ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
*Journal of Korean Society of Water and Wastewater* Vol. 28, No. 4, pp. 377-389, August, 2014
http://dx.doi.org/10.11001/jksww.2014.28.4.377

# Evaluation of short-term water demand forecasting using ensemble model

## 앙상블 모형을 이용한 단기 용수사용량 예측의 적용성 평가

**So, Byung-Jin[1] · Kwon, Hyun-Han[1*] · Gu, Ja-Young[2] · Na, Bong-Kil[3] · Kim, Byung-Seop[4]**

소병진[1] · 권현한[1*] · 구자용[2] · 나봉길[3] · 김병섭[4]

[1]Chonbuk National University · [2]University of Seoul · [3]K-water · [4]LSIS Co., Ltd

[1]전북대학교 · [2]서울시립대학교 · [3]한국수자원공사 · [4]LS산전

**Abstract :** In recent years, Smart Water Grid (SWG) concept has globally emerged over the last decade and also gained significant recognition in South Korea. Especially, there has been growing interest in water demand forecast and this has led to various studies regarding energy saving and improvement of water supply reliability. In this regard, this study aims to develop a nonlinear ensemble model for hourly water demand forecasting which allow us to estimate uncertainties across different model classes. The concepts was demonstrated through application to observed from water plant (A) in the South Korea. Various statistics (e.g. the efficiency coefficient, the correlation coefficient, the root mean square error, and a maximum error rate) were evaluated to investigate model efficiency. The ensemble based model with an cross-validate prediction procedure showed better predictability for water demand forecasting at different temporal resolutions. In particular, the performance of the ensemble model on hourly water demand data showed promising results against other individual prediction schemes.

**Key words : ensemble model, demand forecasting, short-term forecasting, urban water demand**
**주제어 : 앙상블 모형, 수요 예측, 단기 예측, 도시 물수요**

## 1. 서 론

최근 ICT 기술을 접목하여 공급자와 소비자간에 실시간 정보를 양방향으로 교환함으로서 수요량에 따른 공급이 가능한 스마트그리드(smart grid)의 개념이 확대 적용되어지고 있다. 최근 물분야에서도 스마트그리드에 따른 직접적인 효과인 에너지 사용 절감 및 효율향상과 스마트그리드 구축에 따른 일자리 창출 등과 같은 부가적인 장점을 인지하고 물관리에 적용하고자 하는 스마트 워터 그리드(smart water grid, SWG) 개념이 활발하게 연구되어지고 있다. 우리나라의 경우 2020년에 SWG를 적용하기 위해서 연구단이 2012년 출범하였으며 다양한 연구가 활발히 이루어지고 있다. 우리나라 수자원의 대부분이 댐 및 저수지와 같은 저장용량과 더불어 특정 지역까지의 물 공급 소요시간이 필요함에 따라서 양방향의 실시간 물 수요에 대한 즉각적인 대응에 한계가 존재한다. 이러한 문제를 해결하기 위해 가정 및 산업체의 물수요 사용 이력을 바탕으로 향후 수요량을 예측할 수 있는 적절한 예측 기법의 사용은 SWG 기법에 꼭 필요한 기초 연구라 할 수 있다. 적절한 수요예측 기법의 개발은 향후 물공급에 필요한 각종 시설의 운영계획의 기본 자료로서 활용

• Received 16 April 2014, revised 09 June 2014, accepted 30 June 2014.
*Corresponding author: Tel : 063-270-2426 Fax : 063-270-2421 E-mail : hkwon@jbnu.ac.kr

될 수 있으며 에너지 효율 측면에서 여러 장점을 제공한다. 이와 더불어 불필요한 물공급을 줄이고 물공급 부족량을 사전에 파악함으로서 효율적인 물관리 및 물 부족에 따른 적절한 대응책을 마련할 수 있다.

수요량 예측은 수요 시설의 용량 확장, 투자계획, 운용, 관리 등의 목적에 따라서 다양한 시간 해상도에 따른 예측 기법이 적용되어지고 있으며 예측 기간에 따라 장기, 중기, 단기 예측으로 구분되어진다(Alvisi et al., 2007; Ghiassi et al., 2008; Jain et al., 2001). 장기, 중기, 단기예측의 구분은 연구자에 따라 다양하게 구분되어지고 있으나 1년 미만을 예측할 경우 단기예측으로, 1년 ~ 10년 미만 예측시 중기 예측, 10년 이상의 기간을 예측시 장기 예측으로 구분될 수 있다(Gardiner and Herrington, 1990). 또한 단기 예측시 시, 일, 주 단위로 예측이 이루어지며, 중기 예측의 경우 월 및 연단위, 장기 예측의 경우 연단위 이상의 시간 해상도를 갖도록 정의하였다.

예측의 시간 해상도 및 예측 기간의 따라서 다양한 예측 대상자료와 인자들이 사용되어지고 있다. Billings and Jones (2008)에 따르면 일별 최대 수요량, 일별 수요량, 월별 수요량, 인구당 연별 수요량, 소득 계급에 따른 연별 수요량과 같은 인자들이 도시지역의 물 수요 예측 대상 자료로 주로 활용되어지고 있다. 수요량 예측을 위한 예측인자(predictor)로서 시간별 수요량, 주별 최대 수요량, 최대 온도, 강수량 등이 사용되고 있다.

단기 예측의 연구 동향을 살펴보면, Jain and Ormzbee(2001)는 전날 수요량 및 일별 강수량, 최대 온도, 일조시간을 예측인자로 회귀모형과 인공신경망을 이용하여 5일간의 일별 수요량을 예측하였다. Adamowaki and Karapataki(2010)는 전 주의 최대 수요량과 최대 온도를 예측인자로 Levenberg-Marquardt(LM) 인공신경망 기법과 다변량 선형 회귀모형을 통하여 주별 최대 수요량 예측 결과를 비교하였으며 LM 인공신경망의 우수성을 입증하였다. Cutore et al(2008)은 일별 수요량 자료를 기반으로 이전 날의 수요량, 평일 수요량, 일별 수요패턴을 예측인자로 활용하였으며 최적화 기법으로 SCEM(Shuffled Complex Evolution Metropolis) 적용하였으며, 최종적으로 일반적인 인공신경망 기법과 Bayesian 인공신경망 기법을 비교한 결과 두 모형간 유사한 예측능력을 갖고 있음을 확인하였다. Herrera et al(2010)은 1시간과 2시간 전의 수요량과 전 주의 수요량, 온도, 풍속을 예측인자로 인공신경망과 Support Vector Machine(SVM) 등 4개의 모형을 적용 비교한 결과 SVM 모형이 상대적으로 우수한 예측능력이 있음을 확인하였다. Jentgen et al.(2007)은 운영 비용의 최소화를 위한 배수지 펌프 운영에 이용할 수 있는 수요량 예측모형을 검토한 결과 시간단위예측의 경우 인공신경망 기법이 상대적으로 우수하였으며, 일별 예측시에는 다중회귀모형이 우수한 예측능력이 있음을 보여주었다. Bougadis 등(2005)은 최대 수요량에 다양한 지체 시간을 적용함과 더불어 온도, 강우량을 예측인자로 활용하여 선형 회귀모형과 다변량 선형 회귀모형, 인공신경망 모형에 적용한 결과 인공신경망 모형이 주별 최대 수요량 예측에 있어서 가장 우수한 방법임을 입증하였다. 이외에도 Jain and Ormsbee(2002)와 Jain et al(2001)은 각각 일별 수요량과 주별 수요량 예측에 있어서 모형간 예측 결과를 비교한 결과 인공신경망이 가장 우수한 예측력이 있음을 보여주었다.

