4)

도시하였다. 밀양강 N-4 호우사상에 대한 뉴로-퍼지 모형의 적용을 통한 예측 수위와 관측 수위 사이에 180분 선행시간까지 높은 적합도를 보이고 있음을 확인할 수 있다. 금호강 N-4 호우사상의 관측 결과에 대한 적합도 결과 또한 Fig. 7에서 보는 바와 같이 예측 수위와 관측 수위 사이에 높은 적합도를 보이고 있음을 확인할 수 있다. Fig. 5와 마찬가지로 30분 선행시간의 경우에는 관측 수위와 예측 수위 사이에 매우 높은 적합도를 보였고 선행시간이 길어질수록 적합도가 다소 감소하고 있음을 확인할 수 있었으나 대체적으로 각 선행시간별 모의결과가 양호함을 알 수 있었다. 또한 선행시간이 길어질수록 상대적으로 수위가 상승하는 구간에서 예측 수위와 관측 수위의 오차가 커지고 있는 것을 확인할 수 있는데 전체적인 수위의 변동 양상은 상당히 잘 예측하는 것으로 판단된다.

Table 6에서 보는 바와 같이 밀양강 유역의 평균제곱근오차는 30분 선행시간의 경우 0.018m로 상당히 작게 나타나고 있음을 알 수 있고 선행시간이 길어질수록 평균제곱근오차가 조금씩 커지나 선행시간의 길이를 감안한다면 만족할만한 결과로 판단된다. 상관계수와 Nash-Sutcliffe 효율계수의 경우도 모든 검증 사상에 대해 거의 1에 가까운 높은 값을 나타내고 있음을 확인할 수 있었고, 상대첨두 오차 또한 1%내로 잘 일치하는 것을 확인할 수 있었다. 평균절대편차와 오차의 양적 개념인 RAD를 살펴보면 180분 선행시간의 경우에도 0.10m이하의 작은 값을 나타내고 있음을 확인할 수 있다. 금호강 유역 또한 Table 7에서 보는 바와 같이 각 통계지표별로 예측값이 정확함을 확인할 수 있다.

### 3.3.2 다중선형회귀 모형 실시간 홍수예측 결과

구축된 다중선형회귀 모형을 이용하여 뉴로-퍼지 모형과 같은 밀양강 N-4 호우사상과 금호강 N-4 호우사상의 선행시간에 따른 결과를 Fig. 8과 9에 나타내었다. Fig. 8에서 보는 바와 같이 밀양강 유역의 다중선형회귀 모형을 이용한 예측 수위와 관측 수위가 90분 선행시간까지는 초반 수위 상승부를 제외하고 거의 일치하고 있음을 알 수 있고 나머지 선행시간에 대한 예측 수위도 대체적으로 관측 수위와 일치하고 있음을 확인할 수 있다. 물론 선행시간이 길어질수록 전체적인 예측 정확도가 다소 감소하는 양상을 보이고 있으나 선행시간을 감안한다면 예측의 정확도가 우수한 것으로 판단된다. 금호강 유역은 60분 선행시간까지는 예측 수위와 관측 수위가 거의 일치하고 있으며 나머지 선행시간에 대해서도 예측 수위와 관측수위가 대체적으로 일치하고 있음을 확인할 수 있다.

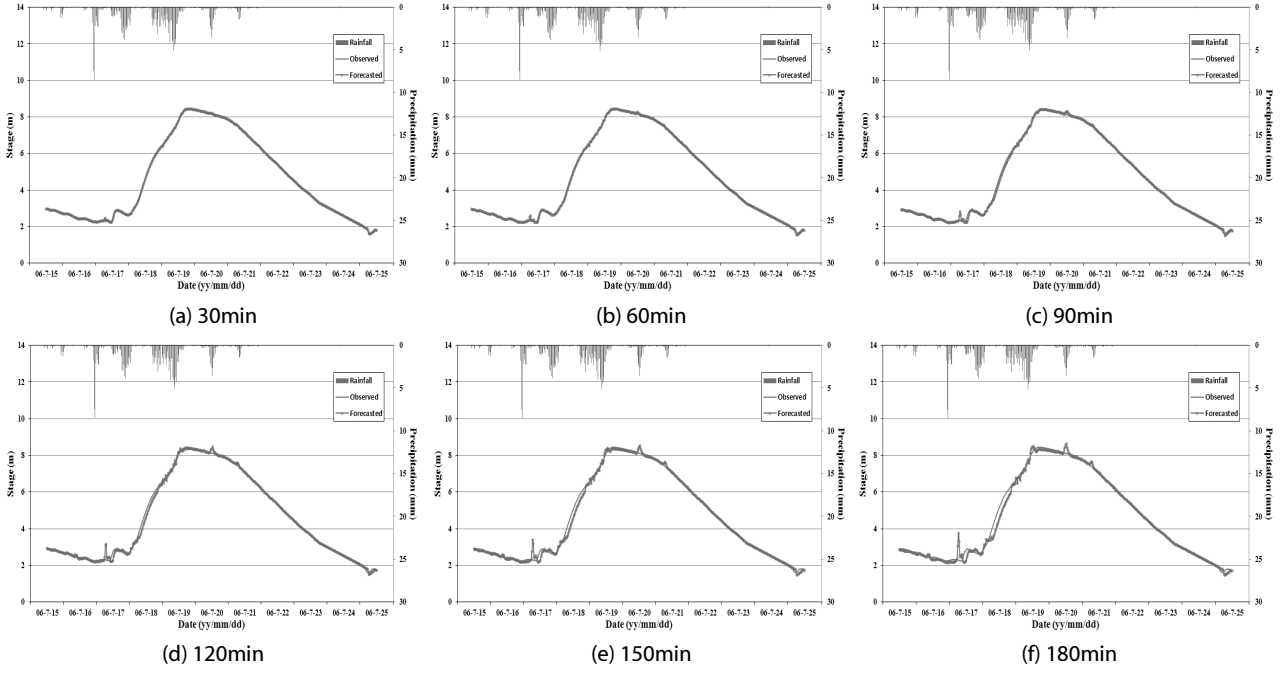
각 선행시간별 예측의 정확도를 좀 더 자세히 알아보기 위해 밀양강 N-4 호우사상의 관측 결과에 대한 적합도 결과를 Fig. 10에 도시하였다. Fig. 10에서 보는 바와 같이 밀양강 N-4 호우사상에 대한 다중선형회귀 모형의 적용을 통한 예측 수위와 관측 수위 사이에 높은 적합도를 보이고 있음을 확인할 수 있다. Fig. 8과 마찬가지로 90분 선행시간까지의 경우에는 관측 수위와 예측 수위 사이에 매우 높은 적합도를 보였고 선행시간이 길어질수록 적합도가 다소 감소하고 있음을 확인할 수 있었으나 대체적으로 각 선행시

