

Estimating Non-revenue Water Ratio Using ANN based on PCA with Data Normalization in Water Distribution Systems

Dong Woo Jang⁺

Department of Civil and Environmental Engineering, Incheon National University, 119 Academy-ro Yeonsu-gu, Incheon, Korea

Abstract

The non-revenue water (NRW) ratio in water distribution systems is an index of the loss of water supply caused by pipe burst, operational loss and physical factors. NRW ratio is a comparative index for city, province and DMA (district metering area) in the domestic water supply maintenance project. An investigation of the factors affecting the NRW as well as its estimation have become increasingly important in an economic sense. In this study, PCA (principal component analysis) and ANN (artificial neural network) are used as statistical methods to estimate the NRW ratio. The normalized data were obtained through the Z-score method, and then the PCA-ANN model was constructed for the NRW ratio estimation. Accuracy assessment was performed to compare the observed NRW ratio with the estimated ratio from the ANN model. The results show that the PCA-ANN model is more accurate than the single ANN and the estimation results differ by the number of neurons in the hidden layer of ANN. As for the six independent variables used in this study, the accuracy of NRW ratio prediction was found highest when 12 neurons were used.

Key words: artificial neural network, non-revenue water ratio, principal component analysis, data normalization, water distribution systems

1. 서론

상수도 내 관망해석은 정수장, 가압장, 배수지, 취수장, 송·배수관로 등 상수도를 구성하는 시설물의 용량 및 운영계획을 수립하기 위해 반드시 수행되어야 하는 절차 중 하나이다. 특히, 안정적인 상수도 공급을 위하여 장래 예측수요량을 기반으로 한 관망해석 모의결과는 관련 실무자와 정책결정자에게 상수도 운영과 시설 확충 계획 결정 시 이용되고 있다(Choi, *et. al.*, 2006).

기존에 수행된 상수관망 연구로는 최적 용수 분배를 위한 관망 모의, 관로 내 신뢰도 향상, 노후도 평가, 관망의 기술 진단에 대한 연구가 우선적으로 수행되었다. 배수시설의 경우 합리적인 관로배치계획과 시간적으로 변동하는 수요량에 대하여 적절한 압력으로 연속적이며 안정적인 공급이 요구된다(Jun, *et. al.*, 2008). 관로 파괴와 같은 사고가 발생했을 경우, 소비자의 피해를 최소화해야 하며 수도공급시설은 높은 안정성을 목표로 구축되어야 한다. 상수관망을 해석하기 위한 연구

⁺ Corresponding author: Dong Woo Jang, Tel. +82-32-835-4760, e-mail, nightray@paran.com

로는 물분배 시스템의 최적 개량계획 수립을 위한 의사결정시스템 구축(Kim, 2004; Goncalves, *et. al.*, 2010), 관로 내 유향추적기법(Koo, 2006), 상수공급 신뢰도를 향상하기 위해 관망의 내구성 향상과 구조개선(Surendran, *et. al.*, 2005; Kim, *et. al.*, 2007) 등의 연구가 활발히 진행되었다. 상수관망의 기술발전을 위해서는 수리, 수질모형에 대한 종합평가기술이 필요하며(Park, 2008), 설계, 운영, 유지관리시스템을 위해 관망진단을 위한 지표, 항목, 조사방법 및 절차의 표준화가 중요하다(Woo, 2008).

상수관로 개량 우선순위의 결정에 있어 관 파손이 전체 관망에 미치는 영향 분석(Park, *et. al.*, 2011)과 광역상수도의 유지관리를 위하여 체계적인 교체 및 보수 계획이 정기적으로 수립되고 있다(Park, 2014). 노후 상수도관의 개량사업은 지속적으로 시행되고 있지만 노후관로 개량사업은 경험적 판단에 의존하는 노후관 평가 및 대안의 선정, 사고 예방을 위한 대응적 차원의 개량 사업을 실시함으로써 인해 경제적 손실은 물론 시스템의 유기적 기능 향상이 이루어지지 않고 있다(Park, *et. al.*, 2006).

따라서 관로 개량사업 및 수도공급시설계획, 관망 유지관리를 위한 지역, 블록별 비교지표가 필요하며 무수율은 전 세계적으로 관망의 물리 및 운영지표로서 사용되고 있다. 무수율(Non-revenue Water ratio)은 급수량(정수장에서 생산해 공급한 수돗물) 중에서 요금수입으로 받아들여지지 않은 수량의 비율로서 관망 내 관로의 파열로 인한 누수 및 불법적인 연결, 측정되지 않은 공공 사용 및 계량기 오류로 인한 물 관련 상업 손실과 같은 물리적 물 손실량을 포함한다(Choi, *et. al.*, 2006). 이러한 무수율은 주요 국가의 경우 5~50%를 보이는데, 칠레(34%)와 멕시코(51%)는 무수율의 비율이 가장 높은 국가 중의 하나이고, 싱가포르, 덴마크, 네덜란드(5~6%)는 무수율이 가장 낮은 국가이다(Wikipedia, 2017). 우리나라의 경우, 상수도통계(2015년)에 따르면 주요 도시의 무수율은 서울이 4.9%로 가장 낮았고 광주가 56.8%로 가장 높았다. 인천은 11.2%로써 전국평균

이 16.3% 보다는 낮은 경향을 보였다(Waterworks Headquarters Incheon Metropolitan City, 2015).

