

신경망의 BPN방법에 의한 소유역 하천유출량 예측

안경수 ^{*}, ○ 윤여승 ^{**}, 나진수 ^{***}, 이정식 ^{***}

1. 서론

수자원관리를 위한 운영방안을 수립하기 위해서는 많은 사전정보를 필요로 하며 이들은 대부분 대상유역의 입수 가능한 많은 양의 자료를 수집하여 수문학적 해석을 통해 얻을 수 있다. 수자원의 효율적인 운영을 위한 정책수립에는 미래의 수문사상에 관한 정보가 운영시스템에 도입되어야 한다. 홍수기의 유출량 예측은 홍수예경보, 홍수조절, 하천유지유량의 확보 등에 기본적으로 이용되며, 정확한 홍수량의 예측은 수자원관리를 위한 의사결정에서 피해를 경감시킬 수 있는 방안을 제시할 수 있을 것이다. 우리나라에서는 집중호우에 의한 홍수피해가 자주 발생하고 있다. 이러한 홍수피해를 줄이기 위해서는 그 유역의 강우-유출과정을 정확히 해석해 홍수유출특성을 파악하는 것이 중요하다. 그러나 자연현상인 강우와 유출관계를 규명하는 일은 매우 복잡하여, 하천에서의 유출량 추정에 많은 어려움을 겪어 왔다. 신경망 모형은 입력과 출력으로 구성되는 하나의 시스템을 병렬적이고 비선형적으로 구축할 수 있다는 장점 때문에 불확실한 강우-유출모형에 적용하기에 유리한 장점을 가지고 있다. 신경회로망은 입력층과 중간층을 거치면서 입력층에서 입력되는 여러 변수들 중에서 출력에 큰 영향을 미치는 변수를 학습규칙에 의하여 선택하여 예측하고 유출과정을 추정하게 된다. 신경회로망을 구성하는 처리소자들은 비선형 전달 함수를 가지고 있어서 시스템의 비선형 특성을 학습할 수 있고 입력과 출력공간의 사상(mapping)특성에 따라 연상기억, 필터, 변환, 인식, 최적화 등의 기능을 수행한다. 신경회로망은 시스템의 입력과 출력에 의하여 연결강도를 최적화함으로써 모형의 구조를 스스로 조작화하기 때문에 유출계의 과정모형(Process Model)을 설정하지 않아도 입력력에 따라 모형의 구조에 적합한 매개변수를 최적 추정할 수 있다. 따라서, 본 연구에서는 강우-유출과정을 유역이라는 시스템 내에서 발생된 신호체계를 패턴으로 간주하고 모형을 구성하는 각 유니트의 매개변수를 인공신경회로망 모형으로 최적화하는 방법을 실제 수문현상에 적용하고 홍수기의 시간유출량 자료를 이용하여 시간유출량 예측을 하였다.

2. 연구개요

이 연구에서는 평창강 유역의 강우량관측소 11개 관측 지점의 강우량 자료 및 평창강 유역의 수위를 시간유출량 자료와 보청천 유역의 12개 지점의 강우량 자료 및 보청천 유역의 수위를 시간유출량 자료를 이용하여 강우량자료를 위상공간에 재구성하여 이를 통하여 신경망에 적용하여 강우사상에 의한 시간유출량을 예측하고 신경망의 학습규칙은 일반화 멜타법칙으로서, 오차역전과 알고리즘을 이용하였으며 연구범위은 다음과 같다.

- 첫째, 위상공간 구성의 주요 성분인 지연시간(τ)은 자기상관함수에 의해 계산된 상관도에 의한 선정과,
- 둘째, 지연시간(τ)의 값을 이용하여 정수형태의 공간구성방법인 Time Delay Embedding Method에 의하여 위상공간을 구성하고,
- 셋째, 위상공간으로의 재구성을 통하여 입력변수의 수를 결정하여 구성한 신경망 Model과 임의로 입력변수를 선정한 신경망 Model을 구성하여 모의하고,

* 인천대학교 공과대학 토목환경시스템공학과 교수

** 인천대학교 공과대학 토목환경시스템공학과 박사과정

*** 인천대학교 공과대학 토목환경시스템공학과 석사과정

넷째, 위상공간으로의 재배치를 통한 Model과 임의로 입력변수를 선정한 Model을 시간강우량의 입력변수, 시간유출량의 입력변수 별로 나누어 통계적으로 비교·검토하였으며, 다섯째, 유역 전체의 시간강우량 자료와 유역의 시간유출량 자료를 학습시켜 유역내 시간강우량에 대하여 모의하고 마지막으로, 미계측 유역에의 적용성을 검토하기 위하여 학습시킨 유역에 동일수계 내 소규모 지역의 시간강우량에 대하여 모의하였다.