2가지 이상의 모형 결합을 통한 단기 예측 연구를 살펴보면 다음과 같다. 시간별 수요량 자료의 계절 성분에 대하여 Fourier 예측기법을 적용하고 계절성분을 제거한 잔차 성분에 대하여 AR(1)모형을 적용하여 시간별 펌프량을 계산할 수 있는 실시간 예측 모형을 개발한 사례가 있다(Alvisi et al., 2007). Aly and Wanakule(2004)은 지수 평활화 모형과 회귀모형을 결합하여 일별 수요량 예측모형을 개발하였

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
*Journal of Korean Society of Water and Wastewater* Vol. 28, No. 4, pp. 377-389, August, 2014

으며, Caiado(2010)은 Holt-Winters, ARIMA, GARCH 모형을 가중평균한 일별 수요량 예측모형을 제안하였다, Gato et al.(2007)은 일단위 예측인자와 계절단위 예측인자를 동시에 활용한 일단위 수요량 예측모형을 제시하였다.

국내의 경우 Kwon and Moon(2004)은 Singular Spectrum Analysis(SSA) 기법을 이용한 일 단위 물수요량 예측모형을 적용하여 기존 모형에 비교하여 우수성을 입증하였다. Gu(1996)은 급수량 단기 수요 예측을 위하여 Multiple Auto Regressive Integrated Moving Average(MARIMA) 모델을 적용하여 계절별 수요량을 예측하였다. Choi et al(2009)은 사천시 물 수요량을 예측하기 위하여 신경회로망과 시간가중계수를 이용한 일별, 시간별 예측모델을 개발하였으며 선형 ARIMA 모델에 비하여 우수한 예측능력이 있음을 확인하였다. Yu et al(2004)은 선형모델과 비선형 모델을 이용한 시간단위 단기 수요량 예측모형을 개발하였으며 선형모형이 비선형 모형에 비하여 안정된 해석결과를 제시하고 있음을 보여주었다.

최근 다양한 분야에서 앙상블(ensemble) 개념을 이용한 시계열 예측기법이 연구되어지고 있으며, 이론적으로 앙상블 모형의 예측 오차가 단일 모형을 통한 예측 오차보다 적음을 여러 연구를 통해서 입증되었다(Krogh and Vedelsby, 1995; Krogh and Sollich, 1997; Naftaly et al., 1997). 앞서 언급되었듯이 국 · 내외적으로 다양한 단기 수요량 예측기법들이 적용되어 왔으나 최적의 모형은 연구에 사용된 자료 및 대상지역에 따라 서로 다르게 결정되는 특성이 있다. 이러한 점에서 본 연구의 목적은 기 개발된 선형 또는 비선형 예측모형을 통합적으로 연계해석 할 수 있는 단기 상수도 앙상블 수요예측모형을 개발하는 것이며, 실제 상수도 수요량 자료를 대상으로 단기 수요 예측에 있어서 개발된 앙상블 모형의 적용성을 평가하는 것이다.

## 2. 대상 자료 및 연구 방법

본 절에서는 단기 수요량 예측시 이용되는 앙상블 모형의 적용과정을 살펴보고 앙상블을 구성하는 단일 예측 모형들에 대한 이론적 배경과 앙상블 모형의 특징을 서술하였다.

### 2.1 연구 방법

본 연구에서는 다양한 시계열 예측모형을 도입하여 연구를 진행하였다. 즉, 본 연구에서는 다중 선형 회귀모형(multiple linear regress model), KNN(k-nearest neighbor), 인공신경망 모형(perceptron radial basis function network, PRBFN), SVM(support vector machine) 모형 등을 개별 예측모형으로 활용하였으며, 개별모형들에 비해 개선된 예측결과를 도출하고자 각 모형에 가중치를 부여할 수 있는 앙상블(ensemble) 예측 모형을 개발하였다. 또한, 개별모형들에 대한 매개변수 추정시 교차검증(cross-validation, CV) 기법을 도입하여 추정 매개변수의 신뢰성을 높이고자 하였다. 본 연구에서 적용한 앙상블 알고리즘의 흐름은 Fig. 1과 같다.

본 연구에서는 2012년 11월 1일부터 2013년 2월 27일 기간에 1분 간격으로 계측된 4개월 기간의 A 정수장 계측자료를 이용하여 앙상블 모형에 적용하였으며, 10분, 60분, 180분 등 3개의 시간 해상도로 재구성하여 모형에 적용성을 평가하였다.

본 연구에서 적용된 앙상블 모형에 적용과정을 요약하면 다음과 같다. 첫 번째 단계는 예측에 이용될 입력 자료 준비와 앙상블모형에 사용될 개별 예측모형의 결정 단계이다. 개별 예측모형과 입력 자료는 앞서 설명되었으며, 첫 단계에서 고려되어야 할 추가사항은 개별 모형에 적용되어질 독립변수의 정의이다. 본 논문에서는 A 정수장의 계측자료만을 이용하여 3개의 독립변수를 설정하도록 정의하였다. 설정된 독립변수들은 이전 시간의 수요량(t-1), 예측하고자 하는

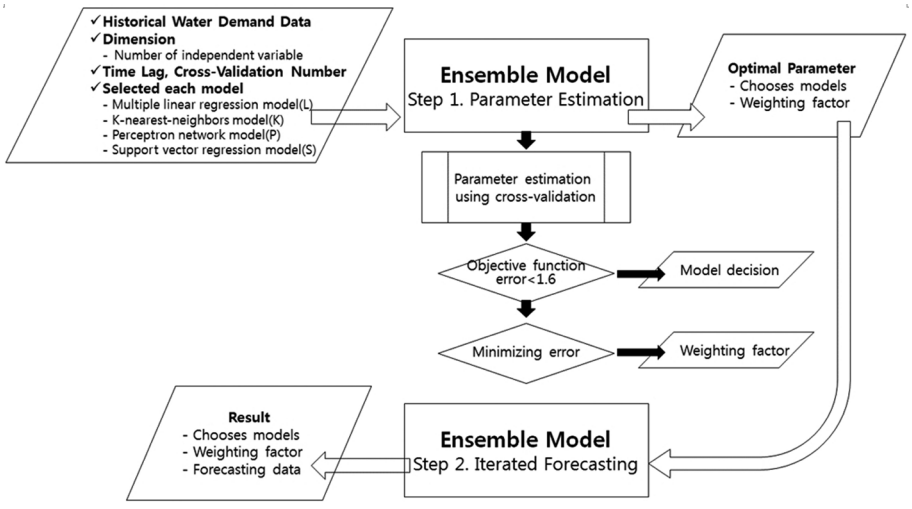


Fig. 1. Flow chart of the proposed ensemble water demand forecasting model.

시간의 전날 수요량(d−1), 예측하고자하는 시간의 전전날 수요량(d−2)을 갖도록 지체시간을 설정하였다.

