

# SOM에 의한 강우-유출 예측모형 개발에 관한 연구

## Development of Rainfall-Runoff Prediction Model for Self Organizing Map

김용구\*, 진영훈\*\*, 이한민\*\*\*, 박성천\*\*\*\*

Yong Gu Kim, Young Hoon Jin, Han Min Lee, Sung Chun Park

### 요 지

본 연구에서는 강우의 시·공간적 분포의 불규칙한 변동성을 고려한 강우-유출예측을 위해 인공신경망(Artificial Neural Networks: ANNs)의 기법의 일종인 자기조직화(Self Organizing Map: SOM) 이론과 역전파 학습 알고리즘(Back Propagation Algorithm: BPA) 이론을 복합적으로 이용하였다.

기존의 인공신경망 연구에서 야기된 저·갈수기의 유출량에 대한 과대평가, 홍수기의 유출량에 대한 과소평가, 예측값이 선행 유출량의 지속성을 갖는 Persistence 현상을 해결하기 위하여 패턴분류 성능을 지닌 SOM 이론을 도입하여 예측모형의 전처리 과정으로 이용하였다. 이는 기존의 인공신경망 모형이 하나의 모형을 구성하여 유출량의 전 범위에 해당하는 자료를 예측하는 방법을 개선한 것으로 SOM에 의해 패턴이 분류된 강우-유출관계의 각 패턴별 예측모형을 통해 분류된 자료들의 예측을 수행하는 방법이다. 이와 같이 SOM을 강우-유출예측모형의 전처리과정으로 이용함으로써 기존의 인공신경망 연구에서 야기된 현상들을 해결할 수 있었고, 예측력 또한 기존의 인공신경망 모형의 결과에 비해 우수하였다.

**핵심용어** : 강우-유출모형, 인공신경망, 자기조직화 이론, 전처리 과정, 역전파 학습알고리즘, Persistence 현상

### 1. 서론

강한 비선형성의 경향을 보이고 있는 강우-유출간의 관계를 모형화하기 위한 연구는 다양한 방법론으로 적용되어 활발히 연구되고 있다. 그 중에서 인공신경망을 이용하여 강우-유출간의 관계를 모형화하기 위한 대부분의 연구들은 역전파 학습 알고리즘을 이용하였다. 이는 강한 비선형성을 나타내는 입·출력간의 관계를 나타내는데 탁월한 성능을 보이고 있는 것으로 알려져 있으며, 또한 자료들의 급격한 변화에 대한 뛰어난 적응성을 보여주고 있다. 이러한 인공신경망의 장점들은 다른 모형들과의 비교에서 우수한 예측력을 보여주는 것으로 평가받고 있다. 그러나 지금까지 적용되어진 기존의 인공신경망을 이용한 대부분의 연구들은 강우-유출해석에 있어 다양한 강우사상을 반영하는데 미진할 뿐만 아니라, 예측의 대상인 유출량의 경우 저·갈수기의 자료에 대해서는 과대평가의 경향이 나타나며, 홍수기의 자료에 대해서는 과소평가되는 경향이 있다. 또한 인공신경망 모형의 구축에 있어 예측 대상자료인 유출량의 선행값을 이용할 경우 그 예측값이 선행 유출량의 지속성을 갖는 Persistence 현상이 유발된다. 또한 강우-유출관계를 규명하기 위한 대부분의 연구들이 홍수기의 강우 및 유출 자료를 이용하기 때문에 개발된 예측모형이 홍수사상에 한정되는 한계를 보여주고 있다. 이에 본 연구에서는 기존의 인공신경망 연구에서 야기된 현상들을 해결하기 위하여 탁월한 패턴분

\* 정회원-한국건설기술연구원 수자원연구부 연구원E-mail : ygkim@kict.re.kr

\*\* 정회원-동신대학교 토목공학과 연구원E-mail : yhjin@dnu.ac.kr

\*\*\* 정회원-광주광역시청 지하철건설본부E-mail : inseo@hotmail.com

\*\*\*\* 정회원-동신대학교 토목공학과 교수E-mail : psc@dnu.ac.kr

류 성능을 지닌 자기조직화 이론을 도입하여 자료의 전처리 과정으로 이용하였다. 또한 강우-유출 예측모형의 전처리 과정으로 도입한 본 연구의 방법과 지금까지 적용되어진 기존의 방법을 비교·검토하기 위하여 대상지점으로 영산강 유역을 대표하는 나주지점을 선정하였다. 나주지점의 유출량을 예측하기 위하여 대상지점에 영향을 미치는 광주, 능주, 동곡지점의 강우자료를 이용하여 강우-유출관계를 모형화하기 위해 본 연구에서 제안한 모형과 기존의 인공신경망 모형을 이용하여 홍수 예·경보발령을 위해 필요한 선행예보 최소시간인 3시간 후의 유출량을 예측하여 그 결과를 비교하였다.

