

# 인공신경망과 중규모기상수치예보를 이용한 강수확률예측

## Predicting Probability of Precipitation Using Artificial Neural Network and Mesoscale Numerical Weather Prediction

강부식\* · 이봉기\*\*

Kang, Boosik · Lee, Bongki

### Abstract

The Artificial Neural Network (ANN) model was suggested for predicting probability of precipitation (PoP) using RDAPS NWP model, observation at AWS and upper-air sounding station. The prediction work was implemented for flood season and the data period is the July, August of 2001 and June of 2002. Neural network input variables (predictors) were composed of geopotential height 500/750/1000 hPa, atmospheric thickness 500-1000 hPa, X & Y-component of wind at 500 hPa, X & Y-component of wind at 750 hPa, wind speed at surface, temperature at 500/750 hPa/surface, mean sea level pressure, 3-hr accumulated precipitation, occurrence of observed precipitation, precipitation accumulated in 6 & 12 hrs previous to RDAPS run, precipitation occurrence in 6 & 12 hrs previous to RDAPS run, relative humidity measured 0 & 12 hrs before RDAPS run, precipitable water measured 0 & 12 hrs before RDAPS run, precipitable water difference in 12 hrs previous to RDAPS run. The suggested ANN has a 3-layer perceptron (multi layer perceptron; MLP) and back-propagation learning algorithm. The result shows that there were 6.8% increase in Hit rate (H), especially 99.2% and 148.1% increase in Threat Score (TS) and Probability of Detection (POD). It illustrates that the suggested ANN model can be a useful tool for predicting rainfall event prediction. The Kuipers Skill Score (KSS) was increased 92.8%, which the ANN model improves the rainfall occurrence prediction over RDAPS.

**Keywords :** Artificial Neural Network, back-propagation, probability of precipitation, RDAPS, Kuipers Skill Score

### 요 지

한반도 영역을 대상으로 RDAPS모형의 수치예보자료, AWS의 관측강수, 상층기상관측(upper-air sounding)의 관측자료를 이용하여 권역별 강수발생확률을 예측할 수 있는 인공신경망 모형을 제시하였다. 사용된 자료의 기간은 2001년 7, 8월과 2002년 6월로 홍수기를 대상으로 하였다. 500/750/1000 hPa에서의 지위고도, 500-1000 hPa에서의 층후(thickness), 500 hPa에서의 X와 Y방향 바람성분, 750 hPa에서의 X와 Y방향 바람성분, 표면풍속, 500/750 hPa/표면에서의 온도, 평균해면 기압, 3시간 누적 강수, AWS관측소에서 관측된 RDAPS모형 실행전의 6시간과 12시간동안의 누적강수, 가강수량, 상대습도 등을 신경망의 예측인자로 사용하였다. 신경망의 구조는 3층 MLP(Multi Layer Perceptron)로 구성하여 역전파알고리즘(Back-propagation)을 학습방법으로 사용하였다. 신경망예측결과 한반도전체에 대한 예측성과의 개선은 H가 6.8%상승하였고, 특히 TS와 POD는 각각 99.2%와 148.1% 상승함으로써 강수예측에 대한 신경망모형이 효과적인 도구가 될 수 있음을 확인하였다. KSS 역시 92.8% 개선됨으로서 RDAPS 예측에 비하여 뚜렷이 개선된 결과를 보여주고 있다.

**핵심용어 :** 인공신경망, 역전파알고리즘, 강수발생확률, RDAPS, Kuipers Skill Score

### 1. 서 론

한반도 지역의 강우는 6월과 9월 사이에 연강수량의 67% 이상이 집중되는 계절성을 가지고 있다(건설교통부, 2006). 홍수관리나 용수공급을 위해서 하천의 유량예측은 수자원 관리의 가장 기본적인 요소이다. 최근 신뢰성 있는 유량예측을 위하여 앙상블 유량예측기법(김영오 등, 2001; 정대일, 김영오, 2002)이 시도되고 있으며 실효성 있는 적용을 위해서는

정확한 강우예측이 필수적으로 요구된다. 수문기상학적 변수들을 예측하기 위하여 수치예보모형(Numerical Weather Prediction model)이 개발되어 왔다. 미국 기상청(NWS; National Weather Service)에서는 이미 강수확률예보를 제공하고 있지만 국내의 경우 정성적 범주예보만이 제공되고 있는 실정이고 정확도 측면에서도 기대에 못 미치는 수준이다(교육과학기술부, 2007). 국내 기상청에서는 RDAPS와 GDAPS모형을 이용하여 중기와 장기에 대한 정량 강수예측

\*정회원 · 교신저자 · 단국대학교 공과대학 토목환경공학전공 조교수 (E-mail : bskang@dankook.ac.kr)

\*\*정회원 · 단국대학교 대학원 토목환경공학전공 석사과정 (E-mail : leebk@dankook.ac.kr)

