

# 가강수량과 인공신경망을 이용한 중규모수치예보의 강수확률예측 개선기법

## Improving Probability of Precipitation of Meso-scale NWP Using Precipitable Water and Artificial Neural Network

강부식\*, 이봉기\*\*  
Kang, Boosik, Lee, Bongki

### 요 지

본 연구는 한반도 영역을 대상으로 2001년 7, 8월과 2002년 6월로 홍수기를 대상으로 RDAPS 모형, AWS, 상층기상관측(upper-air sounding)의 자료를 이용하였다. 또한 수치예보자료를 범주적 예측확률로 변환하고 인공신경망기법(ANN)을 이용하여 강수발생확률의 예측정확성을 향상시키는데 있다. 신경망의 예측인자로 사용된 대기변수는 500/ 750/ 1000hpa에서의 지위고도, 500-1000hpa에서의 층후(thickness), 500hpa에서의 X와 Y의 바람성분, 750hpa에서의 X와 Y의 바람성분, 표면풍속, 500/ 750hpa/ 표면에서의 온도, 평균해면기압, 3시간 누적 강수, AWS관측소에서 관측된 RDAPS모형 실행전의 6시간과 12시간동안의 누적강수, 가강수량, 상대습도이며, 예측변수로는 강수발생확률로 선택하였다. 강우는 다양한 대기변수들의 비선형 조합으로 발생되기 때문에 예측인자와 예측변수 사이의 복잡한 비선형성을 고려하는데 유용한 인공신경망을 사용하였다. 신경망의 구조는 전방향 다층퍼셉트론으로 구성하였으며 역전파알고리즘을 학습방법으로 사용하였다. 강수예측성과의 질을 평가하기 위해서 2×2 분할표를 이용하여 Hit rate, Threat score, Probability of detection, Kuipers Skill Score를 사용하였으며, 신경망 학습후의 강수발생확률은 학습전의 강수발생확률에 비하여 한반도영역에서 평균적으로 Kuipers Skill Score가 0.2231에서 0.4293로 92.39% 상승하였다.

**핵심용어:** 인공신경망기법, 강수발생확률, 역전파알고리즘, Kuipers Skill Score

### 1. 서 론

한반도 지역의 강우는 6월과 9월 사이에 연강수량의 67%이상이 집중되는 계절성을 보인다(건설교통부, 2006). 홍수예방과 안정적인 용수공급을 위해서 하천의 유량예측을 위한 강수예측은 수자원 관리를 위해 필요하다. 미국 기상청(NWS)에서는 강수확률예보를 제공하고 있지만 국내의 경우 정성적 범주예보만이 제공되고 있으며, 정확도 측면에서도 기대에 못 미치는 수준이다(과학기술부, 2007). 국내 기상청에서는 RDAPS 및 GDAPS모형을 이용하여 정량 강수예측을 제공하고 있지만, 유역규모에서의 수문학적 활동을 위해서는 별도의 보정기법이 필요하다. 수치예보모형의 예측값은 초기조건 오차, 모형오차, 예측의 본질적 한계로 인해 정확하지 않다. 이러한 수치예보 모형의 정확도를 향상시키기 위해서 예측인자와 예측변수간의 통계적 관계를 이용하는

\* 정회원-단국대학교 토목환경공학전공 조교수E-mail : [bskang@dankook.ac.kr](mailto:bskang@dankook.ac.kr)  
\*\* 정회원-단국대학교 토목환경공학과 석사과정E-mail : [leebk@dankook.ac.kr](mailto:leebk@dankook.ac.kr)

Perfect-Prog기법(Klein, 1971)과 MOS(Model Output Statistics)기법이 개발되어 왔다. Hall(1997)은 Dalas-Fort Worth지역에 인공신경망을 이용하여 강수발생확률과 정량적 강수예측의 향상을 보인바 있으며, Kuligowski & Barros(1998), Valverde Ramirez(2005)는 다중선형회귀와 ANN을 각각 적용하여 강수결과를 비교함으로써 비선형성을 고려하는 ANN이 신뢰성이 있음을 설명하였다. 강수는 다양한 대기변수들의 비선형 조합으로 발생되기 때문에 많은 예측인자(predictors)와 예측변수(predictands)사이의 복잡한 비선형성을 고려하는데 유용한 ANN기법은 강수예측에 있어서 좋은 보정기법이 될 수 있다.

