

# Neuro-Fuzzy 기법을 이용한 홍수예측

## Flood Estimation Using Neuro-Fuzzy Technique

지정원\*, 최창원\*\*, 이재응\*\*\*

Jungwon Ji, Changwon Choi, Jaeung Yi

### 요 지

물은 생물의 생존을 위해 필수적인 요소로 인류가 시작된 이래로 물을 효율적으로 이용하고 안전하게 관리하기 위한 노력은 계속되어 왔다. 최근 지구 온난화가 주요 원인으로 알려진 국지성 집중호우의 피해는 매우 심각하며, 이로 인해 치수에 대한 중요성은 날로 커지고 있다.

지금까지 사용해 왔던 홍수 예·경보 과정은 특정 지점의 유출량을 예측하기 위해서 강우-유출 모형을 운영하였다. 그러나 물리적 모형의 경우 운영에 필요한 매개변수의 결정과정이 복잡하고, 매개변수 결정을 위해 많은 자료를 필요로 한다. 또한 그 매개변수의 결정과정은 많은 불확실성을 포함하고 있어서 모형의 운영을 위한 전처리과정과 계산과정을 거치는 동안 발생한 오차가 누적되어 결과물 속에는 많은 오차가 포함되어 있다.

본 연구에서는 기존의 홍수 예·경보 시스템의 문제점과 불확실성을 최대한 감소시키고 더 우수한 유출량 예측을 위해 neuro-fuzzy 추론 기법을 이용한 모형인 ANFIS(Adaptive Neuro-Fuzzy Inference System)를 사용하여 하천수위를 예측하였다. ANFIS는 신경회로망과 퍼지이론을 결합한 기법으로 신경회로망의 구조와 학습 능력을 이용하여 제어환경에서 획득한 입·출력 정보로부터 언어변수의 membership 함수와 제어규칙을 제어 대상에 적합하도록 자동으로 조종하는 기법이다. 본 연구에서는 ANFIS를 사용하여 탄천 하류에 위치한 대곡교의 수위를 예측하였다. 분석을 위해 2007년부터 2011년까지의 탄천 유역의 관측 강우자료와 수위자료 중 강우강도와 지속시간, 강우 형태에 따라 7개의 강우사상을 선정하였다. 학습자료 및 보정자료의 변화에 따른 예측 오차를 비교하여 모형의 적용성과 적정성을 평가하였다. 적용결과 입력자료 구성의 경우 해당 시간의 강우량 및 수위자료와 10분 전 강우자료를 이용한 모델이 가장 우수한 예측을 보였고, 학습자료의 경우 자료의 길이가 길고, 최대홍수량이 큰 경우 가장 우수한 예측 결과를 보였다. 본 연구의 적용결과 가장 우수한 모형의 경우 30분 예측 침두수위 오차는 0.32%, RMSE는 0.05m 이고 예측시간이 길어짐에 따라 오차가 비선형적으로 증가하는 경향을 보였다.

**핵심용어** : ANFIS, Neuro-Fuzzy 기법, 홍수예측

### 1. 서론

일반적으로 홍수 예측을 위해 사용되는 물리적 모형에서는 강우-유출 모형의 운영에 필요한 매개변수의 결정과정이 상당히 복잡하고 매개변수의 결정을 위해 많은 자료가 필요하며, 매개변수의 결정과정에 불확실성이 내재되어 있다. 따라서 모형의 운영을 위한 전처리과정과 주 계산과정을 거치는 동안 다양한 오차가 발생하게 되고, 각 과정에서 발생된 오차가 누적되어 모형의 결과물인 예측된 홍수량에는 많은 오차가 포함되어 있다.