## 2. 대상지점 및 입력자료

본 연구의 대상지점인 영산강 유역은 우리나라 서남부에 위치하며, 총 유역면적은 3,455km<sup>2</sup>이고, 본류의 유로연장은 129.5km이며, 본 연구에서는 강우의 시·공간적 분포의 비선형적 변동성을 고려한 강우패턴을 분류하고 강우-유출간의 특성을 분석하기 위해, 그림 1에 나타내었듯이 영산강 본류의 대표지점인 나주수위관측소와 그에 영향을 미치는 상류의 광주, 동곡, 능주강우관측소를 선정하였다. 수위 및 강우자료는 영산강홍수통제소의 30분 자료를 이용하였고, 나주지점의 수위자료를 유출량자료로 환산하기 위해 『영산강 유량측정 보고서』(영산강홍수통제소, 2004)의 나주지점의 수위-유량관계곡선을 이용하였다.

모형의 입력자료 구축을 위해 광주, 동곡, 능주강우관측소의 강우량이 나주지점 유출량에 기여하는 시차를 산정하기 위해 무강우를 제외한 각 강우관측소의 강우자료와 그에 대한 지체시간을 적용한 나주지점의 유출량자료 사이의 상관분석을 실시한 결과 그림 2와 같이 광주, 동곡, 능주강우관측소 모두 7시간 후가 가장 높은 상관성을 갖는 값으로 분석되었다. 그러므로 입력벡터  $x(t)$ 는 나주지점의 유출량과 각 우량관측소의 지체시간을 고려한 강우량에 1시간 30분의 시차를 고려하면 식 1과 같은 12차원 구조의 SOM 훈련 데이터의 구조가 된다.

$$x(t) = (P_{G(t-16)}, P_{D(t-16)}, P_{N(t-16)}, Q_{(t-2)}, P_{G(t-15)}, P_{D(t-15)}, P_{N(t-15)}, Q_{(t-1)}, P_{G(t-14)}, P_{D(t-14)}, P_{N(t-14)}, Q_{(t)}) \quad (1)$$

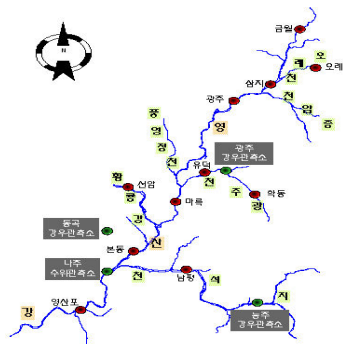
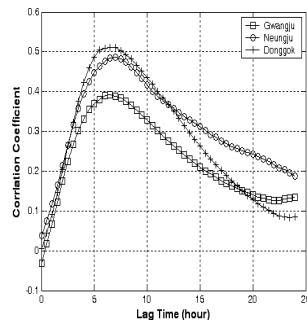


그림 1. 대상지점 유역도 그림 2. 시차결정을 위한 상관분석 결과



## 3. 적용방법 및 결과

### 3.1 SOM 모형의 적용

SOM의 적용을 위해 그림 3의 SOM 적용 흐름도와 같이 광주, 동곡, 능주강우관측소의 강우량과 나주유출량의 30분 자료로부터 2004년 6월 1일부터 2004년 7월 12일까지의 1,999쌍의 자료를 선택하여 각 우량관측소와 나주유출량간의 상관분석으로부터

산정된 시차를 적용하여 1,999×12 행렬의 SOM 입력자료를 구축하였다. 선택된 자료의 정규화 과정을 거친 후 SOM 훈련을 위해 Map 크기를 결정하는데 우선, Map을 구성하는 단위구조의 총 수(M)의 결정은 Garcia 등(2003)에 의해 연구 보고된 식을 이용하였다(식 2).

$$M = 5\sqrt{N} \quad (2)$$

여기서 M은 map을 구성하는 단위구조(unit)의 총 수이며, N은 훈련 자료의 수이다.