따라서 본 연구에서는 상층기상관측자료에서 관측된 가강수량, RDAPS모형의 수치예보자료, AWS에서의 관측자료를 예측인자로 구성하고 강수발생확률을 예측변수로 구성하여 ANN기법을 적용하였으며 RDAPS모형의 예측결과와 신경망 학습후의 예측성과를 비교하였다.

## 2. 인공신경망(Artificial Neural Network)과 예측인자 구성

인공신경망은 고도로 복잡한 인간의 두뇌를 구성하고 있는 기본단위인 뉴런의 개념을 도입한 이론을 기초로 인식과정을 수학적인 모형으로 일반화시키기 위해 개발된 병렬정보처리 시스템이다. 뉴런은 임계값(threshold)을 초과하면 활성화되며, 임계값 이하에서는 활성화 되지 않는다. 이번 연구에서는 일반적으로 사용되는 전방향 다층퍼셉트론구조로 모델링하였으며, 학습방법으로는 역전파알고리즘을 사용하였다. 연구에서 사용된 신경망의 구조와 학습순서는 Fig. 1과 같다.

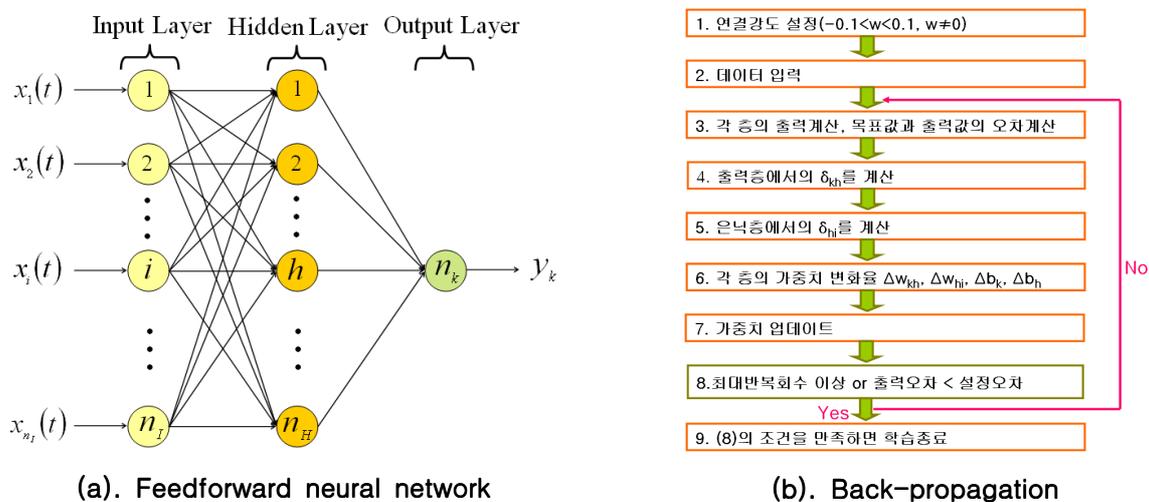


Fig. 1. Feedforward neural network(left) and Back-propagation algorithm

은닉층(Hidden layer)의 활성화함수(activation function)는 비선형성을 고려하기 위해서 비선형 양극(bipolar) 활성화함수를 사용하였다. 신경망의 중요한 특징은 학습을 통하여 예측인자와 예측변수사이의 관계를 학습하고 예측인자와 예측변수의 사이의 관계를 추정할 수 있다는데 있다. 본 연구에서 사용된 신경망의 구조에 따른 수학적 방정식은 Eq. (1)과 같이 표현할 수 있다.

$$y_k = g\left(\sum_{h=0}^{n_H} w_{kh} f\left(\sum_{i=0}^{n_I} w_{hi} x_i(t)\right)\right) \quad (1)$$

여기서,  $g(\cdot)$ 는 출력층에서의 활성화함수,  $f(\cdot)$ 는 은닉층에서의 활성화 함수,  $w_{kh}$ (또는  $w_{hi}$ )는

은닉(또는 입력)뉴런과 출력(또는 은닉)뉴런 사이의 가중치,  $x_i(t)$ 는 입력값,  $y_k$ 는 출력값,  $n_I$ (또는  $n_H$ )는 각각 입력뉴런의 수(또는 은닉뉴런의 수)이다.

