

비선형 모형화를 위한 추계학 및 신경망이론의 통합운영

Integrational Operation of Stochastics and Neural Networks Theory for Nonlinear Modeling

김성원*

Abstract

The goal of this research is to develop and apply the integrational model for the pan evaporation and the alfalfa reference evapotranspiration in Republic of Korea. Since the observed data of the alfalfa reference evapotranspiration using lysimeter have not been measured for a long time in Republic of Korea, PM method is used to assume and estimate the observed alfalfa reference evapotranspiration. The integrational model consists of stochastics and neural networks processes respectively. The stochastics process is applied to extend for the short-term monthly pan evaporation and alfalfa reference evapotranspiration. The extended data of the monthly pan evaporation and alfalfa reference evapotranspiration is used to evaluate for the training performance. For the neural networks process, the generalized regression neural networks model(GRNNM) is applied to evaluate for the testing performance using the observed data respectively. From this research, we evaluate the impact of the limited climatic variables on the accuracy of the integrational operation of stochastics and neural networks processes. We should, furthermore, construct the credible data of the pan evaporation and the alfalfa reference evapotranspiration, and suggest the reference data for irrigation and drainage networks system in Republic of Korea.

Key words : Stochastics, GRNNM, Pan evaporation, Alfalfa reference evapotranspiration

1. 서 론

대기와 육지는 상호 연속적인 물분자의 교환이 있는데도 불구하고 일반적으로 증발 및 증발산의 수문학적 정의는 대기로의 물이송의 순 비율에 한정되어 있다. 증발과 증발산은 제시된 유역에서 습유효율(Moisture efficiency)의 지표적인 변화이고, 그 크기는 하천에서 흥수유출량을 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; Montheith, 1965; Penman, 1948). Jensen 등(1990)은 세계 각 지역의 서로 다른 기후대에 위치한 11개의 지점에서 증발산계로 측정한 증발산량을 20 여개의 서로 다른 경험공식 및 방법론을 이용하여 산정한 값과 비교분석하였다. 그 결과 Penman-Monteith(PM) 공식이 모든 기후대에 대하여 가장 최적의 결과치를 나타내는 것으로 분석되었다. PM 공식을 제외하고는 지역적인 보정방법이나 기후조건에 따라서 순위가 변화하는 것으로 나타났다. 따라서 PM 공식은 증발산량 실측자료가 존재하지 않을 경우에 다른 공식과의 비교를 위한 하나의 기준방법론으로 고려될 수 있다(FAO, 1990). 본 연구의 목적은 추계학적 모형의 하나인 주기성 자기회귀이동평균(Periodic autoregressive moving average, PARMA) 모형과 신경망모형을 연계하여 우리나라에서 결측 및 미계측 월별 증발접시 증발량과 측정 자료가 전무한 월별 알팔파 기준증발산량을 산정하는 모형을 개발 및 통합운영 시스템을 구축하고자 한다.

2. 신경망모형과 유전자 알고리즘

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 구성되었으며, 비선형 회귀이론에 기초를 둔 신경망모형이다. 입력층, 은닉층 및 합산층노드는 완전히 연결된 구조

* 정희원, 공학박사, 수자원개발기술사, 동양대학교 철도토목학과 조교수 ; E-mail : swkim68@phenix.dyu.ac.kr

로 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. GRNNM의 각 출력층노드의 계산은 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 디층 퍼셉트론 신경망모형(MLPNNM)과는 상당히 다른 양상을 가지고 있다. 입력층과 은닉층 사이에서의 훈련과정은 반경기초함수 신경망모형의 경우처럼 자율훈련(Unsupervised training)으로 구성되어서 K-means나 OLS 알고리즘과 같은 특별한 군집알고리즘을 필요로 하며, 군집알고리즘의 반경은 훈련이 시작하기 전에 결정되어 져야 한다. 또한 은닉층과 합산층 사이에서의 훈련과정은 은닉층에서의 출력값에 대하여 평균제곱오차의 최소화과정에 기초를 둔 지도훈련(Supervised training)으로 구성된다(김성원 등, 2001; Kim and Jee, 2006; Specht, 1991; Tsoukalas and Uhrig, 1997; Wasserman, 1993). 본 연구에서 유전자 알고리즘은 각 입력층 노드에 대한 적절한 개별적인 평활인자(Smoothing factor)뿐만 아니라 전반적인 평활인자를 결정하는데 사용된다. 유전자 알고리즘을 이용한 GRNNM에 대한 훈련 수행과정은 크게 두 부분으로 진행되어 진다. 첫 번째는 훈련자료를 이용하여 GRNNM을 훈련시키는 것이며, 두 번째는 첫 번째 과정에서 형성된 GRNNM을 이용하여 테스트자료에 대하여 가장 최적운영을 위한 평활인자의 전 범위를 테스트하는 것이다. 유전자 알고리즘은 각 입력층 노드에 대한 평활인자 승수(Multiplier)를 찾고 있으며, 비록 훈련에 대하여 탐색기법으로 유전자 알고리즘을 사용하지 않은 경우보다 시간은 좀 더 걸리지만 테스트 자료에 대하여 가장 양호하게 운영될 수 있는 GRNNM을 생산할 수 있다(Neuroshell 2, 1993). 본 연구에서는 GRNNM의 훈련을 위하여 유전자 알고리즘을 적용시킨 GRNNM-GA 및 역전파 훈련 알고리즘을 적용시킨 GRNNM-BP를 이용하여 월별 증발접시 증발량 및 알팔파 기준증발산량을 모의발생된 자료와 연계하여 산정하고자 한다. 다음 그림 1은 본 연구에 적용된 광주관측소의 GRNNM 구조를 나타낸 것이다.