본 연구에서의 훈련자료는 1999쌍이며, map 크기는 식2에 의해 산정된 224의 근사치로부터 20×12로 결정하였다. 결정된 map 크기에 대해 최적의 분할구역 구분은 U-matrix method와 K-means algorithm의 방법을 적용하여 구분할 수 있다. U-matrix method에 의해 구분된 결과를 그림 4에 나타내었으며, k-means

algorithm 결과 최적의 분할구역으로 구분된 결과는 그림 5와 같다. 또한 분할구역의 대표적인 강우-유출간의 관계는 그림 6과 같다.

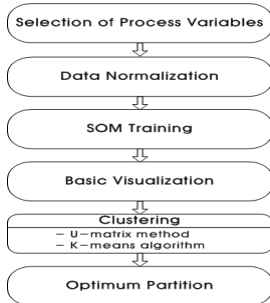


그림 3. SOM 적용 흐름도

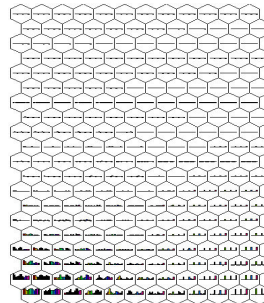


그림 4. U-matrix method 결과

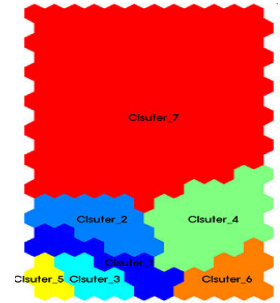


그림 5. k-means algorithm 결과

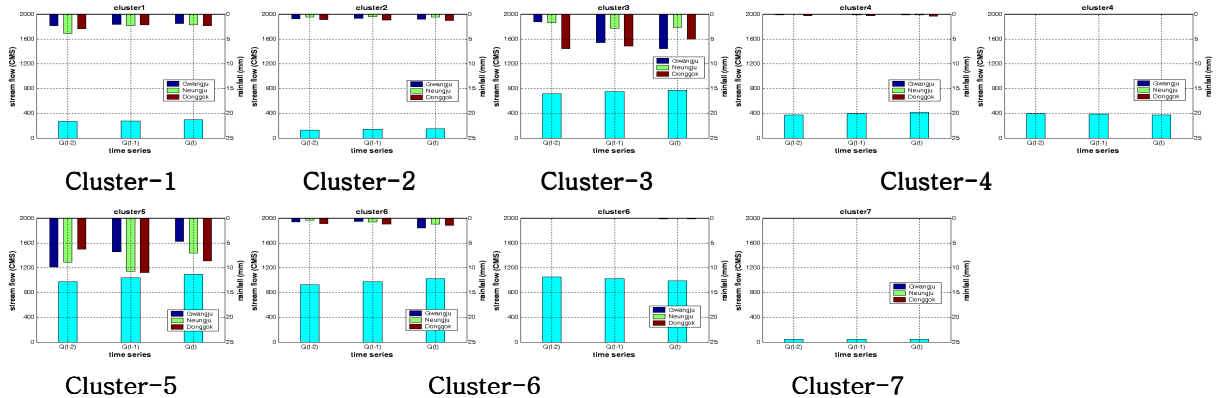


그림 6. 각 분할구역에 대한 대표적인 강우-유출관계

또한 SOM에 의해 구분된 강우-유출관계를 규명하기 위해 그림 7과 같이 도시하였다. 그 결과 Cluster-1은 강우량은 감소하는 경향을 보이는 반면 유출량은 점진적으로 증가하는 수문곡선의 상승부 유출 부분으로 구분되었고, Cluster-2는 수문곡선의 상승기점의 유출부분, Cluster-3은 강우량은 급격히 증가되나 유출량은 미미한 증가를 보이는 유출부분으로 분류되었다. 또한 Cluster-4는 변곡점 아래의 수문곡선 상승부와 감수부, Cluster-5는 수문곡선의 첨두부로 분류되었다. Cluster-6은 변곡점 위의 수문곡선 상승부와 감수부, Cluster-7은 수문곡선의 기저유출 부분으로 분할되었다. 이상의 결과와 같이 강한 비선형성을 가지고 있는 강우-유출관계가 SOM의 적용을 통해 강우-유출의 패턴분류가 가능함을 확인 할 수 있었고, cluster\_4와 cluster\_6은 수문곡선의 상승부와 감수부의 변곡점을 구분지어 주는 것으로 판단된다. 또한 수문곡선의 상승부의 형상은 강우의 특성에 의해 좌우되는 것을 확인 할 수 있었다.

