

SVM-NNM을 이용한 증발접시 증발량자료의 분해기법

Disaggregation Approach of the Pan Evaporation using SVM-NNM

김성원*

Abstract

The goal of this research is to apply the neural networks model for the disaggregation of the pan evaporation (PE) data, Republic of Korea. The neural networks model consists of support vector machine neural networks model (SVM-NNM). The disaggregation means that the yearly PE data divides into the monthly PE data. And, for the performances of the neural networks model, it is 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 SVM-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, Statistical Learning Theory, Stochastic Model, SVM-NNM

1. 서 론

증발과정은 자연적 및 농업적인 환경시스템의 에너지와 물 수지의 주요인자를 나타낸다. 증발과정에 의한 수분손실량의 추정에는 수자원의 모니터링, 조사 및 운영, 관개배수시스템의 설계와 관계계획에 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 진다. 증발량 관측을 위한 직접적인 방법중 하나는 증발접시 증발량이다 (Eslamian et al., 2008). 계획되고 있는 혹은 기존의 저수지를 비롯하여 관개 프로젝트가 있는 모든 지점에 대하여 증발접시를 설치한다는 것은 비현실적이다. 또한 정확한 관측도구가 설치될 수 없거나 혹은 유지·관리가 어려운 접근할 수 없는 지역을 가지는 것 또한 비합리적이다 (Kisi, 2006).

본 연구의 목적은 연 증발접시 증발량의 시간적인 분해를 위하여 SVM-NNM을 개발 및 적용하는데 있다. 그리고 SVM-NNM의 수행평가를 위하여 훈련 및 테스트과정으로 구성되었다. SVM-NNM의 훈련과정을 위하여 실측, 모의 및 혼합자료와 같은 세 가지 형태의 자료가 사용되었으며, 테스트과정을 위해서는 실측자료만 이용되었다. 본 연구에서는 비선형 수문학적 시계열자료의 시간적 분해를 위해서 SVM-NNM의 적용성을 평가하였다. 게다가 연 증발접시 증발량 자료의 시간적 분해로부터 신뢰성있는 월 증발접시 증발량자료를 구축할 수 있을 것이며, 관개배수 네트워크 시스템의 평가를 위한 이용가능한 자료를 제공할 수 있을 것이다.

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

지지벡터기구 신경망모형 (SVM-NNM)은 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 사용한다. 이것은 입력공간으로 부터의 분류의 용량을 분리하고 동시에 양호한 일반화를 제공한다. 이것은 분류를 위한 이상적인 조합이다. 훈련알고리즘은 반경기초함수에 확장된 아다트론 알고리즘 (Adatron algorithm)을 기초로 하고 있다. 아다트론 알고리즘은 핵함수 (Kernel function)에 의하여 내부공간에서 패턴의 내적을 교체함으로써 반경기초함수로서 쉽게 확장되어 질 수 있다 (Vapnik, 1992, 1995; Principe et al., 2000). 본 연구에 적용된 SVM-NNM은 ϵ -SVM 회귀모형이고, kernel 함수는 RBF 함수이다. 또한 단계간격 (Step size)은 0.01, 최대 반복횟수는 10,000회 및 임계치는 0.001로 설정하였다. 다음 Figure 1은 본 연구에 적용된 SVM-NNM 구조를 나타낸 것이다.

∴ 정희원, 공학박사, 수자원개발기술사, 동양대학교 철도토목학과 부교수 ; E-mail : swkim1968@dyu.ac.kr

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. 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. 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., 1980). 본 연구에서는 신경망모형의 훈련자료를 모의발생 하기위하여 PARMA (1,1) 모형을 이용하였다. 그리고 모의발생한 표본은 2개의 표본으로 하였으며, 500년에 해당되는 월별 증발량을 모의발생시켰다. 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. 따라서 제시된 SVM-NNM의 입력층 변수의 모의자료의 경우는 모의발생 시킨 월별 소형 증발량 증발량자료를 합한 결과치에 해당되는 연간 소형 증발접시 증발량자료이고, 출력층 변수는 모의발생시킨 월별 소형 증발접시 증발량자료이다.

5. SVM-NNM의 적용

5.1 훈련과정 (Training Performance)

5.1.1 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 SVM-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 값들을 나타내고 있기 때문에 이러한 입력변수들을 표준화하고 무차원단위에서 입력변수들을 재배치함으로써, 각 변수들의 상사성의 효과를 제거하기 위한 것으로 판단된다 (Kim and Kim, 2008a, b).