**Table 6.** Statistics for Neuro-Fuzzy Simulation Results(miryang, N-4)

| | miryang N-4 | | | | | |
|---|---|---|---|---|---|---|
| | 30min | 60min | 90min | 120min | 150min | 180min |
| RMSE | 0.018 | 0.033 | 0.059 | 0.079 | 0.109 | 0.136 |
| CC | 1.000 | 1.000 | 1.000 | 0.999 | 0.999 | 0.998 |
| NSEC | 1.000 | 1.000 | 0.999 | 0.999 | 0.998 | 0.996 |
| RPE | 0.148 | 0.227 | 0.314 | 0.512 | 0.668 | 0.991 |
| RAD | 0.002 | 0.004 | 0.008 | 0.010 | 0.014 | 0.018 |
| MAD | 0.011 | 0.020 | 0.036 | 0.046 | 0.064 | 0.081 |

**Table 7.** Statistics for Neuro-Fuzzy Simulation Results(Kumho, N-4)

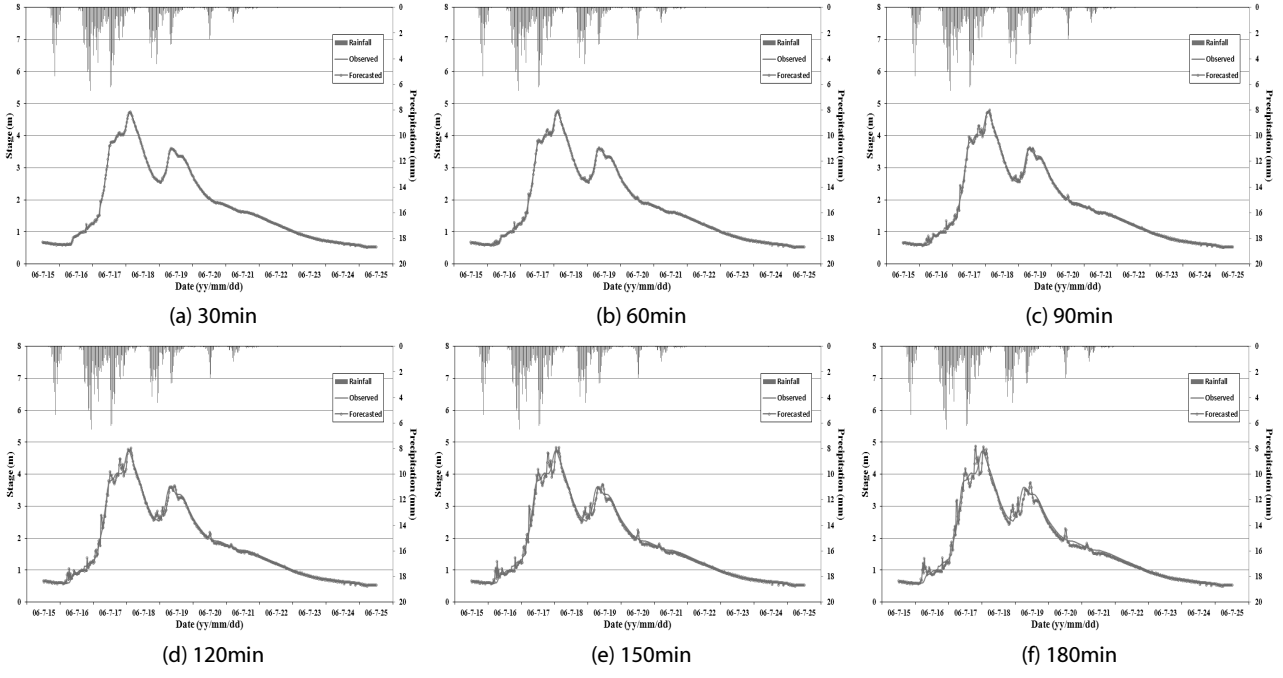
| | Kumho N-4 | | | | | |
|---|---|---|---|---|---|---|
| | 30min | 60min | 90min | 120min | 150min | 180min |
| RMSE | 0.021 | 0.054 | 0.073 | 0.098 | 0.126 | 0.149 |
| CC | 1.000 | 0.999 | 0.998 | 0.997 | 0.994 | 0.992 |
| NSEC | 1.000 | 0.998 | 0.996 | 0.993 | 0.988 | 0.984 |
| RPE | 0.273 | 0.333 | 0.981 | 1.811 | 2.180 | 2.078 |
| RAD | 0.007 | 0.015 | 0.020 | 0.027 | 0.035 | 0.042 |
| MAD | 0.012 | 0.026 | 0.035 | 0.048 | 0.063 | 0.075 |