\* 정회원 · 아주대학교 건설교통공학과 박사과정 · E-mail : [log58@ajou.ac.kr](mailto:log58@ajou.ac.kr)

\*\* 정회원 · 아주대학교 건설교통공학과 박사수료 · E-mail : [itsme99@ajou.ac.kr](mailto:itsme99@ajou.ac.kr)

\*\*\* 정회원 · 아주대학교 환경건설교통공학부 교수 · 공학박사 · E-mail : [jeji@ajou.ac.kr](mailto:jeji@ajou.ac.kr)

본 연구에서는 기존의 홍수 예·경보 시스템의 문제점과 불확실성을 최대한 감소시키고 우수한 홍수량 예측을 위해 Neuro-Fuzzy 추론 기법을 사용하여 하천수위를 예측하였다. 탄천유역의 하류에 위치한 대곡교의 수위를 예측하기 위한 ANFIS를 구축하고 모형구축에 사용된 자료의 변화에 따른 예측 정확도를 비교 분석하였다. 모형의 입력자료는 2007년부터 2011년까지의 강우 및 수위 관측자료 중 약 5 이상의 수위가 발생한 7개의 홍수사상을 추출하여 사용하였다. 10분 간격의 자료를 사용하였으며, 모의 시점으로부터 10분에서 3시간( $t+1 \sim t+18$ )까지의 수위를 모의하였다. 선정된 7개의 홍수사상을 다양한 순서로 조합하여, ANFIS의 학습(training)과 보정(checking), 테스트(testing)를 위한 자료 조합을 만들었으며, 40개 set의 모형과 720개 testing 결과를 첨두수위비와 평균제곱근오차를 통해 비교하였다. 특히, training data와 testing data의 변화에 따른 결과를 비교하였으며, training data가 미치는 영향 대해 중점적으로 검토하였다.

## 2. 퍼지 집합 이론

인간생활이나 자연현상에는 수치만으로 표현할 수 없는 애매모호함과 불확실성이 산재해 있다. 특히 수자원분야에서는 “저수지의 수위가 높다.”, “상류 유역에 큰 강우가 발생하였다.”, “유출량이 많다.” 등의 애매모호한 상황에 대한 정의를 필요로 하는 경우가 많다. 이는 일반적으로 사용하고 있는 언어 표현이지만 수학적 연산이나 컴퓨터 언어들은 모두 이진법(binary system)을 사용하고 있기 때문에 이 같은 표현을 수학적 연산이나 컴퓨터 언어로 사용하기 곤란하다. 특히, 수자원 현상을 다루는 문제에 있어서는 그렇다(yes) 혹은 아니다(no)와 같은 이진법을 사용하여 문제를 풀어가는 데는 상당한 어려움이 따른다. Zadeh(1965)는 이러한 애매함을 정량화하여 수학적 연산이나 컴퓨터 언어를 통해 논리적으로 접근하기 위해 퍼지집합 이론을 개발하였고, Mamdani와 Assilian(1975)이 퍼지 알고리즘과 복잡한 시스템의 언어모델링을 사용하여 제어에 응용하면서 퍼지제어에 본격적으로 이용되기 시작하였다.

ANFIS는 사용자가 자료만 가지고 membership 함수의 형태를 결정하기 어려울 때 입력과 출력 자료에 의해서 membership 함수의 형태를 구축할 수 있는 기법이다. ANFIS는 Jang(1993)이 Sugeno 형태의 FIS와 신경망의 장점을 취합하여 개발하였다. ANFIS 기법은 강우-유출 관계와 같이 고도의 비선형 시스템의 해석에 특히 이상적이다.

## 3. 시험유역 적용

ANFIS를 사용한 수위예측의 적용성을 평가하고, 모형 구축을 위한 자료의 사용에 따른 영향을 분석하기 위해 한강의 지류인 탄천유역의 강우자료와 수위자료를 사용하였다. 탄천은 경기도 용인시에서 발원해서 성남시와 서울특별시 송파구, 강남구를 거쳐 한강으로 흘러드는 한강의 지류로서, 절반이 넘는 약 25km 구간이 성남시의 중심부에 걸쳐 있는 전형적인 도시하천이다.