수돗물 공급과정에서 발생하는 누수는 상수도 경영상의 악화요인으로 작용하고 있으며 무수율에 가장 큰 영향을 미치는 주요 인자이다. 상수관망 정비사업의 우선순위 결정에 대한 의사결정 시 누수 영향 인자를 조사하여 수량, 수압, 수질 요소와 함께 시설개량 비용, 에너지 비용, 유지관리 비용 등의 경제성 인자를 고려한 과학적이며 체계적인 접근방법이 필요하다. 또한, 상수관망 내 누수에 영향을 주는 물리적, 운영적 요소를 파악하고, 이를 이용하여 누수 해결을 위한 방법론을 제시하는 것은 체계적인 관망 운영관리에 있어 필수적인 요소이다(Jo, *et. al.*, 2016).

무수율에 영향을 주는 인자로서 관로 내의 누수와 파열이 가장 큰 영향인자로 여겨지고 있고, 수도공급계획이 있어 수도요금을 산정하는 데에도 영향을 준다(Shilehwa, 2013). Winarni(2009)는 수도공급시스템 내에 시설관리를 위한 운영지표로서 누수지수를 제시하였고, 경제적 손실을 줄이기 위한 누수량, 관로 갱생에 대한 연구도 이루어졌다(Wyatt, 2010). 특히 상수관망내 누수에 영향을 미치는 주요 인자에 관한 연구도 지속적으로 수행되었다(Alegre, *et. al.*, 2000; Lambert, *et. al.*, 1999, Shinde, *et. al.*, 2013).

무수율을 추정하기 위한 방법으로 상수관망 내 물리 요소와 운영요소를 독립변수로 사용하고, 무수율을 종속변수로 하는 다중회귀식을 이용한 방법이 보편적으로 사용되어 왔다(Jo, *et. al.*, 2016, Jo, 2017). 무수율을 추정지역 내에 개발된 다중회귀식이 있을 경우 무수율을 추정하기가 쉽다는 장점이 있지만, 예측 정확도가 낮고, 인자 간의 상호 관련성이 고려하지 못하고 계산된다는 단점도 존재한다.

다양한 통계적인 분석기법 중 인공신경망(Artificial Neural Network)은 독립 변수와 종속 변수와의 인과관계를 학습하고, 통계학적 알고리즘을 통하여 종속변수를 추정하는 모형이다. 모의값과 목표값의 오차를 최소화시키는 방향으로 각 연결함수의 가중치가 조절되

는데 인공지능망은 실측값에 근접한 추정결과를 제시한다는 장점이 있다(National Institute of Environmental, 2011).

인공지능망은 해석에 필요한 충분한 독립 변수 데이터가 종속 변수를 설명할 수 있을 때 사용할 수 있는 최적의 통계적 학습 알고리즘이지만 학습 데이터가 충분히 존재하지 않을 때에는 모의결과를 신뢰할 수 없다는 단점이 있다. 따라서 충분한 관망 요소자료 및 운영자료가 구축되어 있을 경우 인공지능망의 적용이 가능하다. 데이터 취득 및 지역별 주요 영향인자 선정의 어려움으로 인공지능망을 이용한 상수관망의 무수율 추정 연구는 활발히 이루어지지 않고 있고, 무수율 추정에 활용되었던 다중회귀분석 보다 인공지능망을 이용하는 방법이 추정 정확도가 높은 것으로 알려져 있다(Jang and Choi, 2017).

본 연구에서는 주성분분석과 인공지능망을 이용하여 무수율을 추정하는 연구를 수행하였다. 기존 다중회귀분석에 비하여 실측값에 보다 정확한 추정결과를 제시하는 인공지능망의 무수율 추정 정확도를 평가하고, 실제 적용 가능성을 판단하였다. 연구를 수행하기 위하여 상수관망의 블록구축이 이루어져 있고, 인공지능망을 충분히 학습시킬 수 있게 할 수 있는 소블록 자료가 충분한 도시를 대상지역으로 선정하였다.

선정된 대상지역에 대하여 인공지능망의 독립변수로 사용하기 위한 상수관망 내 주요 인자의 소블록 자료

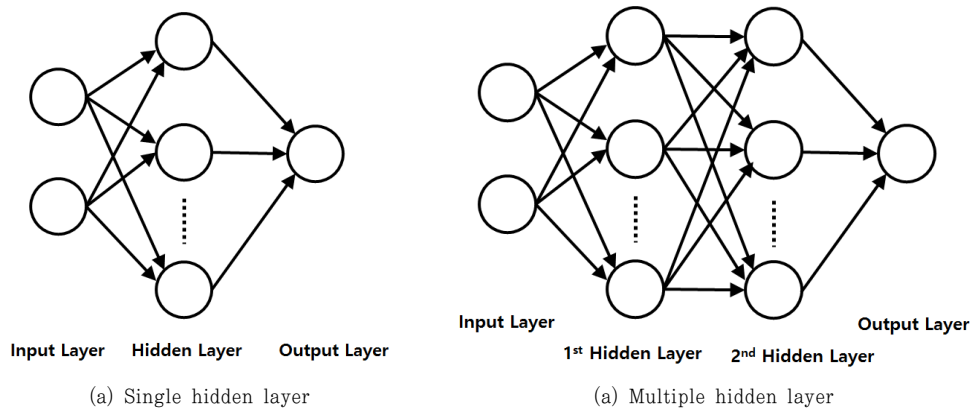
를 구축하고, 무수율 예측을 위한 인공지능망 모형을 구축, 모의하였다. 또한, 인공지능망의 결과예측 정확도를 높이기 위하여 데이터 정규화(Z-score), 주성분분석(Principal components analysis)을 통하여 수집 데이터를 정규화한 후 인공지능망에 적용하였다.