역전파알고리즘은 예측인자와 예측변수사이의 연결가중치를 조절하기 위해서 목표값(예측변수)과 출력값의 오차를 이용하며, 본 연구에서는 오차제곱합함수(Sum-squared error function)을 사용하였다(Eq. (2)). 또한 가중치의 업데이트를 위해서 경사하강(gradient descent)법을 사용하며, 활성화함수는 반드시 미분 가능한 연속함수이어야 한다.

$$E = \sum (d_k - y_k)^2 \quad (2)$$

여기서,  $d_k$ 는 목표값,  $y_k$ 는 출력값이다.

신경망의 예측인자의 구성은 Table. 1과 같으며, 가강수량은 증기압과 일정 고도사이에서의 비습도를 적분하여 Eq. (3)~(5)와 같이 산정할 수 있다(Bras, 1990).

Table 1. Input variables

Input Variables	Variable type	Source
Geopotential height 500 / 750/ 1000 hPa [m]	Continuous	RDAPS
Atmospheric thickness 500 hPa - 1000 hPa [m]	Continuous	RDAPS
X, Y-component of wind at 500 hPa [m/s]	Continuous	RDAPS
X, Y-component of wind at 750 hPa [m/s]	Continuous	RDAPS
Wind speed at surface [m/s]	Continuous	RDAPS
Temperature at 500/ 750 hPa/ surface [°K]	Continuous	RDAPS
Mean sea level pressure [hPa]	Continuous	RDAPS
3-hr accumulated precipitation [mm/(3hr)]	Continuous	RDAPS
Occurrence of observed Precipitation	Categorical	RDAPS
Precipitation accumulated in 6 & 12 hs previous to RDAPS run [mm/(6hr)] [mm/(12hr)]	Continuous	AWS
Precipitation occurrence in 6 & 12 hs previous to RDAPS run	Categorical	AWS
Relative humidity measured 0 & 12 hs before RDAPS run	Continuous	Upper-air sounding
Precipitable water measured 0 & 12 hs before RDAPS run [mm]	Continuous	Upper-air sounding
Precipitable water difference in 12 hs previous to RDAPS run [mm]	Continuous	Upper-air sounding

$$e = 611 \exp \left( \frac{L_v}{R_v} \left( \frac{1}{273.15} - \frac{1}{T_d} \right) \right) \quad (3)$$

$$q_h = \frac{0.622e}{P - 0.378e} \approx 0.622 \frac{e}{P} \quad (4)$$

$$w_p = \frac{1}{g} \int_p^{P_0} q_h dP \quad (5)$$

여기서,  $e$ 는 증기압,  $L_v$ 는 물의 증발과정에서의 잠재열( $2.5 \times 10^6$  J/kg),  $R_v$ 는 수증기기체상수( $461.5$  J/K/kg),  $T_d$ 는 이슬점 온도(K),  $q_h$ 는 비습도,  $P$ 는 대기압,  $p_0$ 와  $p$ 는 고도  $z_1$ 과  $z_2$ 에서의 압력,  $w_p$ 는 가강수량(mm)이다.