3. 실제유역에의 적용

3.1 대상유역 및 자료의 처리

이 연구에서는 평창강 유역과 보청천 유역을 대상유역으로 하였으며 이 유역은 국제수문개발계획의 대표유역으로서 평창강 유역은 강우량 관측소 11개 지점과 수위관측소 6개소가 설치되어 있다. 그리고 보청천 유역은 강우량관측소 12개 지점과 수위관측소 6개소가 설치되어 있다. 분석 시 시간강우량 및 시간수위자료를 사용하였으며 수위유량곡선식을 이용하여 시간유출량으로 환산하여 사용하였다. 이 유역의 1993년에서 1998년간의 평균 총 강우량이 100mm이상인 대표적인 호우사상을 대상으로 하여 흥수기의 시간유출량 예측을 위한 학습자료(Training Set)로 사용하였으며, 이를 이용하여 1999년도 평창강유역의 상안미 지점과 보청천유역의 기대 지점에 대한 시간유출량을 예측하였다. 신경회로망의 학습에는 자료를 정규화(Normalization)하는 전처리 과정이 필요하다. 이것은 처리소자들의 집합으로 구성된 신경회로망의 효율을 높이기 위하여 잡음이나 왜곡에 의해 변형된 것을 바로 잡는 기초적인 연산처리 과정이며 신경회로망에서 출력된 값과 다시 역전환 시켜 처리될 후처리에 대응한다. 본 연구에서는 식(1)을 이용하여 전처리를 시행하였다.

$$Y_i = \frac{1}{X_{\max} - X_{\min}} (X_i - X_{\min}) \quad (1)$$

여기서 Y_i : 정규화된 변수값, X_i : 실제 변수값, X_{\max} : 변수의 최대값, X_{\min} : 변수의 최소값

3.2 자기상관함수의 결정

Time Delay Embedding Method에서 지체시간 τ 가 너무 작으면 재구성 공간내의 자료는 이전의 자료에만 의존하는 극단적인 상관을 띠게 된다. τ 의 최적 결정은 다음 세 가지의 방법을 고려하여 선택하게 된다.

- (1) 1 / (관측된 시계열 자료의 주요한 주기의 수)
- (2) 관측된 시계열자료의 자기상관함수의 값이 극소치를 취할 때까지의 시간간격
- (3) 관측된 시계열자료의 자기상관함수의 값이 최초로 0이 될 때까지의 시간간격

지체시간 τ 는 (1), (2), (3)의 방법 중 가장 작은 값을 선택하며, 본 연구에서는 (3)의 방법에 의하여 지체시간(τ)을 구하였다.

3.3 신경회로망의 입력Set 구성

하천유역을 입력과 출력만을 고려한 Black Box로 가정하고 입력과정을 시계열 분석에 의하여 모형화할 수 있다면 출력으로써의 예측 시간유출량은 시간강우량 및 시간유출량의 선행조건과 밀접한 관계를 갖는다. 따라서 시간강우량 및 시간유출량의 상태에 따라 이를 입력하여 장래의 시간유출량을 예측하게 된다. 이들 입력자료는 자료의 수집여부에 따라 강우량뿐만 아니라 온도, 습도, 토양함수비 등의 여러 가지 다른 인자를 포함할 수 있다.

본 연구에서는 시간강우량 및 시간유출량 자료를 이용하였으며 자료간의 상관성을 해석하기 위하여 그 자료간의 지체시간(τ)을 이용하여 식(2)과 같이 시간유출량을 구하기 위한 모델을 만들었다.

$$Q_k = \Gamma(p_{k-1}, p_{k-2}, \dots, p_{k-r}, q_{k-1}, q_{k-2}, \dots, q_{k-s}) \quad (2)$$

여기서, r : 강우자료의 τ 값, s : 유출자료의 τ 값

다음과 같은 과정에 의하여 시간유출량 예측 신경회로망을 조작할 수 있다.