II. 이론적 접근 방법

상수관망 내 무수율 추정을 위한 통계분석기법으로 인공지능망이 사용하였고, 대상지역의 주요 상수관망 영향인자자료를 수집 후 기초 통계분석 절차로 Z-score와 주성분 분석을 적용하였다. 또한, 인공지능망 분석을 통하여 산출된 모의결과를 실측결과와 비교하기 위하여 정확도 평가를 수행하였다. 2장에서는 사용된 통계기법에 대한 이론과 연구체계를 나타내고 있다.

1. 무수율 추정을 위한 인공지능망 이론

인공지능망의 해석절차는 <Figure 1>과 같이 입력층, 은닉층, 출력층으로 이루어져 있고, 오류역전파알고리즘(Back-propagation algorithm)을 통하여 학습과 출력이 이루어진다. 오류역전파알고리즘은 신경회로망의 입력패턴에 따라 노드의 입력값과 활성화 함수를 통한 출력값을 산출하는 순방향(feed forward)방식과 목표값과 출력값의 차이에 대한 오차를 최소화하려는 역방향(feed back)처리로 이루어져 있다(Lee, 2010). 인공지능 해석을 위하여 입력층과 출력층 사이에 두 개



※ Source: Neural Networks: A Comprehensive Foundation (Haykin, 1994)
Figure 1. Schematic diagram of multi-layer feed forward neural network

이상의 은닉층으로 이루어진 다층퍼셉트론은 오류 역전과 알고리즘을 통해 네트워크 해석능력이 향상된다. 이것은 은닉층이 한 개만 존재하는 단층 신경망의 단점을 보완할 수 있다(Pal and Mitra, 1992; Kim, 2007).

은닉층과 출력층의 뉴런을 시그모이드 활성화 함수라고 하며, 각 은닉층의 뉴런 j 에서 가중치된 입력신호의 합인 x_j ($\sum_{i=1}^n w_{ji}x_i = net_j$, 여기서 w_{ji} 는 입력층의 뉴런 i 와 은닉층의 뉴런 j 사이의 상호 연결된 가중치(연결 강도)이다), 각각의 뉴런으로부터 출력되는 결과 y_j 는 시그모이드 함수인 식 (1)과 같이 나타낼 수 있다(Haykin, 1994).

$$y_j = f\left(\sum_{i=1}^n w_{ji}x_i\right) = \frac{1}{(1 + e^{-net_j})} \quad (1)$$

오류 역전과 알고리즘에 의한 학습은 신경망의 전 방향으로 계산된 값과 학습을 위한 목표값을 비교하여 오차의 제곱 합이 최소로 할 수 있도록 연결 강도를 조정하게 되며 오차가 일정값을 만족시킬 때까지 수행되고, 반복을 마치면 최종값을 출력하게 된다.

2. 주성분 분석을 통한 데이터 변환

주성분 분석은 고차원의 정보를 유지하면서 차원축소를 통하여 데이터의 특징을 밝히거나 변수의 관계를 규명하는 다변량 분석기법이다(Moon & Phillips, 2001). 여러 변수들(x_1, x_2, \dots, x_n)을 상관성이 높은 공동요인으로 묶어 기존 변수들이 가지고 있는 정보를 최대한 확보하면서 적은 수의 새로운 변수를 생성하는 차원 축소방법 중의 하나이다.

'0'에 수렴하는 산술평균 X 라는 데이터 행렬에 대하여 데이터의 특성을 나타내는 p 열과 다른 반복 케이스를 나타내는 n 행을 고려할 수 있다. 수학적으로 p 열벡터의 가중치 세트 또는 하중 ($w_{(k)} = (w_1, \dots, w_p)_{(k)}$), 행방향 벡터($x_{(i)}$)가 새로운 주성분 벡터($t_{(i)} = (t_1, \dots, t_m)_{(i)}$)로 변환되는 것으로 설명할 수 있고, 이것을 정리하면 식 (2)와 같다(Lee, 2010; Jang, 2017).

$$t_{k(i)} = x_{(i)} \cdot w_{(k)} \quad (i=1, \dots, n, \quad k=1, \dots, m) \quad (2)$$

데이터 집합에 대해 t 의 개별 변수는 x 에서 발생할 수 있는 최대 분산을 연속적으로 상속하며, 각 로딩 벡터 w 는 단위 벡터로 제한된다.

3. Z-score를 이용한 데이터 정규화

정규화의 일반적인 의미는 기업, 사용자, 이익 단체, 표준 설정 조직 및 정부를 구성하는 당사자의 합의를 토대로 기술 표준을 구현 및 개발하는 프로세스이다(Xie, et. al., 2016). 통계분야에서 정규화(Z-score)는 관측치 또는 데이터 포인트의 값이 관측되거나 측정된 평균값보다 큰 표준편차의 값으로 부호가 있는 수라고 할 수 있다(Jang, 2017).