(a) 30min (b) 60min (c) 90min

(d) 120min (e) 150min (f) 180min

**Fig. 8.** Forecasting Results of Multiple Linear Regression model(miryang, N-4)



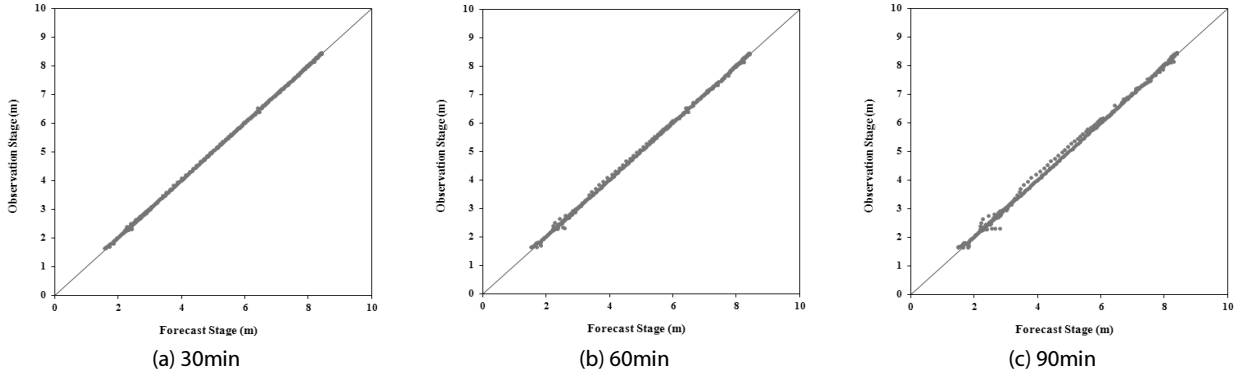

(a) 30min (b) 60min (c) 90min

(d) 120min (e) 150min (f) 180min

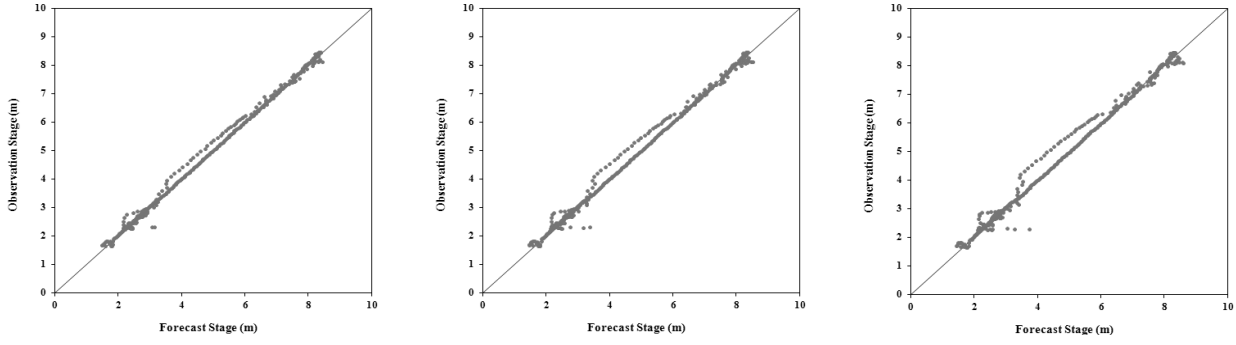
**Fig. 9.** Forecasting Results of Multiple Linear Regression model(Kumho, N-4)

간별 모의결과가 양호함을 알 수 있었다. 또한 선행시간이 길어질수록 상대적으로 수위가 상승하는 구간에서 예측 수위와 관측 수위의 오차가 커지고 있는 것을 확인할 수 있는데 전체적인 수위의 변동 양상은 상당히 잘 예측하는 것으로 판단된다. 금호강 N-4 호우사상의 적합도 결과는 Fig. 11에서 보는 바와 같이 선행시간 60분까지 높은 적합도를 보였으며 나머지 선행시간에 대해서도 대체적으로 양호한 적합도를 나타내었다.

