

# 추계학적 신경망 접근법을 이용한 수문학적 시계열의 모형화

## Modeling of Hydrologic Time Series using Stochastic Neural Networks Approach

김성원\* · 김정현\*\* · 박기범\*\*\*

### Abstract

The goal of this research is to apply the neural networks models for the disaggregation of the pan evaporation (PE) data, Republic of Korea. The neural networks models consist of generalized regression neural networks model (GRNNM) and multilayer perceptron neural networks model (MLP-NNM), respectively. The disaggregation means that the yearly PE data divides into the monthly PE data. And, for the performances of the neural networks models, they are composed of training and test performances, respectively. The training and test performances consist of the historic, the generated, and the mixed data, respectively. From this research, we evaluate the impact of GRNNM and MLP-NNM for the disaggregation of the nonlinear time series data. We should, furthermore, construct the credible data of the monthly PE from the disaggregation of the yearly PE data, and can suggest the methodology for the irrigation and drainage networks system.

**Key words :** Pan Evaporation, Temporal Disaggregation Model, Stochastic Model, MLP-NNM, GRNNM

### 1. 서론

대기와 육지는 상호 연속적인 물분자의 교환이 있는데도 불구하고 일반적으로 증발 및 증발산의 수문학적 정의는 대기의 물이송의 순 비율에 한정되어 있다. 증발과 증발산은 제시된 구역에서 습윤효율 (Moisture efficiency)의 지표적인 변화이고, 그 크기는 하천에서 홍수유출량을 추정하는데 사용되기도 한다. 증발량은 수문학적 순환 중 하나이며, 일반적으로 간접적인 측정을 기초로 하는 질량 이송법, 에너지 수지법 및 물수지법과 같은 여러 방법으로 추정될 수 있다. 직접적인 측정기법 중 하나인 증발접시 증발량은 증발산량의 지표로 사용되고 있으며, 호수 및 저수지의 증발량을 추정하는데 가장 광범위하게 사용되고 있다 (Hargreaves, 1966; Kohler et al., 1955). 식물표면으로부터 증발산량의 추정은 물수지를 계산하고 용수수요와 공급을 추정하는데 있어서 기본적인 도구이다. 세계인구의 절반이상이 관개농업에 의한 생산에 의존하고 있으므로, 농작물 증발산량의 정량화는 매우 중요하다. 증발산량은 증발산계 (Lysimeter)를 이용하여 관측하거나 물수지 기법 혹은 기상변수를 이용하여 추정한다. 그러나 증발산계를 이용한 증발산량의 측정은 필요없이 많은 시간이 요구되는 작업이고, 정확하고 주의깊은 경험이 필요하기 때문에 항상 가능한 방법은 아니다. 따라서 기상자료를 기초로 하는 경험적인 접근법이 증발산량을 추정하는데 일반적으로 사용된다 (Allen et al., 1989; Monthieith, 1965; Penman, 1948).

본 연구의 목적은 연 증발접시 증발량의 시간적인 분해를 위하여 신경망모형을 적용하는데 있다. 신경망모형은 각각 다층 퍼셉트론 신경망모형과 일반화된 회귀신경망모형으로 구성되어 있다. 그리고 신경망모형의 수행평가를 위하여 훈련 및 테스트과정으로 구성되었다. 신경망모형의 훈련과정을 위하여 실측, 모의 및 혼합자료와 같은 세 가지 형태의 자료가 사용되었으며, 테스트과정을 위해서는 실측자료만 이용되었다. 본 연구에서는 비선형 수문학적 시계열자료의 시간적 분해를 위해서 MLP-NNM과 GRNNM의 적용성을 평가하였다. 게다가 연 증발접시 증발량 자료의 시간적 분해로부터 신뢰성 있는 월 증발접시 증발량자료를 구축할 수 있을 것이며, 관계매수 네트워크 시스템의 평가를 위한 이용가능한 자료를 제공할 수 있을 것이다.

### 2. 신경망모형과 유전자 알고리즘

#### 2.1 일반화된 회귀신경망모형 (Generalized Regression Neural Networks Model, GRNNM)

본 연구에 적용된 일반화된 회귀신경망모형 (GRNNM)은 반경기초함수 신경망모형 (Radial basis function neural networks model, RBFNNM)의 변형된 형태의 신경망모형이다. GRNNM은 입력층, 은닉층, 합산층 및 출력층의 4개의 층으로 구성되어 있으며, 비선형 회귀이론에 기초를 둔 신경망모형이다. 입력층, 은닉층 및 합산층노드는 완전히 연결된 구조로 형성되어 있으나, 출력층의 노드는 합산층의 노드 중 몇 개로만 연결이 되어있다. 합산층 (Summation layer)은 여러 개의 합산노드 (Summation node)와 하나의 제법노드 (Division node)의 2종류의 노드로 구성되어 있다. 합산노드수는 출력층 노드수와 동일하다. 제법노드는 어떤 전이함수를 사용하지 않고 은닉층노드의 가중전이값 (Weighted transfer value)을 합한 것과 같다.