5.1.2 모의자료의 이용

연구의 모의자료의 이용을 위한 훈련과정에서 훈련자료는 “5.1.1 실측자료의 이용”에서 선정한 월 증발접시 증발량자료를 모의자료의 구축을 위하여 이용하였다. 따라서 SVM-NNM의 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 SVM-NNM의 훈련과정을 위한 입력층자료는 450개의 모의된 연 증발접시 증발량자료로 구성되어 있다.

5.1.3 혼합자료의 이용

본 연구의 혼합자료의 이용을 위한 훈련과정에서 훈련자료는 “5.1.1 실측자료의 이용”에서 선정한 실측된 월 증발접시 증발량자료와 “5.1.2 모의자료의 이용”에서 선정한 모의된 월 증발접시 증발량자료를 합한 혼합자료로 구성하였다. 따라서 SVM-NNM의 훈련과정을 위한 입력층 자료수는 각각 서울관측소는 536개, 강릉관측소는 538개, 인천관측소는 502개, 부산관측소는 488개, 제주 및 목포관측소는 473개의 혼합된 연 증발접시 증발량자료로 구성되어 있다.

5.2 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 Figure 2(a)-(f)는 서울, 강릉, 인천, 부산, 제주 및 목포관측소의 실측자료를 이용한 훈련과정의 수행결과 산정된 매개변수 및 테스트자료를 이용하여 SVM-NNM의 테스트과정의 수행결과에 대한 월 증발접시 증발량자료의 시계열의 비교를 나타낸 것이다. 또한 다음 Table 1은 실측자료를 이용한 훈련과정의 수행결과 산정된 매개변수를 이용하여 SVM-NNM의 테스트과정의 수행결과를 나타내고 있다.

6. 결 론

본 연구에서는 SVM-NNM을 이용하여 연간 소형 증발접시 증발량자료를 월별 증발접시 증발량자료로의 분리를 시도하였으며, 사용된 자료는 서울, 강릉, 인천, 부산, 제주, 및 목포관측소의 자료를 이용하였다. SVM-NNM의 훈련과정의 수행결과 산정된 각 6종류의 매개변수 군을 이용하여 테스트과정을 수행하였다. 본 연구를 통하여 분석에 이용된 자료의 수는 SVM-NNM의 수행능력에 영향을 끼치는 것으로 판단된다. 또한 본 연구를 통하여 SVM-NNM을 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. Eslamian, S.S., Gohari, S.A., Biabanaki, M., and Malekian, R. (2008). "Estimation of monthly pan evaporation using artificial neural networks and support vector machine." *J. Appl. Sci.*, Vol. 8, No. 19, pp. 3497-3502.
2. Gundekar, H.G., Khodke, U.M., and Sarkar, S. (2008). "Evaluation of pan coefficient for reference crop evapotranspiration for semi-arid region." *Irrig. Sci.*, Vol. 26, pp. 169-175.
3. 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.
4. 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.
5. Kisi, O. (2006). "Daily pan evaporation modeling using a neuro-fuzzy computing technique." *J. Hydro.*, Vol. 329, pp. 636-646.
6. Molina Martinez, J.M., Martinez, A.V., Gonzalez-Real, M.M., and Baille, A. (2005). "A simulation model for predicting hourly pan evaporation from meteorological data." *J. Hydro.*, Vol. 318, pp. 250-261.
7. Principe, J.C., Euliano, N.R., and Lefebvre, W.C. (2000). *Neural and adaptive systems: fundamentals through simulation*, John Wiley & Sons, Inc., NY.
8. Salas, J.D., Delleur, J.R., Yevjevich, V., and Lane, W.L. (1980) *Applied modeling of hydrologic time series*, Water Resor. Pub., Littleton, CO.
9. Vapnik, V.N. (1992). "Principle of risk minimization for learning theory." *Adv. Neural Inf. Processing Syst.* Vol. 4, pp. 831-838.
10. Vapnik, V.N. (1995). *The Nature of statistical learning theory*, Springer Verlag, NY.