### 3. 예측정확성 평가

강수예측성과를 평가하기위해서 2×2 분할표(2×2 contingency table; Table 2)로부터 산정된 스칼라점수를 사용하였다. 본 연구에서는 적중률(H: Hit rate), 위험점수(TS; Threat Score), 탐지확률(POD; Probability Of Detection), Kuipers Skill Score(KSS)를 사용하여 예측자료를 검정하였다(Wilks, 1995). 적중률(H)는 가장 직관적 방법으로써 총 사상에 대하여 예측이 정확한 경우의 비를 나타낸다. 적중률은 0에서 1사이의 값을 가지며, 완벽한 예보의 경우 1값을 가진다. 위험점수(TS)는 적중률의 대안적인 방법으로 종종 사용되며, 위험점수는 0에서 1사이의 값을 갖는데 완벽한 예보의 경우 1값을 가진다. 탐지확률(POD)는 관측강우가 'yes'일 때의 횟수에 대한 강우예측이 'yes'인 횟수의 비이며 완벽한 예측인 경우 1값을 가진다. 마지막으로 Kuipers Skill Score(KSS)는 예측정확성의 질을 판단하기 위한 방법으로써 0에서 1사이의 값이며 완벽한 예측의 경우 1값을 가진다. Fig. 2와 Table. 3는 각 지역별로 강수확률예측정확도에 있어 RDAPS예측에 대한 인공신경망예측의 상대적 개선효과를 보여주고 있다.

Table 2. 2X2 contingency table

		Observation	
		Yes	No
Forecast	Yes	hits (a)	false alarms (b)
	No	misses (c)	correct negative (d)

$$H = \frac{(a + d)}{n}$$

$$TS = \frac{a}{(a + b + c)}$$

$$POD = \frac{a}{(a + c)}$$

$$KSS = \frac{(ad - bc)}{(a + c)(b + d)}$$

$$n = a + b + c + d$$

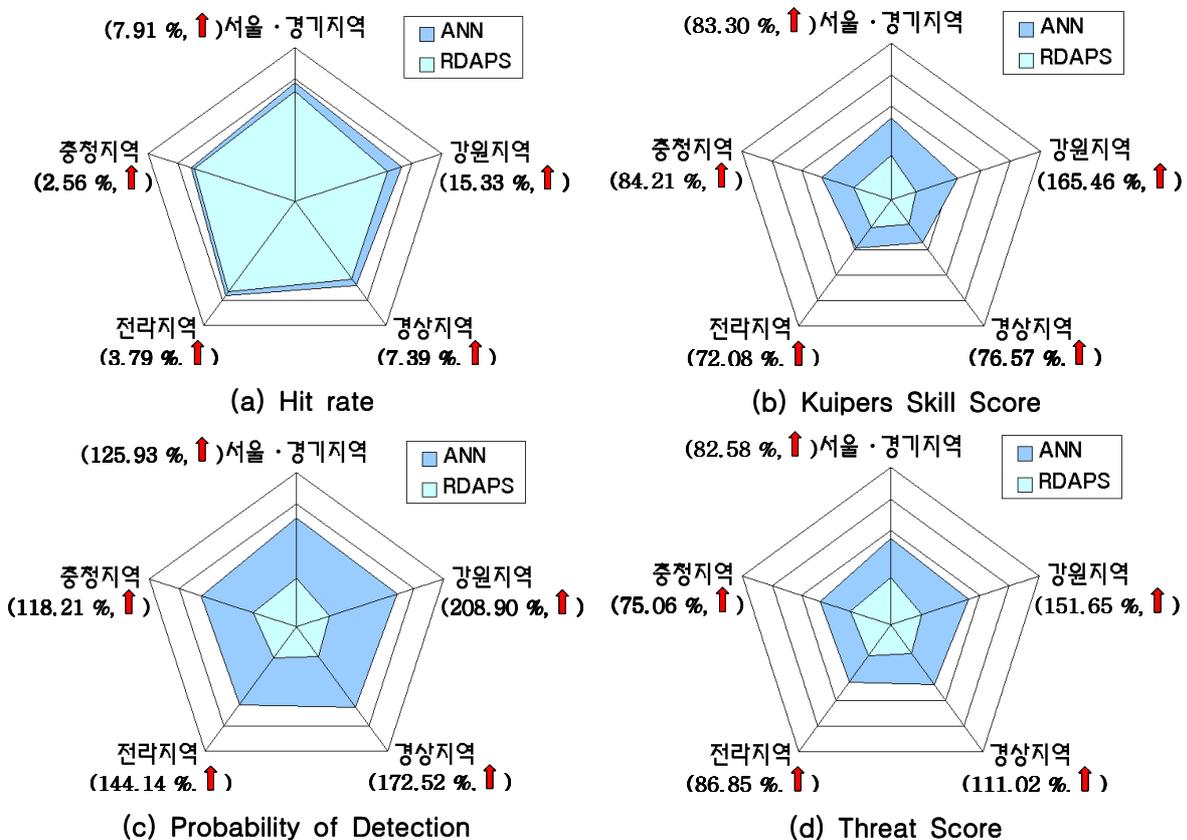


Fig. 2. 각 지역별 예측정확성 평가

Table. 3 Improvement in evaluation score(H, KSS, TS, POD)