을 제공하고 있지만 유역규모의 수문학적 활동을 예측하기 위해서 강수확률이나 정량강수측정에 있어서 별도의 보정기법이 필요하다. 이러한 수치예보모형의 예측값은 초기조건 오차, 모형오차, 예측의 본질적 한계로 인해 많은 부분에서 오차가 발생하기 때문에 수치예보모형의 예측값은 정확하지 않다. 이러한 수치예보모형의 정확도를 향상시키기 위하여 2가지 형태의 통계적 예측기법이 개발되어 왔다. 첫 번째 통계적 접근방법은 수치예보모형으로부터 확정론적 역학예보를 활용하기 위한 Perfect-Prog Forecasts(Perfect-Prog Method; PPM)이다. Perfect-prog 기법은 예측인자와 예측변수간의 통계적 관계를 이용하는 기법이다(Klein, 1971). Perfect-prog 기법은 수치예보모형이 가질 수 있는 오차나 편이의 보정을 고려하지 않고 역학적 수치모형의 예측치가 정확하다고 가정한다(Wilks, 1995). 하지만 실제적으로 예측인자에 대한 모의는 완벽할 수 없기 때문에 이를 보완하기 위하여 Model Output Statistics(MOS)기법을 사용한다. 또한 Perfect-prog 기법은 동일 시간대에서의 예측인자와 예측변수를 이용하여 회귀식에 적용하기 때문에 시간지체를 고려하지 못한다. 두 번째로 Glahn과 Lowry(1972)의 Model Output Statistics(MOS)후처리 기법이다. 이 기법은 관측값뿐 아니라 수치예보를 통계예측모형수립에 활용하는 기법으로서 목표로 하는 예측량과 NWP에 의해 예측된 변수와의 최적 관계를 결정하며 크게 선형적인 방법과 비선형적인 방법이 있다. 선형적인 방법에는 다중선형회귀(MLR; Multi Linear Regression)가 가장 일반적으로 사용 되는데 독립변수와 종속변수사이의 선형적 관계를 가정하기 때문에 두 변수가 관련 있음에도 불구하고 비선형적 관계를 설명하지 못한다(Schoof & Pryor, 2001). 따라서 선형성을 기본으로 하는 MLR은 비선형성이 강한 강수의 보정에 적합하지 않으며 비선형성을 고려하는 신경망기법의 연구가 진행되어 왔다. 국내에서는 강수량의 장기예측기법(김호준 등, 2001), 강수량 규모축소기법(안중배 등, 2003; 차유미, 안중배, 2005)에 인공지능기법을 적용한 바 있고, 김광섭(2006)은 대상유역의 6시간 선행 강우를 예측하기위해서 지속성을 고려한 강우예측, AWS 자료만을 이용한 신경망 모형의 강우예측 그리고 포항 라디오존데로부터 획득한 700 mb에서의 풍향자료와 AWS 자료를 이용하여 신경망학습후의 강우량 예측을 비교한 결과 상층 기상자료(upper-air sounding)를 사용한 경우에 강우량예측에 있어서 개선된 결과를 보였다. 국외에서도 Hall(1997)은 Dalas-Fort Worth지역에 인공지능기법을 이용하여 강수발생확률과 정량적 강수예측의 향상을 보인 바 있으며, Kuligowski와 Barros(1998)는 다중선형회귀와 인공지능기법을 각각 적용하여 강수결과를 비교함으로써 비선형성을 고려한 인공지능기법이 더 뛰어난 설명력을 보였다. Valverde Ramirez(2005)는 브라질 상파울로지역에서 강우예측을 위해 인공지능기법과 다중선형회귀를 비교연구한 결과 인공지능기법이 다중 선형 회귀모형보다 신뢰성이 있음을 나타내었다. 따라서 강수는 다양한 대기변수들의 비선형 조합으로 발생되기 때문에 많은 예측인자(predictors)와 예측변수(predictands)사이의 복잡한 비선형성을 고려하는데 유용한 인공지능기법(ANN; Artificial Neural Network)은 강수예측에 있어서 좋은 보정기법이 될 수 있다.

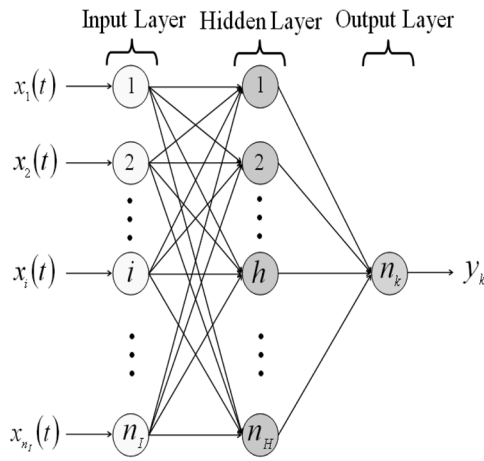
본 연구의 목적은 한반도 영역을 대상으로 RDAPS모형의 수치예보자료, AWS의 관측강수, 상층기상자료를 이용하여 신경망학습을 통해 수치예보자료의 강수발생확률의 예측정확성을 개선시키는데 있다. 자료의 기간은 2001년 7월 1일 00시부터 8월 31일 24시와 2002년 6월 1일 00시부터 6월 30일 24시까지로 홍수기를 대상으로 연구를 수행하였다. 신경망의 예측인자로 사용된 대기변수는 500/750/1000 hPa에서의 지위고도, 500-1000 hPa에서의 층후(thickness), 500 hPa에서의 X와 Y의 바람성분, 750 hPa에서의 X와 Y의 바람성분, 표면풍속, 500/750 hPa/표면에서의 온도, 평균해면 기압(mean sea level pressure), 3시간 누적강수, AWS관측소에서 관측된 RDAPS모형 실행전의 6시간과 12시간동안의 누적강수, 가강수량(precipitable water), 상대습도이며, 예측변수로는 강수발생확률을 선택하였다. 사용한 신경망의 구조는 3층 MLP로 구성하여 역전파알고리즘을 학습방법으로 사용하였다. 따라서 이 연구에서는 비선형성이 강한 수치예보 모형의 출력값을 인공지능기법에 적용시킴으로써 강수발생확률(PoP; Probability of Precipitation)의 정확성을 향상시키는데 있다. 신경망의 임계값(threshold)은 비편이(no bias)일 때의 확률로 산정하였으며, 선택오차(selection error)와 적용오차(test error)와의 차이가 근소할 때 신경망의 일반화가 잘 되었다고 볼 수 있으므로 오차의 차이가 작은 신경망 모형을 선택하였다. 신경망의 학습을 통한 예측정확성의 검정은 Hit rate(H), Threat score(TS), Probability of Detection (POD), Kuipers Skill Score(KSS)를 통하여 검정하였다.

## 2. 인공지능기법(ANN)

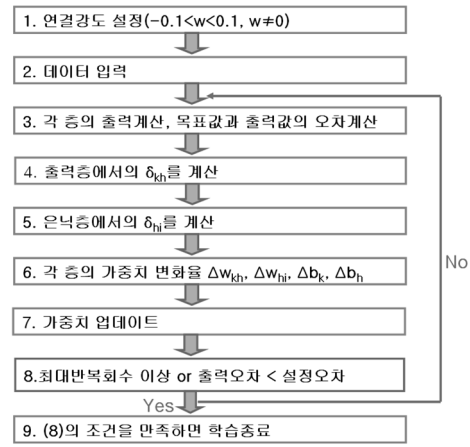
인공지능기법은 고도로 복잡한 인간의 두뇌를 구성하고 있는 기본단위인 뉴런(neuron)의 개념을 도입한 이론을 기초로 인식과정을 수학적 모형으로 일반화시키기 위해 개발된 병렬정보처리 시스템이다. 생물학적으로 각각의 뉴런은 전기적 신호를 발생시키는데 수상돌기(dendrite)로부터 입력을 받아 축색돌기(axon)를 통하여 신호를 보낸다. 각각의 뉴런은 시냅스(synapse)라는 연결물체가 있는데, 이는 자극에 대한 반응을 다른 뉴런에 전달해 주는 매개체 역할을 한다. 또한 각각의 뉴런은 임계값(threshold)을 초과하는 전기적 신호를 받으면 활성화되며, 임계값 이하에서는 활성화 되지 않는다. 따라서 이러한 생물학적 뉴런의 작용원리를 수학적 기법을 통해 인공지능기법이 개발되었다.