각 유역에 대한 다중선형회귀 모형의 홍수 예측에 대한 정량적인

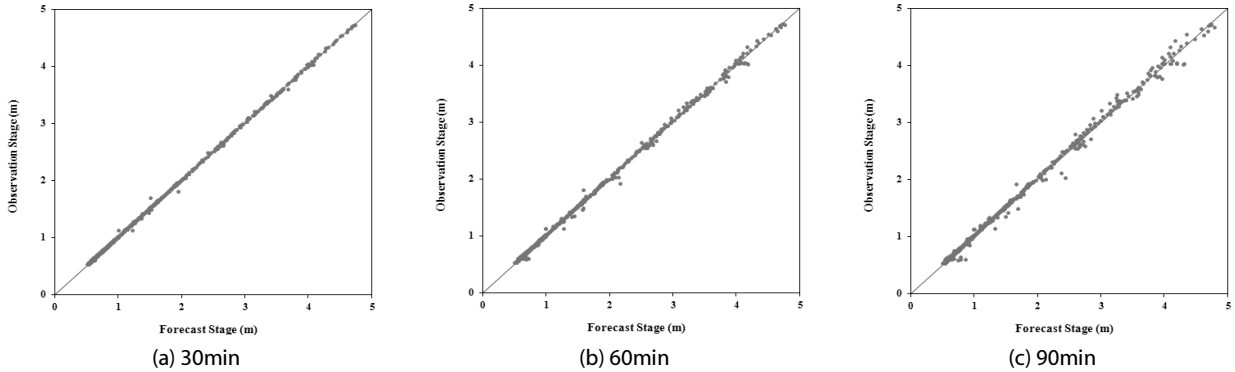



(a) 30min (b) 60min (c) 90min

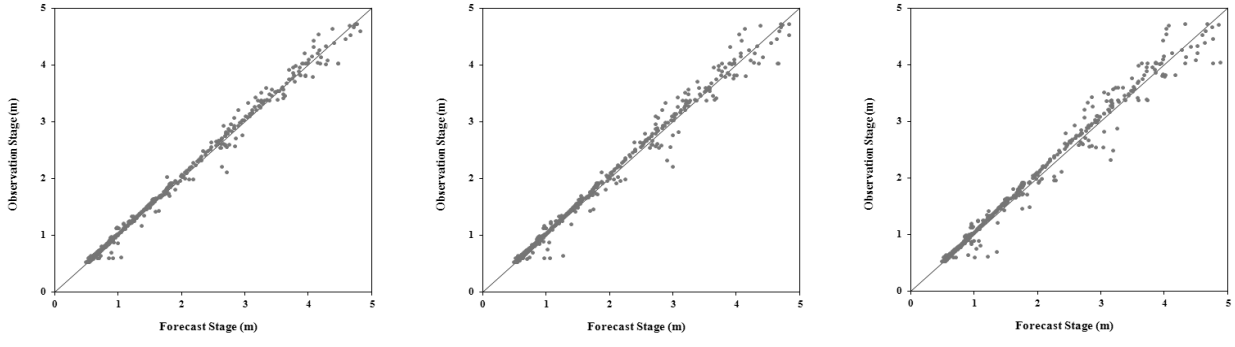



(d) 120min (e) 150min (f) 180min

**FIg. 10.** Suitability Results of Multiple Linear Regression model(miryang, N-4)




(a) 30min (b) 60min (c) 90min




(d) 120min (e) 150min (f) 180min

**FIg. 11.** Suitability Results of Multiple Linear Regression model(Kumho, N-4)

평가를 위하여 뉴로-퍼지와 마찬가지로 통계 지표인 평균제곱근오차, 상관계수, 효율계수, 상대첨두오차, 평균절대편차, RAD들을 사용하여 Table 8 and 9에 각각 나타내었다.

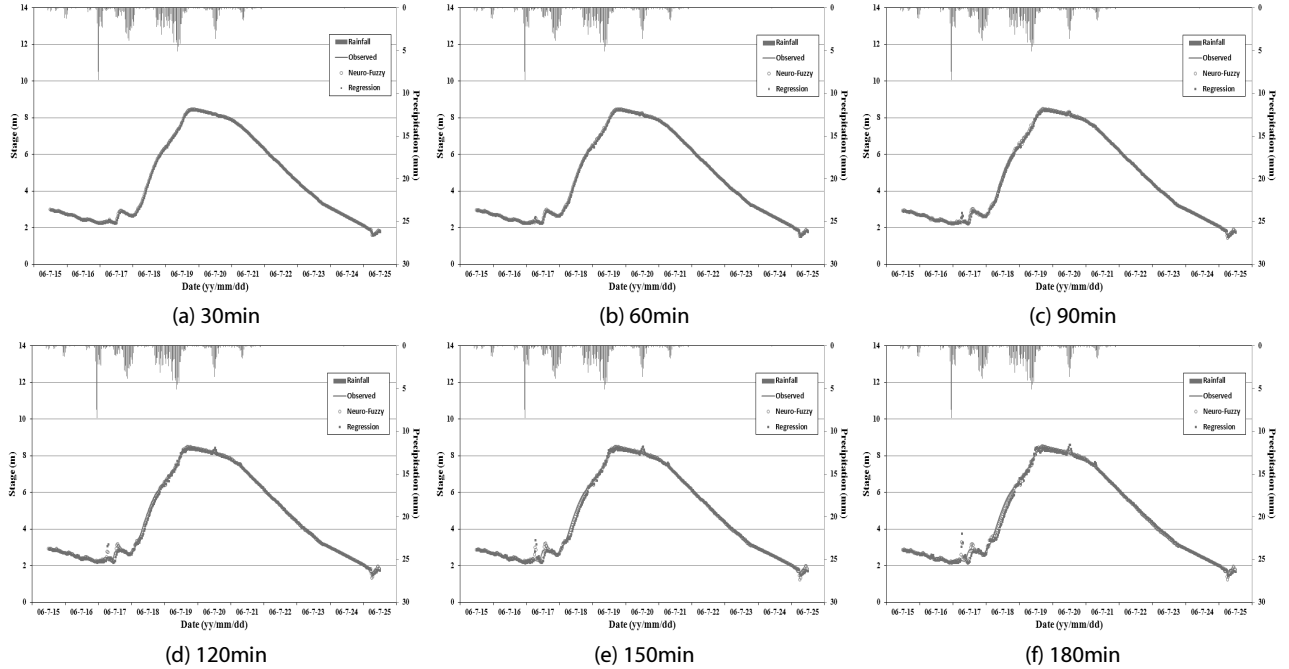
Table 8 and 9에서 보는 바와 같이 평균제곱근오차는 30분 선행시간의 경우 0.020m와 0.019m로 상당히 작게 나타나고 있음을 알 수 있고 선행시간이 길어질수록 평균제곱근오차가 조금씩 커지나 선행시간의 길이를 감안한다면 만족할만한 결과로 판단된다. 상관계수와 Nash-Sutcliffe 효율계수의 경우에도 각 유역의