두 번째 단계는 개별 모형에 따른 매개변수의 추정 단계로서 독립변수와 종속변수로 구성된 입력자료를 앙상블 모형에 적용된 개별모형에 적용하여 개별 모형에 따른 매개변수를 추정한다. 이 때, 5회의 교차검증을 통하여 매개변수가 추정되어진다. 교차검증 과정은 입력자료에 대한 매개변수 추정시 이용되어질 적합(calibration)구간과 추정된 매개변수를 통하여 예측을 실시하는 검정(verification)구간을 교차검증 횟수만큼 임의로 구분하고 반복적으로 매개변수를 재추정하는 과정을 의미한다. 이러한 교차검증 기법은 다양한 자료특성에 따른 모형의 적응력을 높일 수 있는 방안으로서 통계적으로 신뢰성(robust) 있는 매개변수를 추정하는데 활용되는 대표적인 방법이다.

다음 단계는 앙상블 모형의 구성 단계이다. 이 단계에서는 목적함수를 적용하여 교차검증을 통해 산정된 개별모형들을 대상으로 앙상블 모형을 구성할 개별모형을 선별하고 최적화를 통한 개별모형별 가중치를 산정한다.

마지막 단계는 앙상블 모형을 통한 예측 단계로서, 특정 지점에 대하여 구축된 앙상블 모형에 대하여 입력자료를 적용하여 임의의 기간에 대하여 예측할 수 있다.

본 논문에서는 3가지 시간 해상도에 대해서 각각 3개의 독립변수로 구성되는 모형을 구축하였으며 앙상블 모형 결과를 중심으로 평가하였다.

### 2.2 앙상블 모형에 사용된 개별 모형

1) Support Vector Machine(SVM)

Support Vector Machine(SVM)은 시계열 예측을 위해서 사용되어지고 있는 대표적 비선형 회귀분석 모형 중 하나로 SVM 회귀모형은 x∈R에서 추출된 입력항 $\{x_1, x_2, \cdots, x_l\}$과 L차원의 y∈R인 $\{y_1, y_2, \cdots, y_l\}$출력항의 함수적인 의존관계 f(x)를 추정할 수 있다. 여기서 함수 f(x)의 추정은 다음 위험도 함수를 최소화 하는 문제로 귀결될 수 있다. 즉, f(x)는 기하학적으로 다수의 입력벡터와 출력항 중에서 이들 관계를 규정할 수 있는 최적의 회귀함수를 의미한다. 본 연구에서 적용된 SVM 모형에서 각각의 변수들인 x, y, f(x)는 각각 독립변수로 구성된 상수수요량과 종속변수, f(x)는 SVM을 통하여 생성된 예측된 상수수요량을 나타낸다.

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
*Journal of Korean Society of Water and Wastewater* Vol. 28, No. 4, pp. 377-389, August, 2014

$$R[f] = \int \Gamma(x,y,f(X))dP(X,y) \quad (1)$$

$$f(X) = \langle W,X\rangle + b \text{ with } W\in X,\ b\in R \quad (2)$$

$$f(X) = \sum_{j=1}^{K} w_j + x_j + b \quad (3)$$

SVM 회귀분석에서 b는 bias를 w는 Basis 함수를 나타낸다. 〈,〉는 x의 벡터내적(dot product)을 의미하며 K는 입력자료의 차원을 나타낸다. $\Gamma(x,y,f(X))$는 관측치 y와 예측치 f(x)와의 차이 즉, Loss 함수를 나타낸다. 즉, 식 (1)의 오차를 최소화 하는 최적화 과정을 통해서 매개변수들이 결정된다. 결국 SVM 회귀분석은 입력항과 출력항의 비선형 함수 관계를 선형의 조합으로 분류함으로써 가능하며 결국 이러한 문제는 Pattern 분류와 밀접한 관계가 있다. SVM에서 비선형성은 커널함수(kernel function)를 이용하여 훈련자료를 고차원으로 보냄으로써 고려할 수 있으며 훈련자료는 항상 표본들의 내적(inner product)의 형태로 항상 표현된다. 따라서 내적항을 적절하게 선택된 커널함수로 대체함으로써, 부수적인 매개변수의 증가 없이 고차원의 함수로 Mapping할 수 있으며 이를 소위 "커널트릭(Kernel Trick)"이라고 한다. 즉 커널함수를 이용하여 입력항과 출력항의 상호관계를 효과적으로 규명할 수 있다(Kwon and Moon, 2006).

2) 다중 선형 회귀모형(Multiple linear regress model)

단순회귀분석은 실제 상수도 수요량 예측 분야의 인과관계 분석에서 매우 드물게 적용된다. 이는 상수도 수요량과 관련된 독립변수와의 인과관계에서 어떤 결과를 야기하는 원인들은 복수인 경우가 대부분이며, 이 원인들끼리도 서로 얽혀있기 때문이다. 따라서, 거의 대부분의 경우 원인을 나타내는 독립변수가 복수로 구성되는 다중회귀모형을 구성하여 인과관계를 추정하게 된다.

상수도 수요량과 관련된 모든 요인들을 우변에 포함하는 회귀모형을 작성하면 다음 식과 같은 형태를 지니게 될 것이다. 일반적인 관례대로 종속변수는 Y로, 독립변수는 X로 표기한다.

$$Y = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \cdots + u \quad (4)$$

이 모형에서 제일 마지막 항인 $u$는 모든 원인변수들이 다 포함된 뒤에도 남는 것으로 Random Error를 의미한다.

3) K-Nearest Neighbor 모형

시계열의 예측을 위해 주로 사용되고 있고 있는 선형모형의 대안으로서 시계열의 비선형 특성을 고려할 수 있고 단기예측이 우수하며 비매개변수적 방법인 Nearest Neighbor 방법(Casdagli, 1992)이 이용가능하다. Nearest Neighbor 방법과 같은 비선형 방법은 예측 시에 이전에 발생했던 자료의 Pattern을 학습시켜 이전 자료와 가장 유사한 특성을 갖는 자료군을 찾아 예측에 이용한다. 즉, 예측시점 주위에 자료를 이용하여 예측을 수행하기 때문에 자료의 비선형 거동에 효과적으로 대응할 수 있는 장점이 있다.

Nearest Neighbor 방법은 상태-공간 차원에서의 비선형 시계열 분석에 근간을 두고 있으며 비선형 단기 예측을 향상시키기 위해서 공간적인 상관관계를 이용한다. 이 방법의 기본 접근방법은 과거의 발생한 시계열의 한 부분이 유사한 특성을 가지고 미래의 다시 발생한다는 것이다. 따라서 Nearest Neighbor 방법은 예측을 위한 국부적인 정보만을 이용하며 전체적인 함수를 추정하지 않는다.

Nearest Neighbor 방법은 여타의 선형 시계열 모형처럼 정상성(stationarity)의 가정을 요구하지 않으며 국부적인 예측은 단지 이용 가능한 과거 자료주변의 벡터의 거동을 분석함으로서 가능하다. 본 논문에서는 Nearest Neighbor 이론의 배경을 간단히 설명하면 다음과 같다. 다

음 식은 Casdagli(1992)와 Casdagli and Weigend(1994)에 의해서 제안된 Nearest Neighbor 방법을 이용한 상태-공간 모형을 나타낸다.

$$\upsilon_{t+T} = \alpha_0 + \sum_{j=1}^{d} \alpha_j x_{t-(j-1)\tau} + \epsilon_t \quad (5)$$

여기서 $\upsilon_{t+T}$는 Nearest Neighbor 모형을 통해 추정된 예측 수요량으로 T는 예측하고자 하는 시간 간격을 나타내며, $d$는 Embedding Dimension을 $\tau$는 지체시간을 의미한다. 정의된 d와 $\tau$에 따른 독립변수 x에 대한 매개변수 $\hat{\alpha}$, $\hat{\alpha}_1$,……, $\hat{\alpha}_d$를 추정함으로서 모형을 통한 수요량 추정이 가능하다. 이를 계산하기 위한 과정은 다음과 같다.