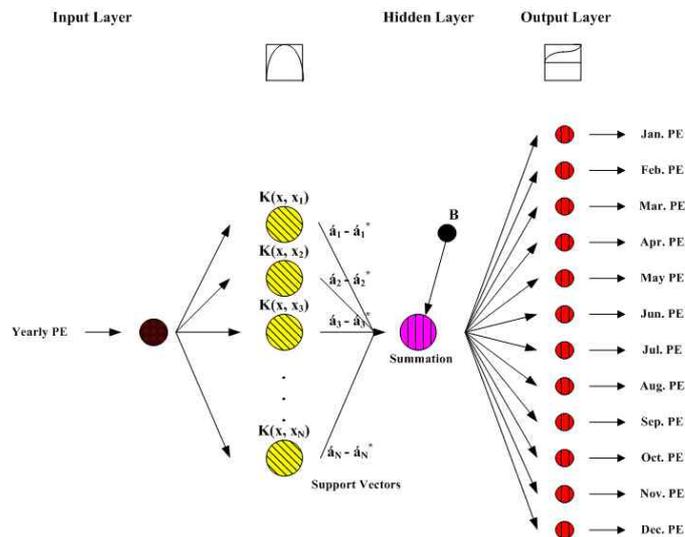


Figure 1. SVM-NNM structure

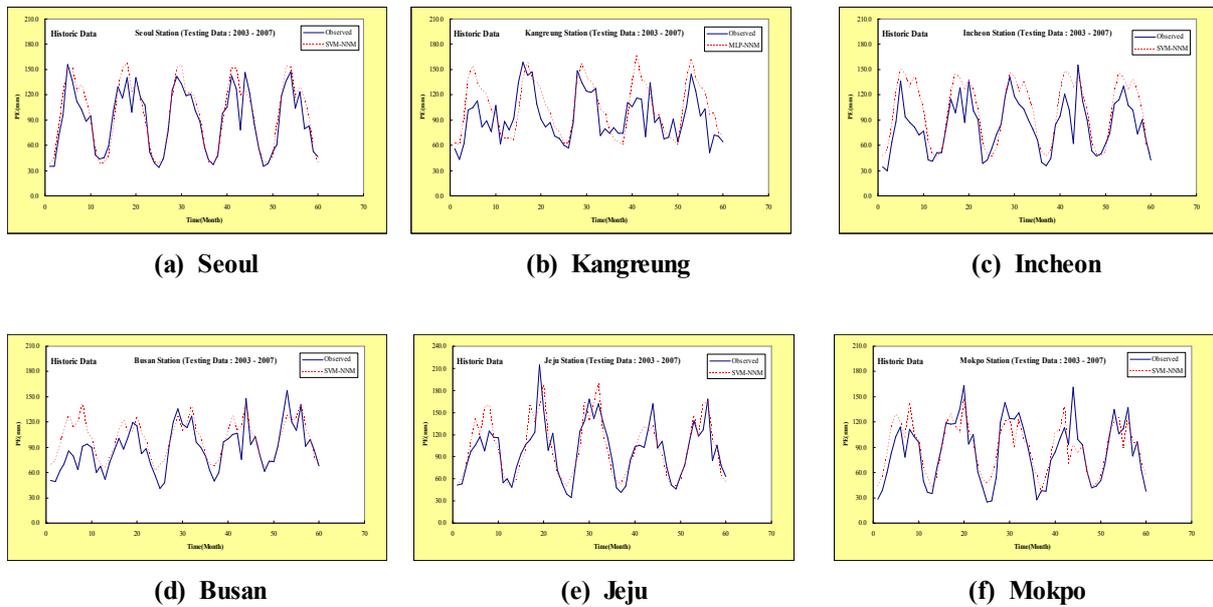


Figure 2. Comparison of the monthly PE for the testing performance (Historic data)

Table 1. Statistical analysis of the monthly PE for the testing performance (Historic data)

Station	Statistical Index	SVM-NNM
Seoul	CC	0.946
	RMSE	14.286
	E	0.852
	AARE	0.0920
Kangreung	CC	0.757
	RMSE	23.734
	E	0.262
	AARE	0.1650
Incheon	CC	0.868
	RMSE	24.954
	E	0.347
	AARE	0.3066
Busan	CC	0.785
	RMSE	19.356
	E	0.450
	AARE	0.1380
Jeju	CC	0.861
	RMSE	20.695
	E	0.703
	AARE	0.0162
Mokpo	CC	0.867
	RMSE	18.436
	E	0.742
	AARE	0.0247