**Table 8.** Statistics for Multiple Linear Regression Simulation Results(miryang, N-4)

| | miryang N-4 | | | | | |
|---|---|---|---|---|---|---|
| | 30min | 60min | 90min | 120min | 150min | 180min |
| RMSE | 0.020 | 0.041 | 0.072 | 0.118 | 0.156 | 0.201 |
| CC | 1.000 | 1.000 | 1.000 | 0.999 | 0.998 | 0.996 |
| NSEC | 1.000 | 1.000 | 0.999 | 0.997 | 0.995 | 0.992 |
| RPE | 0.020 | 0.081 | 0.337 | 0.179 | 0.998 | 2.069 |
| RAD | 0.003 | 0.005 | 0.009 | 0.015 | 0.019 | 0.025 |
| MAD | 0.012 | 0.023 | 0.041 | 0.068 | 0.089 | 0.115 |

**Table 9.** Statistics for Multiple Linear Regression Simulation Results(Kumho, N-4)

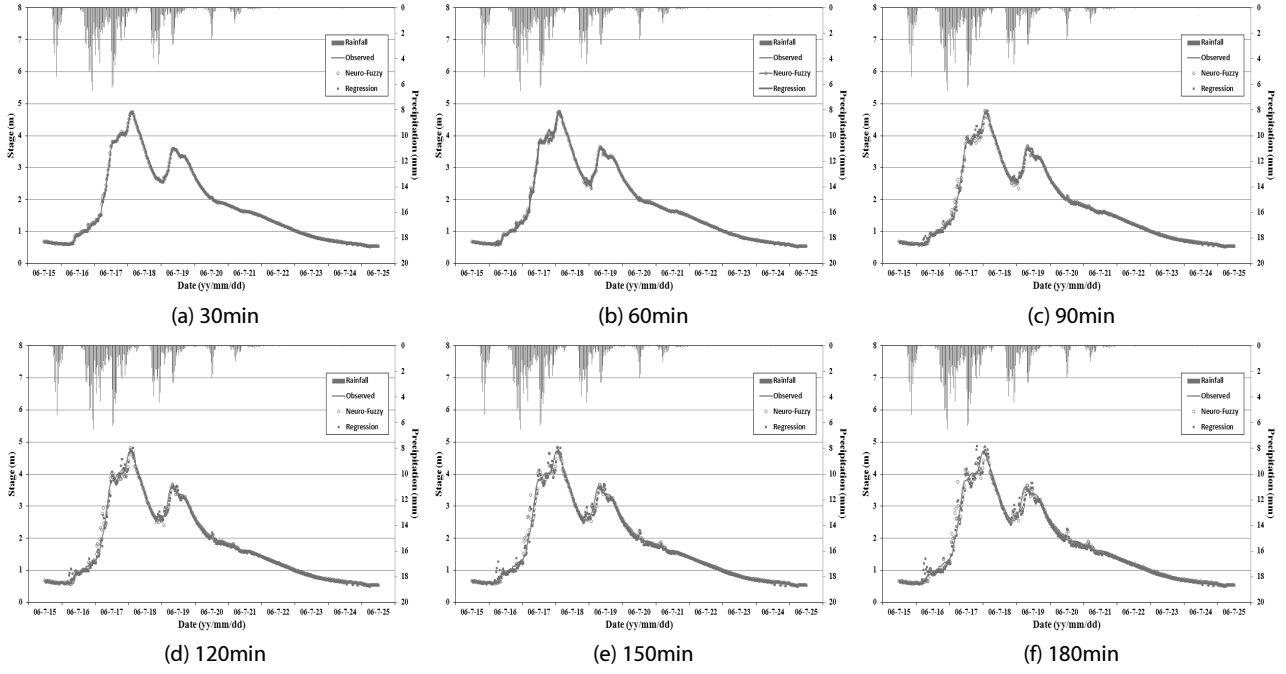
| | Kumho N-4 | | | | | |
|---|---|---|---|---|---|---|
| | 30min | 60min | 90min | 120min | 150min | 180min |
| RMSE | 0.019 | 0.039 | 0.065 | 0.096 | 0.132 | 0.163 |
| CC | 1.000 | 0.999 | 0.998 | 0.997 | 0.994 | 0.991 |
| NSEC | 1.000 | 0.999 | 0.997 | 0.993 | 0.987 | 0.981 |
| RPE | 0.517 | 1.181 | 1.681 | 2.107 | 2.448 | 3.457 |
| RAD | 0.006 | 0.013 | 0.022 | 0.032 | 0.044 | 0.056 |
| MAD | 0.010 | 0.023 | 0.038 | 0.057 | 0.079 | 0.100 |




(a) 30min (b) 60min (c) 90min

(d) 120min (e) 150min (f) 180min

**Fig. 12.** Comparison of Stage Forecasting Results(miryang, N-4)




Flg. 13. Comparison of Stage Forecasting Results(Kumho, N-4)

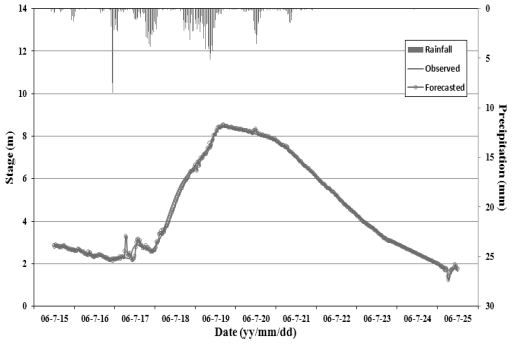
모든 검증 사상에 대해 0.9이상의 높은 값을 나타내고 있음을 확인할 수 있었고, 상대첨두 오차는 선행시간이 길어질수록 오차가 크지만 각각 2~4%내의 값을 보여주었다. 평균절대편차와 오차의 양적 개념인 RAD를 살펴보면 180분 선행시간의 경우에도 0.12m 이하의 작은 값을 나타내고 있음을 확인할 수 있다.