	Hit rate		KSS		TS		POD	
	RDAPS	ANN	RDAPS	ANN	RDAPS	ANN	RDAPS	ANN
서울·경기지역	0.72	0.77	0.28	0.52	0.30	0.55	0.31	0.71
강원지역	0.63	0.73	0.17	0.44	0.21	0.52	0.22	0.69
경상지역	0.63	0.67	0.19	0.34	0.22	0.47	0.24	0.64
전라지역	0.73	0.76	0.22	0.38	0.24	0.45	0.26	0.62
충청지역	0.69	0.70	0.25	0.46	0.27	0.48	0.30	0.64

#### 4. 결 론

다양한 대기변수들의 비선형 조합으로 생성되는 강수는 예측정확성의 질적인 평가에 있어서 RDAPS모형의 수치예보자료보다 ANN기법을 이용한 결과 각 지역별로 적중률, 위험점수, 탐지확률, KSS가 향상 되었다. 개선효과가 가장 큰 스코어는 탐지확률이었으며, 한반도 전체에 대해서는 탐지확률은 평균적으로 0.26에서 0.66으로 150%향상하였으며, 각 지역별로 서울·경기지역 125.93%, 강원지역 208.90%, 경상지역 172.52%, 전라지역 144.14%, 충청지역 118.21% 개선되었다. 예측정확성 질의 평가결과, 강원지역에서의 예측정확도는 적중률 15.33%, KSS 165.46%, 위험점수 151.65%, 탐지확률이 208.90%로 한반도영역 중 가장 많이 개선되었다. 강원지역에서의 RDAPS모형의 예측확률은 다른 지역에 비하여 예측확률이 다소 낮았으나, 인공신경망기법을 이용한 결과 강수발생확률이 개선되었음을 확인할 수 있었다. 따라서 가강수량, AWS, RDAPS모형의 자료를 예측인자로 구성한 ANN기법은 강수발생확률 예측에 있어서 기존의 RDAPS모형의 예측값보다 강수발생확률이 향상되었음을 알 수 있었다.

#### 참 고 문 헌

1. 과학기술부(2007). **유역 물 관리 운영 기술개발 1, 21세기 프로티어 「수자원의 지속적 확보 기술개발 사업」**, 한국수자원공사
2. 건설교통부(2006). **수자원장기종합계획(2006-2020) 보고서**, 한국수자원공사
3. Bras, R.L.(1990). *Hydrology: An Introduction to Hydrologic Science*. Addison-Wesley publishing company, pp. 82-92.
4. Hall, T., Brooks, H.E., and Doswell, III C.A.(1999). Precipitation Forecasting Using a Neural Network. *Wea. Forecasting*, vol. 14, issue 3, pp. 338-345.
5. Klein, W.H.(1971). Computer Prediction of Precipitation Probability in the United States. *J. Appl. Meteor.*, vol. 10, issue 5, pp. 903-915.
6. Kuligowski, R.J., Barros, A.P. and Ferreira, N.J.(1998). Experiments in Short-Term Precipitation Using Artificial Neural Networks. *Mon. Wea. Rev.*, vol. 126, issue 2, pp. 1194-1204.
7. Valverde Ramirez, M.C., de Campos Velho, H.F.(2005). Artificial neural network technique for rainfall forecasting applied to the São Paulo region. *J. Hydrol.*, Vol. 301, issue 1-4, pp. 146-162.
8. Wilks, D.S.(1995). *Statistical methods in the Atmospheric Sciences*. Academic Press, pp. 220-250.