

신경망기법을 이용한 수문학적 분해모형

김 성 원**

Hydrologic Disaggregation Model using Neural Networks Technique

Sungwon Kim**

요약 : 본 연구의 목적은 연 증발접시 증발량의 수문학적 분해를 위하여 신경망모형을 적용하는데 있다. 신경망모형은 각각 다층 퍼셉트론 신경망모형(MLP-NNM)과 지지벡터기구 신경망모형(SVM-NNM)으로 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. 평가를 위하여 4가지의 통계학적 지표(CC, RMSE, E, AARE)가 각각 제시되었으며, ANOVA 및 Mann-Whitney U 검증을 이용하여 실측 및 계산된 월 증발접시 증발량자료에 동질성검증을 실시하였다. 본 연구를 통하여 비선형 시계열자료의 수문학적 분해를 위해서 MLP-NNM과 SVM-NNM의 적용성을 평가하였다. 게다가 연 증발접시 증발량 자료의 수문학적 분해로부터 신뢰성있는 월 증발접시 증발량자료를 구축할 수 있을 것이며, 관계배수 네트워크 시스템의 평가를 위한 이용가능한 자료를 제공할 수 있을 것이다.

핵심용어 : 증발접시 증발량, 수문학적 분해모형, 추계학적 모형, MLP-NNM, SVM-NNM

Abstract : The purpose of this research is to apply the neural networks models for the hydrologic disaggregation of the yearly pan evaporation(PE) data in Republic of Korea. The neural networks models consist of multilayer perceptron neural networks model(MLP-NNM) and support vector machine neural networks model(SVM-NNM), respectively. And, for the evaluation of the neural networks models, they are composed of training and test performances, respectively. The three types of data such as the historic, the generated, and the mixed data are used for the training performance. The only historic data, however, is used for the testing performance. The application of MLP-NNM and SVM-NNM for the hydrologic disaggregation of nonlinear time series data is evaluated from results of this research. Four kinds of the statistical index for the evaluation are suggested; CC, RMSE, E, and AARE, respectively. Homogeneity test using ANOVA and Mann-Whitney U test, furthermore, is carried out for the observed and calculated monthly PE data. We can construct the credible monthly PE data from the hydrologic disaggregation of the yearly PE data, and the available data for the evaluation of irrigation and drainage networks system can be suggested.

Keywords : Pan Evaporation, Hydrologic Disaggregation Model, Stochastic Model, MLP-NNM, SVM-NNM

1. 서 론

증발과정은 자연적 및 농업적인 환경시스템의 에너지와 물 수지의 주요인자를 나타내며, 증발과정에 의한 수분손실량의 추정은 수자원의 모니터

링, 조사 및 운영, 관계배수시스템의 설계와 관계 계획에 있어서 주요한 인자이다(Molina Martinez et al., 2005; Gundekar et al., 2008). 증발과정은 일반적으로 질량전이(Mass transfer)와 물 수지 방법과 같은 간접적인 방법에 의하여 관측되어

+ Corresponding author : swkim1968@dyu.ac.kr

* 교신저자 · 정회원 · 동양대학교 철도토목학과 부교수 · 공학박사 · 수자원개발기술사

Corresponding Author, Associate Professor, Ph.D./P.E., Dept. of Rail. and Civil Engr., Dongyang University, Yeongju, South Korea

지며, 증발량 관측을 위한 직접적인 방법중 하나는 증발접시 증발량이다(Eslamian et al., 2008). 계획되고 있는 혹은 기존의 저수지를 비롯하여 관계 프로젝트가 있는 모든 지점에 대하여 증발접시를 설치한다는 것은 비현실적이며, 정확한 관측도구가 설치될 수 없거나 혹은 유지관리가 어려운 접근할 수 없는 지역을 가지는 것 또한 비합리적이다(Kisi, 2006). 증발접시의 이용은 도구적인 제한과 관측 오차, 유수의 교란, 조류 혹은 다른 동물들에 의한 유수의 이용뿐만 아니라, 증발량 관측의 정도를 감소시킬 수 있는 여러 가지 유지문제와 같은 실제적인 문제에 의하여 영향을 받는다(Jensen, et al., 1990). 증발접시가 이용될 수 없는 곳에서의 증발접시 증발량의 추정에 대한 연구는 수문학자, 농학자 및 기상학자들에게는 중요한 의미를 가지고 있다(Kisi, 2006). 증발과정은 자연현상에서 매우 심한 비선형관계이며, 증발접시 증발량의 추정에 대한 연구는 진행과정에 있어서 내재적인 비선형성을 나타낼 수 있는 모형을 통하여 접근 할 수 있다(Kisi, 2006; Eslamian et al., 2008).

최근에 많은 연구자들은 증발접시 증발량의 산정에 있어서 신경망이론을 이용한 모형화 기법의 적용성을 증명하여 왔다(Keskin and Terzi, 2006; Deswal and Pal, 2008; Rahimi Khoob, 2009). Sudheer et al.(2002)은 신경망이론을 이용하여 Class A Pan 증발접시 증발량의 예측을 검토하였으며, 예측 결과치는 기존의 공식 중에서 Stephens and Stewart 공식과 비교되었다. 그들은 이용할 수 있는 기상인자로부터 증발과정을 모형화 하는데 신경망기법의 이용을 제시하였으며, 온도, 상대습도, 일조시간 및 풍속과 같은 관측된 기상인자들의 적절한 조합을 이용하였다. Bruton et al.(2000)은 입력자료로서 관측된 기상인자들을 이용하여 일 증발접시 증발량을 산정하는데 신경망 기법을 이용하였다. 또한 일 증발접시 증발량을 다중선형회귀분석 및 Priestly-Taylor 공식을 이용하여 산정하였으며, 신경망모형의 결과치와 비교하였다. 그들은 강우, 온도, 태양열 방사 및 풍속과 같은 관측된 기상인자들의 적절한 조합을 이용하였

다. Kisi(2006)는 관측된 기상학적 변수를 이용하여 일 증발접시 증발량을 추정하는데 뉴로-퍼지모형을 이용하였다. 그는 뉴로-퍼지모형을 위하여 온도, 태양열방사, 풍속, 대기압 및 상대습도와 같은 관측된 기상인자들의 적절한 조합을 이용하였다. 그는 이용할 수 있는 기상학적 자료로부터 증발과정을 모형화하는데 있어서 뉴로-퍼지모형이 성공적으로 이용될 수 있다는 것을 나타내었다. 또한 Kim and Kim(2008)은 우리나라에서 일 증발접시 증발량과 알팔파 기준증발산량의 모형화를 위하여 유전자알고리즘이 내재된 신경망모형을 이용하였으며, 입력층노드의 불확실성 분석을 통하여 최적모형을 선정하여, 기뮴해석 및 관계배수 시스템해석을 위한 참고자료를 제공할 수 있는 증발접시 증발량 및 알팔파 기준증발산량의 지도를 구축하였다.

증발접시 증발량의 산정에 있어서 추계학적 모형과 신경망모형을 연계한 연구성과로는 김성원과 김형수(2008)가 월 증발접시 증발량과 월 알팔파 기준증발산량의 모형화를 위하여 추계학적 모형과 신경망모형을 연계한 통합운영방법을 개발하여 적용하였다. 그 결과 통합운영방법을 이용하여 최소비용과 노력으로 월 증발접시 증발량과 월 알팔파 기준증발산량을 동시에 모형화가 가능하게 되었다. 또한 강우 및 유출량과 같은 비선형 시계열자료에 대한 수문학적 분해모형의 개발 및 적용에 대한 연구(Burian et al., 2000, 2001; Gutierrez-Magness and McCuen, 2004; Tan et al., 2007; Zhang et al., 2008; Choi et al., 2008)는 수행된 경험이 있으며, 증발접시 증발량에 대한 수문학적 분해에 대한 연구로는 김성원 등(2010)이 MLP-NNM과 GRNNM을 이용하여 연 증발접시 증발량을 월 증발접시 증발량으로 분해를 시도하였다. 그러나 SVM-NNM을 이용한 연 증발접시 증발량을 월 증발접시 증발량으로 분해를 시도한 연구사례는 아직 없다.

본 연구의 목적은 우리나라 주요 기상관측소에서 관측된 연 증발접시 증발량자료를 제시된 두 가지의 신경망모형인 MLP-NNM과 SVM-NNM을 이용하여 월 증발접시 증발량자료로 분리하고자 한

다. 신경망모형을 적용한 수문학적 분해모형의 훈련과정 수행을 위해 구성된 훈련자료는 세 가지의 종류로 이루어지며, 각각 실측자료, 모의자료(추계학적 모형을 이용하여 모의발생) 및 혼합자료(실측자료와 추계학적 모형을 이용하여 모의발생시킨 모의자료를 혼합)로 각각 구성된다. 그리고 훈련과정의 수행결과 선정된 최적 매개변수를 이용하여 실측자료로서 테스트과정을 실시하고자 한다. 또한 본 연구에서는 신경망모형을 이용하여 신뢰성있는 월 증발접시 증발량을 산정할 수 있는 수문학적 분해모형의 개발을 제시하고자 한다.