일반적으로 가장 많이 사용되는 신경망은 다층퍼셉트론(MLP)이다. MLP는 입력층과 출력층 사이에 하나 이상의 은닉층(hidden layer)이 존재하는 형태의 신경망의 구조이다. 입력층은 각각의 입력변수들이 입력되는 층을 말하며, 은닉층(출력층)은 입력층(은닉층)의 입력변수들이 활성화함수(activation function)를 통하여 출력되는 층을 말한다. 본 연구에서 사용한 전방향 구조(feedforward)의 3층 신경망은 Fig. 1(a)와 같으며, 역전파알고리즘의 학습순서는 Fig. 1(b)와 같다.

은닉층의 활성화 함수(activation function or sigmoidal function)는 신경망에 비선형성을 고려하기위해서 비선형 양극(bipolar) 시그모이드함수(sigmoidal function)를 사용하였으며 활성화함수에서 계산된 값의 범위는 -1과 1사이의 값을



(a) Feedforward Neural Network



(b) Backpropagation algorithm(정슬, 2004)

Fig. 1 Feedforward neural network and back-propagation algorithm

가진다(Gupta, 2003). 신경망의 중요한 특징은 학습(training 또는 learning)을 통하여 예측인자(predictor)와 예측변수(predictand)사이의 관계를 학습하고 예측인자와 예측변수 사이의 관계를 추정한다. 이 연구에서 사용한 신경망은 출력층에서 임계값(threshold)을 고려하여 강수발생확률을 계산하며 수학적 방정식은 Eq. (1)과 같이 표현할 수 있다.

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

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),  $\theta_h$ 는 은닉층과 출력층사이의 임계값이다.

역전파알고리즘은 예측인자와 예측변수사이의 연결가중치를 조절하기 위해서 목표값(예측변수)과 출력값의 오차를 이용한다. 본 연구에서는 오차제곱합함수(Sum-squared error function)를 사용하였다. 또한 가중치의 업데이트를 위해 경사하강(gradient descent 또는 Delta-rule)법을 사용하며 이때 사용되는 활성화함수는 반드시 미분이 가능한 연속함수여야 한다. 이번 연구에서 사용된 오차함수는 Eq. (2)와 같다.

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

여기서,  $d_k$ 는 목표값,  $y_k$ 는 출력값이다.

### 2.1 각 지역별 신경망의 구조

은닉층의 뉴런의 수가 증가 할수록 일반적으로 예측인자와

예측변수사이의 더 많은 패턴을 인식하고 추정할 수 있지만, 과다학습(over training)을 피하기 위해서 학습성과(train performance)가 선택성과(selection performance)보다는 커야 하고, 선택오차(selection error)와 적용오차(test error)와의 차이가 근소할수록 신경망은 일반화가 잘 되었다고 볼 수 있다. 따라서 각 지역별로 오차의 차이가 적은 신경망의 구조를 선택하였다. 서울·경기지역의 경우 오차의 차이가 0.001이며 상대오차 비율은 -0.3%, 강원지역 2.0%, 전라지역 -3.2%, 경상지역 14.6%, 충청지역은 3.5%의 상대오차 비율을 보이고 있다. 각 지역별 신경망의 구조와 상대적 오차는 Table 1과 같다. 한반도영역에서 각 지역별 강수발생확률을 예측하기위한 예측인자의 구성을 하기위해 민감도분석을 실시하여, 민감도 분석결과를 토대로 각 지역별로 신경망 예측인자로 사용하였으며 2.3절의 민감도 분석에서 설명하였다.

### 2.2 신경망 입력데이터의 구성

이 연구에서는 한국기상청에서 제공 되어진 RDAPS모형의 수치예보자료, AWS의 관측강수, 상층기상관측에서 관측된 가능강수량과 상대습도를 사용한다. 사용되어진 자료의 기간은 2001년 7, 8월과 2002년 6월(홍수기)에 대해서 신경망의 입력변수(predictors)로 구성하였다. RDAPS모형의 예측자료는 30×30 km 격자로 구성된 대기변수들로 구성되어있으며 Fig. 2는 RDAPS모형의 출력격자와 상층기상관측소(upper-air sounding)의 위치를 설명하고 있다. RDAPS모형의 수치예보 자료는 매일 12 UTC와 12UTC에 48시간까지의 대기예측정보가 3시간 단위로 생산된다. 상층기상관측(백령도, 오산, 속

Table 1. The artificial neural network structure and relative errors

지역	신경망구조	선택오차	적용오차	상대오차 비율
서울·경기지역	19:19-15-1:1	0.299	0.298	-0.3%
강원지역	17:17-14-1:1	0.293	0.299	2.0%
전라지역	22:22-13-1:1	0.291	0.282	-3.2%
경상지역	21:21-19-1:1	0.258	0.302	14.6%
충청지역	19:19-16-1:1	0.307	0.318	3.5%

※aa:bb-cc-dd:ee : aa와 bb는 신경망의 입력변수의 개수와 입력층에서의 입력뉴런의 수, cc는 하나의 은닉층에서의 은닉뉴런의 개수, dd와 ee는 하나의 출력층에 존재하는 출력뉴런의 수이다.

초, 포항, 광주)에서도 매일 00 UTC와 12 UTC에 상대습도와 이슬점온도 등을 관측하여 가강수량(precipitable water)과 상대습도를 계산하는데 사용된다. 경상지역 연안과 전라지역 연안의 가강수량을 RDAPS격자점에 일치시켜 내삽하기 위해서 임의의 격자점에 광주상층기상관측소의 가강수량값과 같다고 가정하여 크리깅기법을 사용하였다. 가강수량은 증기압(vapor pressure)과 일정 고도사이에서의 비습도(specific humidity)를 적분하여 Eq. (3)~(5)를 통하여 산정된다(Bras, 1990). 또한 기상청에서 운영하는 AWS관측소에서 관측되는 강수는 1시간단위 강수이기 때문에 3시간단위 누적강수로 변환하고, RDAPS모형의 수치예보값과 예측 선행시간을 일치시킨 후, 신경망모형의 입력데이터로 구성하였다.

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

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

$$w_p = \frac{1}{g} \int_p^{p_0} q_h dP \quad (5)$$

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.