- (1) 입출력 패턴을 구성하기 위하여 시간강우량 및 시간유출량 자료로부터 지체시간 τ 값을 구한다. 이 때 지체시간 τ 값은 시간강우량 및 시간유출량 자료의 연속적 형태의 내부적 관계를 파악하기 위하여 각 자료의 Correlation Integral을 Graph로 도시하여 처음 0과 만나는 시점의 시간을 정수로 하여 사용한다.
- (2) 입력 Set를 구성하는 방법은 Time Delay Embedding Method이다. 여기서 임의로 선정한 입력 Set를 구성하여 모델 1-1~모델 6-1로 하였으며, (1)에 의하여 얻어진 강우량과 유출량의 지연시간 τ 값을 이용하여 입력 Set를 구성하였다.
- (3) 입출력 패턴을 구성하는 각각의 처리소자는 서로 다중결선구조로 연결하고 입력자료를 받는 입력층과 예측 시간유출량을 출력하는 출력층 사이에 중간층을 두어 입력정보를 저장한다.
- (4) 입력층과 중간층 사이에는 비선형 전달함수인 시그모이드(Sigmoid) 함수를 두며 중간층에서의 계산 결과를 출력층으로 전진(Feedforward) 시킨다.
- (5) 중간층과 출력층 사이에도 전달함수를 두어 예측 유출량과 관측 유출량의 차이가 최소화 될 때까지 엘타법칙에 의하여 매개변수를 최적화하도록 조정한다.
- (6) 매개변수가 최적화 될 때까지 (4), (5)의 과정을 반복하여 최적 중간층 처리소자의 수를 결정하여 연결강도를 추정한다.

3.4 모형의 구성

이 연구에서는 시간유출량 예측을 위한 신경회로망의 구조를 입력자료에 의한 시계열모형으로 가정하였으며, 평창강 유역의 시간강우량 자료와 평창강 유역의 시간유출량 자료를 위상공간 구성방법(Time Delay Embedding Method)에 의하여 신경회로망의 입력 Set를 구성하였다. Time Delay Embedding Method의 가장 주요한 변수인 지연시간(τ)은 시계열자료가 가지는 내부적인 반복간격을 말하는 것으로 n 차 계열공분산을 0차 계열공분산으로 나눈값을 도시하여 처음으로 0에 접하는 점의 시간을 정수화하는 것이다.

본 연구에서는 모델 1-1에서 모델 6-1까지 총 21개의 모델을 구성하였으며 각 모델의 입력변수는 식(3.1)~식(3.7)과 같다.

$$\text{Model 1-1} : Q_k = \Gamma(p_{k-1}, q_{k-1}) \quad (3.1)$$

$$\text{Model 2-1} : Q_k = \Gamma(p_{k-1}, p_{k-2}, q_{k-1}) \quad (3.2)$$

$$\text{Model 1-2} : Q_k = \Gamma(p_{k-1}, q_{k-1}, q_{k-2}) \quad (3.3)$$

$$\text{Model 3-1} : Q_k = \Gamma(p_{k-1}, p_{k-2}, p_{k-3}, q_{k-1}) \quad (3.4)$$

$$\text{Model 3-2} : Q_k = \Gamma(p_{k-1}, p_{k-2}, p_{k-3}, q_{k-1}, q_{k-2}) \quad (3.5)$$

$$\text{Model 3-3} : Q_k = \Gamma(p_{k-1}, p_{k-2}, p_{k-3}, q_{k-1}, q_{k-2}, q_{k-3}) \quad (3.6)$$

$$\text{Model 6-1} : Q_k = \Gamma(p_{k-1}, p_{k-2}, p_{k-3}, p_{k-4}, p_{k-5}, p_{k-6}, q_{k-1}) \quad (3.7)$$