분석하고자 하는 데이터에 대하여 관측값의 평균 이상으로 변환된 값은 양의 표준 점수를 가지며, 평균보다 낮은 값은 음의 표준 점수를 갖는다. 표준 점수는 개별 원시 점수에서 모집단 평균을 뺀 다음 모집단 표준 편차로 그 차이를 나눈 값이다. 이 변환 과정을 표준화 또는 정규화라고 한다(Wikipedia, 2017).

Z-score 방법은 결과 분석을 수행할 때 사용된 데이터의 차이와 분포를 구별하는 데 사용할 수 있다. 평균 및 표준 편차는 표준 편차가 단위로 취해질 때 데이터가 평균으로부터 얼마나 벗어나는지를 결정하는 데 사용된다. Z-score의 계산 방법은 식 (3)과 같이 표시할 수 있다(Kreyszig, 1979).

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

여기에서, μ 는 모집단의 평균이고, σ 는 모집단의 표준편차이다.

상수관망 시스템의 매개 변수의 경우 사용되는 단위 수와 값의 범위는 블록(District Metered Area)규모에 따라 다르다. 따라서 Z-score는 사용된 매개 변수의 범위를 통합하고 이들 간의 비교를 용이하게 하는 이점이

있다. 이는 주성분 분석(PCA)을 수행하기 전에 필요한 통계적 절차이기도 하다.

4. 추정결과의 정확도 검증

인공신경망에 의해 산출된 무수율 모의값의 정확성을 평가하기 위하여 실제값과 모의값과의 차이를 평가하는 오차율 분석을 수행하였다. 예측값의 정확성을 평가하는 평균절대오차(Mean Absolute Error) 및 평균제곱오차(Mean Squared Error), 추정결과의 편향 정도를 평가하는 평균편향오차(Mean BIAS Error)와 예측의 효율도(Effectiveness)를 평가하는 G값(Goodness of prediction)을 이용하였고, 각 평가인자를 계산하는 방법은 다음의 식(4)~(7)과 같다. 실제값과 모의값은 회귀분석을 통해 추가로 비교할 수 있다.

$$MAE = \frac{1}{n} \sum_{i=1}^n [|z(x_i) - \hat{z}(x_i)|] \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n [z(x_i) - \hat{z}(x_i)]^2 \quad (5)$$

$$BIAS = \frac{1}{n} \sum_{i=1}^n (z(x_i) - \hat{z}(x_i)) \quad (6)$$

$$G = \left(1 - \frac{\sum_{i=1}^n [z(x_i) - \hat{z}(x_i)]^2}{\sum_{i=1}^n [z(x_i) - \bar{z}]^2}\right) \times 100 \quad (7)$$

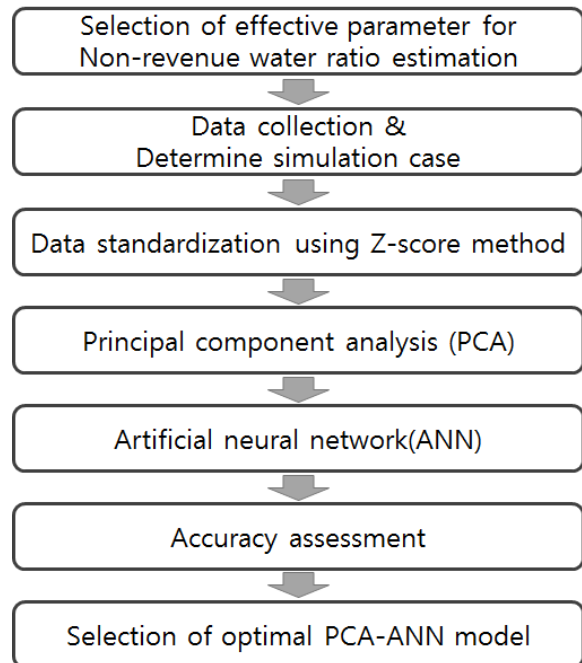
여기에서, $\hat{z}(x_i)$ 는 위치 i 에서 추정된 값, \bar{z} 은 조사 자료의 평균이다. 평균절대오차(MAE), 평균제곱오차(MSE)는 계산된 결과 값이 작을수록 추정된 값이 정확하다는 것을 의미하며, 평균편향오차(BIAS)는 0 값에 가까울수록 추정결과가 편향되지 않은 것으로 판단한다. 또한, G값이 '100'일 경우 완벽한 추정이라고 여겨지며, G값이 음수인 경우에는 추정자(Predictor)로써 사용된 자료 값의 평균을 사용하는 것보다 신뢰성이 없다는 것을 나타낸다(Kim, *et. al.*, 2010).

5. 연구 방법

무수율 추정을 위하여 대상지역 선정, 관련자료 수집, 기초통계분석, 인공신경망 시뮬레이션, 정확도 평가의 순으로 연구를 수행하였다. 기초통계분석으로 데이터 정규화, 주성분 분석을 수행하였고, 무수율 추정을 위한 인공신경망의 경우 은닉층을 단층과 다층레이어로 구분하여 적용하였다. 상수관망 주요 인자 데이터를 토대로 학습된 인공신경망 모형을 통한 무수율 추정 결과는 실측 무수율과 비교하여 최적 인공신경망 모형을 선정하였다. 주요 연구체계는 <Figure 2>와 같다.