첫째, $\upsilon_t$추정하고자 하는 $d$는 Embedding Dimension과 지체시간 $\tau$를 갖는 상태-공간 $\upsilon_t=(x_t, x_{t-\tau}\cdots, x_{t-(d-1)\tau})$으로 최대의 예측 결과를 도출하도록 구성한다. 일반적으로 지체시간 $\tau$는 가능한 작은 값을 취해야 하지만 상태-공간 내에서 연속된 점들 사이에 자기상관성을 최소화하기 위해서는 충분하게 큰 지체시간이 요구되기도 한다.

둘째, 이들 구성된 상태-공간 내에서 $\upsilon_t$에 가장 유사한 특징을 갖는 Nearest Neighbor를 추정할 수 있다. 여기서 Nearest Neighbor의 거리(distance)를 추정하기 위한 많은 방법들이 있으나 가장 보편적으로 사용되는 Euclidean 거리를 이용하였으며 $X=x_1, x_2, x_3, \cdots, x_n$, $Y=y_1, y_2, \cdots, y_n$을 갖는 시계열을 가정하면 Euclidean 거리는 다음의 식 (6)과 같다.

$$D = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2} \quad (6)$$

셋째, 식 (5)의 각 매개변수는 단계 2에서 Nearest Neighbor를 이용하여 추정한다.

넷째, $\upsilon_t$의 i-단계 예측은 다음 식 (7)로 추정한다.

$$\widehat{\upsilon_{t+i}} = \hat{\alpha}_0 + \sum_{j=1}^{d} \alpha_j \upsilon_{t-(j-1)\tau} \quad (7)$$

4) Perceptron radial basis function network

Perceptron Radial Basis Function Network(PRBFN)는 Neural Networks 모형과 Radial Basis Function(RBF) 방법의 확장형으로서 PRBFN 모형은 RBF 모형과 은닉층(hidden layer)안에서 Sigmoid 함수를 결합시킨 모형으로서 다양하고 복잡한 훈련과정을 포함한 비선형 예측모형이다(Cohen and Intrator, 2001). 은닉층의 수 및 중심들은 모형 훈련과정에서 유도되며 관련 매개변수들은 경사하강법(gradient descent)에 의해서 추정된다.

인공신경망 알고리즘은 신경전달 과정을 단순화 하고 이를 수학적으로 해석한 모델로서, 복잡하게 얽혀있는 뉴런을 통과시켜가면서 뉴런끼리의 연결강도를 조절하는 일종의 학습과정을 통해 문제를 분석한다. 이러한 과정은 사람이 학습하고 기억하는 과정과 유사하며, 이를 통해 추론, 분류, 예측 등을 수행할 수 있다. 현재 신경망은 최적화 문제, 예측문제 등에 많이 활용되고 있다.

본 연구에서 상수도 수요량 예측을 위한 신경망모형은 다중 퍼셉트론의 feedforward 구조를 가지고 있다. 신호는 뉴런입력으로부터 은닉층을 통하여 전방으로 흐르게 되고 출력뉴런에 이르게 되는 구조를 가지고 있다. 뉴런은 층으로 배열되어 있으며, 입력층은 단순히 입력변수의 값을 가지며, 은닉층과 출력층의 뉴런은 선행층의 모든 요소들과 연결되어 있다. 기본방정식은 다음과 같다.

$$y_k(t) = g\left(\sum_{h=0}^{n_H} w_{kh}^0 f\left(\sum_{i=0}^{n_I} w_{h_i}^h x_i(t) - \theta_i\right) - \theta_h\right) \quad (8)$$

여기서 $x_i$는 입력층 요소 i로의 입력, $y_{k(\cdot)}$는 출력층 요소 k로의 출력을 의미하며, $n_I$는 입력요소 수, $n_H$는 은닉 요소 수, $n_k$는 출력 요소 수를 의미한다. $w_{h_i}^h$는 입력요소 i와 은닉요소 h 사이의 연결강도를 조절하는 매개변수 또는 가중

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
*Journal of Korean Society of Water and Wastewater* Vol. 28, No. 4, pp. 377-389, August, 2014

치를 나타내며 $\theta_i$와 $\theta_h$는 기준값을 의미한다. $w_{kh}^0$는 은닉요소 $h$와 출력요소 $k$사이의 연결강도를 조절하는 매개변수 또는 가중치를 의미하며 $f$는 Logistic 함수형태의 활성함수, $g$는 $g(x)=x$형태의 Identify Function을 나타낸다.

5) Ensemble Model

최근에는 하나의 모형이 아닌 여러 모형을 조합하는 앙상블 개념의 예측모형 개발이 이루어지고 있다. 예를 들어 다수의 예측모형으로 통해 추정된 각 모형별 예측치를 $f_i(x)$라고 한다면 다음과 같이 앙상블 모형을 나타낼 수 있다.

$$\hat{f}_i = \sum_{i=1}^{K} w_i f_i(x) \tag{9}$$

여기서 가중치 $w_i$의 합은$\sum_{i=1}^{K} w_i=1$로 가정한다. 앙상블 개념의 가장 큰 특징은 예측의 일반화 능력(generalization ability)이라 할 수 있다. 앙상블 모형을 통한 일반적 예측 오차는 단일 모형을 통한 일반적 예측 오차보다 작으며(Krogh and Vedelsby, 1995) 여러 연구를 통해서 입증된바 있다(Krogh and Sollich, 1997; Naftaly et al., 1997). 이러한 점은 일반적 예측오차의 기댓값을 평가해보면 쉽게 인지할 수 있다. 일반적 예측오차의 기댓값은 다음과 같이 나타낼 수 있다.

$$Error(x) = \| \hat{f}(x) - y \|^2 \tag{10}$$

위의 기댓값에 대한 편의 및 분산의 분해과정은 다음과 같이 나타낼 수 있다(Geman et al., 1992).

$$Error(x) = \sigma^2 + (\mathrm{bias}(\hat{f}(x)))^2 \; Var(\hat{f}(x)) \tag{11}$$

식(11)에서 $\sigma^2$은 주어진 입력벡터 $x$에 따른 $y$의 분산을 나타내며 분산 $Var(\hat{f}(x))$는 다음과 같이 분해가 가능하다.

$$Var(\hat{f}) = \sum_{i=1}^{K} w_i^2 (E[f^2{}_i] - E^2[f_i]) + 2\sum_{i<j} w_i w_j (E[f_i f_j] - E[f_i]E[f_j]) \tag{12}$$

여기서 기댓값은 사용된 자료로부터 추정된다. 식(12)의 첫 번째 항은 앙상블 분산의 최소구간을 나타내고 앙상블 멤버들의 가중평균을 의미한다. 두 번째 항은 앙상블 멤버들의 교차상관성을 나타내는 항으로서 만약 앙상블 멤버들 간의 상관성이 존재하지 않는다면 사라지는 항이라 할 수 있다(Krogh and Sollich, 1997).