3.5 평창강 유역의 신경회로망 적용 결과

그림 1은 평창강 유역의 방림 지점의 1998년 8월 8일 시간강우량 자료에 대한 n 차 계열공분산을 0차 계열공분산으로 나눈 값을 나타낸 그림이다. 이 그림의 계열상관도에서 처음으로 0에 접해 있는 점은 4.4051이다. 따라서 τ 값은 정수가 되어야 함으로 4가 된다. 그림 2는 평창강 유역의 방림 지점의 1998년 8월 8일의 시간유출량 자료에 대한 계열상관도를 나타낸 그림이다. 이 그림에서 보면 처음으로 0에 닿는 점은 3.3982이다. 따라서 τ 값은 정수로써 3가 된다. τ 의 결정은 중간층의 변화와는 상관이 없이 모두 동일한 지체시간(τ)을 갖는다. 그림 3은 중간층의 수가 $2N+1$ 인 경우의 입력층의 처리소자의 수가 7인 경우이며 모델이며, 예측결과를 실측치와 비교하여 보면, 다른 처리소자의 모델과 비교해서 시간유출량 곡선의 상승부분에서 실측치와 유사한 결과를 나타내고 있으며, 특히 모델 4-3이 다른 모델에 비해서 실측치와 비교해서 가장 유사한 예측을 하고 있다. 이러한 학습의 결과를 가지고 1999년 평창강의 상안미

지점에 대해 검증하여 보면 다음과 같은 결과를 얻을 수가 있다. 평창강 유역의 상안미 지점의 1999년 시간강우량 자료에 대하여 계열상관도를 플롯트(plot)하여 n 차 계열공분산을 0차 계열공분산으로 나눈 값(τ)을 구해보면 계열상관도에서 처음으로 0에 접해 있는 점은 3.4426이다.

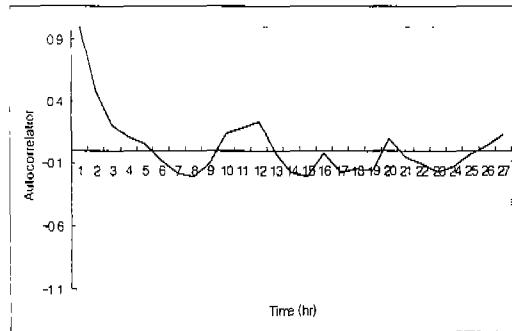


그림 1. 평창강유역의 방류지점의
시간강우량 계열상관도('98.8.8)

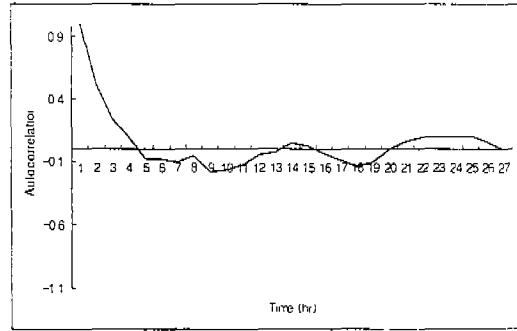


그림 2. 평창강유역의 방류지점의
시간유출량 계열상관도('98.8.8)

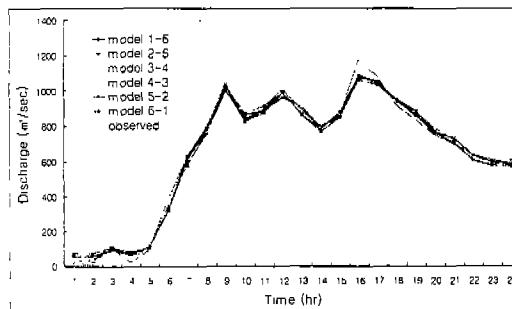


그림 3. 평창강유역 방류지점의 시간유량의
신경회로망 예측결과('98-2N+1-7)

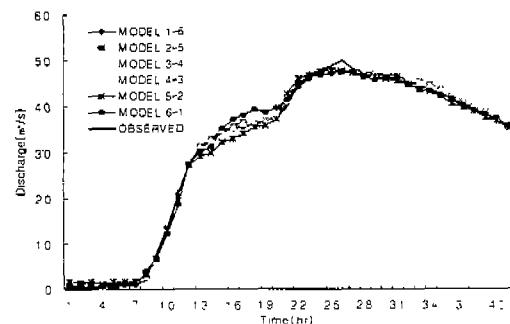


그림 4. 평창강유역 상안미지점의 시간유출량의
신경회로망 예측결과('99-2N+1-7)