\* 정회원, 공학박사, 수자원개발기술사, 동양대학교 철도토목학과 부교수 ; E-mail : [swkim1968@dyu.ac.kr](mailto:swkim1968@dyu.ac.kr)

\*\* 정회원, 공학사, 동양대학교 대학원 석사과정 ; E-mail : [idiunu@hanmail.net](mailto:idiunu@hanmail.net)

\*\*\* 정회원, 공학박사, 동양대학교 철도토목학과 겸임교수 ; E-mail : [pkb5032@naver.com](mailto:pkb5032@naver.com)

GRNNM의 각 출력층노드는 합산층의 합산노드와 제법노드로 연결되어 있으며, 합산층과 출력층 사이는 연결강도로 구성되어 있지 않다. 각 출력층노드의 계산은 합산층의 합산노드로부터 계산된 값을 제법노드로부터 계산된 값으로 나누어 산정한다. GRNNM의 훈련과정은 다층 퍼셉트론 신경망모형 (MLP-NNM)과는 상당히 다른 양상을 가지고 있다. 입력층과 은닉층 사이에서의 훈련과정은 반경기조함수 신경망모형의 경우처럼 자율훈련 (Unsupervised training)으로 구성되어서 K-means나 OLS 알고리즘과 같은 특별한 군집알고리즘을 필요로 하며, 군집알고리즘의 반영은 훈련이 시작하기 전에 결정되어 져야 한다. 또한 은닉층과 합산층 사이에서의 훈련과정은 은닉층에서의 출력값에 대하여 평균제곱오차의 최소화 과정에 기초를 둔 지도훈련 (Supervised training)으로 구성된다 (김성원 등, 2001; Kim and Jee, 2006; Specht, 1991; Tsoukalas and Uhrig, 1997; Wasserman, 1993). 본 연구에서 GRNNM은 은닉층 1개, 클러스터 센터는 20개, 은닉층 노드수는 12개, 합산층 노드수는 13개, 출력층 노드수는 12개, 전이함수는 tansigmoid, 최대 반복횟수는 10,000회 및 오차임계치는 0.001로 설정하였다. 또한 본 연구에서는 QuickProp 역전파 훈련알고리즘을 이용하였다. 다음 Figure 1은 본 연구에 적용된 GRNNM 구조를 나타낸 것이다.

## 2.2 다층퍼셉트론 신경망모형 (Multilayer Perceptron Neural Networks Model, MLP-NNM)

일반적으로 MLP-NNM은 입력층, 출력층 및 그 사이에 1개 혹은 그 이상의 은닉층으로 구성되어 있으며, 각 층에서의 노드는 그 다음 층의 노드와 상호 긴밀하게 연결되어 있다. 또한 본 연구에서 MLP-NNM을 적용하는 과정은 훈련 및 테스트과정으로 구성되어 있다. 일반적으로 훈련과정을 통하여 MLP-NNM의 최적 매개변수를 결정하며, 결정된 매개변수를 이용하여 MLP-NNM의 테스트과정을 수행한다 (Haykin 2003). 본 연구에서 MLP-NNM은 은닉층 1개, 은닉층 노드수는 12개, 출력층 노드수는 12개, 전이함수는 tansigmoid, 최대 반복횟수는 10,000회 및 오차임계치는 0.001로 설정하였다. 또한 본 연구에서는 QuickProp 역전파 훈련알고리즘을 이용하였다. 다음 Figure 2는 본 연구에 적용된 MLP-NNM 구조를 나타낸 것이다.

## 2.3 유전자 알고리즘 (Genetic Algorithm, GA)

본 연구에서 유전자 알고리즘 (GA)은 첫 번째는 훈련자료 및 교차검증자료를 이용하여 GRNNM을 훈련시키는 것이며, 두 번째는 첫 번째 과정에서 형성된 GRNNM을 이용하여 테스트자료에 대한 최적운명을 위하여 전 범위를 검토하는 것이다. GA는 각 입력층 노드에 대한 평활인자 승수 (Multiplier)를 찾고 있으며, 비록 훈련과정에 대하여 탐색기법으로 GA를 사용하지 않은 경우보다 시간은 좀 더 걸리지만 테스트자료에 대하여 가장 양호하게 운영될 수 있는 GRNNM을 구성할 수 있다. 본 연구에서는 GA의 적용을 위하여 세대수 (Generation)는 100, 염색체수 (Chromosome)는 50 및 최대 반복횟수는 10,000회로 설정하였다.