### 2.3 입력변수의 민감도 분석

각 지역별로 강수발생확률에 영향을 미치는 변수를 알아보기 위하여 민감도 분석을 실시하였다. 예측인자로 Table 2의 대기변수들을 기준으로 입력변수를 하나씩 제거하면서 출력변수에 미치는 영향을 파악하였으며 민감도 분석에서의 비율(Ratio)은 예측인자를 제거했을때의 목표값과 신경망 출력변수 사이의 오차와 예측인자를 사용했을때의 목표값과 출력변수값의 오차에 대한 상대적비율이다(Eq. (6)). 따라서 R

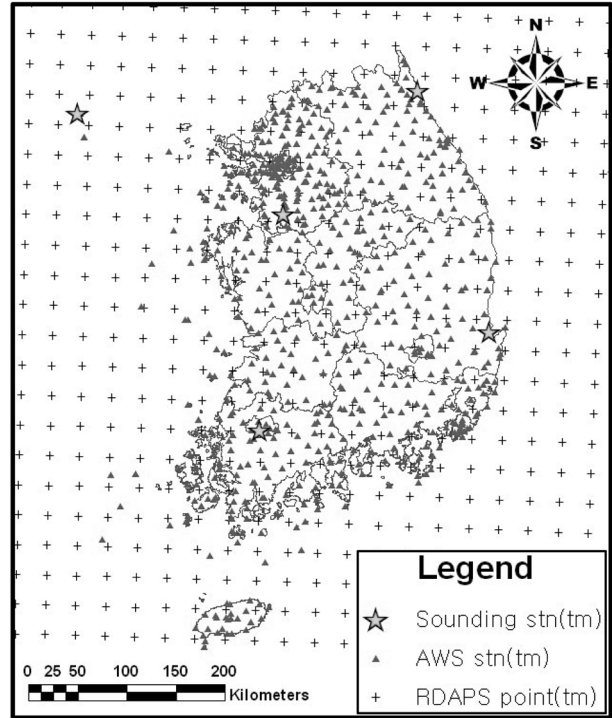


Fig. 2 Map for distribution of RDAPS grid, AWS and sounding stations

이 1보다 작을 경우 이는 예측인자가 신경망의 예측성능에 기여하지 못하고 있다는 것을 의미하며, 반대로 R이 1보다 크다면 예측인자의 기여가 인정된다고 볼 수 있다. 그러나 신경망모형의 통계모형적 특성상 R이 1보다 작다고 해도 예측변수의 물리적의미를 무시해서는 안된다. 다만, 물리적으로 상관성이 큰 변수들은 R값도 대부분 1보다 큰 경향을 보인다.

$$R = \frac{\sum \epsilon_s}{\sum \epsilon_o} \quad (6)$$

Table 2. List of predictors for the artificial neural network model

예측인자	변수명	형식	자료원
Geopotential height 500/750/1000 hPa [m]	RDAPSheight500/RDAPSheight750/ RDAPSheight500	Continuous	RDAPS
Atmospheric thickness 500 hPa-1000 hPa [m]	RDAPSthk1000_500	Continuous	RDAPS
X, Y-component of wind at 500 hPa [m/s]	RDAPSwd500_Xcomp/ RDAPSwd500_Ycomp	Continuous	RDAPS
X, Y-component of wind at 750 hPa [m/s]	RDAPSwd750_Xcomp/ RDAPSwd750_Ycomp	Continuous	RDAPS
Wind speed at surface [m/s]	RDAPSwspdsfc	Continuous	RDAPS
Temperature at 500/750 hPa/surface [°K]	RDAPStmp500/RDAPStmp750/ RDAPStmpsfsc	Continuous	RDAPS
Mean sea level pressure [hPa]	RDAPSmslp	Continuous	RDAPS
3-hr accumulated precipitation [mm/(3hr)]	RDAPSapep	Continuous	RDAPS
Precipitation Occurrence	RDAPSOcc	Categorical	RDAPS
Precipitation accumulated in 6 & 12 hrs previous to RDAPS run [mm/(6 hr)] [mm/(12 hr)]	Prev6Prc/Prev12Prc	Continuous	AWS
Precipitation occurrence in 6 & 12 hrs previous to RDAPS run	Prev6Occ	Categorical	AWS
Relative humidity measured 0 & 12 hrs before RDAPS run	RH0/RH_12	Continuous	Upper-air sounding
Precipitable water measured 0 & 12 hrs before RDAPS run [mm]	PrW0/PrW12	Continuous	Upper-air sounding
Precipitable water difference in 12 hrs previous to RDAPS run [mm]	DiffPrW	Continuous	Upper-air sounding