III. 적용 대상

인공신경망을 이용한 무수율 산정기법을 적용하기 위해서 상수관망내 블록시스템(District Metered Area)이 구축되어있고, 인공신경망을 충분한 학습시키기 위한 데이터 확보가 가능한 지역을 선정하였다. Jo, *et. al.*(2016)은 인천시 소블록 데이터를 수집하여 다중회귀분석을 통한 무수율 산출 공식을 제안하였다. 본



※ Source: Estimation of Leakage Ratio using Principal Component Analysis and Artificial Neural Network in Water Distribution Systems(Jang, *et. al.*, 2018)

Figure 2. Research flowchart for Non-revenue water ratio estimation using PCA-ANN

연구에서는 비교적로 블록시스템이 잘 구축되어있고, 무수율의 통계분석 연구 사례가 있는 인천시를 대상으로 인공신경망을 적용하였다. 또한, 기존 Jo(2017)의 연구에서 추가로 인천시 소블록 자료를 수집하여 총 164개 소블록에 대한 무수율 영향인자 데이터를 활용하여 인공신경망 및 기본통계분석을 하였다.

자료수집은 인천시 행정구역 중 섬 지역인 강화군과 옹진군을 제외한 도시지역을 대상으로 하였다. 본 연구에 필요한 사항으로 지역 현황, 상수도 시설 및 운영 현황, 2015년에 수립된 인천 수도정비기본계획의 급수 지표를 조사하였다. 또한 배수관망 해석을 위한 자료구축과 관망해석 모의를 통하여 상수관망 주요 인자에 대한 정보를 수집하였다.

1. 대상지역의 상수도 현황

인천의 총인구는 2,851,491명이고, 급수보급률은 98.3%이다(Waterworks Headquarters Incheon Metropolitan City, 2015). 인천시의 1인 1일 급수량은 343 LPCD 이고, 급수구역은 9개의 정수 권역과 급수권역으로 나누어져 있다.

인천시는 총 24개소의 배수지가 있으며 유인배수지는 1개소, 무인배수지는 23개소이다(2011년 기준). 1940년 부평가압장의 설치를 시작으로 1990년대에는 50여개소의 무인가압장이 설치되었으며 2011년 도화 가압장이 폐쇄되어 현재 무인 68개소의 가압장이 운영되고 있다. 상수관망의 경우 배수관로는 3,634km, 급수관로는 1,925km가 매설되어 있다.

인천시는 전체 급수지역을 정수장에서 직접 공급하는 방식 보다는 배수지를 통하여 각각의 단위 급수구역으로 세분화하여 간접 배수방식을 이용하는 블록시스템 송·배수방식으로 수도공급을 하고 있다. 단위배수지 급수구역도 다수의 소급수구역으로 분할하는 구역개량 사업이 진행 중이다. 인천광역시 상수관망의 블록구축 현황은 정수장권역의 6개 대블록, 배수지권역의 32개 중블록과 367개 소블록으로 구성되어 있다(Waterworks Headquarters Incheon Metropolitan City, 2015).

2. 무수율 추정인자의 선정

2015년 수립한 인천시 상수도 관망 기술진단 결과 관로노후도 인자가 인천시 소블록의 무수율과 상관도가 높은 것으로 나타났다(Waterworks Headquarters Incheon Metropolitan City, 2015). 또한, 기존 연구에서 제시한 무수율 영향 인자는 배수관연장, 평균관경, 급수전수, 급수량, 누수발생 횟수, 용도지역, 관망구성 형태와 잔류수압에 따른 에너지 항목도 포함되었다(Jo, 2017; Jang, 2017). 급수에너지비는 절점에서 수요량을 공급하기 데 필요한 공급에너지와 수용가에서 필요로 하는 적정수압에너지의 비를 나타내며 급수 에너지비가 클수록 수요압력보다 공급압력이 높기 때문에 관망 내 누수량이 증가할 가능성이 높다(Jo, 2017).

본 연구에서 무수율 추정을 위하여 선정된 인자는 관로노후도, 급수전당 공급량, 급수전당 배관연장, 평균관경, 누수발생횟수, 급수에너지비이며 기존 상수관망 주요 영향인자 선정연구의 일환으로 통계분석을 통하여 선정된 인자이다(Jang, 2017).

3. 방법 적용 데이터

총 6개의 무수율 영향인자를 본 연구에 적용하였고, 2015년에 수립된 인천시 상수관망 기술진단 보고서 중 소블록별 자료를 사용하였다. 인천시 전체 367개의 소블록 중 미구축 블록, 블록 해제 운영 블록, 비정상 운영 블록, 가압장 블록, 미계측 블록을 제외한 총 164개의 소블록 데이터를 이용하여 연구를 수행하였다. 연구에 사용된 무수율과 6개 인자의 소블록별 데이터 현황은 <Figure 3>과 같다.

인공신경망 분석을 위하여 무수율은 종속변수로서 사용되었고, 독립변수로는 평균관경, 급수전당공급량, 급수전당 관로길이, 급수에너지비, 관로노후도, 누수발생횟수 인자가 사용되었다. 특히 독립변수로 사용된 6개 인자는 주성분분석을 통하여 6개의 주성분인자 데이터로 재생성하여 인공신경망에 적용하였고, 원 데이터와 주성분 데이터로 구분하여 모의하였다. 또한 실측