따라서 τ 값은 정수가 되어야 함으로 3이 된다. 또한 평창강 유역의 상안미 지점의 1999년의 시간유출량 자료에 대한 계열상관도에서 처음으로 0에 닿는 점은 4.4789이다. 따라서 τ 값은 정수로써 4가 된다. τ 의 결정은 중간층의 변화와는 상관이 없이 모두 동일한 지체시간(τ)을 갖는다. 그림 4는 중간층의 수가 $2N+1$ 인 경우의 입력층의 처리소자의 수가 7인 경우이며 모델이며, 예측결과를 실측치와 비교하여 보면, 다른 처리소자의 모델과 비교해서 시간유출량 곡선의 상승부분에서 실측치와 유사한 결과를 나타내고 있으며, 특히 모델 3-4가 다른 모델에 비해서 실측치와 가장 유사한 유출량을 예측하였다.

3.6 보청천 유역의 신경회로망 적용 결과

그림 5는 보청천 유역의 산계교 지점의 1996년 6월 16일의 시간강우량 자료에 대한 n 차 계열공분산을 0차 계열공분산으로 나눈 값을 나타낸 그림이다. 이 그림의 계열상관도에서 처음으로 0에 접해 있는 점은 6.4371이다. 따라서 τ 값은 정수가 되어야 함으로 6이 된다. 그림 6은 보청천 유역의 산계교 지점의 1998년의 시간유출량 자료에 대한 계열상관도를 나타낸 그림이다. 이 그림에서 보면 처음으로 0에 닿는 점은 1.4235이다. 따라서 τ 값은 정수로써 1이 된다. τ 의 결정은 중간층의 변화와는 상관이 없이 모두 동일한 지체시간(τ)을 갖는다. 그림 7은 중간층의 수가 $2N+1$ 인 경우의 입력층의 처리소자의 수가 7인 경우이며 모델이며, 예측결과를 실측치와 비교하여 보면, 다른 처리소자의 모델과 비교해서 시간유출량 곡선의 상승부분에서 실측치와 유사한 결과를 나타내고 있으며, 특히 모델 6-1이 다른 모델에 비해서 실측치와 비교해서 가장 유사한 예측을 하고 있다. 이러한 학습의 결과를 가지고 1999년 보청천의 기대 지점에 대해 검증하여 보면 다음과 같은 결과를 얻을 수가 있다. 보청천 유역의 기대 지점의 1999년 시간강우

량 자료에 대하여 계열상관도를 플롯트(plot)하여 n 차 계열공분산을 0차 계열공분산으로 나눈 값(τ)을 구해보면, 계열상관도에서 처음으로 0에 접해 있는 점은 1.3793이다.

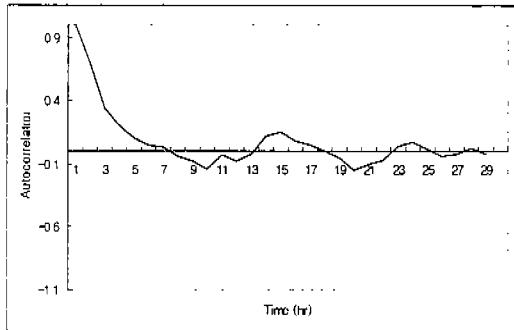


그림 5. 보청천유역 산계교지점의
시간강우량 상관계열도('96.6.16)

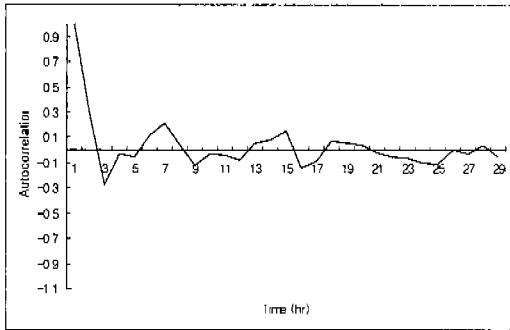


그림 6. 보청천유역 산계교지점의
시간유출량 상관계열도('96.6.16)

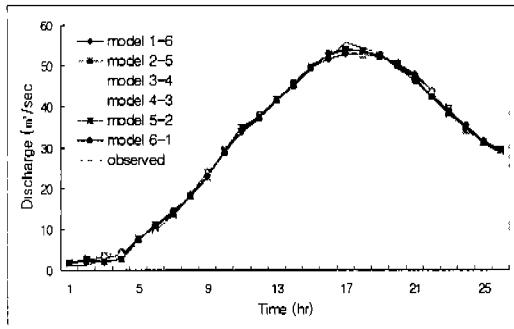


그림 7. 보청천유역 산계교지점의 시간유출량의
신경회로망의 예측결과 ('96-2N+1-7)