2. 신경망모형과 유전자 알고리즘

2.1 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, 2009). 본 연구에서 적용된 MLP-NNM은 은닉층 1개, 은닉층 노드수는 12개, 출력층 노드수는 12개, 최대 반복횟수는 10,000회, 오차 임계치는 0.001로 설정하였다. 또한 본 연구에서는 QuickProp 역전파 훈련알고리즘을 이용하였다. 따라서 본 연구에서 적용된 MLP-NNM의 출력층 노드중에서 1월, 6월 및 12월의 노드는 다음 식(1a)-(1c)와 같이 나타낼 수 있으며, 다음 그림 1은 본 연구에 적용된 MLP-NNM 구조를 나타낸 것이다.

$$PE_1(t) = \Phi_2 \left[\sum_{j=1}^n W_{1j} \cdot \Phi_1 \left(\sum_{j=1}^{12} W_{ji} \cdot PE_y(t) + B_1 \right) + B_2 \right] \quad (1a)$$

$$PE_6(t) = \Phi_2 \left[\sum_{j=1}^n W_{6j} \cdot \Phi_1 \left(\sum_{j=1}^{12} W_{ji} \cdot PE_y(t) + B_1 \right) + B_2 \right] \quad (1b)$$

$$PE_{12}(t) = \Phi_2 \left[\sum_{j=1}^n W_{12j} \cdot \Phi_1 \left(\sum_{j=1}^{12} W_{ji} \cdot PE_y(t) + B_1 \right) + B_2 \right] \quad (1c)$$

여기서 i, j, k 는 입력층, 은닉층 및 출력층을 나타내고 있으며, $PE_1(t)$, $PE_6(t)$, $PE_{12}(t)$ 는 계산된 1월, 6월 및 12월 증발접시 증발량, $\Phi_1(\cdot)$ 는 은닉층의 시그모이드 전이함수, $\Phi_2(\cdot)$ 는 출력층의 시그모이드 전이함수, W_{1j} , W_{6j} , W_{12j} 는 1월, 6월 및 12월의 은닉층과 출력층사이의 연결강도, W_{ji} 는 입력층과 은닉층사이의 연결강도, $PE_y(t)$ 는 연 증발접시 증발량의 시계열, B_1 은 은닉층에서의 편차, B_2 는 출력층에서의 편차를 나타낸 것이다.

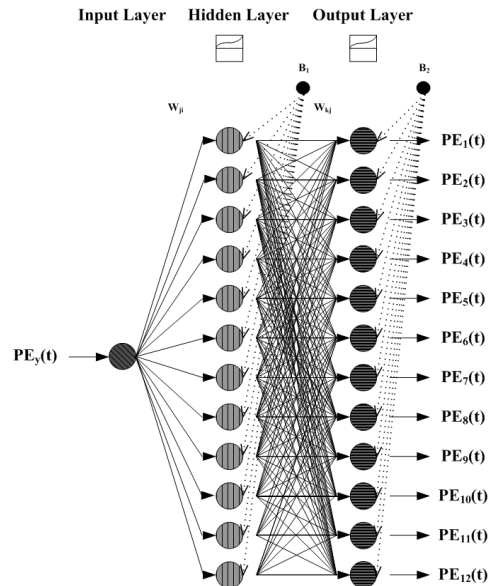


Fig. 1 The developed architecture of MLP-NNM

2.2 지지벡터기구 신경망모형

(Support Vector Machine Neural Networks Model, SVM-NNM)

지지벡터기구 신경망모형(SVM-NNM)은 패턴인식, 회귀, 멀티미디어, 생물정보학 및 인공지능을 포함하여 여러분야에서 광범위하게 적용되어 왔으며, 최근에는 수문학에서의 적용성을 검증하였다(Dibike et al., 2001; Khadam and Kaluarachchi, 2004). SVM-NNM은 경험적인 위험도 최소화보다는 구조적인 위험도 최소화에 기초를 두기 때문에 주어진 훈련자료에 대하여 보편적인 신경망모형과 비교하여 양호한 일반화 오차를 제공한다. 따라서 보편적인 신경망모형은 국부적인 최적 해에 수렴될 가능성을 가지고 있지만 SVM-NNM은 광역적인 최적 해를 보장한다(Haykin, 2009). SVM-NNM은 다음 두 가지 개념에 의하여 운영되는 새로운 개념의 분석기법이다. 첫 번째는 고차원공간으로의 자료의 전환이며, 이것은 복잡한 결정 평면들을 가진 복잡한 문제를 선형 판별식의 함수로 사용할 수 있는 단순한 문제로 전이 할 수 있다. 두 번째는 SVM-NNM의 경우 분류에 대한 좀 더 많은 정보를 제공할 수 있기 때문에 결정 평면 근처에 있는 단지 그들 입력을 이용하고, 훈련의 개념에 의하여 동기를 가지게 된다. SVM-NNM 적용에 있어서 첫 번째의 단계는 고차원공간으로 전이하는 것이다. 일반적으로 이것은 각 자료에 가우시안을 설정하는 반경기초함수(Radial Basis Function, RBF)를 사용하여 이루어진다. 그리하여 특성공간(Feature space)은 자료의 수만큼 증가하게 된다. 그러나 일반적으로 반경기초함수는 마지막 결과치를 계산하기 위하여 가우시안의 선형조합을 훈련시키기 위하여 역전파 훈련알고리즘(BackPropagation training algorithm)을 이용한다. 그러나 본 연구에서 SVM-NNM은 훈련과정을 위하여 대규모의 극한 분류(Margin classifier)의 개념을 사용하며, 입력 공간으로 부터의 분류의 용량을 분리하고 동시에 양호한 일반화를 제공한다. 이러한 방법론은 분류

를 위한 이상적인 조합이다. 훈련알고리즘은 반경기초함수에 확장된 아다트론 알고리즘(Adatron algorithm)을 기초로 하고 있으며 아다트론 알고리즘은 핵함수(Kernel function)에 의하여 내부공간에서 패턴의 내적을 교체함으로써 반경기초함수로서 쉽게 확장되어 질 수 있다(Vapnik, 1992, 1995; Principe et al., 2000; Tripathi et al., 2006).

SVM-NNM의 입력층과 은닉층사이의 연결함수는 입력벡터인자 x_i 에서 중심인 u_{ji} 를 제한 값의 평방함에 대한 제곱근으로 나타내고 있으며, 다음 식 (2)와 같이 나타낼 수 있다.

$$R_j = \sqrt{\sum_{i=1}^m (x_i - u_{ji})^2} \quad (2)$$

여기서 아랫첨자 i, j 는 각각 입력층과 은닉층을 나타내고 있으며, R_j 는 입력벡터들 사이의 반경거리를 나타내며 식(2)는 다음 식 (3)과 같이 나타낼 수 있다.

$$R_j = \|X - U^{(j)}\| \quad (3)$$

여기서 $\|\cdot\|$ 은 유클리드 거리 (Euclidean length) 이다. R_j 를 은닉층의 전이함수 $\phi_1(\cdot)$ 에 적용하면 다음 식 (4)와 같이 나타낼 수 있다.

$$K(x, x_j) = \phi_1(R_j) \quad (4)$$

여기서 $K(x, x_j)$ 는 내적 핵함수이고, 은닉층의 전이함수 $\phi_1(\cdot)$ 는 핵함수로서 정의된다. 본 연구에서는 가우시안 핵함수(GKF)가 사용되었으며, 입력벡터값이 0일 경우에 최대값은 1이고 입력벡터와 중심사이의 거리가 감소함에 따라서 출력벡터의 값은 증가한다. GKF는 다음 식 (5)와 같이 나타낼 수 있다.

$$\phi_1 = \exp(-B_1 R_j^2) = \exp\left[-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right] \quad (5)$$

여기서 B_1 는 $\frac{1}{2\sigma^2}$ 으로 나타내고 보통 일정한 값을 가지며, σ 는 반경기초함수의 폭을 나타내고 있다. 반경기초함수의 연산결과를 받는 출력층노드에 대한 함수는 다음 식 (6)과 같이 나타낼 수 있다.

$$G_k = \left[\sum_{j=1}^n (\alpha_j - \alpha_j^*) \cdot K(x, x_j) \right] + B \quad (6)$$

여기서 k 는 출력층, G_k 는 출력층노드의 연산값, α_j 와 α_j^* 는 양의 실수형 상수값을 가지는 라그랑지 승수(Lagrange multipliers)이고 B 는 출력층에서의 편차이다. $(\alpha_j - \alpha_j^*)$ 에 대한 0이 아닌 값에 해당되는 데이터를 지지벡터(Support vector)라고 한다. 본 연구에서 1월, 6월 및 12월에 해당되는 수문학적 분해모형은 다음 식 (7a)-(7c)와 같이 나타낼 수 있으며, 다음 그림 2는 본 연구에 적용된 SVM-NNM 구조를 나타낸 것이다.

$$PE_1(t) = W_1 \cdot \Phi_2(G_k) = W_1 \cdot \Phi_2 \left[\sum_{j=1}^{12} (\alpha_j - \alpha_j^*) \cdot K(x, x_j) \right] + B \quad (7a)$$

$$PE_6(t) = W_6 \cdot \Phi_2(G_k) = W_6 \cdot \Phi_2 \left[\sum_{j=1}^{12} (\alpha_j - \alpha_j^*) \cdot K(x, x_j) \right] + B \quad (7b)$$

$$PE_{12}(t) = W_{12} \cdot \Phi_2(G_k) = W_{12} \cdot \Phi_2 \left[\sum_{j=1}^{12} (\alpha_j - \alpha_j^*) \cdot K(x, x_j) \right] + B \quad (7c)$$

여기서 $\Phi_2(\cdot)$ 는 출력층의 시그모이드 전이함수, W_1, W_6, W_{12} 는 1월, 6월 및 12월에 해당하는 연결강도(Connection weights)이고, $PE_1(t), PE_6(t), PE_{12}(t)$ 는 계산된 1월, 6월 및 12월 증발접시 증발량이다. 본 연구에서 적용된 SVM-NNM은 단계크기(Step size)는 0.01, 최대 반복횟수는 10,000회, 오차 임계치는 0.001 및 테스트자료의 결과치가 최적치를 유도할 수 있도록 설정하였다.