3. 주기성 자기회귀이동평균(PARMA) 모형

3.1 주기성 자기회귀이동평균 모형의 개념

주기성 자기회귀이동평균 모형은 주기성 자기회귀(Periodic Autoregression, PAR)모형에 주기성의 이동평균 매개변수를 포함하여 확장하며 일반적으로 PARMA(p,q)로 나타낸다. 저차수의 PARMA 모형은 주기적인 수문시계열의 모형화에 사용된다. 예를들면 PARMA(1,1) 모형은 다음 식(1)과 같이 나타낼 수 있다.

$$y_{v,\tau} = \mu_\tau + \Phi_{1,\tau}(y_{v,\tau-1} - \mu_{\tau-1}) + \varepsilon_{v,\tau} - \Theta_{1,\tau}\varepsilon_{v,\tau-1} \quad (1)$$

여기서 v 는 년(year), τ 는 계절(Season)이고 $\tau = 1, 2, \dots, \omega$ 이며 계절수를 나타내고 있다. 식(1)과 같은 모형은 주로 월 유출량계열에 적용된다(Salas et al., 1993). 게다가 PARMA(2,1), PARMA(2,2) 및 좀 더 복잡한 형태의 PARMA 모형은 계절 및 연간 통계치의 유지가 필요할 경우에 하천유출량의 모의발생에 필요하다.

3.2 주기성 자기회귀이동평균 모형의 구성

본 연구에서는 GRNNM의 훈련에 부족한 실측 기상자료를 모의발생 하기위하여 PARMA 모형을 이용하였다. 사용된 모형은 PARMA(1,1) 모형으로 구성하였으며, 계절의 수는 12계절 즉 12달로 하였다. 그리고 모의발생한 표본은 2개의 표본으로 하였으며, 각 표본마다 100년, 500년 및 1,000년씩의 월평균 평균기온, 최대기온, 최소기온, 평균 이슬점온도, 최소 상대습도, 평균 상대습도, 평균풍속, 최대풍속 및 일조시간을 모의발생시키고, 또한 월별 총 증발접시 증발량 및 알팔파 기준증발산량을 모의발생하였다. 모의된 자료 중에서 첫 번째 표본자료는 선택하지 않았으며, 두 번째 표본자료를 선택하였다. 또한 두 번째 표본자료에서 초기에 발생되는 편차(Bias)를 제거하기 위하여 초기의 50년 자료는 제거하였다. 각 PARMA 모형의 매개변수는 근사 최소자승법(Method of approximate least square)을 이용하여 산정하였다.

4. GRNNM-GA 및 GRNNM-BP의 적용

4.1 훈련과정

본 연구에서는 광주관측소의 월평균 평균기온, 최대기온, 최소기온, 평균 이슬점온도, 최소 상대습도, 평균풍속, 최대풍속 및 일조시간과 월별 총 증발접시 증발량 및 알팔파 기준증발산량 자료를 이용하여 PARMA(1,1) 모형에 의해 모의발생된 자료를 GRNNM-GA 및 GRNNM-BP의 훈련자료로 구성하였다. 따라서 훈련에 사용된 자료는 모두 50년, 450년 및 950년으로 구성되어, 훈련자료는 600개, 5,400개 및 11,400개의 자료시계열로 구성되어 있다. 그리고 GRNNM-GA 및 GRNNM-BP의 훈련에 따른 선행과정으로서 모든 입력층의 입력변수들은 표준화하여 적용하였다. 입력변수들을 표준화하는 주요한 이유 중 하나는 각 입력변수들이 서로 다른 단위로서 관측된 값들을 나타내고 있기 때문에 이러한 입력변수들을 표준화하고 무차원단위에서 입력변수들을 재배치함으로써, 각 변수들의 상사성의 효과를 제거하기 위한 것으로 판단된다.