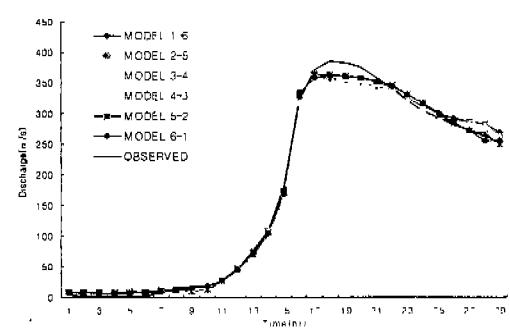


그림 8. 보청천유역 기대지점의 시간유출량의
신경회로망 예측결과('99-2N+1-7)

따라서 τ 값은 정수가 되어야 함으로 1이 된다. 또한 보청천 유역의 기대 지점의 1999년의 시간유출량 자료에 대한 계열상관도에서 처음으로 0에 닿는 점은 6.3887이다. 따라서 τ 값은 정수로써 6이 된다. τ 의 결정은 중간층의 변화와는 상관이 없이 모두 동일한 지체시간(τ)을 갖는다. 그림 8은 중간층의 수가 $2N+1$ 인 경우의 입력층의 처리소자의 수가 7인 경우이며 모델이며, 예측결과를 실측치와 비교하여 보면, 다른 처리소자의 모델과 비교해서 시간유출량 곡선의 상승부분에서 실측치와 유사한 결과를 나타내고 있으며, 특히 모델 1-6이 다른 모델에 비해서 실측치와 가장 유사한 유출량을 예측하였다.

4. 분석 및 고찰

4.1 평창강 유역의 분석 결과의 고찰

IHP 대표유역 중 하나인 평창강 유역의 방류 지점에 대한 '95년도 7월, '95년도 8월, '98년도 홍수기의 평균 총강우량이 100mm인 호우사상의 시간강우량 및 시간유출량 자료를 사용하여 신경회로망 이론에 의해 시간유출량을 예측한 후 그 적합성을 검정하였다. 관측 시간유출량자료와 예측되어진 시간유출량을 함께 플롯트(plot)하여 그림으로 비교하고, 또한 통계치(예측오차, 분산오차, 평균제곱오차, 평균절대백분비 오차, 시계열 오차, 결정계수)를 구하여 적합성을 분석하였다. 유출량 예측을 위한 모델 구성시 사용되는 처리소자의 개수는 지체시간(τ)에 의해 결정하였으며 지체시간(τ)은 Chaos이론의 도입에 의한 계열상관도에 의해 구하였다. 또한 구성된 모델의 적합성을 확인한 후 처리소자의 개수별로 구성된 6개 모델을 이용하여 미계측 지점의 시간유출량을 예측하고 그 적용성 검토를 위하여 '99년도 상안미 지점을 검정대상 지점으로 선정하였다. '95~'98년도의 방류 지점에 적용한 신경회로망의 연결강도를 '99년도 상안미 지점에 신경회로망을 적용하여

예측된 시간유출량 자료와 관측한 시간유출량 자료와의 일치여부를 비교·분석한 내용을 고찰하였다. '98년도의 경우 계열상관도에 의해 구한 지체시간(τ)이 시간강우량일 경우 4이고, 시간유출량일 경우는 3이었다. 계열상관도에 의해 구한 지체시간(τ)을 신경회로망(BPN)에 적용시켜 예측한 시간유출량 모델 중 처리소자의 개수가 7일 때 가장 좋은 결과를 나타냈으며, 중간총이 $2N+1$ 인 모델 4-3의 경우 0.9925으로 결정계수의 값이 가장 높게 나타나 그 적합성이 인정되었다. '99년도의 상안미 지점의 계열상관도에 의해 구한 지체시간(τ)은 시간강우량일 때 3이었고, 시간유출량은 4였다. 계열상관도에 의해 구한 지체시간을 신경회로망(BPN)에 적용시켜 분석한 결과를 그림으로 나타내었으며, 이 그림에서 나타난 결과로 처리소자의 개수가 7인 경우가 가장 좋은 예측결과를 나타내었다. 그리고 중간총이 $2N+1$ 로 구성한 모델 3-4는 0.9893로 결정계수의 값이 가장 높게 나타나 그 적합성이 인정되었다.