유역 평균강우량 산정을 위한 강우관측소는 남한산성, 과천, 대장, 낙생, 구미, 산성, 성남의 7개 강우관측소를 선정하여 그 영향을 thiessen weighting method를 사용하여 산정하였다. 수위예측을 위한 수위관측소는 대곡교에 위치한 성남수위관측소의 자료를 사용하였다.

### 3.1 모형 구성

ANFIS 모형 구축을 위해서 탄천 유역의 2007년부터 2011년까지의 강우 및 수위자료 중 비교적 큰 유출이 발생한 7개 자료를 data1 ~ data7로 선정하였다. 모형을 학습시키기 위해 필요한

training data에 따른 모형의 예측 결과를 비교하기 위해, 각 자료는 자료의 길이와 최고수위, 홍수의 양상을 나타낼 수 있는 peak의 수가 다른 자료들로 선정하였다.

모형의 입력자료에 사용될 강우자료와 수위자료는 10분 간격의 자료를 사용하였으며, 입력자료의 구성에 따른 결과를 비교하기 위해 Table 4에 나타난 것과 같이 현재시간을 기준으로  $t \sim t-3$  시간(현재로부터 30분전 시간)의 강우자료와  $t \sim t-2$  시간(현재로부터 20분전 시간)의 수위자료를 사용하여 5개의 Model을 구성하였다(Table 1). Height error와 RMSE(Root Mean Square Error)를 사용하여 결과를 비교하였다(Eq. 1 and Eq. 2).

Table 1. Characteristics of Selected Data

	Obs. Precipitation	Obs. Water level	Est. Water level
Model A	$P_{t-1}$	$H_t$	$H_{t+1} \sim H_{t+18}$
Model B	$P_t, P_{t-1}$	$H_t, H_{t-1}$	$H_{t+1} \sim H_{t+18}$
Model C	$P_t, P_{t-1}, P_{t-2}$	$H_t, H_{t-1}$	$H_{t+1} \sim H_{t+18}$
Model D	$P_t, P_{t-1}, P_{t-2}$	$H_t, H_{t-1}, H_{t-2}$	$H_{t+1} \sim H_{t+18}$
Model E	$P_t, P_{t-1}, P_{t-2}, P_{t-3}$	$H_t, H_{t-1}, H_{t-2}$	$H_{t+1} \sim H_{t+18}$

$$HE = \frac{|HP_{obs.} - HP_{est.}|}{HP_{obs.}} \times 100(\%) \quad (1)$$

여기서  $HE$ 는 %로 계산된 height error ratio,  $HP_{obs.}$ 는 실측침두수위,  $HP_{est.}$ 는 예측침두수위를 나타낸다.

$$RMSE = \frac{(H_{est.} - H_{obs.})^2}{n-1} \quad (2)$$

여기서  $H_{est.}$ 는 예측수위자료,  $H_{obs.}$ 는 실측수위자료,  $n$ 은 자료의 수(길이)를 나타낸다.

### 3.2 모형의 입력자료 구성에 따른 결과 비교

모형의 입력자료 구성에 따른 결과를 비교해본 결과 height error에 따른 비교에서는 특별한 경향이 나타나지 않았지만 RMSE의 Average 비교에서는 Model A가 우수한 예측을 하는 것으로 나타났다(Fig 1, Fig 2). 이것은 탄천유역의 크기가  $302 \text{ km}^2$ 으로 비교적 크지 않기 때문에 10~20분 이내에 관측된 강우와 수위의 영향이 유출에 가장 큰 영향을 미치기 때문인 것으로 생각된다. ANFIS를 통해 생성한 전체 모형의 height error는 평균 24.48 %, RMSE는 평균 0.367 m 나타났다.