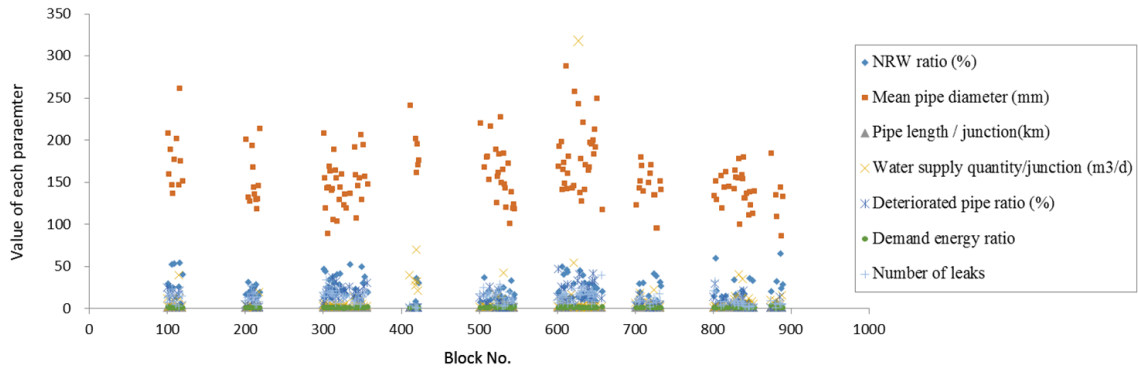


Figure 3. Data status of Non-revenue water ratio and main parameters in district metered area of Incheon

무수율은 인공신경망에 의하여 모의된 무수율과 정확도 평가를 위해 사용되었다.

IV. 적용 및 결과 분석

1. 방법 적용

2장의 <Figure 2>에서는 상수관망 주요인자를 이용하여 무수율을 추정하기 위한 통계분석 절차를 제시하였다. 방법론을 적용하기 위하여 소블릭 데이터 취득이 가능한 대상지역을 인천시로 선정하였고, <Figure 3>과 같이 무수율을 비롯한 7개의 인자데이터를 수집하였다. 인공신경망을 이용한 무수율 추정을 위하여 실측 무수율을 독립변수로, 6개의 상수관망 주요인자를 독립변수로 하는 인공신경망 모형을 구축하였다. 또한 사용된 독립변수의 주성분분석을 통한 변환 데이터를 인공신경망 모형에 적용하여 무수율 예측 정확도를 높이고자 하였다.

4장에서는 주성분분석과 인공신경망 레이어 수에 따른 무수율 예측정확도를 판별하고자 3가지의 인공신경망 시뮬레이션 조건을 제시하였다. 이에 따라 Z-score를 이용한 데이터 정규화, 주성분분석을 통한 데이터 변환과 인공신경망 모의조건에 대한 무수율 추정 결과를 제시하였다. 또한 인공신경망을 통하여 모의된 무수율 추정 결과를 실측 무수율과 비교함으로써 최적 인공신경망 시뮬레이션 조건을 결정하였다.

2. 시뮬레이션 모형 구축

<Table 1>은 무수율 추정을 위한 인공신경망 모의 조건이다. 모의조건의 분류기준은 주성분 분석 여부와 인공신경망의 은닉층 개수, 은닉층 내의 뉴런수로 3가지 분류기준으로 시뮬레이션을 구분하였다. 모의 조건은 PCA와 ANN을 이용한 누수율 추정연구의 모의 조건(Jang, *et. al.*, 2018)과 동일하게 설정하여 시뮬레이션을 하였다.

주성분 분석 여부에 따른 무수율 산출 정확도의 비교 및 은닉층의 개수, 은닉층내의 사용된 뉴런수에 따라 무수율 추정 정확도가 다르기 때문에 이와 같은 모의조건을 선정하였다. 사용뉴런의 수는 Heaton(2005)이 제시한 독립변수의 두 배 이하에서 최적 시뮬레이션 결과를 제시한다는 연구를 토대로 사용된 독립변수와 동일한 6개의 뉴런, 그리고 독립변수의 두 배에 해당하는 12개의 뉴런을 적용하는 조건으로 구성하였다.

은닉층이 2개인 조건에서는 뉴런수를 첫 번째 은닉층에서 6개 12개인 조건, 두 번째 은닉층에서도 동일하게 6개, 12개로 구분하여 뉴런수에 따른 무수율 추정정확도를 산출할 수 있도록 하였다. 주성분인자 적용조건에서는 주성분 분석을 통하여 생성된 전체 6개 주성분을 사용하는 조건과 85%의 고유값을 나타내는 4개 주성분을 사용하는 조건으로 구분하였다.

인공신경망 모형구축 시 전체 데이터의 80%는 학습을 20%는 검증하도록 하여 구축된 신경망을 모의하도록 하였다. 활성화 함수는 쌍곡 탄젠트 함수를 이용하였으며 최적화 알고리즘으로는 켈레기울기법을 사용하

Table 1. Modeling simulation cases

Classification	Hidden layer of ANN	Number of neurons	Modeling case
6 Original parameters usage case	Single layer	6	Case 1-1-1
		12	Case 1-1-2
	Multiple layer	12 (6 and 6)	Case 1-2-1
		18 (6 and 12)	Case 1-2-2
		18 (12 and 6)	Case 1-2-3
		24 (12 and 12)	Case 1-2-4
6 Principal component parameters usage case	Single layer	6	Case 2-1-1
		12	Case 2-1-2
	Multiple layer	12 (6 and 6)	Case 2-2-1
		18 (6 and 12)	Case 2-2-2
		18 (12 and 6)	Case 2-2-3
		24 (12 and 12)	Case 2-2-4
4 Principal component parameters usage case	Single layer	6	Case 3-1-1
		12	Case 3-1-2
	Multiple layer	12 (6 and 6)	Case 3-2-1
		18 (6 and 12)	Case 3-2-2
		18 (12 and 6)	Case 3-2-3
		24 (12 and 12)	Case 3-2-4

※ Source: Estimation of Leakage Ratio using Principal Component Analysis and Artificial Neural Network in Water Distribution Systems (Jang, *et. al.*, 2018)

였다. 학습오차의 최소화량은 0.0001, 학습오차의 최소상대변화량은 0.001로 하여 오차를 최소화할 때까지 반복 계산하여 결과를 산출하도록 하였다.