4.2 보청천 유역의 분석 결과의 고찰

'96년도의 경우 계열상관도에 의해 구한 지체시간(τ)은 시간강우량은 6이고, 시간유출량은 1이었다. 계열상관도에 의해 구한 지체시간(τ)을 신경회로망(BPN)에 적용시켜 예측한 시간유출량 모델 중 처리소자의 개수가 7일 때 가장 좋은 결과를 나타냈으며, 중간총이 $2N+1$ 인 모델 6-1의 경우 0.9987로 결정계수의 값이 가장 높게 나타나 그 적합성이 인정되었다. 미계측 유역의 시간유출량에 예측한 신경회로망(BPN)의 적용성을 검토하기 위하여 '93년도, '96년도, '97년도의 산계교 지점을 신경회로망의 학습유역으로 선정하고, 시간유출량을 예측한 결과와 학습한 유역의 동일한 연결강도를 '99년도의 기대 지점의 계열상관도에 의해 구한 지체시간(τ)은 시간강우량일 때 1이었고, 시간유출량은 6이었다. 계열상관도에 의해 구한 지체시간을 신경회로망(BPN)에 적용시켜 분석된 결과를 그림으로 나타내었으며, 이 그림에서 나타난 결과로 처리소자의 개수가 7인 경우가 가장 좋은 예측결과를 나타내었다. 그리고 중간총이 $2N+1$ 로 구성한 모델 1-6은 0.9980로 결정계수의 값이 가장 높게 나타나 그 적합성이 인정되었다.

5. 결론

IHP 대표유역인 보청천의 산계교 지점과 평창강의 방류 지점을 신경회로망에 적용하여 예측한 시간유출량 자료에 대한 적용성 검토를 하였으며, 이때 신경회로망(BPN)의 모델 결정은 계열상관도를 이용한 지체시간(τ)에 의해 예측모델을 선정하였다. 신경회로망 적용에 의한 시간유출량을 예측할 때 사용한 모델에서 지체시간에 의해 결정된 모델과 지체시간에 의해서 결정되지 않은 다른 모델과의 적합성을 상호 비교하였다.

학습된 연결강도의 적용성을 비교하기 위하여 보청천의 기대지점과 평창강의 상안미 지점을 설정 대상지점으로 선정하여 시간유출량 자료를 비교·분석한 결과 다음과 같은 결론을 얻을 수 있었다.

1. 시간유출량 예측에 신경망 이론을 적용한 결과 관측치에 유사한 예측을 하였으며 결정계수의 값도 0.998 이상인 것으로 나타났다. 또한 유역내 시간강우량 및 시간유출량의 학습에 의해 예측된 시간유출량 자료가 결정대상 지점에 대하여도 적합한 것으로 나타나 동일 유역내 시간강우량 및 시간유출량 자료에 의한 신경회로망의 적용으로 미계측 지점의 시간유출량 자료를 예측할 경우 정도가 높은 결과를 얻을 수 있을 것으로 사료된다.
2. 유역내의 복잡한 유출시스템을 강우-유출사상만을 고려한 신경망의 BPN 해석방법에 의한 입·출력 패턴의 학습기법을 이용하여 최적의 유출량을 위한 모델의 구성시 사용되는 처리소자의 개수를 시간지체방법(Time Delay Embedding Method)에 의해 결정하였으며, 이 때의 지체시간(τ)은 Chaos 이론의 도입에 의한 계열상관도에 의해서 구성된 모델에서 가장 좋은 예측결과를 얻을 수 있었다.

6. 참고 문헌

1. 김주환(1993), "신경회로망을 이용한 하천유출량의 수문학적 예측에 관한 연구", 박사학위 논문, 인하대학교.
2. 박종훈(1996), "신경회로망을 이용한 하천유출량의 Chaos적 해석에 관한 연구", 석사학위 논문, 인천대학교.
3. R. Rojas.(1996), "Neural Networks", Springer.
4. 合原・辛, 銅谷賢治, 松葉育雄, 市橋秀友, 林勲, 德永隆治(1994), "뉴트로 퍼지 카오스", 대광서림