이러한 점에서 앙상블 예측값의 분산 감소 여부는 각각의 모형의 독립성과 관계가 깊다(Naftaly et al., 1997). 앙상블 멤버들의 예측값에 대한 독립성을 확보하기 위해서 다양한 방안들이 도입되어 왔다. 가장 일반적인 방안은 입력자료의 일부분만을 선택하여 모형을 훈련시켜 모형을 적합시키는 방안(Krogh and Vedelsby, 1995)과 초기조건을 무작위로 선택하는 훈련 알고리즘이라 할 수 있다(Naftaly et al., 1997; Breiman, 1996). 다른 방안으로 여러 모형의 결과를 범주화하고 이를 통해 대표모형을 선택하는 앙상블 개념이 도입되고 있다(Bakker and Heskes, 2003).

본 연구에서는 2가지 방안을 결합한 앙상블 예측기법을 제안하고 상수도 예측모형에 활용하고자 한다. 즉 교차검증(cross validation)을 통한 모형 매개변수의 신뢰성 개선과 함께 각 모형과의 독립성을 확보하며 여러 예측모형을 앙상블로 결합하는 통합모형을 구축하고자 한다.

본 연구에서 반복예측기법을 통하여 모형을 구축하였다. $n$개의 자료 수를 가진 $x_v$, $v=1,\cdots,n$의 시계열이 있다면 $d$차원을 가진 상태공간벡터(state space vector)

$$\vec{x}_n = (x_{(n-\lambda(d-1))}, x_{(n-\lambda(d-2))}), \cdots, x_n \tag{13}$$

여기서 λ는 시간지체(time lag)를 나타낸다. 반복 예측을 위한 "one-step ahead" 예측 모형 $f(\vec{x}_n)$은 다음과 같이 나타낼 수 있다.

$$f : \mathrm{R}^d \rightarrow \mathrm{R} \tag{14}$$

$$f(\vec{x}_n) = x_{n+1} \tag{15}$$

여기서 예측값 $x_{n+1}$은 다음 시간의 $x_{n+2}$를 예측하기 위한 상태공간벡터의 입력자료 $\vec{x}_{n+1}$로 이용된다. 비선형 예측을 위한 모형 구축시 선행 예측 시간에 따라서 "one-step ahead" 모형은 $f^1$, "two-step ahead" 모형은 $f^2$, 마지막으로 "n-step ahead" 모형은 $f^n$ 등과 같이 정의할 수 있다.

본 연구에서 제시되는 방법론을 수행하는데 있어서 가장 어려운 점은 첫째 모델의 선택 및 매개변수의 추정이다. 예를 들어 Neural Network 모형의 Neurons의 수를 결정하는 과정이 필요하다. 이를 위해서 본 연구에서는 K-fold 교차검증 방법을 도입하였으며 여기서 자료는 K개의 부분으로 나누어져 매개변수의 검보정시 이용된다. 즉 선택된 부분자료를 대상으로 여러 개의 독립된 예측모형을 훈련시키고 매개변수를 추정하는 과정을 K번 실시하여 신뢰성 있는 모델 매개변수를 추정하게 된다. 훈련이 끝난 모형을 검증하는 과정에서 최적의 성능을 발휘하는 모형만이 앙상블 멤버로 선택된다.

훈련기간 동안 입력자료 $(\vec{x}_i, y_i)$의 M개 Set이 구축된다. 이들 입력 자료는 상태공간벡터의 형태로 식 (13)과 "one-step ahead" 예측값인 $y_i = x_{i+1}$으로 구성된다. 훈련기간의 입력자료 Set $M_{train}$은 "one-step ahead" 의 예측오차를 최소화하는데 이용되며 다음과 같이 나타낼 수 있다.

$$E_{train} = \sum_i = (y_i - f^1(\vec{x}_i))^2 \quad (16)$$

여기서 $E_{train}$은 모든 앙상블 멤버를 대상으로 계산이 이루어지며 훈련자료에 대해서 오차를 최소화하는 과정을 통해 모형을 적합시킨다. 특히 모형의 적합과정에서 과적합(over-fitting)을 피하기 위해서 교차 검증시 인접한 자료만을 추출하는 방안을 도입하였으며 이를 통해 최적의 모형을 구축하는 과정을 수행하였다. 모든 훈련과정에서 "n-step ahead" 예측모형의 교차검증을 실시하고 평균제곱오차(mean squared error)를 계산하였다.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - f^i)^2 \quad (17)$$

여기서 $f^i$는 "i-step ahead" 예측을 나타내며 검증자료를 대상으로 계산이 이루어진다. 이러한 과정을 통해 가장 작은 MSE를 나타내는 모형이 선택되며 최종적으로 앙상블 멤버로 활용된다. 이러한 과정을 K번 반복하면 모든 자료구간에 대해서 다양한 형태의 결과를 얻을 수 있으며 모형의 일반화 관점에서 가장 우수한 성능을 가지는 모형의 구축이 가능하다. 최종적으로 앙상블 모형은 다음과 같다.

$$\hat{f}(x) = \frac{1}{K} \sum_{i=1}^{K} f_i(x) \quad (18)$$

본 연구에서 제안하는 방법은 일종의 적응(adaptive) 모델로서 인식될 수 있으며 다양한 모형집단에서 최적의 성능을 발휘하는 모형을 선택하는데 목표를 두고 있다.

## 3. 적용 및 분석

본 연구에서는 10분, 60분, 180분에 시간 해상도을 갖는 A정수장 자료를 대상으로 연구를 진행하였다. 3가지 경우의 시간간격 자료의 사용 목적은 10분 자료의 경우 초단기 예측을 수행하기 위하여 설정되었으며, 60분과 180분 자료의 사용은 단기예측을 통한 펌프운영계획 수립을 목적으로 설정하였다. 측정된 A정수장의 2012년 11월 1일부터 2013년 2월 27일의 자료 중 2013년 2월 26일 ~ 2월 27일에 자료를 검정구간으로 2012년 11월 27일 ~ 2013년 2월 25일 동안의 자료를 매개변수 추정 구간으로 적용하였다. 매개변수 추정시 자료 길이는 계절성을 고려하기 위하여 추정일로부터 약 3개월 자료만을 사용하였다.

앞 절에서 언급하였듯이 앙상블 모형에 독립변수는 3개로 정의되며 추정시점으로부터 t-1, d-1, d-2일에 자료가 적용되어졌으며, 교차검증에 횟수는 5회로 설정하고 교차검증시 적

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
*Journal of Korean Society of Water and Wastewater* Vol. 28, No. 4, pp. 377-389, August, 2014

용되는 검정자료의 비율은 전체 자료의 20%가 적용되었다. 앙상블 모형에 적용되어진 개별모형은 다중 선형 회귀모형(L), KNN(K), perceptron(P), Support Vector Machine(S) 등 4개 모형이 적용되어졌다.

Table 1, Table 2, Table 3은 10분, 60분, 180분 자료에 대한 앙상블 모형 적용결과를 나타낸다. 이와 더불어 앙상블 구성에 이용된 개별모형과 앙상블 모형을 통한 예측 결과에 대한 각종 검정통계량 값들을 보여주고 있다. 각 Table에서 MME(Multi Model Ensemble)는 로 개별모형들이 조합된 앙상블 모형에 최종결과를 나타내어주며 P, S, L, K는 각각 perceptron, SVM, 다중선형 회귀모형, KNN 모형을 나타내어 준다. 예를 들어, Table 1에서 MME 산출에 이용되어진 모형들은 P1, S1, P2로 구성되어지며 2개의 perceptron 모형과 1개의 SVM 모형의 결합으로 이루어졌음을 확인할 수 있다.