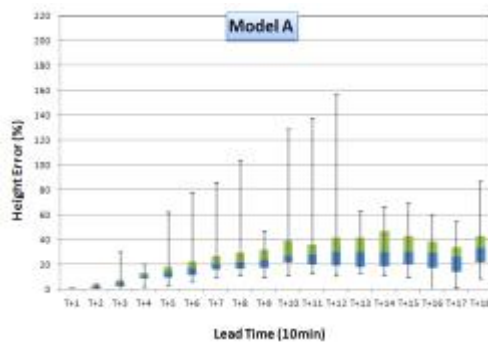


Fig 1. Height Error Comparison

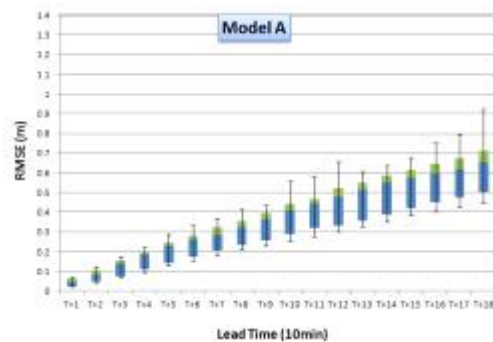


Fig 2. RMSE Comparison

### 3.3 Training data의 사용에 따른 결과 비교

모형의 구성을 위한 자료 중 training data의 변화에 따른 영향을 비교하였다. Training data에 따른 수위예측 결과를 height error를 통해 비교하면, training data로 data7을 사용한 경우가 height error가 average 12.2%, median 12.5%로 가장 작게 나타났으며, median으로부터 편차범위도 가장 좁게 나타났다. data1과 data2를 비교해보면, lead time이 t+14까지는 data2를 training data로 사용한 경우에 height error가 더 낮은 것으로 나타났으며, t+15부터 t+18까지는 data1을 training data로 사용한 경우가 더 우수한 결과를 나타냈다. median으로부터의 편차는 전체적으로 data1의 경우가 더 좁게 분포된 것으로 나타났으며 data2를 training data로 사용한 경우는 편차의 전체적인 분포가 상당히 넓게 나타났다.

RMSE를 통해 비교해 보면, height error의 경우와 유사하게, training data로 data7을 사용한 모형에서 각 lead time에서 전체적으로 근사하지만 가장 우수한 결과를 나타내었고, 전체 lead time의 average 및 median에서도 가장 좋은 결과를 보였다. 또한, 편차의 범위에 있어서도 data1이나 data2를 training data로 사용한 경우에 비해 편차의 범위가 좁게 나타났다. data1과 data2를 비교해보면, 각 lead time별 median은 data2가 더 우수한 것으로 나타났고, median으로부터의 상하 25% 편차는 data2가 더 좁은 범위에 분포된 것으로 나타났으나 전체적인 편차의 범위는 data2가 더 크게 나타났다.

Training data의 사용에 따른 홍수예측 정확도를 전체적으로 정리해보면, training data는 자료의 길이가 짧을 경우 예측값의 편차가 크게 나타날 수 있고, 최대홍수량이 큰 것이 비교적 예측 정확도를 높이는 데 도움이 되며, peak가 많이 발생하여 수위의 증감이 많은 것이 우수한 모형을 만드는데 도움이 되는 것으로 판단된다.

### 3.4 Testing data의 사용에 따른 결과 비교

Testing data로 사용된 세 개의 data를 상호 비교해보면, 최고 수위는 data4, data5, data6 순으로 증가하며, data5는 비교적 자료길이가 짧고 단순한 형태이고, data4는 peak가 3개 있는 복잡한 형태이지만 최고수위가 상대적으로 낮은 편이다. data7은 자료의 길이가 가장 길고, 형태도 복잡하며, 첨두수위가 가장 크다. 특히 data7의 첨두수위는 다른 모든 자료보다 커서 training data와 checking data에 사용된 수위보다 큰 값을 갖는다(Table 2).