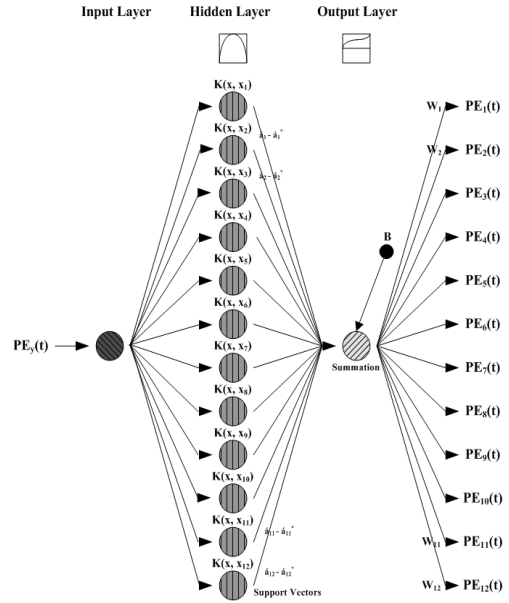


Fig. 2 The developed architecture of SVM-NNM

3. 추계학적 모형

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. 예를 들면 PARMA (1,1) 모형은 다음 식(8)과 같이 나타낼 수 있다.

$$y_{v,\tau} = \mu_\tau + \Phi_{1,\tau}(y_{v,\tau-1} - \mu_{\tau-1}) + \epsilon_{v,\tau} - \theta_{1,\tau} \epsilon_{v,\tau-1} \quad (8)$$

여기서 v 는 년(year), τ 는 계절(Season)이고 $\tau = 1, 2, \dots, \omega$ 이며 계절수를 나타내고 있다. 식 (8)과 같은 모형은 주로 월 유출량계열에 적용된다 (Salas et al., 1980). 본 연구에서는 선행연구(김성원과 김형수, 2008)에서 월 증발접시 증발량 및 PM 공식에 의해 산정된 월 알팔과 기준증발산량 훈련자료를 모의발생시켜 가장 양호한 결과를 유도

한 PARMA (1,1) 모형을 이용하였다. 그리고 모의 발생한 표본은 2개의 표본으로 하였으며, 500년에 해당되는 월 증발접시 증발량을 모의발생 시켰다. 모의된 자료 중에서 첫 번째 표본자료는 선택하지 않았으며, 두 번째 표본자료를 선택하였다. 또한 두 번째 표본자료에서 초기에 발생하는 편차(Bias)를 제거하기 위하여 초기의 50년 자료는 제거하였으며, 각 PARMA (1,1) 모형의 매개변수는 근사 최소자승법(Method of approximate least square)을 이용하여 산정하였다(Kim, 2004).

4. 연구대상의 선정 및 자료

본 연구에서는 기상청의 관리 하에 있는 76개의 기상관측소 중에서 기상관측 자료의 기록기간이 장기, 중기 및 단기간의 자료를 보유하고 있어야 하며, 각 권역을 대표할 수 있는 도시와 도서지역인 제주도에 소속된 기상관측소를 포함하는 것을 기준으로 하여 선정하였다. 따라서 연구대상 기상관측소는 서울, 강릉, 인천, 부산, 제주 및 목포관측소로 선정하였다. 본 연구에 필요한 기상자료는 국토해양부의 관리하에 있는 국가 수자원관리종합정보시스템(Water Management Information System, WAMIS)의 홈페이지(www.wamis.go.kr)의 무료 공개자료로 부터 필요한 자료를 이용하였다. 본 연구에서 장기간 자료를 보유하고 있는 기상관측소로는 서울과 강릉관측소를 선정하였으며, 중기간 자료를 보유하고 있는 기상관측소로는 인천과 부산관측소를 선정하였다. 마지막으로 단기간 자료를 보유하고 있는 기상관측소로는 제주와 목포관측소를 선정하였다.

5. 훈련과정의 수행

MLP-NNM과 SVM-NNM의 훈련과정에 따른 선행과정으로 입력층자료인 연 증발접시 증발량자료를 표준화하여 적용하였다. 일반적으로 입력층자료들을 표준화하는 주요한 이유 중 하나는 각 입력층자료들이 서로 다른 단위로서 관측된 값들을 나

타내고 있기 때문에 이러한 입력층자료들을 표준화하고 무차원단위에서 입력층자료들을 재배치함으로써, 각 입력층자료들의 상사성의 효과를 제거하기 위한 것으로 판단된다(Kim and Kim, 2008; Kim et al., 2009). MLP-NNM과 SVM-NNM을 적용한 수문학적 분해모형의 훈련과정 수행을 위해 구성된 훈련자료는 세 가지인 실측자료, 모의자료(추계학적 모형을 이용하여 모의발생) 및 혼합자료(실측자료와 추계학적 모형을 이용하여 모의발생시킨 모의자료를 혼합)로 각각 구성되었으므로 모형화에 따른 훈련결과치에 대한 통계분석을 실시한다는 것은 중요한 의미가 없으므로 훈련과정의 결과치에 대한 통계분석은 실시하지 않았다. 따라서 테스트과정의 결과치만 통계분석을 실시하였다.

5.1 최적모형선정을 위한 통계지표

본 연구에서 적용된 MLP-NNM 및 SVM-NNM의 훈련 및 테스트과정의 수행결과를 실측된 월 증발접시 증발량과 MLP-NNM 및 SVM-NNM에 의해 산정된 월 증발접시 증발량의 통계분석에 의하여 평가되도록 하였다. 본 연구에서 선정된 통계학적 지표는 상관계수(Correlation coefficient, CC), 평균제곱오차의 평방근(Root mean square error, RMSE), Nash-Sutcliffe 계수(Nash-Sutcliffe coefficient, E) 및 평균절대 상대오차(Average absolute relative error, AARE)이다. 다음 표 1은 MLP-NNM 및 SVM-NNM의 수행결과를 평가하기 위한 통계학적 지표를 나타낸 것이다. 여기서 $\bar{y}_i(x)$ = 산정된 월 증발접시 증발량(mm), $y_i(x)$ = 실측된 월 증발접시 증발량(mm), \bar{u}_y = 산정된 월 증발접시 증발량의 평균(mm), u_y = 실측된 월 증발접시 증발량의 평균(mm) 및 n = 고려된 월 증발접시 증발량의 총수이다. AARE의 통계학적 지표는 훈련된 모형으로부터 정확하게 월 증발접시 증발량을 모형화하는데 있어서 모형의 유효성(Effectiveness)을 평가하며, 그 외의 CC, RMSE 및 E의 통계학적 지표는 복잡하고 비선형의 자연 현상을 기록할 수 있는 모형의 효율(Efficiency)을

정량화한다(Kim and Kim, 2008). 수문학적 분해 모형에 있어서 유효성이 있고, 여러 입력 및 출력 변수 사이에서 복잡한 현상을 기록하는데 효율적인 모형이 최적의 모형으로 고려된다.

Table 1. Summary of statistical index

Index	Equation	Model Evaluation
CC	$\frac{\frac{1}{n} \sum_{i=1}^n [y_i(x) - u_y][\bar{y}_i(x) - \bar{u}_y]}{\sqrt{\frac{1}{n} \sum_{i=1}^n [y_i(x) - u_y]^2} \sqrt{\frac{1}{n} \sum_{i=1}^n [\bar{y}_i(x) - \bar{u}_y]^2}}$	Efficiency
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n [\bar{y}_i(x) - y_i(x)]^2}$	Efficiency
E	$1 - \frac{\sum_{i=1}^n [y_i(x) - \bar{y}_i(x)]^2}{\sum_{i=1}^n [y_i(x) - u_y]^2}$	Efficiency
AARE	$\frac{1}{n} \sum_{i=1}^n \left \frac{\bar{y}_i(x) - y_i(x)}{y_i(x)} \right \times 100\%$	Effectiveness

5.2 실측자료의 이용

본 연구의 실측자료의 이용을 위한 훈련자료는 다음과 같다. 장기간의 자료로는 서울관측소의 경우는 1908년부터 2002년까지의 결측기간을 제외한 86년의 자료를 선택하였으며, 강릉관측소의 경우는 1912년부터 2002년까지의 결측기간을 제외한 88년의 자료를 선택하였다. 중기간의 자료로는 인천관측소의 경우는 1949년부터 2002년까지의 결측기간을 제외한 52년의 자료를 선택하였다. 부산관측소의 경우는 1965년부터 2002년까지의 38년의 자료를 선택하였다. 그리고 단기간의 자료로는 제주 및 목포관측소의 경우는 1980년부터 2002년까지의 23년의 자료를 선택하였다.