3. 인공신경망을 이용한 무수율 모의 결과

1) Z-score 분석 결과

주성분 분석과 인공신경망 적용 시 사용데이터의 통계적 적합성을 검토하기 위하여 Z-score를 이용한 정

규화된 데이터를 생산하였다. 전체 164개 소블록의 각 인자 데이터를 이용한 정규화 분석을 수행하였고, 그 결과는 <Figure 4>와 같다.

분석결과 총 164개의 소블록 중 11개 소블록에서 사용된 7개 인자의 Z-score 값이 ±3 이상을 나타냈고, 886번 소블록의 경우 무수율과 급수에너지비 두 항목이 Z-score 값을 초과하였다. 따라서 총 10개의 소블록이 Z-score 기준을 만족하지 않는 것으로 나타났다.

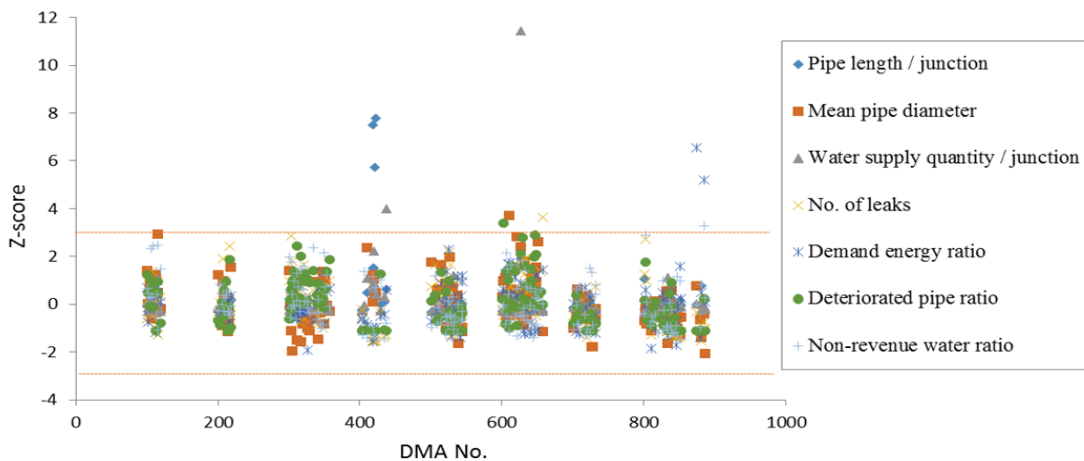


Figure 4. Data conversion by Z-score in each district metered area of Incheon

〈Figure 4〉에서 보는바와 같이 Z-score 기준을 초과하는 소블럭은 400번대와 900번대 블록에서 주로 발생하였고, 3 이상의 값을 나타내고 있다.

Z-score를 만족하지 않는 블록의 수는 급수전당 관로길이가 3개 블록, 평균관경이 1개 블록, 급수전당 공급량이 2개 블록, 누수발생횟수가 1개 블록, 급수에너지비가 2개 블록, 관로노후도가 1개 블록, 무수율이 1개 블록으로 나타났다. Z-score의 특정범위를 초과하는 데이터는 통계분석 시 이상치로 간주되어 전체 분석결과에 영향을 끼칠 수 있으므로 기준값을 초과하는 값을 제거한 후 시뮬레이션을 수행하였다.

2) 주성분 분석 결과

무수율 주요 영향인자로 선별된 관로노후도, 급수전당 공급량, 급수전당 배관연장, 평균관경, 누수발생횟수, 급수에너지비 인자에 대하여 각 선정된 소블록 데이터를 이용하여 주성분 분석을 수행하였다.

〈Table 2〉는 사용된 6개 인자 데이터에 대한 주성분 분석 결과이다. 주성분 분석을 통하여 가장 많이 생성된 성분 1은 전체 6개 인자 특성에 38.8%를 나타내고 있고, 성분 2는 22.6%를 나타내고 있다. 성분 1과 성분 2가 전체 인자의 61.5%를 담당하고 있으며 성분 1~4가 전체 87.7%의 고유값(eigenvalue)을 가지고 있는 것으로 나타났다 (Jang, et. al., 2018).

선형대수학에서, 선형 변환의 고유벡터(eigenvector)는 그 선형 변환이 일어난 후에도 방향이 변하지 않는, 영벡터가 아닌 벡터이다. 고유 벡터의 길이가 변하는 배

Table 2. Results of principal component analysis (PCA)

Component	Initial eigenvalues		
	Total	Variance (%)	Cumulative (%)
1	2,332	38,868	38.9
2	1,359	22,643	61.5
3	0,885	14,747	76.3
4	0,685	11,413	87.7
5	0,471	7,846	95.6
6	0,269	4,483	100.0

※ Source: Estimation of Leakage Ratio using Principal Component Analysis and Artificial Neural Network in Water Distribution Systems (Jang, et. al., 2018)

수를 선형 변환의 그 고유 벡터에 대응하는 고유값이라고 하는데 선형 변환은 대개 고유 벡터와 그 고유값만으로 완전히 설명할 수 있다고 한다(Wikipedia, 2017).