### 3.3.3 실시간 홍수예측 결과 비교·검토

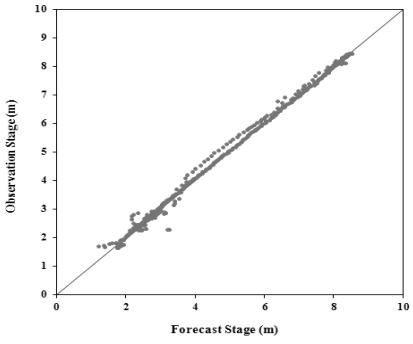
강우와 수위의 최적 입력자료 조합을 바탕으로 뉴로-퍼지 모형과 다중선형회귀모형을 구축하여 밀양강 및 금호강 유역에 대해 30, 60, 90, 120, 150, 180분 선행시간에 대해 홍수예측을 수행하였으며, 모의결과에 대해 그림과 통계지표를 통해 홍수예측 결과의 적용성을 검증하였다. 뉴로-퍼지 모형과 다중선형회귀 모형을 비교해 본 결과 두 모형 모두 예측 수위와 실측 수위가 상당히 잘 일치하고 있음을 할 수 있었다. 그러나 예측 선행시간이 길어질수록 대체적으로 뉴로-퍼지 모형이 다중선형회귀 모형에 비해 조금 더 우수한 결과를 보여주었다. Figs. 12 and 13은 밀양강과 금호강의 N-4 호우사상에 대한 뉴로-퍼지 모형과 다중선형회귀 모형의 선행시간별 수위예측 결과 비교를 나타낸 것이다. 아래의 그림을 살펴보면 선행시간 60분까지는 두 모형 모두 예측 수위와 실측 수위가 상당히 잘 일치하고 있으나 그 이후 선행시간에서는 다중선형회귀 모형의 경우 예측 수위가 뉴로-퍼지 모형에 비해 좀 더 큰 값을 나타내는 것을 확인 할 수 있다. 이를 통해 뉴로-퍼지 모형의 결과가 다중선형회귀 모형의 결과에 비해 수위예측 능력이 우수한 것을 확인할 수 있다.

또한, 서론에서 언급한 바와 같이 사람들이 대가올 홍수에 준비할 수 있도록 충분한 예측 선행시간이 확보되어야 하므로 선행시간 180분의 경우에 대하여 뉴로-퍼지 모형과 다중선형회귀 모형의 홍수예측 결과를 비교해 보았다. 선행시간이 길어질수록 예측 수위의 정확도가 떨어지나 Figs. 14 and 15에 나타나 있듯이 뉴로-퍼지 모형이 다중선형회귀 모형에 비해 수위예측 결과 및 적합도에서 더 우수함을 확인할 수 있다.

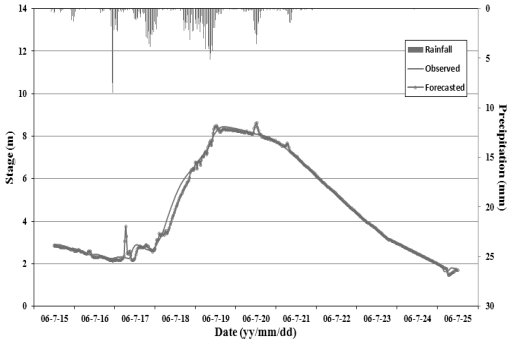
N-5와 N-6 사상에 대해서도 N-4 사상과 일치하는 결과를 나타내었다. 선행시간이 60분 이하일 때에는 두 모형에서 예측 수위가 일치하고 있으나 선행시간이 길어질수록 뉴로-퍼지 모형의 예측 결과가 우수한 결과를 나타내었다. 또한 다중선형회귀의 경우 과거 자료가 일관성이 유지되어야 할뿐만 아니라 오랜 기간에 대해 신뢰도 높은 자료가 요구된다. 낙동강 본류의 경우 4대강 사업으로 인해 다중선형회귀 방법을 적용하기 위해서는 향후 오랜 기간 동안의 자료 축적이 요구되지만 뉴로-퍼지 모형의 적용을 위해서는 과거 유사한 강우-유출 사상이 확보된다면 예측이 가능하므로 뉴로-퍼지 모형의 적용성이 더욱 우수한 것으로 판단된다.