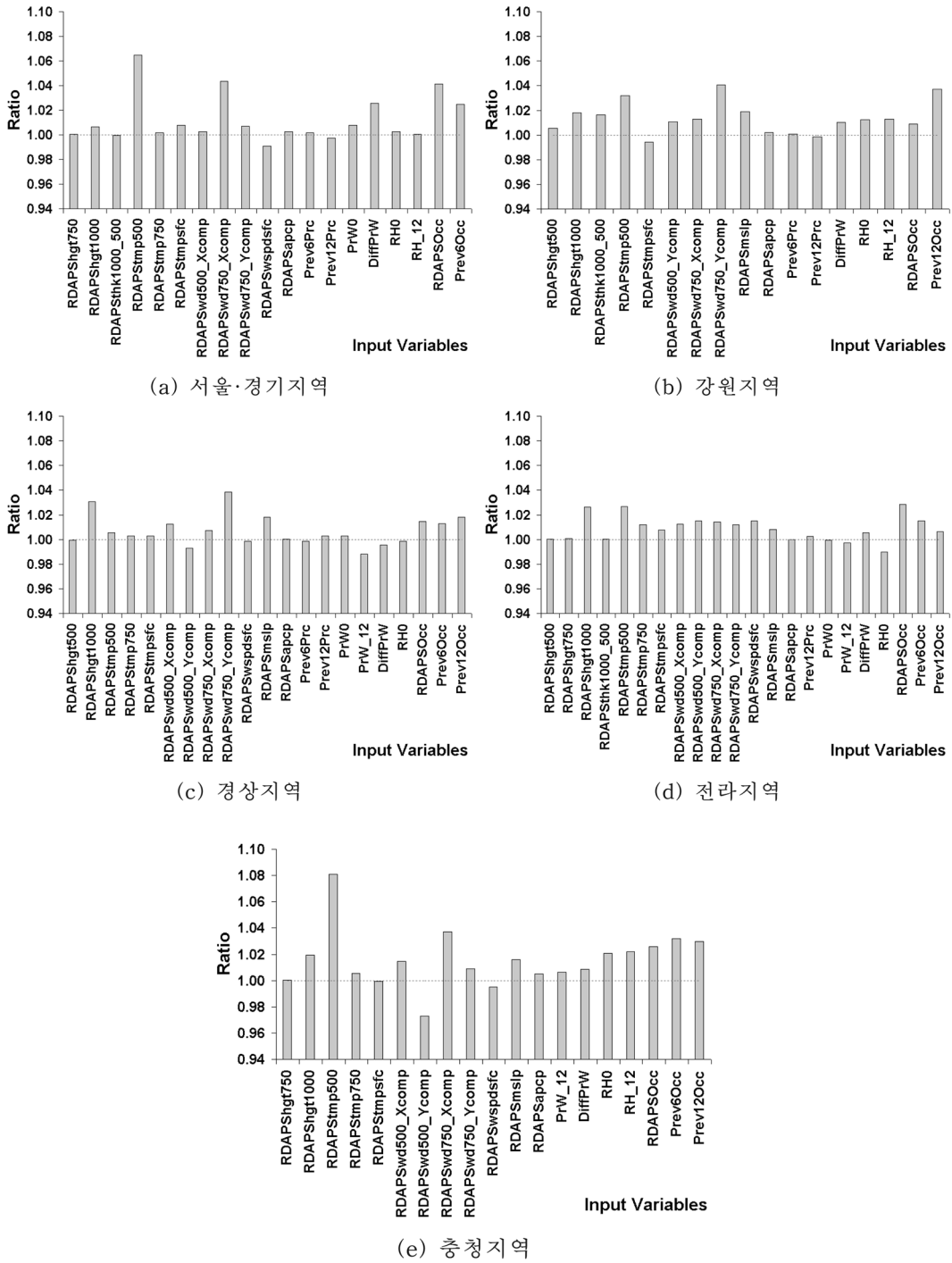


Fig. 3 Sensitivity ratios for RDAPS and observation variables

여기서, R: 민감도 비율(sensitivity ratio),  $\varepsilon_s$ : 예측인자 제거 시 예측오차,  $\varepsilon_0$ : 예측인자 포함시 예측오차

Fig. 3은 우리나라 강수발생확률을 개선하기위해 사용된 예측인자로 사용된 입력변수의 민감도분석 결과를 보여주고 있다. 서울·경기지역에서 강수발생확률에 가장 크게 영향을 미치는 변수는 500 hPa에서의 온도와 750 hPa에서의 X방향 바람성분(X component of wind at 750 hPa)이며 오차의 비율은 각각 1.06과 1.04로 강수발생확률에 입력된 예측인자 중 가장 크게 영향을 미치고 있다. 또한 강원지역에서의 오차의 비율은 750 hPa에서의 Y방향 바람성분이 1.04, RDAPS모형 실행전의 12시간동안의 누적강수의 강수발생의 유무를 나타내는 범주형 변수가 1.04로 강수발생확률에 영향을 미치고 있다. 충청지역에서의 오차비율은 500 hPa에서의

온도가 1.08, 750 hPa에서의 X방향 바람성분이 1.04로 나타났다. 강원지역에서의 오차비율은 750 hPa에서의 Y방향 바람성분이 1.04, 1000 hPa에서의 지위고도가 1.03로 나타났다. 전라지역에서의 오차비율은 RDAPS의 강수발생과 1000 hPa에서의 지위고도가 각각 1.03으로 나타났다. 자료기간동안 한반도영역에서 강수발생확률에 가장 크게 영향을 미치는 변수는 750 hPa에서의 Y방향 바람성분이 3회로 관찰되었다. 민감도분석결과 신경망구성에 사용된 변수들은 Table 3에 정리되어 있다.

앞서 언급하였듯이 본 연구에 사용된 자료의 기간은 2001년 7월 1일 00시부터 8월 31일 24시과 2002년 6월 1일 00시부터 6월 30일 24시까지로서 홍수기를 대상으로 연구를 수행하였다. 자료에 사용된 RDAPS의 경우 3시간 단위로

Table 3. Variables used for neural network modeling selected through the sensitivity analysis

변수명	지역명				
	서울·경기	강원	경상	전라	충청
RDAPSheight500	×	○	○	○	×
RDAPSheight750	○	×	×	○	○
RDAPSheight1000	○	○	○	○	○
RDAPSthk1000_500	○	○	×	○	×
RDAPSwd500_Xcomp	○	×	○	○	○
RDAPSwd500_Ycomp	×	○	○	○	○
RDAPSwd750_Xcomp	○	○	○	○	○
RDAPSwd750_Ycomp	○	○	○	○	○
RDAPSwspdsfc	○	×	○	○	○
RDAPStmp500	○	○	○	○	○
RDAPStmp750	○	×	○	○	○
RDAPStmpsf	○	○	○	○	○
RDAPSmstp	×	○	○	○	○
RDAPSapcp	○	○	○	○	○
RDAPSOcc	○	○	○	○	○
Prev6Prc	○	○	○	×	×
Prev12Prc	○	○	○	○	×
Prev6Occ	○	×	○	○	○
Prev12Occ	×	○	○	○	○
RH0	○	○	○	○	○
RH_12	○	○	×	×	○
PrW0	○	×	○	○	×
PrW12	×	×	○	○	○
DiffPrW	○	○	○	○	○

48시간까지의 예측자료를 생산하며, 하루에 2회(00 UTC와 12 UTC)에 예측하게되어 총 2892개의 데이터(케이스)를 가지고 있다. 전체 자료는 학습(training), 검증(selection/verification), 적용(test) 세트로 나누어 분석을 진행하는데 과학습(over-learning)을 피하기 위하여 각각 2:1:1의 비율로 자료를 나누어 수행하였다. 따라서 학습과정에서는 50%의 자료를 무작위로 추출하여 신경망모형내부 은닉뉴런에서의 연결강도를 결정하였으며, 검증과정에서 나머지 50%중 25%를 무작위로 추출하여 예측결과를 검증하였다. 적용단계에서는 남은 25%에 대하여 예측을 수행하였다. 구성된 신경망이 학습에 사용된 자료이외에 새로운 자료에 대해서도 적합(fitting)을 잘 수행할 수 있는 능력을 일반화(generalization)능력이라고 표현하며, 이는 검증과 적용과정에서의 오차를 비교하여 판단하며 이값들의 차가 적을수록 일반화가 잘된 모형으로 판단한다. 검증과 적용과정에서의 오차는 Table 1에 제시되어 있는데, 경상지역에서의 상대오차비율이 14.6%로서 다소 큰 편이지만 기타지역에서는 상대오차의 절대값이 0.3~3.5%로서 모형의 일반화는 무리 없이 구성되었다고 판단된다.