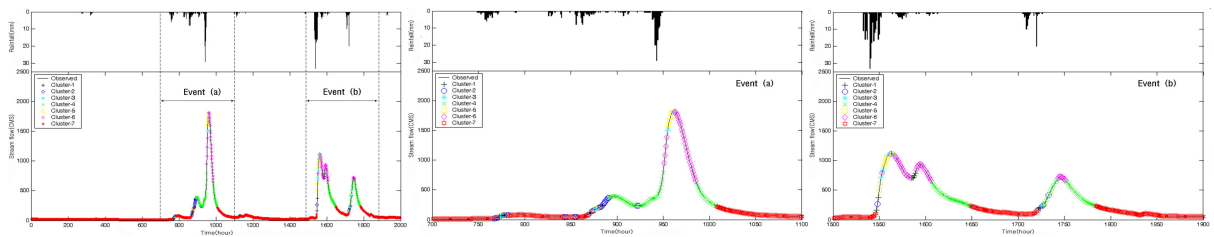


그림 7. SOM에 의해 구분된 강우-유출관계 : 강우주상도 및 유량도

### 3.2 예측모형의 비교 및 검토

본 절에서는 강우-유출관계의 패턴분류를 수행하지 않은 기존의 인공신경망 예측모형과 본 연구에서 적용한 SOM에 의해 패턴분류를 수행한 후 예측모형에 적용한 결과를 상호비교를 통하여 앞서 기술한 기존 인공신경망 모형의 연구에 대한 문제점들을 검토하기 위해 역전과 학습 알고리즘을 이용하여 모형을 구축하였다. 모형의 구성은 SOM의 입력자료와 동일하게 모형을 구축하였다. 구축 모형에 대하여 은닉층의 노드의 수는 기존의 연구결과를 토대로 우수한 예측력을 보이는 것으로 알려진  $3n=36$ 개( $n$ 은 입력자료로 본 연구에서는 12개)를 선택하였고 훈련 수는 20,000회로 동일 조건하에서 각각의 모형에 대해 홍수예경보 발령을 위해 필요한 선행예보 최소 시간인 3시간 후 유출량을 예측하여 이를 검토하였다.

SOM을 적용하지 않은 모형과 적용한 모형 사이의 비교를 위하여 훈련에 대하여 패턴분류를 수행하지 않은 입력자료를 SOM에 의해 패턴이 구분된 자료군과 동일하게 분할구역별로 추출하여 관측값과 각각의 예측값을 그림 8에 도시하였고, 그림 9에는 각 예측값의 오차를 도시하였다.