5.3 모의자료의 이용

본 연구의 모의자료의 이용을 위한 훈련자료는 “5.2 실측자료의 이용”에서 실측된 월 증발접시 증발량자료를 모의자료의 구축을 위하여 이용하였

다. 따라서 MLP-NNM 및 SVM-NNM의 모의자료의 이용을 위한 훈련자료는 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 연 증발접시 증발량은 각 월 증발접시 증발량자료를 합하여 선정하였다. 그러므로 MLP-NNM 및 SVM-NNM의 모의자료의 이용을 위한 훈련자료는 450개의 모의된 연 증발접시 증발량자료로 구성되어 있다.

5.4 혼합자료의 이용

본 연구의 혼합자료의 이용을 위한 훈련자료는 “5.2 실측자료의 이용”에서 실측된 월 증발접시 증발량자료와 “5.3 모의자료의 이용”에서 산정된 월 증발접시 증발량자료를 합한 혼합자료로 구성하였다. 따라서 MLP-NNM 및 SVM-NNM의 혼합자료의 이용을 위한 훈련자료는 각각 서울관측소는 536개, 강릉관측소는 538개, 인천관측소는 502개, 부산관측소는 488개, 제주 및 목포관측소는 473개의 혼합된 연 증발접시 증발량자료로 구성되어 있다.

6. 테스트과정의 수행

테스트과정은 몇 개의 범주 이내에서 신경망모형이 모형화의 목적을 충분히 충족시킬 수 있을지에 대하여 검증하는 과정을 의미하고 있다. 따라서 신경망모형이 훈련과정을 통하여 선정된 각각의 매개변수를 이용하여 새로운 자료인 테스트자료로서 검증하는 과정이다. 본 연구에서 MLP-NNM 및 SVM-NNM의 경우 실측자료, 모의자료 및 혼합자료를 이용하여 훈련과정이 이루어 졌으므로 총 6 종류의 매개변수 군을 훈련과정을 통하여 산정할 수 있다. 또한 훈련과정의 경우 각각 실측자료, 모의자료 및 혼합자료를 이용하였으므로, 그 결과치를 비교한다는 것은 큰 의미가 없다. 그러나 테스트

트과정의 경우 6개 기상관측소에서 동일한 시계열을 이용함으로써 어떤 종류의 신경망모형과 훈련자료가 가장 양호한 것인지를 판단할 수 있는 기준을 제시할 수 있을 것으로 판단된다.

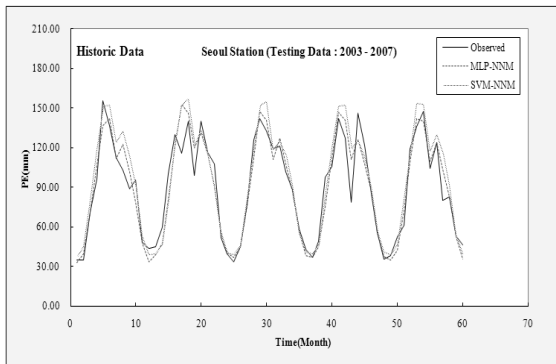
6.1 실측자료의 이용

본 연구에서는 실측자료를 이용한 훈련과정을 통하여 선정된 MLP-NNM과 SVM-NNM의 매개변수를 이용하여 테스트과정을 실시하였다. 테스트과정에서 테스트자료는 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 동일한 시계열인 2003년부터 2007년까지의 5년간의 자료를 선택하였으며, 5개의 자료로 구성되어 있다. 다음 표 2는 실측자료를 이용한 훈련과정의 수행결과 산정된 매개변수를 이용하여 MLP-NNM과 SVM-NNM의 테스트과정의 수행결과를 나타내고 있다. 표 2에 의하면 MLP-NNM의 테스트과정에 대한 통계분석 결과 CC가 0.789~0.952, RMSE가 11.800~20.015(mm), E가 0.574~0.899 및 AARE가 0.0004~0.0312(%)이고, SVM-NNM의 테스트과정에 대한 통계분석 결과 CC가 0.757~0.946, RMSE가 14.286~24.954(mm), E가 0.262~0.852 및 AARE가 0.0162~0.3065(%)으로 분석되었다. 따라서 선택된 6개의 기상관측소에서 실측자료를 이용한 경우 목포관측소를 제외하고는 MLP-NNM의 테스트과정의 수행결과가 SVM-NNM의 테스트과정의 수행결과보다 양호한 것으로 분석되었다. 다음 그림 3(a)-(f)는 서울, 강릉, 인천, 부산, 제주 및 목포관측소의 실측

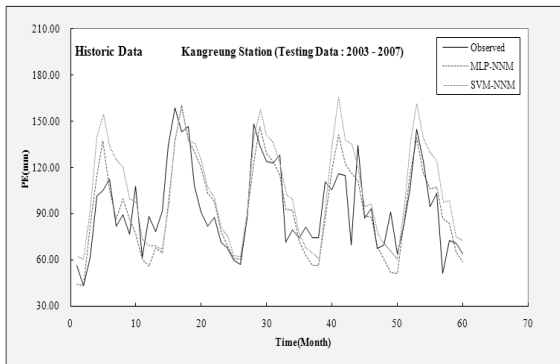
자료를 이용한 훈련과정의 수행결과를 이용하여 MLP-NNM과 SVM-NNM의 테스트과정의 수행결과에 대한 월 증발접시 증발량자료의 시계열의 비교를 나타낸 것이다.

Table 2. Statistical analysis of the monthly PE for the testing performance(Historic data)

Station	Statistical Index	Neural Networks Model	
		MLP-NNM	SVM-NNM
Seoul	CC	0.952	0.946
	RMSE	11.800	14.286
	E	0.899	0.852
	AARE	0.0004	0.0920
Kangreung	CC	0.801	0.757
	RMSE	18.037	23.734
	E	0.574	0.262
	AARE	0.0189	0.1650
Incheon	CC	0.825	0.868
	RMSE	18.452	24.954
	E	0.643	0.347
	AARE	0.0066	0.3065
Busan	CC	0.789	0.785
	RMSE	16.195	19.356
	E	0.615	0.450
	AARE	0.0115	0.1380
Jeju	CC	0.875	0.861
	RMSE	18.504	20.695
	E	0.763	0.703
	AARE	0.0306	0.0162
Mokpo	CC	0.837	0.867
	RMSE	20.015	18.436
	E	0.695	0.741
	AARE	0.0312	0.0247



(a) Seoul



(b) Kangreung

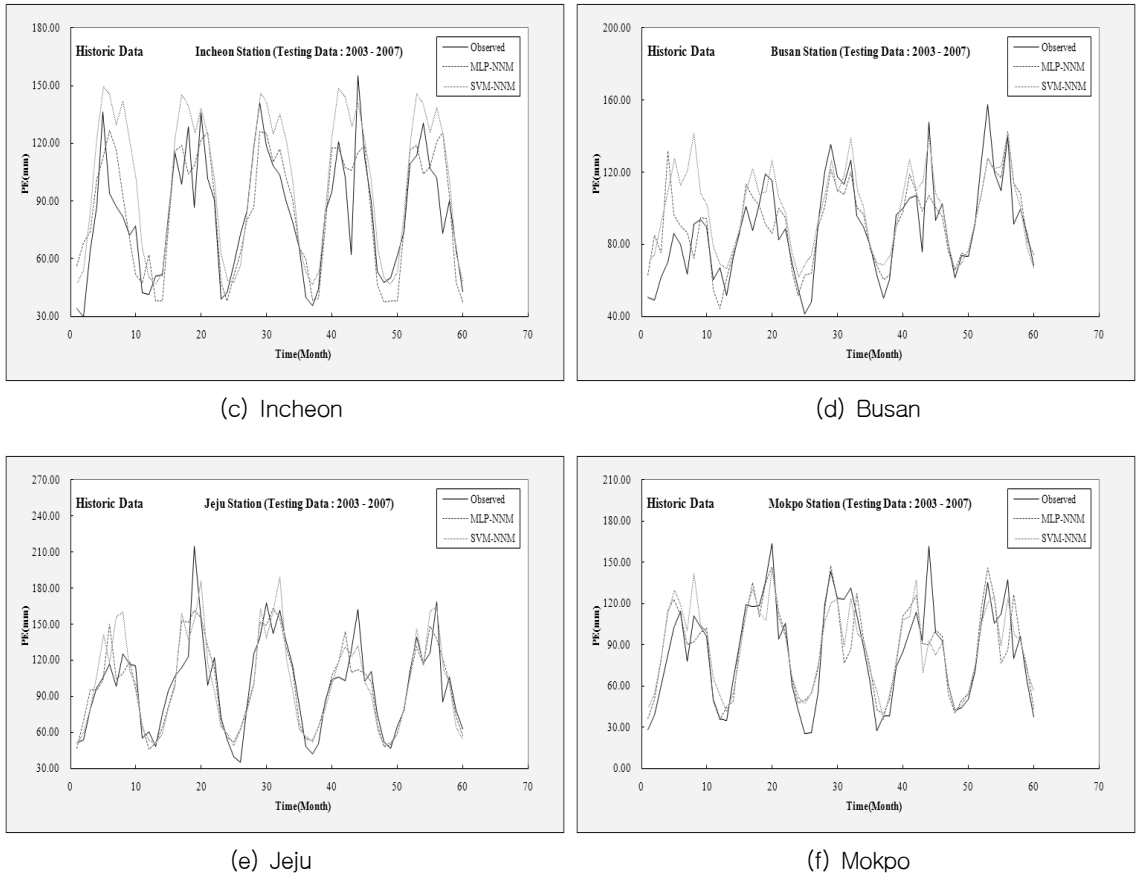
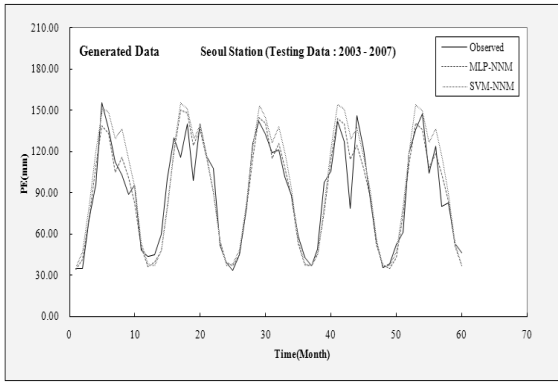