4.2 테스트과정

본 연구에서는 광주관측소의 1985년부터 1990년의 6년 동안의 실측 월평균 평균기온, 최대기온, 최소기온, 평균 이슬점온도, 최소 상대습도, 평균 상대습도, 평균풍속, 최대풍속 및 일조시간의 기상자료와 월별 총 증발접시 증발량 및 PM 공식에 의해 산정된 월별 알팔파 기준증발산량을 GRNNM-GA와 GRNNM-BP의 테스트자료로 구성하였으며, 총 72개의 자료시계열로 구성되어 있다. 본 연구의 테스트과정은 GRNNM-GA와 GRNNM-BP의 과적합문제(Overfitting problem)

를 극복하기 위하여 교차검증법(Cross-validation method)을 적용하였다. 교차검증법은 GRNNM-GA와 GRNNM-BP가 최소 RMSE에 도달할 때까지 모든 훈련자료를 훈련시키게 하는 것이 아니고, 각 훈련단계의 마지막에 테스트자료로서 교차검증하는 방법이다. 만약 과접합문제가 발생되면 테스트자료의 평균제곱오차의 수렴과정은 훈련자료가 여전히 훈련되고 있는 경우에도 감소하지 않고 증가하는 경향을 나타낼 것이다(Bishop, 1994; Haykin, 1994; 김성원, 2005). 다음 그림 2의 (a)-(b)는 광주관측소에 대한 PARMA(1,1)에 의해 1,000년동안 모의발생시킨 자료를 훈련자료로 이용할 경우의 훈련 및 테스트결과에 따른 월별 총 증발접시 증발량 및 알팔파 기준증발산량의 비교결과를 나타낸 것이다. 또한 다음 표 1은 광주관측소에 대한 총 증발접시 증발량의 경우 PARMA(1,1)에 의해 100년, 500년 및 1,000년동안 GRNNM의 훈련과정에 의해 모의발생시켰을 경우에 테스트자료에 대한 GRNNM-GA와 GRNNM-BP의 통계분석 결과를 나타내고 있다.

5. 결 론

본 연구에서는 GRNNM-GA와 GRNNM-BP를 주기성의 추계학적 모형을 연계하여 월별 총 증발접시 증발량과 알팔파 기준증발산량의 값을 산정할 수 있었다. 또한 본 연구에서는 광주관측소의 경우 모의발생된 자료의 수가 1,000년일 때 모형출력값의 정도는 양호한 것으로 나타났으며, GRNNM-GA가 GRNNM-BP보다 양호한 것으로 나타났다. 본 연구를 통하여 미계측 기상자료가 많이 존재하는 기상관측소의 경우 적은 수의 실측 기상자료를 이용하여 양호한 기상자료의 확충을 위한 모의발생이 가능하며, 또한 확충된 기상자료를 이용하여 우리나라 관개배수 시스템 구성을 위한 기준자료를 제공할 수 있을 것으로 판단된다.

감 사 의 글

본 연구에서는 2006년 동양대학교 연구비지원(Contract : 2006-1012)에 의해 수행되었습니다. 본 연구를 가능하게 한 동양대학교에 감사드립니다.

참 고 문 헌

1. 김성원(2005). “신경망모형에 의한 흥수위예측의 신뢰성분석 1. 모형의 개발 및 적용.” *대한토목학회 논문집*, 대한토목학회, 제 25권, 제 6B호, pp. 473-482.
2. 김성원, 이순탁, 조정석(2001). “중소하천유역에서 Hybrid Neural Networks에 의한 수문학적 예측.” *한국수자원학회 논문집*, 한국수자원학회, 제 34권, 제 4호, pp. 303-316.
3. 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.
4. Bishop, C.M.(1994). “Neural networks and their applications.” *Rev. Scien. Instru.* Vol. 65, pp. 1803-1832.
5. Food and Agriculture Organization(FAO).(1990) *Report on the expert consultation on revision of FAO methodologies for crop water requirement*, Land and Water Devel. Div., Rome, Italy.
6. Hargreaves, G.H.(1966). “Consumptive use computations from evaporation pan data.”, *In Irrigation and Drainage Special Conference*, ASCE, Las Vegas, NV, pp. 35-62.
7. Haykin, S.(1994). *Neural networks : A comprehensive foundation*, Macmillan College Pub. Comp., Inc., MA.
8. Jensen, M.E., Burman, R.D., and Allen, R.G.(1990). *Evapotranspiration and irrigation water requirements*, ASCE Manual and Report on Engineering Practice No. 70, ASCE, NY.
9. Kim, S., and Jee, H.(2006). “An expansion of the ungaged 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]
10. 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.
11. 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.
12. Neuroshell 2(1993). Ward systems group, Inc., MD.
13. Penman, H.L.(1948). “Natural evaporation from open water, bare soil and grass.” *Proc. R. Soc. London*, 193, pp. 120-146.
14. Salas, J.D.(1993). Chapter 19. Analysis and modeling of hydrologic time series, *Handbook of hydrology*, Maidment, D.R, ed., McGraw-hill Inc, New York, pp. 1-72.
15. Specht, D.F.(1991). “A general regression neural network.” *IEEE Trans. on Neural Networks*, Vol. 2, No. 6, pp. 568-576.
16. Tsoukalas, L.H. and Uhrig, R.E.(1997). *Fuzzy and neural approaches in engineering*, John Wiley & Sons Incorporated, New York.
17. Wasserman, P.D.(1993). *Advanced methods in neural computing*, Van Nostrand Reinhold, New York.