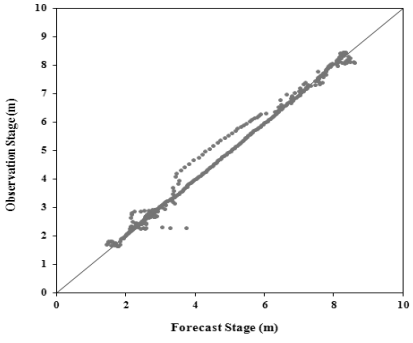

(a) Result of Nuero-Fuzzy




(b) Suitability of Nuero-Fuzzy

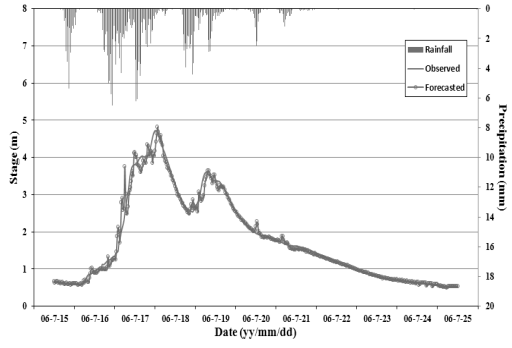



(c) Results of Regression

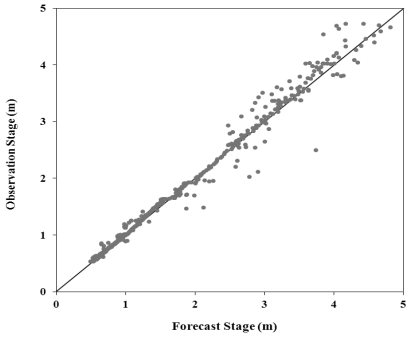



(d) Suitability of Regression

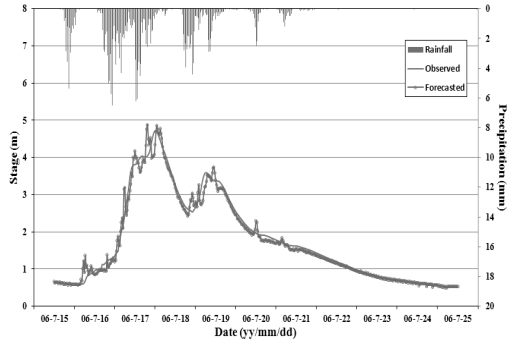
**FIg. 14.** Comparison of Nuero-Fuzzy and Regression Results(miryang, N-4)



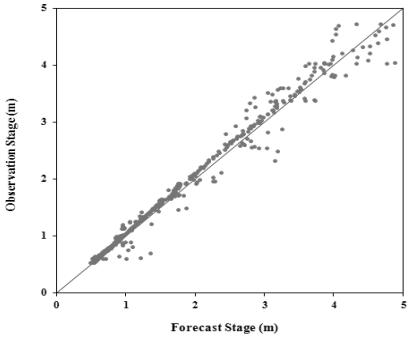

(a) Result of Nuero-Fuzzy




(b) Suitability of Nuero-Fuzzy




(c) Results of Regression




(d) Suitability of Regression

**FIg. 15.** Comparison of Nuero-Fuzzy and Regression Results(Kumho, N-4)

## 4. 결 론

본 연구에서는 정확한 실시간 홍수예측을 위해 간단한 입력자료만으로도 홍수예측이 가능한 뉴로-퍼지 모형과 다중선형회귀 모형을 구축하여 낙동강에서의 실시간 홍수예측 결과를 비교해보았다. 본 연구의 주요 연구결과는 다음과 같이 요약할 수 있다.

(1) 최적 입력 자료 조합을 근거로 밀양강과 금호강에 대해 6개의 홍수사상을 이용하여 훈련과 확인 사상을 선정하였고 또한 각 유역별로 subtractive clustering기법을 통하여 홍수예측 선행시간별 소속 함수의 개수와 매개변수의 산정을 통한 뉴로-퍼지 홍수예측 모형을 구축하였다. 실시간 홍수예측 결과 매우 정확도가 높음을 확인할 수 있었고 180분 선행시간의 경우에도 만족할만한 결과를 보여주었다. 또한 모든 통계지표분석에서도 우수한 예측결과를 보이고 있었다. 이를 통해 정확도 높은 3시간의 예측 선행시간을 확보할 수 있었다.

(2) 뉴로-퍼지 모형과 같이 객관적인 자료를 바탕으로 수위를 예측할 수 있는 다중선형회귀 모형을 구축하였다. 회귀계수 추정기법에는 최소 자승법, 가중 최소 자승법, 단계별 선택법이 있으며 본 연구에서는 단계별 선택법으로 다중선형회귀 홍수예측 모형을 구축하였다. 뉴로-퍼지 모형과 실시간 홍수예측 결과를 비교하기 위해 같은 입력자료를 사용하여 모형을 구축하였으며 수위예측 결과 및 적합도에서 우수함을 확인할 수 있었다.

(3) 그래프 및 통계지표를 통해 같은 유역의 같은 사상에 대해 뉴로-퍼지 모형과 다중선형회귀 모형을 비교해 본 결과 대체적으로 다중선형회귀 모형의 예측결과는 뉴로-퍼지모형의 예측결과와 비교해 진동이 조금 더 많이 나타났고 최고수위예측에 있어 상대적으로 부정확한 값을 나타냈다. 하지만 그 차이는 통계지표를 통한 수치로 확인해 보아도 미세하였고 두 모형 모두 홍수예측에 있어 우수한 결과를 나타냈다.

(4) 이런 비교를 통해 구축된 낙동강유역의 홍수 예측 모형은 기존의 물리적 모형들과 비교해 홍수 예·경보에 있어 만족할만한 3시간 선행시간을 확보할 수 있었다. 앞으로 더 정확한 예측을 위해서는 관측자료의 정확성을 높이고 실측자료에 대한 보완대책을 마련한다면 모형구축과 더불어 예측결과의 정확성을 더욱 높일 수 있을 것으로 기대된다. 또한 모형구축에 있어 시간상의 물리적 제한으로 과거의 모든 사상을 적용해 구축해보지는 못하였으나 추후 모든 사상에 대해 적용해본다면 더욱 우수한 결과를 도출 할 것으로 사료된다. 그리고 방류량을 포함한 조합의 댐 직 하류 지점에서의 예측 정확도를 검증한다면 방류량을 포함한 조합의 신뢰도를 더욱 높일 수 있을 것이다.