Table 1부터 Table 3의 모형들 뒤에 표기된 _10 m, _60 m, _180 m은 Table에서 모형간에 혼돈을 피하기 위하여 표기하였으며, 사용된 자료의 시간 해상도를 의미한다. Table 1부터 Table 3에 나타낸 검정통계량의 계산식은 Table 4와 같다. Table 4에 포함된 통계량은 각각 평균 오차율, 최대 오차율, 평균 제곱근 오차, 평균 절대 오차, 효율성 계수, 상관계수, 최적 가중치 설정시 이용되어진 목적함수를 나타낸다. 앙상블 모형에 개별 모형별 가중치 산정에 이용된 목적 함수에 경우 자료의 평균에 대한 오차비를 적용하였으며 0에 가까워질수록 평균 모형의 오차를 기준으로 모형이 개선되어졌음을 나타낸다.

10분 자료에 대한 예측결과를 보여주는 Table 1에서 MME에 대한 오차가 0.0283으로 개별모형들에 비하여 개선된 결과를 보여줌으로서 앙상블 모형에 가중치가 목적함수에 따라 효과적으로 도출되었음을 확인할 수 있다. 앙상블 모형의 구성은 2개의 perceptron 모형과 1개의 SVM 모형으로 구성되었으며 SVM이 30%, perceptron1이 18%, perceptron2가 52%에 비율로 구성되어졌으며, 최대 오차율, RMSE, COE, 상관계수, 오차값에서 개별모형들에 비하여 개선된 결과가 나타났다. 평균 오차율에 경우 SVM모형에서 0.5%로 가장적은 오차율을 보여주었지만 최대 오차율에서 SVM 모형에 결과가 가장 크게 나타났다. 이러한 결과는 SVM을 통한 예측 결과가 평균적으로 과소 추정되고 있음을 나타낸다 하겠다. 앙상블 모형의 경우 평균 오차율과 평균 절대 편차에서 SVM 모형에 결과에 비하여 크게 나타나고 있지만 이는 SVM 모형과 perceptron 모형과의 가중치를 통한 최적의 예측결과 도출시 분산 특성을 반영하면서 나타난 현상으로 볼 수 있다.

60분 간격으로 구성된 수요량 자료에 대한 앙상블 모형 적용결과를 보여주는 Table 2에서 앙상블 구성비를 살펴보면 SVM이 48%, 2개의 perceptron 모형이 52%에 비율을 나타내고 있다. 또한, 10분 자료의 결과 양상과 동일하게 앙상블 모형이 최대 오차율, RMSE, COE, 상관계수, 목적함수에 결과에서 개별모형들보다 개선된 결과를 나타내었으며, 평균오차율과 MAE에서는 SVM 모형에 비하여 크게 나타났다.

180분 수요량 자료에 대한 결과 앙상블의 구성에 있어서 3개의 perceptron 모형과 1개의 SVM과 선형회귀모형으로 구성되어졌으며, SVM 60%, 선형회귀모형 2%, perceptron이 38% 비율을 나타내고 있다. 2% 선형회귀모형에 비율을 제외하면 10분과 60분 자료와 마찬가지로 SVM과 perceptron 모형이 앙상블을 구성하고 있음을 확인할 수 있다. 앞의 결과와 동일하게 앙상블 모형이 최대 오차율, RMSE, COE, 상관계수, 목적함수에서 개별모형에 비하여 개선된 결과를 보여주고 있음을 확인할 수 있다.

다음으로 A지점의 2012.11.27 ~ 2013.2.25 상수 수요량 자료를 통하여 추정된 앙상블 모형을 적용하여 2013.2.26 ~ 2013.2.27일 기간에 대한 반복 예측(iterated forecasting)을 실시

Table 1. Various test statistic of training data: temporal resolution for 10 minutes(multi-step forecasting)

| Test Statistic | MME | P1_10m | S1_10m | P2_10m |
|---|---|---|---|---|
| Weight | | 0.1809 | 0.3000 | 0.5191 |
| Mean error rate | 0.9106 | 1.624 | -0.4738 | 1.462 |
| Maximum error rate | 74.4514 | 79.7474 | 90.303 | 79.6789 |
| RMSE[1)] | 62.3662 | 62.6653 | 63.3141 | 62.458 |
| MAE[2)] | 33.9319 | 34.9729 | 33.0003 | 34.8197 |
| COE[3)] | 0.7072 | 0.7044 | 0.6982 | 0.7064 |
| Correlation coefficient | 0.8412 | 0.8394 | 0.8389 | 0.8405 |
| Objective function | 0.0283 | 0.0286 | 0.0292 | 0.0284 |

[1)] Root mean square error, [2)] Mean absolute error, [3)] Coefficient of efficiency

Table 2. Various test statistic of training data: temporal resolution for 60 minutes(multi-step forecasting)

| Test Statistic | MME | P1_60m | S1_60m | P2_60m |
|---|---|---|---|---|
| Weight | | 0.4118 | 0.4756 | 0.1126 |
| Mean error rate | 0.49 | 1.2774 | -0.4221 | 1.4642 |
| Maximum error rate | 72.7159 | 70.0528 | 71.2896 | 88.4866 |
| RMSE[1)] | 298.6445 | 300.7468 | 302.2418 | 304.4099 |
| MAE[2)] | 207.2702 | 214.3883 | 203.1563 | 215.2649 |
| COE[3)] | 0.734 | 0.7303 | 0.7276 | 0.7237 |
| Correlation coefficient | 0.8574 | 0.8546 | 0.856 | 0.8508 |
| Objective function | 0.025 | 0.0253 | 0.0256 | 0.0259 |

[1)] Root mean square error, [2)] Mean absolute error,[3)] Coefficient of efficiency

Table 4. Calculation formula of the test statistic

| Test Statistic | Formula |
|---|---|
| Mean error rate | $\frac{1}{N}\sum_{i=1}^{n}\frac{Y_{sim,i}-Y_{obs,i}}{Y_{obs,i}}\times 100$ |
| Maximum error rate | $\max_{i=1}^{n}\left\{\frac{Y_{\sim,i}-Y_{obs,i}}{Y_{obs,i}}\times 100\right\}$ |
| Root mean square error | $\sqrt{\frac{\sum_{i=1}^{n}(Y_{sim,i}-Y_{obs,i})^2}{N}}$ |
| Mean absolute error | $\frac{1}{N}\sum_{i=1}^{n}\frac{\lvert Y_{sim,i}-Y_{obs,i}\rvert}{Y_{obs,i}}$ |
| Coefficient of efficiency | $1-\frac{\sum_{i=1}^{n}(Y_{sim,i}-Y_{obs,i})^2}{\sum_{i=1}^{n}(Y_{obs,i}-\overline{Y})^2}$ |
| Correlation coefficient | $\frac{\sum_{i=1}^{n}(Y_{sim,i}-\overline{Y_{sim}})(Y_{obs,i}-\overline{Y_{obs}})}{\sqrt{\sum_{i=1}^{n}(Y_{sim,i}-\overline{Y_{sim}})^2}\sqrt{\sum_{i=1}^{n}(Y_{obs,i}-\overline{Y_{obs}})^2}}$ |
| Objective function | $\frac{\sum_{i=1}^{n}(Y_{sim,i}-Y_{obs,i})^2}{\sum_{i=1}^{n}(Y_{obs,i}-\overline{Y})^2}$ |