### 3. 예측정확성 평가

예측의 정확성은 예측값과 관측값이 얼마나 일치하는지에

Table 4. 2×2 contingency table for evaluating predictability of the artificial neural network model

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

관한 것이며, 예측의 검증(verification)은 예측값의 질을 결정하는 과정이다. 이번 연구에서는 RDAPS모형의 수치예보값과 신경망에 의해 생성된 강수발생확률에 대해서 검증하였다. RDAPS모형의 강수데이터는 0.1 mm/3 hr을 기준값(threshold)로 사용하여 강수가 0.1 mm/3 hr보다 크다면 강수발생으로 구분하여 1값을 부여하고 0.1 mm/3 hr보다 작다면 무강수로 구분하여 0값을 부여하여 2가지 형태의 범주형태 데이터로 전환된다. 강수예측성결과를 평가하기위해서 2×2 분할표(2×2 contingency table; Table 4)로부터 산정된 스칼라점수를 사용하였다. 본 연구에서는 적중률(H; Hit rate), 위험점수(TS; Threat Score), 탐지확률(POD; Probability Of Detection), 편이(B; Bias), Kuipers Skill Score(KSS)등을 사용하여 예측자료를 검증하였다(Wilks, 1995). 적중률(H)은 가장 직관적 방법으로써 총 사상에 대하여 강수와 무강수를 포함한 모든 예측이 정확한 경우의 비율을 나타낸다(Eq.

(7)). 이 경우 강수사상과 무강수사상에 대한 비중을 동등하게 적용하게 된다. H는 0에서 1사이의 값을 가지며, 완벽한 예보의 경우 1값을 가진다. 하지만 강수사상에 비하여 무강수사상이 월등히 많을 경우와 무강수보다는 강수사상에 대한 예측이 더욱 중요할 경우 전체 예측성과중에서 강수사상만을 고려한 성과지수가 필요하다. 이 경우 위험점수(TS)가 적중률의 대안적인 방법으로 사용될 수 있으며, 이는 관측강수사상과 예측강수사상의 합에 대한 예측강수적중사상의 비율로 나타낸다(Eq. (8)). TS는 0에서 1사이의 값을 갖는데 완벽한 예보의 경우 1값을 가진다. 적중률에 대한 또다른 대안으로는 탐지확률(POD)이 있는데, 이는 총관측강수사상에 대한 예측강수적중사상의 비로 나타낸다(Eq. (9)). POD 역시 완벽한 예측인 경우 1값을 가진다. 편이율(Bias)은 관측강수사상과 예측강수사상의 비를 나타내는데(Eq. (10)), 이는 예보의 정확성을 나타내는 지표라기보다는 강수예측이 지나치게 보수적인지 낙관적인지 여부를 판단하는 지표로 활용된다.

$$H = \frac{(a+d)}{n} \quad (7)$$

$$TS = \frac{a}{(a+b+c)} \quad (8)$$

$$POD = \frac{a}{(a+c)} \quad (9)$$

$$B = \frac{a+b}{a+c} \quad (10)$$

여기서,  $n=a+b+c+d$  이다.

Skill Score는 특정한 예측기법의 상대적 평가기술이다. 이는 완벽한 예측과 기준예측사이의 상대적 예측정확도에 대한 평가기법으로서 일반적으로는 Eq. (11)과 같이 표현된다.  $A_{ref}$ 는 기준예측의 정확도를 의미하는데, 강수예측의 경우 지역의 기후적 강수확률값을 사용한다. 예측정확도로서 적중률(H)을 사용하고 Eq. (11)의 분모에서  $A_{ref}$  계산시 비편이확률을 사용하게 되면, Table 4와 같은 분할표에 대한 Kuipers Skill Score는 Eq. (12)와 같은 형태가 된다. 결과적으로 KSS는 0에서 1사이의 값을 갖는데 예측값이 기준예측과 같다면 0을, 완벽한 예측의 경우 1값을 가진다.

$$SS_{ref} = \frac{A - A_{ref}}{A_{perf} - A_{ref}} \times 100\% \quad (11)$$

여기서, A: Accuracy이다.

$$KSS = \frac{(ad-bc)}{(a+c)(b+d)} \quad (12)$$

신경망 학습후의 강수발생확률이 RDAPS모형의 예측값과 비교한 정확도의 개선은 Fig. 4를 통해 제시되어 있다. 서울·경기 지역에서 H와 TS는 상대적으로 RDAPS에 비하여 각각 6.9%와 83.3%가 예측정확성이 상승하였으며, POD는 129.0%상승하였다. 강원, 전라, 경상, 충청지역에서의 H는