(5) 향후 수위예측-유량변환, 유량예측-수위변환 관계의 정확성 비교와 입력 자료의 시간간격에 따른 예측의 정확성 비교, 방류량을 고려한 모형구축과 수위예측 지점선정 사이의 상관관계, 낙동강 본류 내 지점의 수위 및 유량 예측 정확성 검증, 나아가 Grid partition 과 sub. clustering의 결과를 비교하는 연구를 통해 본 연구 주제에 관해 더욱 심도 깊은 결과를 얻을 수 있을 것으로 기대된다.

## 감사의 글

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

## References

Brown, M. and Harris, C. (1994). *Neuro-fuzzy adaptive modeling and control,* Prentice Hall International (UK) Ltd. Hertfordshire, UK, p. 508.

Carpenter, T. M., Georgakakos, K. P. and Sperfslage, J. A. (2001). "On the parametric and Nexrad-Radar sensitivities of a distributed hydrologic model suitable for operational use." *Journal of Hydrology*, Vol. 253, pp. 169-193.

Chaoulakou, A., Assimacopoulos, D. and Lekkas, T. (1999). "Forecasting daily maximum ozon concentration in the Athens basin." *Environmental Monitoring and Assessment*, Vol 56, pp. 97-112.

Choi, S.-Y. (2011). *Real-time flood forecasting and inundation analysis in medium and small streams*. Doctor Dissertation, Kyungpook National University (in Korean).

Gautam, D. K. and Holz, K. P. (2001). "Rainfall-runoff modeling using adaptive neuro-fuzzy systems." *Journal of Hydroinformatics*, pp. 3-10.

Imrie, C. E., Durucan, S. and Korre, A. (2000). "River flow prediction using artificial neural networks: generalisation beyond the calibration range." *Journal of Hydrology*, Vol. 233, pp. 138-153.

Jain, A., Sudheer, K. P. and Srinivasulu, S. (2004). "Identification of physical processes inherent in artificial neural network rainfall-runoff models." *Hydrologic Process*, Vol. 118, pp. 571-581.

Jang, J.-S. (1992). "Self-learning fuzzy controllers based on temporal backpropagation." *IEEE Trans Neural Netw*, Vol. 3, No. 5, pp. 714-723.

Jeong, D.-K. and Lee, B.-H. (2009). "Development of urban flood water level forecasting model using regression method." *Journal of Korea Water Resources Association*, Vol. 43, No. 2, pp. 221-231 (in Korean).

Kim, K.-T., Kim, J.-H. and Choi, Y.-S. (2006). "Study of flood warning and forecasting in small to medium scale watershed." *Korea Water Resources Association Conf.*, pp. 1126-1130 (in

Korean).

Ko, Y.-J. (2001). *The application of fuzzy neural network on the hourly river flow forecating*. Master Dissertation, Chonnam National University. pp. 1-6, pp. 14-17 (in Korean).

Kurtulus, B. and Razack, M. (2009). "Modeling daily discharge responses of a large karstic aquifer using soft computing methods artificial neural network and neuro fuzzy." *Journal of Hydrology*, Vol. 375, pp. 146-162.

Lohani, A. K., Kumar, R. and Singh, R. D. (2012). "Hydrological time series modeling: A Comparison Between Adaptive Neuro-Fuzzy, Neural Network And Autoregressive Techniques." *Journal of Hydrology*, Vol. 442-443, pp. 23-35.

Luk, K. C., Ball, J. E. and Sharma, A. (2001). "An application of artificial neural networks for rainfall forecasting." *Math Computer Model*, Vol. 33, pp. 683-693.

Nayak, P. C., Sudheer, K. P., Rangan, D. M. and Ramasastri, K. S. (2005). "Short-term flood forecasting with a neurofuzzy model." *Water Resources Research*, Vol. 41, No. 4, W04004.

Ramirez, M. C. P., Velho, H. F. C. and Ferreira, N. J. (2005). "Artificial neural network technique for rainfall forecasting applied to the sao paulo region." *Journal of Hydrology*, Vol. 301, pp. 146-162.

Schilling, K. E. and Wolter, C. F (2005). "Estimation of streamflow, baseflow and nitrate-nitrogen loads in Iwoa using multiple regression models." *Journal of American Water Resources Association*, Vol. 41, No. 6, pp. 1333-1346.

Shin, S.-I. (2002). *Study on forecasting flood discharge using neural network and neuro-fuzzy,* Master Dissertation, Kyungil University (in Korean).

Smith, J. and Eli, R. N. (1995). "Neural network models of the rainfall-runoff process." *Journal of Water Resources Planning and Management,* ASCE, Vol. 121, pp. 499-508.

Sung, J.-Y. and Heo, J.-H. (2009). "Tributary flood forecasting using statistical analysis method." *Korea Water Resources Association Conf*., pp. 1524-1527 (in Korean).

Talei, A., Chua, L. H. C. and Wong, S. W. (2010). "Evaluation of rainfall and discharge inputs used by adaptive network-based fuzzy inference systems (ANFIS) in rainfall-runoff modeling." *Journal of Hydrology*, Vol. 391, Issues 3-4, pp. 248-262.

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.

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.

Yarar, M., Onucyildiz, M. and Copty, N. K. (2009). "Modelling level change in lakes using neuro fuzzy and artificial neural networks." *Journal of Hydrology*, Vol. 365, pp. 329-334.

Yoon, Y.-N. and Wone, S.-Y. (1991). "A multiple regression model for the estimation of monthly runoff from ungaged watersheds." *Journal of Korean Association of Hydrological Sciences*, Vol. 24, No. 3, pp.71-82 (in Korean).

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.