## 3. 연구범위 및 자료

본 연구에서는 제주도를 포함한 기상청의 관리하에 있는 71개의 기상관측소 중에서 서울, 강릉, 인천, 부산, 제주 및 목포관측소를 선정하였다. 서울관측소의 경우 1907년 10월에 기상관측을 실시한 이후 1919년, 1950-1953년 및 1991-1994년의 자료가 결측이 되어 있으나, 자료의 신뢰성이 아주 양호한 상태이다. 강릉관측소의 경우 1912년 1월에 기상관측을 실시한 이후에 1919년, 1950-1951년 자료가 결측이 되어 있으나, 서울관측소의 경우처럼 자료의 신뢰성이 아주 양호한 상태이다. 인천관측소는 1949년 1월에 기상관측을 실시한 이후에 1950-1951년 자료가 결측이 되어 있으나, 자료의 신뢰성은 양호하다. 그리고 부산은 1965년 1월부터, 제주 및 목포는 1980년 1월부터 기상관측을 실시하였고 결측자료는 없는 상태이다.

## 4. GRNNM 및 MLP-NNM의 적용

### 4.1 훈련과정 (Training Performance)

#### 4.1.1 실측자료의 이용

본 연구의 훈련과정에서 훈련자료는 서울관측소의 경우는 1908년부터 2002년까지의 결측기간을 제외한 86년의 자료를 선택하였으며, 강릉관측소의 경우는 1912년부터 2002년까지의 결측기간을 제외한 88년의 자료를 선택하였다. 인천관측소의 경우는 1949년부터 2002년까지의 결측기간을 제외한 52년의 자료를 선택하였다. 부산관측소의 경우는 1965년부터 2002년까지의 38년의 자료를 선택하였다. 그리고 제주 및 목포관측소는 1980년부터 2002년까지의 23년의 자료를 선택하였다. 그리고 GRNNM 와 MLP-NNM의 훈련에 따른 선행과정으로서 입력층의 입력변수인 연간 소형 증발접시 증발량자료를 표준화하여 적용하였다. 일반적으로 입력변수들을 표준화하는 주요한 이유 중 하나는 각 입력변수들이 서로 다른 단위로서 관측된 값을 나타내고 있기 때문에 이러한 입력변수들을 표준화하고 무차원단위에서 입력변수들을 재배치함으로써, 각 변수들의 상사성의 효과를 제거하기 위한 것으로 판단된다 (Kim and Kim, 2008a, b).

#### 4.1.2 모의자료의 이용

연구의 모의자료의 이용을 위한 훈련과정에서 훈련자료는 “4.1.1 실측자료의 이용” 에서 선정한 월 증발접시 증발량자료

를 모의자료의 구축을 위하여 이용하였다. 따라서 MLP-NNM 및 GRNNM의 훈련과정을 위한 입력층자료는 본 연구에서 선정한 추계학적 모형인 PARMA (1,1)을 이용하여 500년의 자료를 모의발생시켰다. 모의발생된 월 증발접시 증발량자료 중에서 초기에 발생하는 편차(Bias)를 제거하기 위하여 초기의 50년 자료는 제거하였으며, 나머지 450년의 월 증발접시 증발량자료를 이용하였다. 따라서 입력층자료에 해당하는 연 증발접시 증발량은 각 월 증발접시 증발량자료를 합하여 선정하였다. 그러므로 MLP-NNM 및 GRNNM의 훈련과정을 위한 입력층자료는 450개의 모의된 연 증발접시 증발량자료로 구성되어 있다.

#### 4.1.3 혼합자료의 이용

본 연구의 혼합자료의 이용을 위한 훈련과정에서 훈련자료는 “4.1.1 실측자료의 이용”에서 선정한 실측된 월 증발접시 증발량자료와 “4.1.2 모의자료의 이용”에서 선정한 모의된 월 증발접시 증발량자료를 합한 혼합자료로 구성하였다. 따라서 MLP-NNM 및 GRNNM의 훈련과정을 위한 입력층 자료수는 각각 서울관측소는 536개, 강릉관측소는 538개, 인천관측소는 502개, 부산관측소는 488개, 제주 및 목포관측소는 473개의 혼합된 연 증발접시 증발량자료로 구성되어 있다.