Table 5. Test statistic of iterated forecasting for each temporal resolution

| Test Statistic | 10m | 60m | 180m |
|---|---|---|---|
| Mean error rate | 5.9462 | 0.2726 | 0.3842 |
| Maximum error rate | 43.3800 | 22.9482 | 14.8838 |
| RMSE[1)] | 80.2242 | 284.8526 | 540.7623 |
| MAE[2)] | 57.1892 | 206.1847 | 417.0571 |
| COE[3)] | 0.4033 | 0.7092 | 0.8406 |
| Correlation coefficient | 0.6749 | 0.8432 | 0.9178 |

[1)] Root mean square error, [2)] Mean absolute error, [3)] Coefficient of efficiency

Table 3. Various test statistic of training data: temporal resolution for 180 minutes(multi-step forecasting)

| Test Statistic | MME | P1_180m | S1_180m | P2_180m | L1_180m | P3_180m |
|---|---|---|---|---|---|---|
| Weight | | 0.1385 | 0.6006 | 0.2190 | 0.0224 | 0.0195 |
| Mean error rate | 0.2901 | 0.3619 | 0.1223 | 0.6459 | 0.9004 | 0.2488 |
| Maximum error rate | 67.8487 | 65.7012 | 68.1863 | 68.3464 | 67.8776 | 67.0824 |
| RMSE[1)] | 555.492 | 559.9402 | 560.9723 | 562.1335 | 602.9809 | 565.3807 |
| MAE[2)] | 407.5306 | 419.3535 | 402.8858 | 422.5541 | 453.266 | 422.7365 |
| COE[3)] | 0.8616 | 0.8594 | 0.8589 | 0.8583 | 0.8369 | 0.8566 |
| Correlation coefficient | 0.9283 | 0.9271 | 0.927 | 0.9264 | 0.9149 | 0.9257 |
| Objective function | 0.021 | 0.0214 | 0.0215 | 0.0215 | 0.0248 | 0.0218 |

[1)] Root mean square error, [2)] Mean absolute error, [3)] Coefficient of efficiency

ISSN (Print) 1225-7672
ISSN (Online) 2287-822X
*Journal of Korean Society of Water and Wastewater* Vol. 28, No. 4, pp. 377-389, August, 2014

하고 실측치와의 비교를 통한 검증을 실시하였다. 반복예측 기법은 모형을 통해 예측된 예측결과를 다음 예측을 위한 독립변수로 이용하여 예측하는 기법으로 실제 수요량 예측에서 앙상블 모형이 적용되어질 때 이와 같은 Multistep 예측이 이루어진다.

Table 5는 10분, 60분, 180분 시간단위에 대한 2.26 ~ 2.27일에 수요량 예측결과와 실제 관측값을 비교하기 위한 여러 통계량을 보여주고 있으며, 평균 오차율에 경우 10분 자료에 경우 6%, 60분과 180분 자료에 경우 1% 미만에 평균 오차율이 나타났으며 최대 오차율에 경우 3개의 시간단위 모두 45% 미만의 최대 오차율을 보여주었다. 시간간격이 커질수록 최대 오차율이 크게 감소하였다. 모형의 효율성 계수인 COE에 경우 10분 자료에 경우 0.4, 60분과 180분에 경우 각각 0.71과 0.84로 나타났으며, 상관계수의 경우 10분, 60분, 180분 자료에 경우 각각 0.67, 0.84, 0.92에 상관계수를 보여주었다. 10분의 경우 상대적으로 낮은 상관계수를 나타내고 있는데 이는 Fig. 2와 같이 실측치가 갖는 큰 폭의 변동성을 예측값이 효과적으로 재현해 주지 못하는데 기인한다. 60분과 180분에 경우 0.7이상의 COE와 0.84이상의 상관계수, 20% 내외의 최대 오차율 및 1% 미만의 평균 오차율을 나타냄으로서 상대적으로 우수한 예측 결과를 보여주었다.

Fig. 2에서 Fig. 4는 각각 10분, 60분, 180분 자료에 대한 2월 26 ~ 27일에 대한 예측결과를 도시한 결과를 나타낸다. 여기서 실선은 실제 관측된 수요량을 의미하며, 점선은 앙상블 예측결과를 나타내며, 회색 영역은 개별모형들에 대한 최대, 최소값으로 구성된 불확실성 구간을 의미한다.

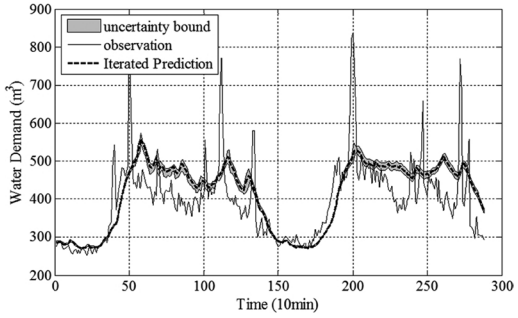


Fig. 2. Results of iterated forecasting: temporal resolution for 10 minutes.

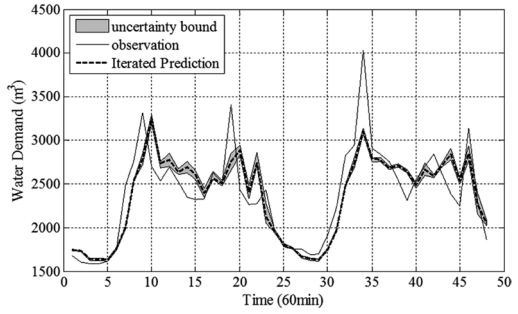


Fig. 3. Results of iterated forecasting: temporal resolution for 60 minutes.

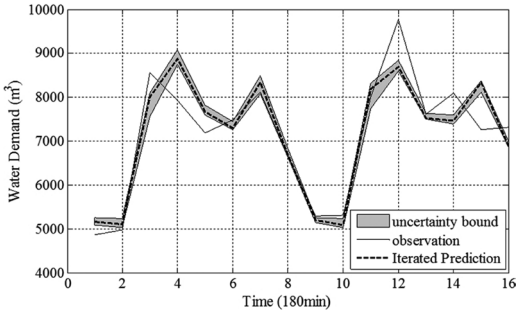


Fig. 4. Results of iterated forecasting: temporal resolution for 180 minutes.

## 4. 결 론

본 연구에서는 기 개발된 선형 및 비선형 예측모형들을 효율적으로 통합할 수 있는 상수도 단기앙상블예측기법을 적용하였다. 즉, 개별 모형들 사이의 가중치를 적용하여 최적의 예측결과를 도출할 수 있는 앙상블 모형을 개발하였으며 실제 상수도 수요량 자료에 대해서 적합성을 검

증하였다. 또한, 교차검증 기법을 적용하여 자료의 변화에 따른 신뢰성 있는 매개변수가 추정될 수 있도록 하였다.

다양한 통계량의 비교 결과 앙상블 모형을 통한 예측결과가 개별모형들에 개선됨에 따라서 예측의 정확성이 향상되었음을 확인할 수 있었다. 또한, 앙상블 모형을 구성하는 SVM의 가중치 증가는 시간간격이 커질수록 강하게 나타나는 수요량 자료의 수요패턴을 반영하는 결과로 판단 할 수 있다. 즉, 대표적인 패턴인식 기법인 SVM 모형이 시간단위가 커질수록 자료를 잘 적합시켜주고 있음을 확인 할 수 있었다. 또한, 자료의 시간단위가 180분으로 커짐에 따라서 앙상블 모형을 이루는 개별 모형의 수가 증가하고 있음을 확인할 수 있는데 이는 자료의 시간 간격이 증가함에 따라 상수도 수요량 자료의 패턴 및 주기성이 강해진 결과로 판단된다. 즉, 10분, 60분 자료에 대해서 상대적으로 예측성이 낮았던 선형모형 등도 수요패턴을 잘 모사하고 있음을 의미한다.