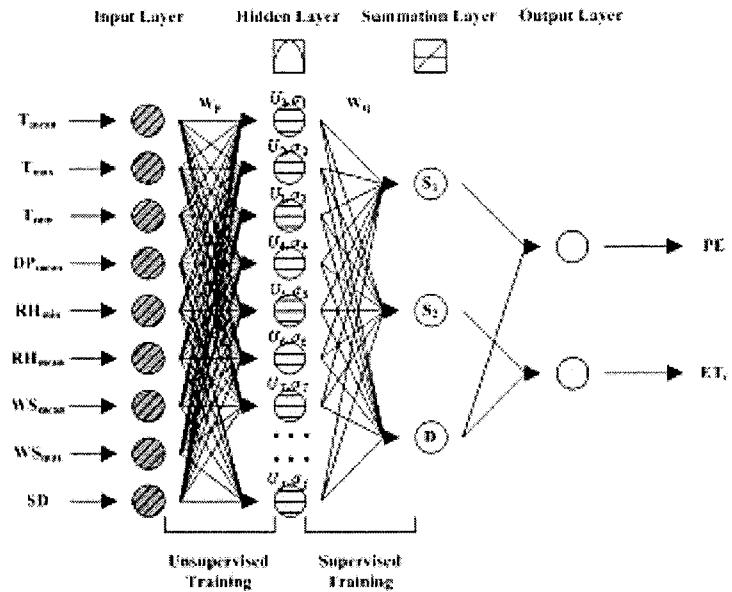


그림 1. GRNNM 의 구조 (광주관측소)

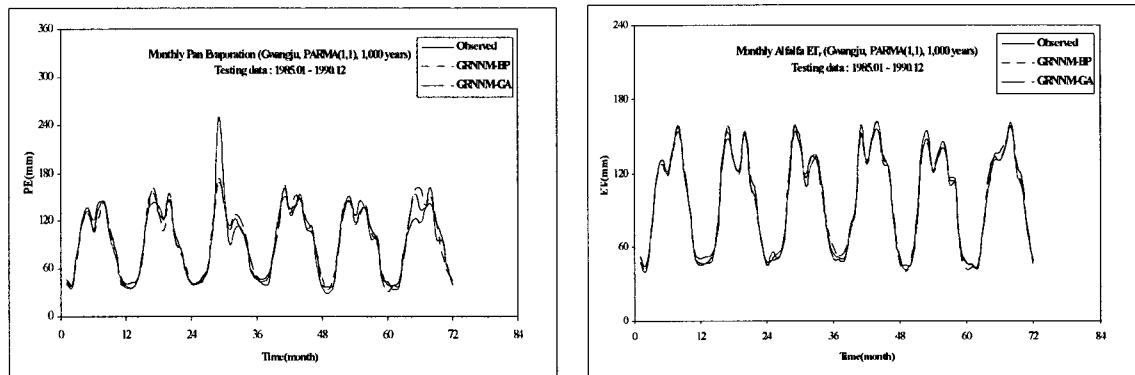


그림 2(a) 증발접시 증발량의 비교(광주)

그림 2(b) 알팔파 기준증발산량의 비교(광주)

표 1. 추계학 및 신경망모형을 연계한 통계분석 결과 (광주관측소, 증발접시 증발량)

지점	통계특성치	100 year		500 year		1,000 year	
		GRNNM-BP	GRNNM-GA	GRNNM-BP	GRNNM-GA	GRNNM-BP	GRNNM-GA
광주	CC	0.916	0.954	0.935	0.955	0.941	0.957
	RMSE(mm)	17.618	13.359	15.625	13.052	14.863	12.790
	E	0.839	0.908	0.873	0.912	0.886	0.915
	MAE(mm)	11.375	7.906	10.145	7.954	9.680	7.761