#### 4.2 테스트과정 (Test Performance)

본 연구의 테스트과정에서 테스트자료는 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 동일한 시계열인 2003년부터 2007년까지의 5년간의 자료를 선택하였다. 테스트과정은 훈련과정을 통하여 산정된 매개변수를 이용하여 테스트자료를 검증하였다. 다음 그림 3(a)-(f)는 서울, 강릉, 인천, 부산, 제주 및 목포관측소의 실측자료를 이용한 훈련과정의 수행결과 산정된 매개변수 및 테스트자료를 이용하여 MLP-NNM과 GRNNM의 테스트과정의 수행결과에 대한 월 증발접시 증발량 자료의 시계열의 비교를 나타낸 것이다.

### 5. 결 론

본 연구에서는 GRNNM과 MLP-NNM을 이용하여 연간 소형 증발접시 증발량자료를 월별 증발접시 증발량자료로의 분리를 시도하였으며, 사용된 자료는 서울, 강릉, 인천, 부산, 제주, 및 목포관측소의 자료를 이용하였다. MLP-NNM 및 GRNNM의 훈련과정의 수행결과 MLP-NNM의 경우 실측자료를 이용한 수행결과가 모의자료 및 혼합자료를 이용한 수행결과보다 양호한 것으로 분석되었고, GRNNM의 경우 목포관측소를 제외하고는 모의자료 및 혼합자료를 이용한 수행결과가 실측자료를 이용한 수행결과보다 양호한 것으로 분석되었다. MLP-NNM 및 GRNNM의 훈련과정의 수행결과 산정된 각 6종류의 매개변수 군을 이용하여 테스트과정을 수행하였다. 그 결과 모의자료 및 혼합자료의 훈련과정의 수행결과 산정된 매개변수를 이용한 테스트과정의 수행결과가 실측자료의 훈련과정의 수행결과 산정된 매개변수를 이용한 테스트과정의 수행결과보다 양호한 것으로 분석되었다. 따라서 본 연구를 통하여 분석에 이용된 자료의 수는 신경망모형의 수행 능력에 영향을 끼치는 것으로 판단된다. 본 연구를 통하여 신경망모형을 이용하여 연 증발접시 증발량자료에 대한 시간적 분해모형의 구축에 대한 적용성을 검토하였으며, 그에 따른 잠재성을 분석하였다. 그러나 시간적 분해모형의 보편적인 해석과 결론을 유도하기 위해서는 예를 들면 강우 및 유출량과 같은 더욱 다양한 시계열자료를 이용한 검증과정이 필요하며, 많은 종류의 신경망모형을 이용한 지속적인 연구가 필요한 것으로 판단된다.

#### 참고문헌

1. 김성원, 이순탁, 조정석 (2001). “중소하천유역에서 Hybrid Neural Networks에 의한 수문학적 예측.” **한국수자원학회 논문집**, 한국수자원학회, 제 34권, 제 4호, pp. 303-316.
2. Allen, R.G., Jensen, M.E., Wright, J.L., and Burman, R.D. (1989). “Operational estimates of reference evapotranspiration.” *Agrono. J.*, Vol. 81, No. 4, pp. 650-662.
3. Hargreaves, G.H. (1966). “Consumptive use computations from evaporation pan data.” *In Irrigation and Drainage Special Conference*, ASCE, Las Vegas, NV, pp. 35-62.
4. Haykin, S. (2003). *Neural networks : A comprehensive foundation*, 4th Indian Reprint, Pearson Education, Singapore.
5. Kim, S. and Jee, H. (2006). “An expansion of the ungedged pan evaporation using neural networks model in rural regions, South Korea.” *Proc. World Environmental & Water Resources Congress 2006*, ASCE/EWRI, Omaha, NE. [Printed in CD]
6. Kim, S. and Kim, H.S. (2008a). “Uncertainty reduction of the flood stage forecasting using neural networks model.” *J. Americ. Water Resor. Associ.*, Vol. 44, No. 1, pp. 148-165.
7. Kim, S. and Kim, H.S. (2008b). “Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling.” *J. Hydro.*, Vol. 351, pp. 299-317.
8. Kohler, M.A., Nordenson, T.J., and Fox, W.E. (1955). *Evaporation from pans on lakes*, US Department of Commerce, Weather Bureau Research Paper 38, Washington, DC.
9. Monteith, J.L. (1965). “The state and movement of water in living organism.” *Proc., Evaporation and Environment, XIXth Symp.*, soc. For Exp. Biol., Swansea, Cambridge Univ. Press, NY, pp. 205-234.