Fig. 3. Comparison of the monthly PE for the testing performance (Historic data)

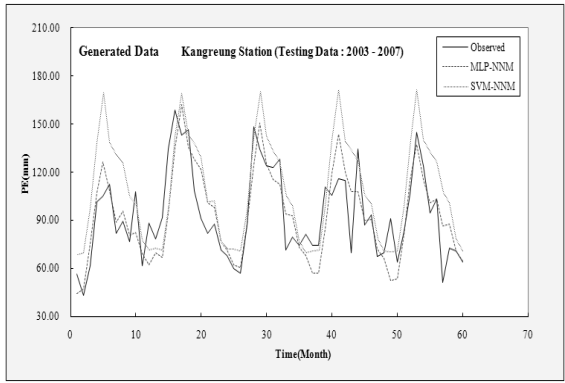
6.2 모의자료의 이용

본 연구에서는 모의자료를 이용한 훈련과정을 통하여 선정된 MLP-NNM과 SVM-NNM의 매개변수를 이용하여 테스트과정을 실시하였다. 테스트과정에서 테스트자료는 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 동일한 시계열인 2003년부터 2007년까지의 5년간의 자료를 선택하였으며, 5개의 자료로 구성되어 있다. 다음 표 3은 모의자료를 이용한 훈련과정의 수행결과 산정된 매개변수를 이용하여 MLP-NNM과 SVM-NNM의 테스트과정의 수행결과를 나타내고 있다. 표 3에 의하면 MLP-NNM의 테스트과정에 대한 통계분석 결과 CC가 0.811~0.952, RMSE가 11.733~16.999

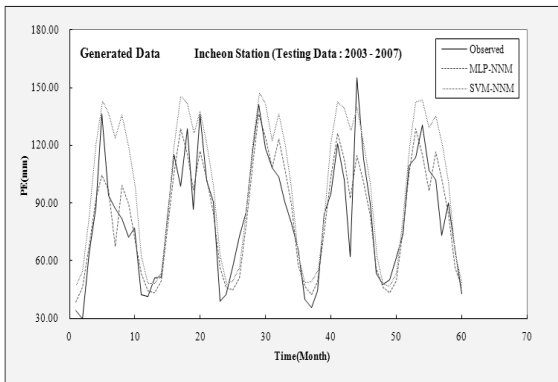
(mm), E가 0.622~0.900 및 AARE가 0.0065~0.0509(%)이고, SVM-NNM의 테스트과정에 대한 통계분석 결과 CC가 0.756~0.945, RMSE가 15.576~26.463(mm), E가 0.083~0.824 및 AARE가 0.0430~0.2866(%)으로 분석되었다. 따라서 선택된 6개의 기상관측소에서 모의자료를 이용한 경우 MLP-NNM의 테스트과정의 수행결과가 SVM-NNM의 테스트과정의 수행결과보다 양호한 것으로 분석되었다. 다음 그림 4(a)-(f)는 서울, 강릉, 인천, 부산, 제주 및 목포관측소의 모의자료를 이용한 훈련과정의 수행결과를 이용하여 MLP-NNM과 SVM-NNM의 테스트과정의 수행결과에 대한 월 증발접시 증발량자료의 시계열의 비교를 나타낸 것이다.



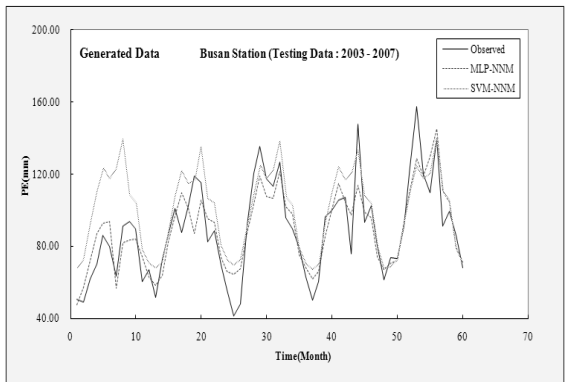
(a) Seoul



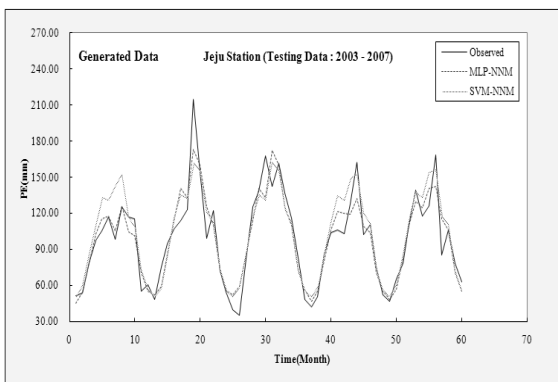
(b) Kangreung



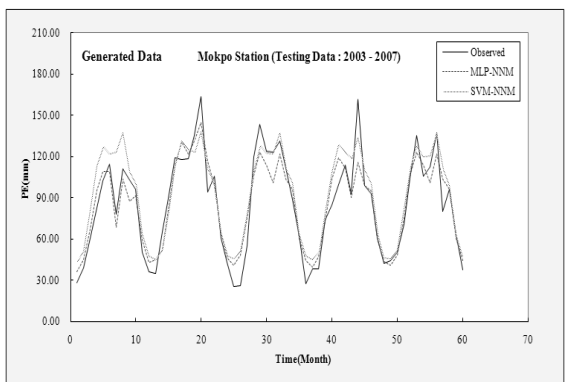
(c) Incheon



(d) Busan



(e) Jeju



(f) Mokpo

Fig. 4. Comparison of the monthly PE for the testing performance (Generated data)

Table 3. Statistical analysis of the monthly PE for the testing performance(Generated data)

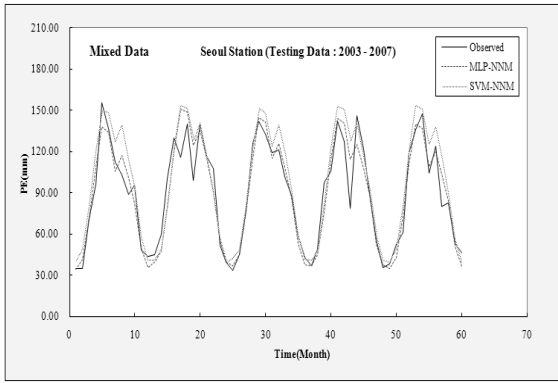
Station	Statistical Index	Neural Networks Model	
		MLP-NNM	SVM-NNM
Seoul	CC	0.952	0.945
	RMSE	11.733	15.590
	E	0.900	0.824
	AARE	0.0065	0.1265
Kangreung	CC	0.811	0.756
	RMSE	16.999	26.463
	E	0.622	0.083
	AARE	0.0253	0.2418
Incheon	CC	0.896	0.879
	RMSE	13.762	23.462
	E	0.801	0.423
	AARE	0.0272	0.2866
Busan	CC	0.885	0.783
	RMSE	12.314	19.791
	E	0.777	0.425
	AARE	0.0291	0.1490
Jeju	CC	0.928	0.907
	RMSE	14.134	16.807
	E	0.862	0.804
	AARE	0.0318	0.0430
Mokpo	CC	0.945	0.929
	RMSE	12.550	15.576
	E	0.880	0.816
	AARE	0.0509	0.0866

6.3 혼합자료의 이용

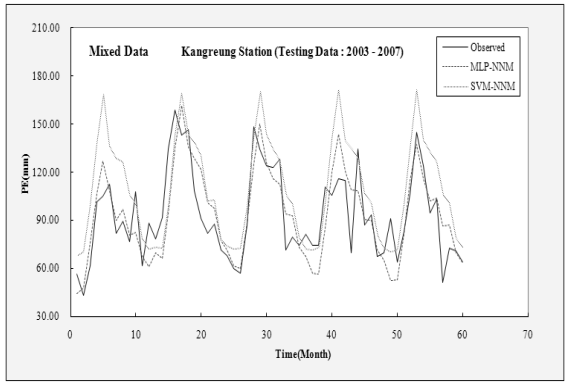
본 연구에서는 혼합자료를 이용한 훈련과정을 통하여 선정된 MLP-NNM과 SVM-NNM의 매개변수를 이용하여 테스트과정을 실시하였다. 테스트과정에서 테스트자료는 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 동일한 시계열인 2003년