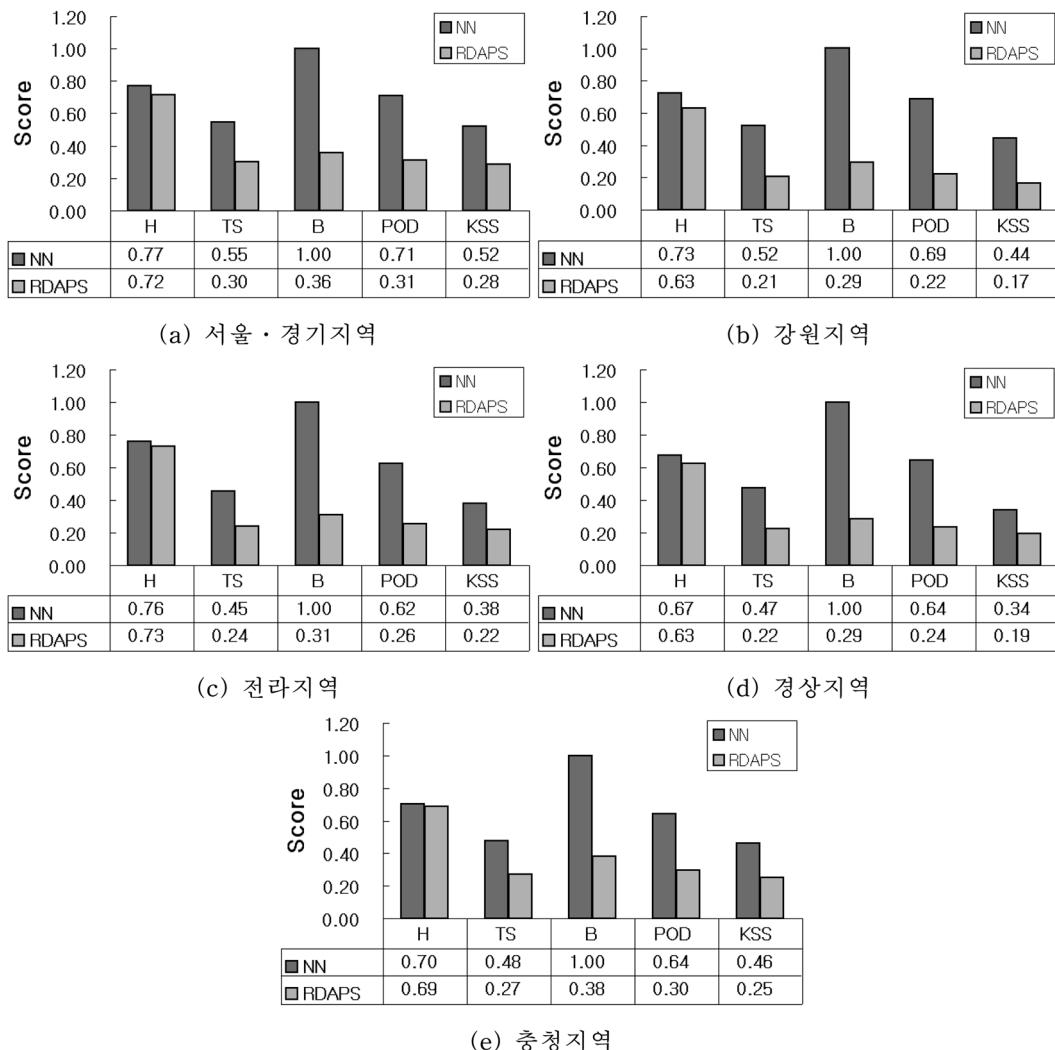


Fig. 4 Comparison of the evaluation score (H, TS, B, POD, KSS) for the RDAPS and the artificial neural network

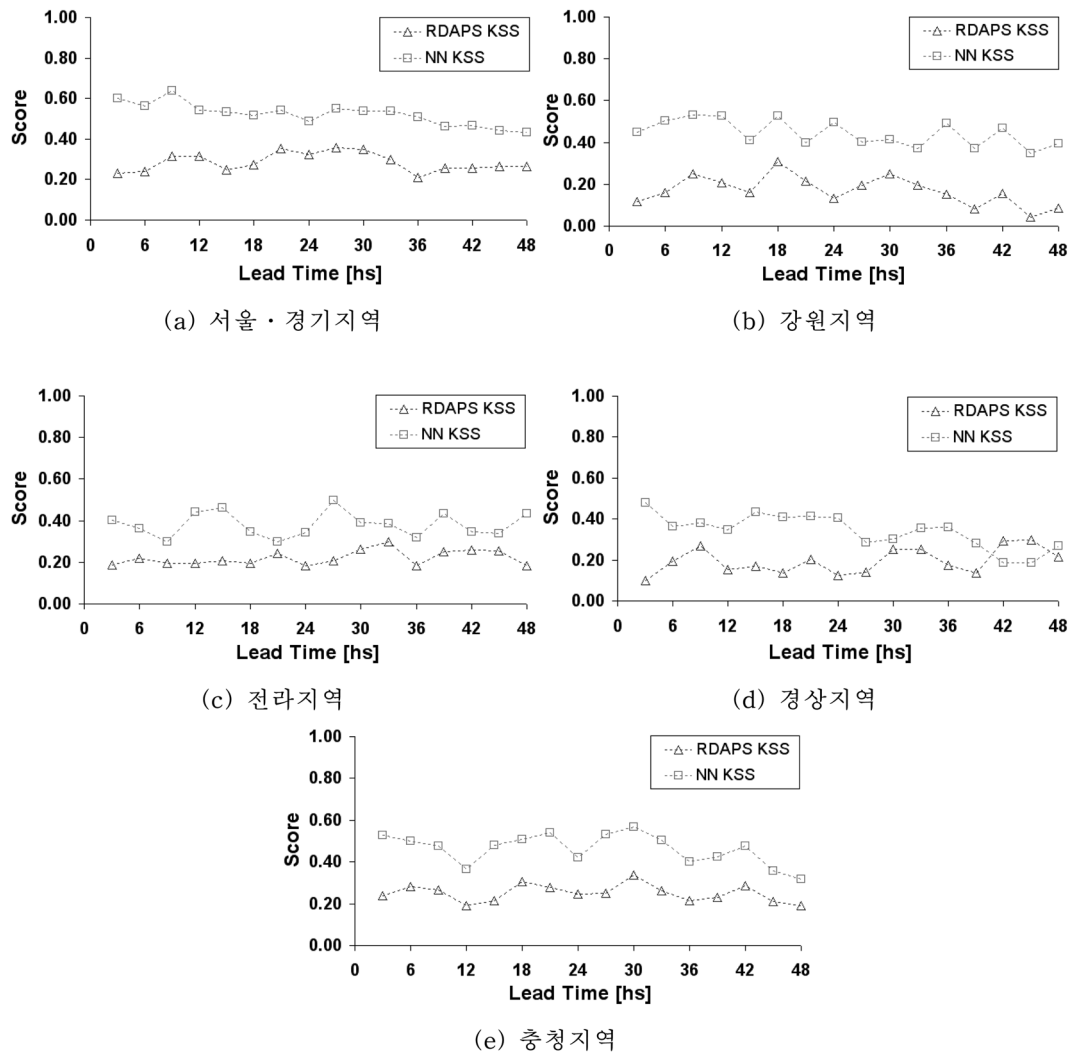


Fig. 5 Regional Kuipers Skill Scores over various lead times

각각 15.9%, 4.1%, 6.3%, 1.4% 상승하였으며, TS는 각각 147.6%, 87.5%, 113.6%, 77.8%의 예측정확성이 향상 되었다. POD는 각각 213.6%, 138.5%, 166.7%, 113.3% 상승 하였다. H에 비하여 TS와 POD값의 향상이 두드러진 것을 확인할 수 있는데, 이는 신경망모형이 무강수보다 강수예측에 있어서 특히 효과적이었다는 것을 의미한다.