Table 2. Characteristics of Selected Data

	data length(day)	max. water level (m)	number of peak
data 4	4	4.92	3
data 5	1.5	5.73	1
data 7	5.5	6.13	6

Testing data의 적용에 따른 수위예측 결과를 height error를 통해 비교하면, 전체적으로 data4의 결과가 가장 우수하게 나타났다(Table 3). Median으로부터 25% 편차의 분포가 비교적 좁은 범위 안에 밀집되어있고, 전체 편차 역시 가장 적은 값을 나타냈다. data5의 경우 T+14 ~ T+18에서는 가장 우수한 median을 나타내었고, 25% 편차 분포도 양호하지만 전체 편차는 가장 넓게 분포하였다. data7의 경우 median이 data4, data5에 비해 상당히 크게 나타났고, 편차의 분포도 비교적 넓게 분포되어있다.

RMSE를 통해 비교해 보면, 가장 좋은 결과를 보인 경우는 data4를 testing data로 사용한

경우로 height error를 통한 비교와 동일했다. 25% 편차의 분포는 상당히 밀집되어있어 높은 정확도를 보였다. data5와 data7의 median을 비교할 경우 height error의 경우와 달리 data7이 data5보다 우수한 예측 결과를 보였다. 편차의 분포 역시 data7의 경우 상대적으로 밀집된 분포를 보였다.

Testing data의 적용에 따른 홍수예측 정확도를 전체적으로 정리해보면, 모형을 training data 및 checking data보다 큰 값을 갖는 testing data에 적용하게 될 경우 정확도가 낮아지는 것으로 판단된다. 따라서 모형을 구축할 경우 관측된 강우 및 홍수사상 중 가장 큰 자료를 사용하여 모형을 구축하는 것이 이후 홍수량 예측에서 정확도를 높일 수 있는 방법이라 판단된다.

Table 3. Comparison as Applying of Testing Data

	Height Error (%)		RMSE (m)	
	Average	Median	Average	Median
Data 4 for Testing	18.50	17.59	0.28	0.26
Data 5 for Testing	22.14	19.25	0.44	0.42
Data 7 for Testing	32.81	30.91	0.40	0.38

#### 4. 결론

본 연구에서는 관측 강우자료와 수위자료를 Neuro-Fuzzy 기법에 적용하여 탄천유역의 하류에 위치한 대곡교의 수위를 예측하였다. 모형의 입력자료는 2007년부터 2011년까지의 관측자료 중 7개의 홍수사상을 추출하여 사용하였으며, 모형운행을 위해서 10분 간격의 자료를 사용하고,  $t+1 \sim t+18$ 까지의 수위를 모의하였다.

모형의 입력자료 구성에 사용된 강우자료와 수위자료의 시간에 따른 비교에서는 두드러진 특징은 나타나지 않았으나 RMSE의 비교에서 Model A ( $P_{t-1}, H_t$  자료 사용)의 경우 비교적 가장 작은 오차를 보였다. 이는 탄천유역의 크기가 비교적 크지 않기 때문에 10 ~ 20분 이전에 관측된 강우와 수위의 영향이 유출에 가장 큰 영향을 미치기 때문인 것으로 생각된다.

Training data로 사용된 자료의 길이가 너무 짧을 경우 예측값의 편차가 크게 발생할 가능성이 있으며, testing data의 적용에 따른 홍수예측 정확도를 통해서도 나타났듯이 최대홍수량이 큰 자료를 training data로 사용하는 것이 예측 정확도를 높이는 데 도움이 된다. 또, peak가 많이 발생하여 수위의 증감이 많은 것이 우수한 모형을 만드는데 도움이 되는 것으로 판단된다. 따라서 ANFIS 모형을 구축할 경우 기존 관측 자료 중 가장 크고 복잡한 자료를 사용하여 모형을 구축하는 것이 홍수량의 예측 정확도를 높일 수 있을 것으로 판단된다.

#### 감 사 의 글

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2011-0029851).

#### 참 고 문 헌

1. Zadeh, L.A., Fuzzy Sets. Information and Control, pp. 338-353, 1965.
2. Mamdani, E.H. & Assilian, S., An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. International Journal of Man-Machine Studies, 7(1), pp. 1-13, 1975.
3. Jang, J.S., ANFIS: Adaptive-Networkbased Fuzzy Inference Systems, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 3, pp. 665-685, 1993