부터 2007년까지의 5년간의 자료를 선택하였으며, 5개의 자료로 구성되어 있다. 다음 표 4는 혼합자료를 이용한 훈련과정의 수행결과 산정된 매개변수를 이용하여 MLP-NNM과 SVM-NNM의 테스트과정의 수행결과를 나타내고 있다. 표 4에 의하면 MLP-NNM의 테스트과정에 대한 통계분석 결과 CC가 0.811~0.952, RMSE가 11.808~17.086(mm), E가 0.618~0.899 및 AARE가 0.0003~0.0320(%)이고, SVM-NNM의 테스트과정에 대한 통계분석 결과 CC가 0.761~0.946, RMSE가 15.500~26.191(mm), E가 0.102~0.826 및 AARE가 0.0444~0.2897(%)으로 분석되었다. 따라서 선택된 6개의 기상관측소에서 혼합자료를 이용한 경우 MLP-NNM의 테스트과정의 수행결과가 SVM-NNM의 테스트과정의 수행결과보다 양호한 것으로 분석되었다. 다음 그림 5(a)-(f)는 서울, 강릉, 인천, 부산, 제주 및 목포관측소의 혼합자료를 이용한 훈련과정의 수행결과를 이용하여 MLP-NNM과 SVM-NNM의 테스트과정의 수행결과에 대한 월 증발점시 증발량자료의 시계열의 비교를 나타낸 것이다.

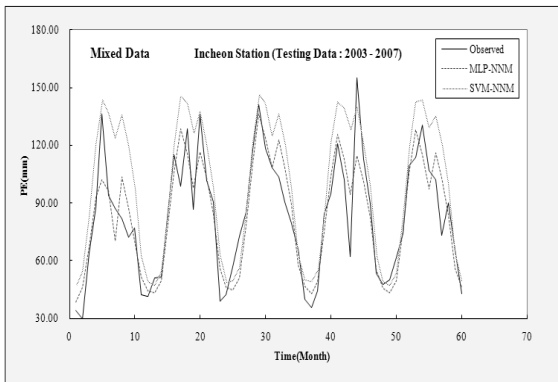
본 연구에서 실측자료, 모의자료 및 혼합자료를 이용한 MLP-NNM과 SVM-NNM의 테스트과정의 수행결과를 분석하면 다음과 같다. MLP-NNM의 경우 서울관측소를 제외하고는 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과가 실측자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과보다 양호한 것으로 분석되었다. 또한 SVM-NNM의 경우 제주 및 목포관측소에서 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과가 실측자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과보다 양호한 것으로 분석되었으며, 나머지 4개 관측소는 유사한 결과를 나타내는 것으로 분석되었다.



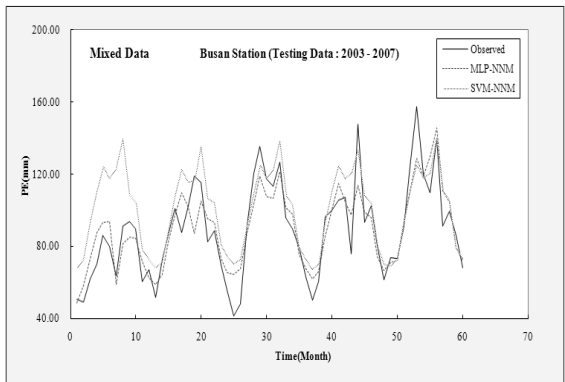
(a) Seoul



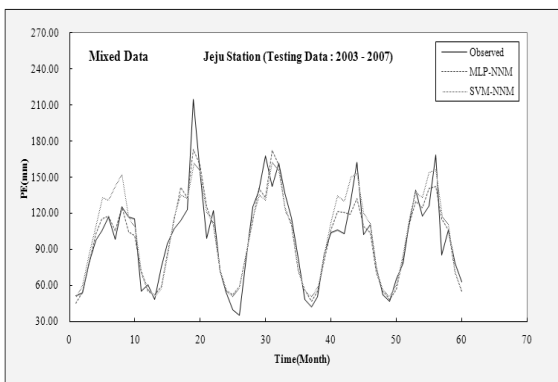
(b) Kangreung



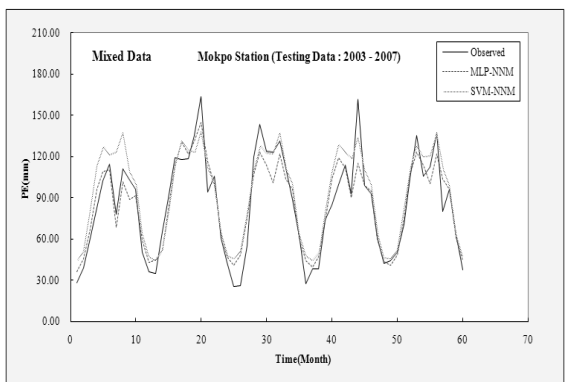
(c) Incheon



(d) Busan



(e) Jeju



(f) Mokpo

Fig. 5. Comparison of the monthly PE for the testing performance (Mixed data)

Table 4. Statistical analysis of the monthly PE for the testing performance(Mixed data)

Station	Statistical Index	Neural Networks Model	
		MLP-NNM	SVM-NNM
Seoul	CC	0.952	0.946
	RMSE	11.808	15.500
	E	0.899	0.826
	AARE	0.0055	0.1350
Kangreung	CC	0.811	0.761
	RMSE	17.086	26.191
	E	0.618	0.102
	AARE	0.0242	0.2439
Incheon	CC	0.894	0.878
	RMSE	13.926	23.589
	E	0.797	0.416
	AARE	0.0268	0.2897
Busan	CC	0.887	0.783
	RMSE	12.263	19.954
	E	0.779	0.415
	AARE	0.0280	0.1540
Jeju	CC	0.928	0.907
	RMSE	14.151	16.870
	E	0.861	0.803
	AARE	0.0320	0.0444
Mokpo	CC	0.940	0.929
	RMSE	11.994	15.606
	E	0.883	0.815
	AARE	0.0003	0.0875

7. 자료의 동질성 검증

본 연구에서는 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서의 테스트자료인 실측된 월 증발접시 증발량자료와 실측자료, 모의자료 및 혼합자료를 이용한 훈련과정을 통하여 산정된 MLP-NNM과 SVM-NNM의 각 매개변수를 이용하여 테스트과정 수행결과 산정된 월 증발접시 증발량자료를 비교하여 동일 모집단에서 추출된 자료인지를 검토하는 동질성검증(Homogeneity test)을 실시하였다.

동질성검증은 서울, 강릉, 인천, 부산, 제주 및 목포관측소의 월 증발접시 증발량 자료에 대한 평균과 분산에 대한 일원분산분석(One-way analysis of variance, ANOVA) 및 Mann-Whitney U 검증(Test)을 실시하였다(McCuen 1993; Salas et al., 2001).

7.1 일원분산분석

실측된 월 증발접시 증발량자료와 MLP-NNM과 SVM-NNM을 이용해서 산정된 월 증발접시 증발량자료에 대하여 먼저 평균에 대한 일원분산분석을 실시하였다. 분석은 2표본 t-검증(Two-sample t-test)을 이용하여 검증통계량(Test statistics) 및 자유도(Degrees of freedom)를 산정하였으며, 유의수준(Level of significance) 5% 및 1%에서 임계 검증통계량(Critical test statistics)을 산정하여 두 표본에 대한 모집단의 평균은 동일하다는 귀무가설(Null hypothesis)의 채택 (Accept) 혹은 기각 (Reject)하는 과정을 실시하였다. 다음 표 5는 월 증발접시 증발량자료의 평균에 대한 동질성 검증결과를 나타내고 있다. 표 5에 의하면 실측된 월 증발접시 증발량자료와 MLP-NNM을 이용해서 실측자료, 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용하여 테스트과정 수행결과 산정된 월 증발접시 증발량자료의 평균에 대한 일원분산분석의 결과 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 5% 및 1%의 유의수준에서 모집단의 평균이 동일하다는 귀무가설이 채택되었다. 또한 실측된 월 증발접시 증발량자료와 SVM-NNM을 이용해서 실측자료, 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용하여 테스트과정 수행결과 산정된 월 증발접시 증발량자료의 평균에 대한 일원분산분석의 결과 서울, 제주 및 목포에서는 5% 및 1%의 유의수준에서 모집단의 평균이 동일하다는 귀무가설이 채택되었다. 그러나 강릉의 경우 실측자료는 5%에서는 기각, 1%의 유의수준에서 모집단의 평균이 동일하다는 귀무가설이 채택되었고, 모의자료 및 혼합자료는 5% 및 1%에서 기각되었다. 인천의 경우 실측자료는