10. Penman, H.L. (1948). "Natural evaporation from open water, bare soil and grass." *Proc. R. Soc. London*, 193, pp. 120-146.
11. Specht, D.F. (1991). "A general regression neural network." *IEEE Trans. on Neural Networks*, Vol. 2, No. 6, pp. 568-576.
12. Tsoukalas, L.H. and Uhrig, R.E. (1997). *Fuzzy and neural approaches in engineering*, John Wiley & Sons Incorporated, New York.
13. Wasserman, P.D. (1993). *Advanced methods in neural computing*, Van Nostrand Reinhold, New York.

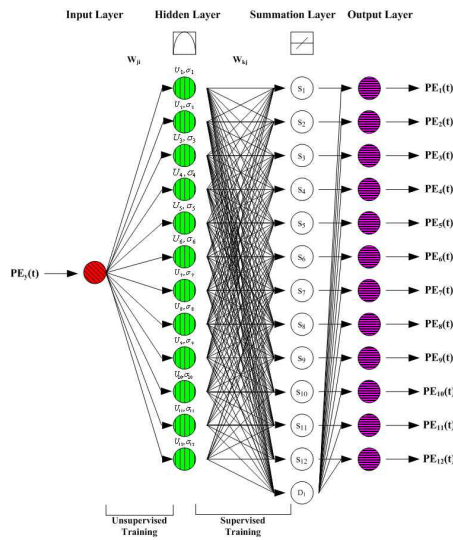


Figure 1. GRNNM structure

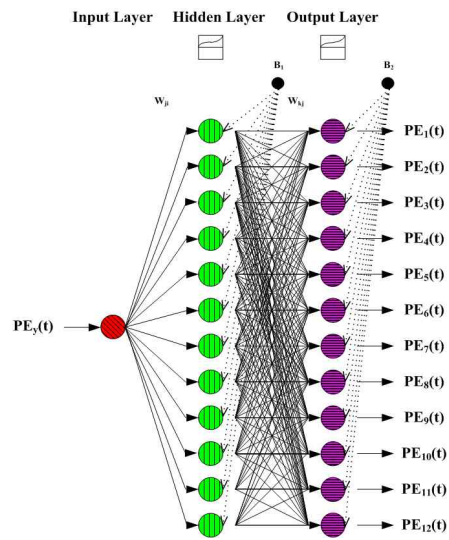


Figure 2. MLP-NNM structure

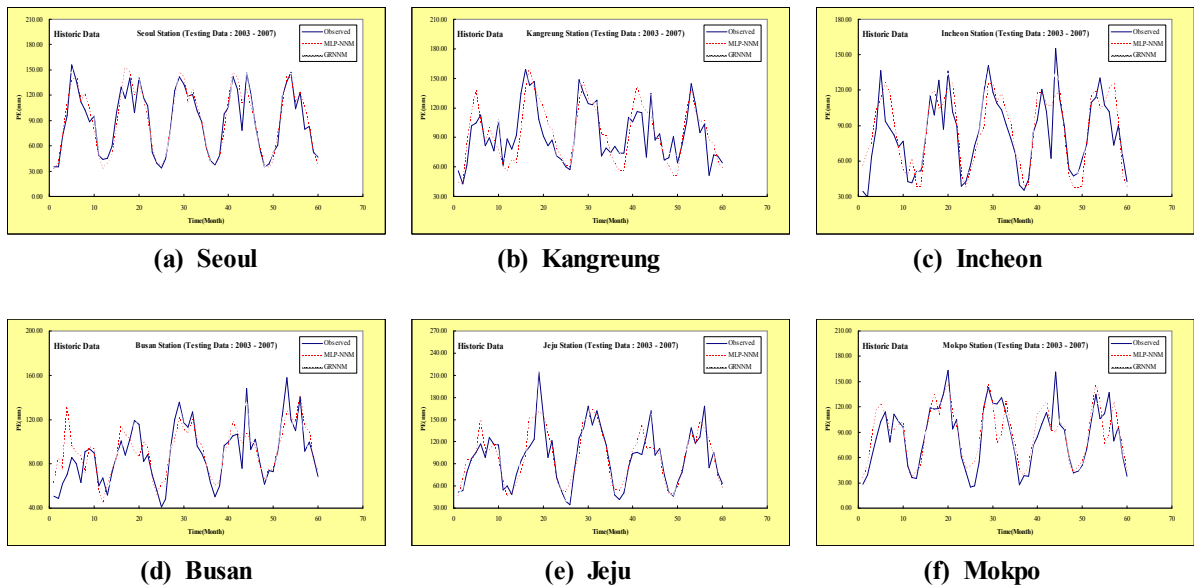


Figure 3. Comparison of the monthly PE for the testing performance (Historic data)