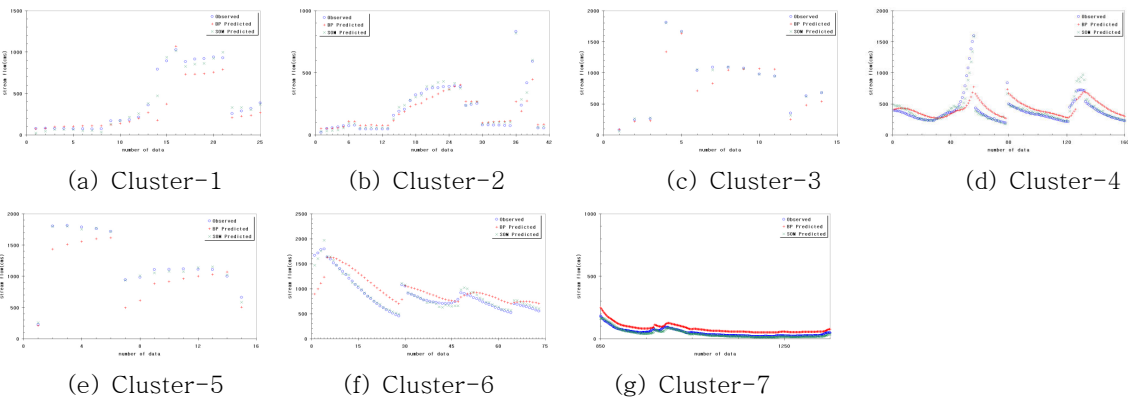


그림 8. 분할구역별 예측값과 관측값의 비교

패턴분류를 수행하지 않은 기존의 방법의 예측값을 BP Predicted, SOM에 의해 패턴분류를 수행한 예측값을 SOM Predicted라고 명명할 때, 저갈수기 자료와 홍수기 자료로 구분된 cluster-7과 cluster-5의 기저유출부와 첨두부를 그림 8과 9의 결과에서 살펴보면 기저유출부(cluster-7)에서는 기존의 방법인 BP Predicted는 과대추정, 홍수기자료로 구분된 cluster-5의 첨두부에서는 과소추정의 경향을 나타내고 있다. 반면 SOM Predicted는 관측값에 대응하여 보다 정확한 예측값을 나타내고 있다. 전반적으로 BP Predicted가 SOM Predicted보다 더 큰 오차를 수반하고 있으며, 전 범위의 유출량 예측의  $R^2$ 과  $RMSE$ 의 비교에서도 SOM Predicted는 각각 0.9799와 0.0624로 BP Predicted의 0.8942와 0.076보다 더 우수한 예측력을 보여주었다.

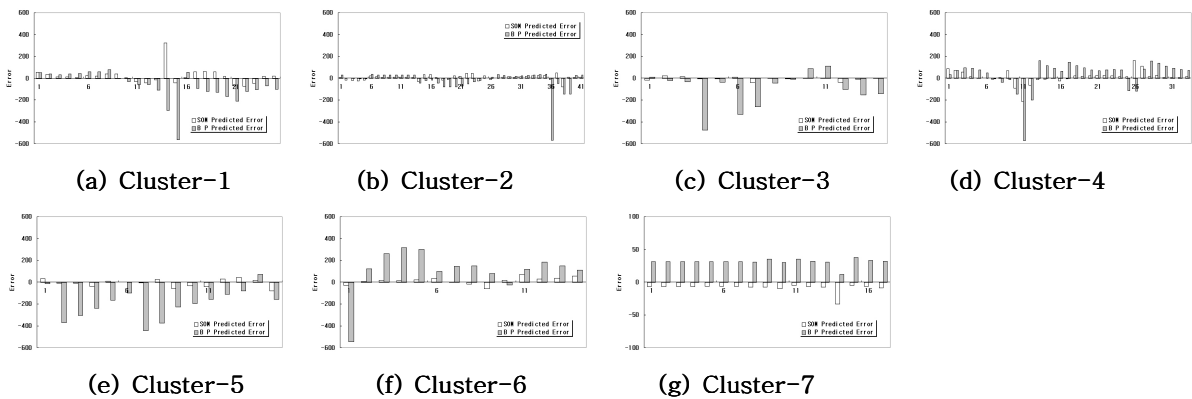


그림 9. 분할구역별 예측값과 관측값의 오차 분석

#### 4. 모형의 검증

검증 방법으로는 검증기간의 자료를 SOM으로 패턴분류를 실시하고, 훈련 과정에서 각 분류별로 결정되어진 매개변수를 이용하여 예측값을 계산하였다. 그 결과를 그림 10에 나타내었으며, 그림에서 보듯이 기존의 적용방법인 BP predicted는 앞서 기술한 저·갈수기의 자료의 과대평가, 홍수기의 과소평가되는 문제점과 예측값이 선행 유출량의 지속성을 갖는 Persistence 현상의 문제점을 수반하였다. 반면 SOM predicted는 예측값이 관측값에 대한 보다 정확한 예측을 해주고 있다. 그림 11과 12는 관측값과 예측값의 산포도를 나타낸 결과로 그림 11에서와 같이 기존의 인공신경망의 경우 수문곡선의 상승부는 산포도의 45° 선상의 아래에 분포하는 과소평가가 되고 있는 반면에 수문곡선의 하강부는 산포도의 45° 선상의 위에 분포하는 과대평가로 나타나고 있으며, 반시계방향으로 전이되는 Persistence 현상이 명확하게 나타나고 있으나, 그림 12에서와 같이 SOM을 전처리 과정으로 이용한 인공신경망의 경우는 산포도의 45도 상에 집중 분포되어 있으며 전이현상이 나타나지 않아 이전단계의 예측값에 지속성을 갖는 Persistence 현상이 제거되었음을 확인할 수 있었다. 또한  $R^2$ 와  $RMSE$ 의 비교에서도 SOM Predicted는 각각 0.9971과 0.0778로 BP Predicted의 0.8957과 0.09217보다 더 우수한 예측력을 보여주었다.