Kuipers Skill Score (KSS)의 경우 서울·경기지역의 경우 상대적으로 85.7% 향상되었으며, 강원, 전라, 경상, 충청지역에서는 각각 상대적으로 158.8%, 72.7%, 78.9%, 84.0%가 향상되었다. 선행시간별로는 경상지역의 42 hr과 45 hr를 제외하고는 RDAPS모형의 수치예보자료에 비하여 두드러지게 향상되었다(Fig. 5). 따라서 신경망모형이 기후값에 대한 확률예측보다 개선효과가 뚜렷함을 알 수 있다.

#### 4. 결 론

강우의 발생확률의 예측정확성을 향상시키기 위해서는 예측인자의 구성과 자료의 기간이 중요한 역할을 한다. 구성된 신경망의 검증오차와 적용오차가 서로 비교함으로써 신경망이 적절히 일반화되었음을 확인할 수 있었다. 전체 변수목록 중 결과에 유의미한 영향을 미치는 변수를 선택적으로 신경망구성에 활용함으로써 신경망에서 각 지역별로 적합한 예측인자를 효과적으로 구성 할 수 있었다. 구성된 변수들에

대한 민감도분석을 통하여 변수들의 상대적 상관성을 확인할 수 있었다. 다양한 대기변수들의 비선형 조합으로 생성되는 경우는 신경망학습을 통하여 RDAPS모형의 수치예보자료보다 모든 점수에서 정확성이 향상되었다. KSS를 통한 예측정확성은 지역별로 72.7~158.8%의 개선도를 보여주고 있으며, 예측선행시간별로는 경상지역에서 42 hr과 45 hr를 제외하고 모든 지역에서 눈에 띄게 향상되었다. 본 연구의 성과는 다음과 같이 요약될 수 있다.

1. 신경망구성에 사용된 변수들의 민감도분석결과 1000 hPa 지위고도(1000 hPa geopotential height), 750 hPa 풍속(X, Y-성분), 500 hPa 기온, RDAPS 3시간 누적강우, RDAPS 강수발생, 0시간전 상대습도 등의 변수가 민감도가 높은 것으로 확인되었다.
2. 강수발생확률을 예측하기 위하여 RDAPS 중규모수치예보 모형 예측값, AWS 관측자료, 상층기상관측소(upper-air sounding station)의 자료등 다양한 자료원을 이용하여 신경망모형을 구성함으로써 모형의 예측능력을 개선하고자 시도하였다.
3. 2001년 7, 8월과 2002년 6월의 RDAPS 수치예보에 대한 신경망예측결과, 한반도전체에 대한 예측성과의 개선은 H가 6.8%상승하였고, 특히 TS와 POD는 각각 99.2%와 148.1% 상승함으로써 강수예측에 대한 신경망모형이 효과적인 도구가 될 수 있음을 확인하였다. KSS 역시 92.8%



개선됨으로서 기준예측에 대한 개선이 RDAPS 예측에 비하여 뚜렷이 개선된 결과를 보여주고 있다.

4. QPF(Quantitative Precipitation Forecast) 신경망을 구성하고, PoP(Probability of Precipitation) 신경망과의 연계를 시도한다면 보다 신뢰성있는 정량강수예측이 가능할 것으로 판단된다.

## 감사의 글

본 연구는 2006학년도 단국대학교 대학연구비 지원으로 연구되었음.

## 참고문헌

건설교통부(2006) 수자원장기종합계획(2006-2020) 보고서, 한국수자원공사  
과학기술부(2007) 유역 물 관리 운영 기술개발 1, 21세기 프론티어 「수자원의 지속적 확보기술개발 사업」, 한국수자원공사  
김광섭(2006) 상층기상자료와 신경망기법을 이용한 면적강우 예측. *한국수자원학회논문집*, 한국수자원학회, 제39권, 제8호, pp. 717-726.  
김영오, 정대일, 김형섭, 이길성(2001) 앙상블 예측을 통한 물공급전망 개선방안. *한국수자원학회 학술발표회 논문집*, 한국수자원학회, pp. 304-308.  
김호준, 백희정, 권원태, 최병철(2001) 구간 연산 신경망을 이용한 강수량 장기예측 기법. *한국기상학회지*, 한국기상학회, 제37권, 제5호, pp. 443-452.  
안중배, 박정규, 임은순, 차유미(2003) 인공신경망 모형을 이용한 기온과 강수량 규모축소 연구. *한국기상학회지*, 한국기상학회, 제13권, 제1호, pp. 476-477.  
정대일, 김영오(2002) 앙상블 예측을 이용한 충주댐 월 유입량 예측. *대한토목학회논문집*, 대한토목학회, 제22권, 제3-B호,

pp. 321-331.

정 슬(2004) *인공지능시스템 I(신경회로망의 구조 및 사용법)*. 충남대학교 출판부. pp. 76-106.  
차유미, 안중배(2005) 역학적으로 규모축소된 남한의 여름철 강수에 대한 인공신경망 보정 능력평가. *한국기상학회지*, 한국기상학회, 제41권, 제6호, pp. 1125-1135.  
Bras, R.L. (1990) *Hydrology: An Introduction to Hydrologic Science*. Addison-Wesley publishing company, pp. 82-92.  
Glahn, H.R. and Lowry, D.A. (1972) The use of model output statistics (MOS) in Objective Weather Forecasts. *J. Appl. Meteor.*, Vol. 11, issue 8, pp. 1203-1211.  
Gupta, M.M. (2003) *Static and Dynamic Neural Networks from Fundamentals to Advanced Theory*, Wiley-Interscience, IEEE Press. pp. 80-118.  
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.  
Klein, W.H. (1971) Computer prediction of precipitation probability in the united states. *J. Appl. Meteor.*, Vol. 10, issue 5, pp. 903-915.  
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.  
Schoof, J.T. and Pryor, S.C. (2001) Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *Int. J. Climatol.*, Vol. 21, issue 7, pp. 773-790.  
Valverde Ramrez, M.C., de Campos Velho, H.F. (2005) Artificial neural network technique for rainfall forecasting applied to the so paulo region. *J. Hydrol.*, Vol. 301, issue 1-4, pp. 146-162.  
Wilks, D.S. (1995) *Statistical methods in the Atmospheric Sciences*. Academic Press, pp. 220-250.

(접수일: 2008.3.19/심사일: 2008.6.30/심사완료일: 2008.8.29)