Table 5 Results of the ANOVA test on the mean

Model	Stations	Level of Significance	Mean						
			Two - sample t test						
			Critical t statistic	Computed t statistic			Null hypothesis		
				Historic	Generated	Mixed	Historic	Generated	Mixed
MLP-NNM	Seoul	0.05/0.01	1.981/2.621	-0.0109	0.0117	0.0101	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Kangreung	0.05/0.01	1.981/2.621	-0.0118	0.0038	0.0004	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Incheon	0.05/0.01	1.981/2.621	-0.4080	-0.0606	-0.0661	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Busan	0.05/0.01	1.981/2.621	-0.4604	-0.1146	-0.1327	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Jeju	0.05/0.01	1.981/2.621	-0.0970	-0.0074	-0.0064	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Mokpo	0.05/0.01	1.981/2.621	-0.3401	-0.0293	-0.0315	Accept/ Accept	Accept/ Accept	Accept/ Accept
SVM-NNM	Seoul	0.05/0.01	1.981/2.621	-0.7221	-0.9356	-1.0904	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Kangreung	0.05/0.01	1.981/2.621	-2.0478	-2.8565	-2.9151	Reject/ Accept	Reject/ Reject	Reject/ Reject
	Incheon	0.05/0.01	1.981/2.621	-2.7162	-2.5855	-2.6183	Reject/ Reject	Reject/ Accept	Reject/ Accept
	Busan	0.05/0.01	1.981/2.621	-2.3322	-2.4836	-2.5702	Reject/ Accept	Reject/ Accept	Reject/ Accept
	Jeju	0.05/0.01	1.981/2.621	-0.3570	-0.7163	-0.7306	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Mokpo	0.05/0.01	1.981/2.621	-0.5122	-1.2248	-1.2335	Accept/ Accept	Accept/ Accept	Accept/ Accept

5% 및 1%의 유의수준에서 모집단의 평균이 동일하다는 귀무가설이 기각되었고, 모의자료 및 혼합자료는 5%에서는 기각, 1%에서는 채택되었다. 마지막으로 부산의 경우 실측자료, 모의자료 및 혼합자료에서 5%에서는 기각, 1%의 유의수준에서 모집단의 평균이 동일하다는 귀무가설이 채택되었다.

두 번째로 분산에 대한 일원분산분석의 경우는 5% 및 1%의 유의수준에서 F-검증통계치(F-test statistics)을 계산하였으며, 임계 F-검증통계치와의 비교를 통하여 모집단의 분산이 동일하다는 귀무가설을 채택 혹은 기각하는 과정을 실시하였다. 다음 표 6은 월 증발점시 증발량자료의 분산에 대한 동질성 검증결과치를 나타내고 있다. 표 6에 의하면 실측된 월 증발점시 증발량자료와 MLP-

NNM을 이용해서 실측자료, 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용하여 테스트과정 수행결과 산정된 월 증발점시 증발량자료의 분산에 대한 일원분산분석의 결과 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 5% 및 1%의 유의수준에서 모집단의 분산이 동일하다는 귀무가설이 채택되었다. 또한 실측된 월 증발점시 증발량자료와 SVM-NNM을 이용해서 실측자료, 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용하여 테스트과정 수행결과 산정된 월 증발점시 증발량자료의 분산에 대한 일원분산분석의 결과 서울, 강릉, 인천, 부산, 제주 및 목포에서 5% 및 1%의 유의수준에서 모집단의 분산이 동일하다는 귀무가설이 채택되었다.

Table 6 Results of the ANOVA test on the variance

Model	Stations	Level of Significance	Variance						
			F-test statistics						
			Critical F statistic	Computed F statistic			Null hypothesis		
				Historic	Generated	Mixed	Historic	Generated	Mixed
MLP-NNM	Seoul	0.05/0.01	1.981/2.621	1.088	1.057	1.065	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Kangreung	0.05/0.01	1.981/2.621	1.130	1.007	1.022	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Incheon	0.05/0.01	1.981/2.621	1.012	1.146	1.141	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Busan	0.05/0.01	1.981/2.621	1.478	1.536	1.548	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Jeju	0.05/0.01	1.981/2.621	1.161	1.197	1.199	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Mokpo	0.05/0.01	1.981/2.621	1.286	1.435	1.438	Accept/ Accept	Accept/ Accept	Accept/ Accept
SVM-NNM	Seoul	0.05/0.01	1.981/2.621	1.208	1.289	1.215	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Kangreung	0.05/0.01	1.981/2.621	1.300	1.346	1.305	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Incheon	0.05/0.01	1.981/2.621	1.417	1.358	1.141	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Busan	0.05/0.01	1.981/2.621	1.343	1.351	1.378	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Jeju	0.05/0.01	1.981/2.621	1.093	1.102	1.099	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Mokpo	0.05/0.01	1.981/2.621	1.529	1.237	1.438	Accept/ Accept	Accept/ Accept	Accept/ Accept

7.2 Mann-Whitney U 검증

Mann-Whitney U 검증은 두 개의 독립적인 표본의 2표본 t-검증(Two-sample t-test)에 대한 비모수적인 대응법의 하나이며, 동일한 모집단으로부터 두 개의 독립적인 표본을 선정할 수 있는지를 검증하기 위하여 사용되어 진다. 또한 세 개 이상의 독립적인 표본에 대해서는 Kruskal-Wallis 검증을 사용한다. 다음 표 7은 월 증발접시 증발량자료의 Mann-Whitney U 검증을 이용한 동질성검증에 대한 결과치를 나타내고 있다. 표 7에 의하면 실측된 월 증발접시 증발량자료와 MLP-NNM을 이용해서 실측자료, 모의자료 및 혼합자료의 훈련 과정을 통하여 선정된 매개변수를 이용하여 테스트

과정 수행결과 산정된 월 증발접시 증발량자료의 Mann-Whitney U 검증 결과 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 5% 및 1%의 유의수준에서 두 개의 표본자료가 동일 모집단에서 추출되었다는 귀무가설이 채택되었다. 또한 실측된 월 증발접시 증발량자료와 SVM-NNM을 이용해서 실측자료, 모의자료 및 혼합자료의 훈련 과정을 통하여 선정된 매개변수를 이용하여 테스트과정 수행결과 산정된 월 증발접시 증발량자료의 Mann-Whitney U 검증 결과 서울, 제주 및 목포에서는 5% 및 1%의 유의수준에서 두 개의 표본자료가 동일 모집단에서 추출되었다는 귀무가설이 채택되었다. 그러나 강릉의 경우 실측자료는 5% 및 1%의 유의수준에서 두 개의 표본자료가 동일 모집

Table 7 Results of the Mann-Whitney U test

Model	Stations	Level of Significance	Mann-Whitney U test						
			Critical Z	Computed Z			Null hypothesis		
				Historic	Generated	Mixed	Historic	Generated	Mixed
MLP-NNM	Seoul	0.05/0.01	1.960/2.575	-0.010	-0.063	-0.058	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Kangreung	0.05/0.01	1.960/2.575	-0.016	-0.100	-0.068	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Incheon	0.05/0.01	1.960/2.575	-0.588	-0.310	-0.299	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Busan	0.05/0.01	1.960/2.575	-0.777	-0.346	-0.346	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Jeju	0.05/0.01	1.960/2.575	-0.189	-0.331	-0.315	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Mokpo	0.05/0.01	1.960/2.575	-0.357	-0.210	-0.205	Accept/ Accept	Accept/ Accept	Accept/ Accept
SVM-NNM	Seoul	0.05/0.01	1.960/2.575	-0.751	-1.018	-1.233	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Kangreung	0.05/0.01	1.960/2.575	-1.827	-2.446	-2.567	Accept/ Accept	Reject/ Accept	Reject/ Accept
	Incheon	0.05/0.01	1.960/2.575	-2.698	-2.635	-2.640	Reject/ Reject	Reject/ Reject	Reject/ Reject
	Busan	0.05/0.01	1.960/2.575	-2.420	-2.551	-2.635	Reject/ Accept	Reject/ Accept	Reject/ Reject
	Jeju	0.05/0.01	1.960/2.575	-0.315	-1.123	-1.134	Accept/ Accept	Accept/ Accept	Accept/ Accept
	Mokpo	0.05/0.01	1.960/2.575	-0.577	-1.601	-1.611	Accept/ Accept	Accept/ Accept	Accept/ Accept

단에서 추출되었다는 귀무가설이 채택되었고, 모의 자료와 혼합자료는 5%에서는 기각, 1%의 유의수준에서는 채택되었다. 인천의 경우 실측자료, 모의 자료 및 혼합자료는 5% 및 1%의 유의수준에서 두 개의 표본자료가 동일 모집단에서 추출되었다는 귀무가설이 기각되었다. 마지막으로 부산의 경우 실측자료 및 모의자료는 5%에서는 기각, 1%에서는 채택되었고, 혼합자료는 5% 및 1%의 유의수준에서 두 개의 표본자료가 동일 모집단에서 추출되었다는 귀무가설이 기각되었다. 따라서 본 연구에서 MLP-NNM을 이용하여 산정된 월 증발접시 증발량자료는 실측된 월 증발접시 증발량자료와 동질성 검증을 통하여 동질의 모집단을 가지는 자료군으로

채택되어 질 수 있는 것으로 판단되나, SVM-NNM의 경우는 관측소별로 차이가 있는 것으로 판단된다. 따라서 보편적으로 MLP-NNM의 수행결과가 SVM-NNM의 수행결과보다 양호하다고 판단된다.