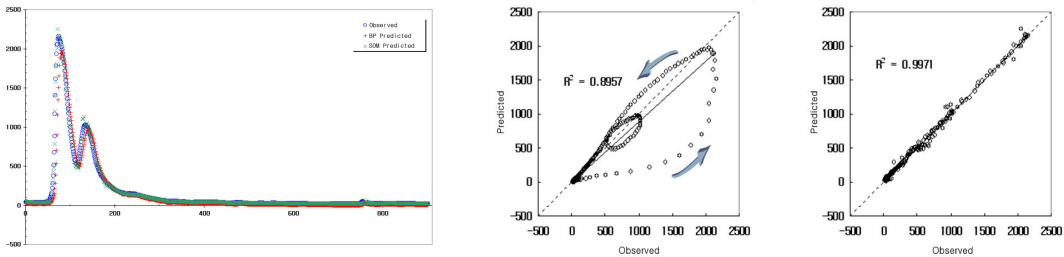


그림 10. 모형 검증 결과 그림 11. BP Predicted와 관측값의 산포도 그림 12. SOM Predicted와 관측값의 산포도

#### 5. 결론

SOM으로 구분된 분할구역(Cluster-1 ~ Cluster-7)과 유출과의 관계를 비교 검토한 결과 Cluster-1은 강우량은 감소하는 경향을 보이는 반면 유출량은 점진적으로 증가하는 수문곡선의 상승부 유출부분으로 구분되었고, Cluster-2는 수문곡선의 상승기점의 유출부분, Cluster-3은 강우량은 급격히 증가되나 유출량은 미미한 증가를 보이는 유출부분으로 분류되었다. 또한 Cluster-4는 변곡점 아래의 수문곡선 상승부와 감수부, Cluster-5는 수문곡선의 침두부로 분류되었다. Cluster-6은 변곡점 위의 수문곡선 상승부와 감수부, Cluster-7은 수문곡선의 기저유출 부분으로 분할되었다. 또한 기존의 인공신경망 모형과 SOM을 전처리 과정으로 이용한 인공신경망 모형을 상호 비교한 결과 훈련과정에 대한  $R^2$ 은 0.8942와 0.9799,  $RMSE$ 는 0.076과 0.0624로 각각 나타났으며, 검증과정의  $R^2$ 은 0.8957에서 0.9971,  $RMSE$ 는 0.09217에서 0.0778로 향상되어 과소 및 과대평가되는 현상을 개선할 수 있었고, 관측값과 예측값을 나타내는 산포도에 의하면 수문곡선의 상승부는 과소평가가 하강부의 과대평가로 반시계방향으로 전이되는 Persistence 현상이 명확하게 나타나고 있으나, SOM을 전처리 과정으로 이용한 인공신경망의 경우는 이상적인 적합을 나타내는 45도 선상에 집중 분포되어 있으며 전이현상이 나타나지 않아 Persistence 현상이 제거되었다. 마지막으로, SOM 적용의 심화 연구를 위해 강우자료 또는 유출량자료만을 이용한 단일변량의 패턴분류가 가능할 것으로 판단되며, 이는 각 변량의 본질적인 특성을 파악할 수 있을 것으로 기대된다.

#### 참고문헌

- 김용구(2006) 자기조직화 이론을 적용한 강우-유출 예측모형 개발, 박사학위논문, 동신대학교.
- 김주환(1993) 신경회로망을 이용한 하천 유출량의 수문학적 예측에 관한 연구, 박사학위논문, 인하대학교.

- 건설교통부 영산강홍수통계소(2004) **영산강 유량측정 보고서**.
- 김용구, 진영훈, 박성천(2006) 강우-유출특성 분석을 위한 자기조직화방법의 적용, **대한토목학회 논문집**, 제 26권, 제1B호, pp. 61-67.
- 이관수, 박성천, 이한민, 진영훈(2000) 인공신경망의 이론의 B.P.알고리즘을 적용한 영산강의 유출량 예측에 관한 연구, **대한토목학회 논문집**, 제 20권, 제5B호, pp. 679-688.
- Garcia, H. L. and Gonzalez, I. M.(2004) Self-organizing map and clustering for wastewater treatment monitoring, *Engineering Applications of Artificial Intelligence*, Vol. 17, pp. 215-225.
- Hsu, K. L., Gupta, H. V., Gao, X., Soroochian, S. and Imam, B.(2002) Self-organizing linear output map(SOLO): An artificial neural network suitable for hydrologic modeling and analysis, *Water Resources Research*, Vol. 38, No. 12, pp. 38-1-38-17.