전반적으로 앙상블 모형은 개별모형들의 특징을 효과적으로 반영하여 개별 모형들보다 증진된 예측결과를 보여주고 있으며, 검정구간에 대한 앙상블 예측결과 10분, 60분, 180분 자료에 대하여 평균 오차율이 각각 6%, 0.3%, 0.4%로 7%이내의 뛰어난 예측력을 나타났으며, 최대 오차율에 경우 10분 자료를 이용하였을 경우 43%, 60분과 180분 자료에 경우 각각 23%, 15%로 모든 경우에서 50% 미만에 최대 오차율을 보여주었다.

결론적으로 본 연구에서 개발된 앙상블 모형은 예측 능력의 향상과 더불어 개별 모형의 추가에 따른 모형의 확장성과 불확실성 구간 정량화에 따른 예측 결과의 신뢰성을 제시할 수 있는 장점을 지니고 있다 하겠다. 추후 연구로서 다양한 수요 관측 자료를 통하여 앙상블 모형의 효율성을 점검하고 실제 관측지점에 대하여 앙상블 모형에 실시간 예측능력을 평가할 필요가 있으며, 보다 다양한 변량을 이용한 독립변수를 구성함과 동시에 다양한 개별예측 모형을 추가하여 확장하여 모형의 적응성을 높이는 것도 필요한 연구라고 판단된다.

## 사 사

본 연구는 환경부 “차세대 에코이노베이션 기술개발사업 (GT-11-G-02-001-5)”으로 지원받은 과제입니다.

## 참고문헌

Gu, J.Y. (1996) Seasonal Prediction Model for Urban Water Demand, *Journal of the Korean Society of Water and Wastewater.*, **23(6)**, pp.36-46.

Kwon, H.H. and Moon, Y.I. (2004) A Study of Short Term Forecasting of Daily Water Demand Using SSA, *Korean Society of Water and Wastewater.*, **18(6)**, pp.758-769.

Kwon, H.H. and Moon, Y.I. (2006) Dynamic Nonlinear Prediction Model of Univariate Hydrologic Time Series Using the Support Vector Machine and State-Space Model, *Journal of the Korean Society of Civil Engineers.*, **26(3)**, pp.279-289.

Yu, M.J., Gu, J.Y., Gu, Y.H. and Kim, S.G. (2004) Forecasting Hourly Water Demand Using Linear and Non-Linear Model, *Journal of Korean Society of Environmental Engineers.*, **26(3)**, pp.277-283.

Choi, G.S., Choi, Y.H., Choi, S.K., Lee, S.S. and Chun, M.G. (2009) Short-term Water Demand Forecasting with NN-TWI, *Journal of Korean Institute of Information Technology.*, **7(5)**, pp.9-16.

Adamowski, J. and Karapataki, C. (2010) Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms, *Journal of Hydrologic Engineering.*, **15(10)**, pp.729 - 743.

Alvisi, S., Franchini, M. and Marinelli, A. (2007) A short-term, attern-based model for water-demand forecasting, *Journal of*

*Hydroinformatics.*, **9(1)**, pp.39 – 50.

Aly, A. and Wanakule, N. (2004) Short-term forecasting for urban water consumption, *Journal of Water Resources Planning and Management-ASCE.*, **130(5)**, pp.405 – 410.

Bakker, B. and Heskes, T. (2003) Clustering ensembles of neural network models, *Neural networks.*, **16(2)**, pp.261-269.

Billings, B. and Jones, C. (2008) *Forecasting Urban Water Demand (2nd ed.).*, American Waterworks Association.

Bougadis, J., Adamowski, K. and Diduch, R. (2005) Short-term municipal water demand forecasting, *Hydrological Processes.*, **19(1)**, pp.137 – 148.

Breiman, L. (1996) Bagging Predictors, *Machine Learning.*, **24(2)**, pp.123 – 140.

Caiado, J. (2010) Performance of combined double seasonal univariate time series models for forecasting water demand, *Journal of Hydrologic Engineering.*, **15(3)**, pp.215 – 222.

Casdagli, M. (1992) Chaos and deterministic versus stochastic non-linear modelling, *Journal of the Royal Statistical Society. Series B (Methodological).*, **54(2)**, pp.303-328.

Casdagli, M. and Weigend, A.S. (1994) Exploring the continuum between deterministic and stochastic modelling, *Time Series Prediction: Forecasting the Future and Understanding the Past.*, pp.347-366.

Cohen, S. and Intrator, N. (2001) Automatic model selection in a hybrid perceptron/radial network, *In Multiple Classifier Systems.*, Springer Berlin Heidelberg, pp.440-454.

Cutore, P., Campisano, A., Kapelan, Z., Modica, C. and Savic, D. (2008) Probabilistic prediction of urban water consumption using the SCEM-UA algorithm, *Urban Water Journal.*, **5(2)**, pp.125 – 132.

Gardiner, V. and Herrington, P. (1990) *Water Demand Forecasting (1st ed.).*, Spon Press.

Gato, S., Jayasuriya, N. and Roberts, P. (2007) Forecasting residential water demand: Case study, *Journal of Water Resources Planning and Management-ASCE.*, **133(4)**, pp.309 – 319.

Geman, S., Bienenstock, E. and Doursat, R. (1992) Neural networks and the bias/variance dilemma, *Neural Computation.*, **4**, pp.1-58.

Ghiassi, M., Zimbra, D. and Saidane, H. (2008) Urban water demand forecasting with a dynamic artificial neural network model, *Journal of Water Resources Planning and Management-ASCE.*, **134(2)**, pp.138 – 146.

Herrera, M., Torgo, L., Izquierdo, J. and Perez-Garcia, R. (2010) Predictive models for forecasting hourly urban water demand, *Journal of Hydrology.*, **387(1-2)**, pp.141 – 150.

Jain, A. and Ormsbee, L. (2001) A decision support system for drought characterization and management, *Civil Engineering and Environmental Systems.*, **18(2)**, pp.105 – 140.

Jain, A. and Ormsbee, L. (2002) Short-term water demand forecast modeling techniques -Conventional methods versus AI, *Journal American Water Works Association.*, **94(7)**, pp.64 – 72.

Jain, A., Varshney, A. and Joshi, U. (2001) Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks, *Water Resources Management.*, **15(5)**, pp.299 – 321.

Jentgen, L., Kidder, H., Hill, R. and Conrad, S. (2007) Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations, *Journal American Water Works Association.*, **99(6)**, pp.86 – 94.

Krogh, A. and Sollich, P. (1997) Statistical mechanics of ensemble learning, *Physical Review E.*, **55(1)**, pp.811.

Krogh, A. and Vedelsby, J. (1995) Neural network ensembles, cross validation, and active learning, *Advances in neural information processing systems.*, **7**, pp.231-238.

Naftaly, U., Intrator, N. and Horn, D. (1997) Optimal ensemble averaging of neural networks, *Network: Computation in Neural Systems.*, **8(3)**, pp.283-296.