8. 결론 및 추후 연구과제

본 연구에서는 우리나라 주요 기상관측소에서 관측된 연 증발접시 증발량자료를 제시된 MLP-NNM 및 SVM-NNM을 이용하여 월 증발접시 증발량자료로 분리하는데 있어서 각 신경망모형의 잠재성을 검토하였다.

1. MLP-NNM 및 SVM-NNM의 훈련과정의 수행결과 산정된 각 6종류의 매개변수 군을 이용하여 테스트과정을 수행하였다. 그 결과 선택된 6개의 기상관측소에서 실측자료를 이용한 경우 목포관측소를 제외하고는 MLP-NNM의 테스트과정의 수행결과가 SVM-NNM의 테스트과정의 수행결과보다 양호한 것으로 분석되었다. 또한 선택된 6개의 기상관측소에서 모의자료 및 혼합자료를 이용한 경우 MLP-NNM의 테스트과정의 수행결과가 SVM-NNM의 테스트과정의 수행결과보다 양호한 것으로 분석되었다.
2. MLP-NNM의 경우 서울관측소를 제외하고는 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과가 실측자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과보다 양호한 것으로 분석되었다.
3. SVM-NNM의 경우 제주 및 목포관측소에서 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과가 실측자료의 훈련과정을 통하여 선정된 매개변수를 이용한 테스트과정의 수행결과보다 양호한 것으로 분석되었으며, 나머지 4개 관측소는 유사한 결과를 나타내는 것으로 분석되었다.
4. 서울, 강릉, 인천, 부산, 제주 및 목포관측소에서 실측된 월 증발집시 증발량자료와 MLP-NNM과 SVM-NNM을 이용해서 실측자료, 모의자료 및 혼합자료의 훈련과정을 통하여 선정된 매개변수를 이용하여 테스트과정 수행결과 산정된 월 증발집시 증발량자료에 대하여 동질성검증을 위하여 평균과 분산에 대한 일원분산분석 및 Mann-Whitney U 검증을 실시하였다. 그 결과 MLP-NNM을 이용하여 산정된 월 증발집시 증발량자료는 실측된 월 증발집시 증발량자료와 동질성검증을 통하여 동질의 모집단을 가지는 자료군으로 채택되어 질 수 있는 것으로 판단되나,

SVM-NNM의 경우는 관측소별로 차이가 있는 것으로 판단된다. 따라서 보편적으로 MLP-NNM의 수행결과가 SVM-NNM의 수행결과보다 양호하다고 판단된다.

참 고 문 헌

- 국토해양부(2007) 수자원 관리 종합정보 시스템 홈페이지 <http://www.wamis.go.kr>
- 김성원, 김정현, 박기범, 김형수(2010) “비선형 증발집시 증발량 산정을 위한 시간적 분해모형” **대한토목학회 논문집**, 대한토목학회, 제 30권, 제4B호, pp. 399-412.
- 김성원, 김형수(2008) “증발집시 증발량과 알팔과 기준증발산량의 모형화를 위한 통합운영방법” **대한토목학회 논문집**, 대한토목학회, 제 28권, 제 2B호, pp. 199-213.
- Bruton, J.M., McClendon, R.W., and Hoogenboom, G.(2000) Estimating daily pan evaporation with artificial neural networks. *Transaction of the ASAE*, ASAE, Vol. 43, No. 2, pp. 491-496.
- Burian, S.J., Durrans, S.R., Nix, S.J., and Pitt, R.E.(2001) Training artificial neural networks to perform rainfall disaggregation. *Journal of Hydrologic Engineering*, ASCE, Vol. 6, No. 1, pp. 43-51.
- Burian, S.J., Durrans, S.R., Tomic, S., Pimmel, R.L., and Wai, C.N.(2000) Rainfall disaggregation using artificial neural networks. *Journal of Hydrologic Engineering*, ASCE, Vol. 5, No. 3, pp. 299-307.
- Choi, J., Socolofsky, S.A., and Olivera, F.(2008) Hourly disaggregation of daily rainfall in Texas using measured hourly precipitation at other locations. *Journal of Hydrologic Engineering*, ASCE, Vol. 13, No. 6, pp. 476-487.
- Deswal, S., and Pal, M.(2008) Artificial neural

- network based modeling of evaporation losses in reservoirs. *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 29, pp. 279-283.
- Dibike, Y.B., Velickov, S., Solomatine, D., and Abbott, M.B.(2001) Model induction with support vector machines: introductions and applications. *Journal of Computing in Civil Engineering*, ASCE, Vol. 15, pp. 208-216.
- Eslamian, S.S., Gohari, S.A., Biabanaki, M., and Malekian, R.(2008) Estimation of monthly pan evaporation using artificial neural networks and support vector machines. *Journal of Applied Sciences*, Vol. 8, No. 19, pp. 3497-3502.
- Gundekar, H.G., Khodke, U.M., and Sarkar, S.(2008) Evaluation of pan coefficient for reference crop evapotranspiration for semi-arid region. *Irrigation Science*, Vol. 26, pp. 169-175.
- Gutierrez-Magness, A.L., and McCuen, R.H.(2004) Accuracy evaluation of rainfall disaggregation methods. *Journal of Hydrologic Engineering*, ASCE, Vol. 9, No. 2, pp. 71-78.
- Haykin, S. (2009). *Neural networks and learning machines*, 3rd Edition, Pearson Education Inc., NJ, USA.
- 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, pp. 332.
- Khadam, I.M., and Kaluarachchi, J.J.(2004) Use of soft information to describe the relative uncertainty of calibration data in hydrologic models. *Water Resources Research*, Vol. 40, No. 11, W11505.
- Keskin, M.E., and Terzi, O.(2006) Artificial neural networks models of daily pan evaporation. *Journal of Hydrologic Engineering*, ASCE, Vol. 11, No. 1, pp. 65-70.
- Kim, S.(2004) Neural Networks Model and Embedded Stochastic Processes for Hydrological Analysis in South Korea. *KSCE Journal of Civil Engineers*, KSCE, Vol.8, No.1, pp. 141-148.
- Kim, S., and Kim, H.S.(2008) Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. *Journal of Hydrology*, Vol. 351, pp. 299-317.
- Kim, S., Kim, J.H., and Park, K.B.(2009) Statistical learning theory for the disaggregation of the climatic data. *Proceedings of 33rd IAHR Congress 2009*, IAHR/AIRH, Vancouver, British Columbia, Canada, PP. 1154-1162.
- Kisi, O.(2006) Daily pan evaporation modeling using a neuro-fuzzy computing technique. *Journal of Hydrology*, Vol. 329, pp. 636-646.
- McCuen, R.H.(1993) *Microcomputer applications in statistical hydrology*, Prentice Hall, NJ, USA.
- Molina Martinez, J.M., Martinez Alvarez, V., Gonzalez-Real, M.M., and Baille, A.(2005) A simulation model for predicting hourly pan evaporation for meteorological data. *Journal of Hydrology*, Vol. 318, pp. 250-261.
- Principe, J.C., Euliano, N.R., and Lefebvre, W.C.(2000) *Neural and adaptive systems: fundamentals through simulation*, John Wiley & Sons, New York, USA.
- Rahimi Khoob, A.(2009) Estimating daily pan

- evaporation using artificial neural network in a semi-arid environment. *Theoretical and Applied Climatology*, Doi : 10.1007/s00704-008-0096-3.
- Salas, J.D., Delleur, J.R., Yevjevich, V., and Lane, W.L.(1980) *Applied modeling of hydrologic timese ries*, Water Resources Publication, Littleton, CO, USA.
- Salas, J.D., Smith, R.A., Tabios III, G.Q., and Heo, J.H.(2001) *Statistical computing techniques in water resources and environmental engineering*, Unpublished book in CE622, Colorado State University, Fort Collins, CO, USA.
- Sudheer, K.P., Gosain, A.K., Rangan, D.M., and Saheb, S.M.(2002) Modeling evaporation using an artificial neural network algorithm. *Hydrological Processes*, Vol. 16, pp. 3189-3202.
- Tan, K.S., Chiew, F.H.S., and Grayson, R.B.(2007) A steepness index unit volume flood hydrograph approach for sub-daily flow disaggregation. *Hydrological Processes*, Vol. 21, pp. 2807-2816.
- Tripathi, S., Srinivas, V.V., and Nanjundish, R.S.(2006) Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology*, Vol. 330, pp. 621-640.
- Vapnik, V.N.(1992) Principles of risk minimization for learning theory. *Advances in Neural Information Processing Systems* Vol. 4, pp. 831-838.
- Vapnik, V.N.(1995) *The nature of statistical learning theory*, Springer Verlag, New York, NY, USA.
- Wasserman, P.D.(1993) *Advanced methods in neural computing*, Van Nostrand Reinhold, New York, NY, USA.
- Zhang, J., Murch, R.R., Ross, M.A., Ganguly, A.R., and Nachabe, M.(2008) Evaluation of statistical rainfall disaggregation methods using rain-gauge information for west-central florida. *Journal of Hydrologic Engineering*, ASCE, Vol. 13, No. 12, pp. 1158-1169.

○논문접수일 : 10년 07월 21일

○심사의뢰일 : 10년 07월 29일

○심사완료일 : 10년 12월 24일