〈Table 3〉은 주성분 분석을 통해서 생성된 6개 주성분이 사용된 6개 인자와의 관련성을 보여주는 가중치계수를 나타낸 것이다 (Jang, et. al., 2018). 성분 1의 경우 급수전당 배관연장, 급수전당 공급량과 양의 부호를 갖고, 관로노후도와 누수발생횟수와는 음의 관계를 갖는 성분으로 관망의 물리인자와 관련되는 성분으로 볼 수 있다. 성분 2는 평균관경과 가장 밀접한 관계를 가지고 있는 성분으로 나타났다. 성분 3은 에너지비와 높은 양의 관계를 갖는 성분으로 다른 인자와는 반비례하는 성분을 가진다. 성분 4는 평균관경과 반비례하며 급수전당 관경과 비례하는 성분으로 관로의 물리인자와 관련되는 것으로 나타났다. 성분 5와 6은 사용된 6개 인자 중 특정 성분과 관계가 성분 1~4에 비하여 크지 않아, 고유값(eigenvalue)이 상대적으로 낮게 나타났다.

Table 3. Weight coefficient of each principal component

Classification	Component					
	1	2	3	4	5	6
Mean pipe diameter	0.126	0.795	-0.229	-0.534	0.032	0.116
Pipe length per junction	0.809	0.222	-0.019	0.420	0.109	0.328
Water supply quantity per junction	0.795	0.426	-0.082	0.191	-0.007	-0.379
Deteriorated pipe ratio	-0.649	0.451	-0.226	0.346	-0.452	0.032
Demand energy ratio	-0.261	0.464	0.844	0.051	0.038	-0.006
Number of leaks	-0.736	0.278	-0.249	0.254	0.502	-0.055

※ Source: Estimation of Leakage Ratio using Principal Component Analysis and Artificial Neural Network in Water Distribution Systems (Jang, et. al., 2018)

3) 인공신경망 모의 결과

모의조건 1에 대한 무수율 모의결과와 실측 무수율을 비교하였다. <Figure 5>는 무수율 실측값과 모의값의 비교 그래프이다. Case 1은 선정된 블록의 6개 인자의 원 데이터를 이용하여 무수율을 추정한 결과로써 x축은 실측 무수율, y축은 인공신경망에 의해 모의된 무수율이다.

<Figure 5>의 (a), (b)는 인공신경망 내 은닉층이 1개인 조건에서 은닉층 내 뉴런의 수가 6개와 12개인 조건이다. 결정계수(R^2)의 값이 뉴런이 12개인 조건이 6개인 조건보다 높은 것으로 나타났고, $y = x$ 그래프에 비교적 수렴하는 것으로 나타났다. 따라서 은닉층에 12개의 뉴런이 적용된 조건이 6개 조건보다 무수율을 잘 추정하는 것으로 나타났다. <Figure 5>의 (c)~(f)는 인공신경망 내 2개의 은닉층이 사용된 조건으로 순서대로 6-6, 6-12, 12-6, 12-12 (은닉층 1 - 은닉층 2)의 뉴런이 사용된 무수율 산정 결과이다. 2개의 은닉층이 사용된 조건에서는 은닉층 1에서 12개의 뉴런이, 은닉층 2에서 6개의 뉴런이 사용된 조건에서 결정계수(R^2)가 0.49로 가장 높게 나타났고, 은닉층 1개가 사용된 조건보다 실측 무수율을 더욱 잘 추정하는 것으로 나타났다.

<Figure 6>은 주성분 분석과 인공신경망을 통해 산출된 무수율과 실측 무수율의 비교 결과로 주성분 분석을 통해 생성된 6개 주성분을 모두 인공신경망에 사용한 조건이다.

<Figure 6>의 (a), (b)는 인공신경망 내 은닉층이 1개인 조건에서 은닉층 내 뉴런의 수가 6개와 12개인 조건이다. 결정계수(R^2)의 값은 뉴런이 12개인 조건이 6개인 조건보다 다소 높은 것으로 나타났다. 원 데이터 이용조건과 비교시 차이는 크지 않은 것으로 나타났다. <Figure 6>의 (c)~(f)는 인공신경망 내 2개의 은닉층이 사용된 조건으로 순서대로 6-6, 6-12, 12-6, 12-12 (은닉층 1 - 은닉층 2)의 뉴런이 사용된 무수율 산정 결과이다. 2개의 은닉층이 사용된 조건에서는 은닉층 1에서 12개의 뉴런이, 은닉층 2에서 6개의 뉴런이 사용된 조건에서 결정계수(R^2)가 0.56으로 가장 높게 나타났고, 은닉층 1과 2에서 12개가 사용된 조건도 결정계수가 높은 것으로 나타났다. 원데이터를 이용하는 조건과 같이 은닉층 1개가 사용된 조건보다 실측 무수율을 더욱 잘 모의하는 것으로 나타났다.

<Figure 7>은 선정된 블록의 6개 인자를 주성분 분석을 수행한 후 영향이 가장 큰 4개의 주성분을 선별하

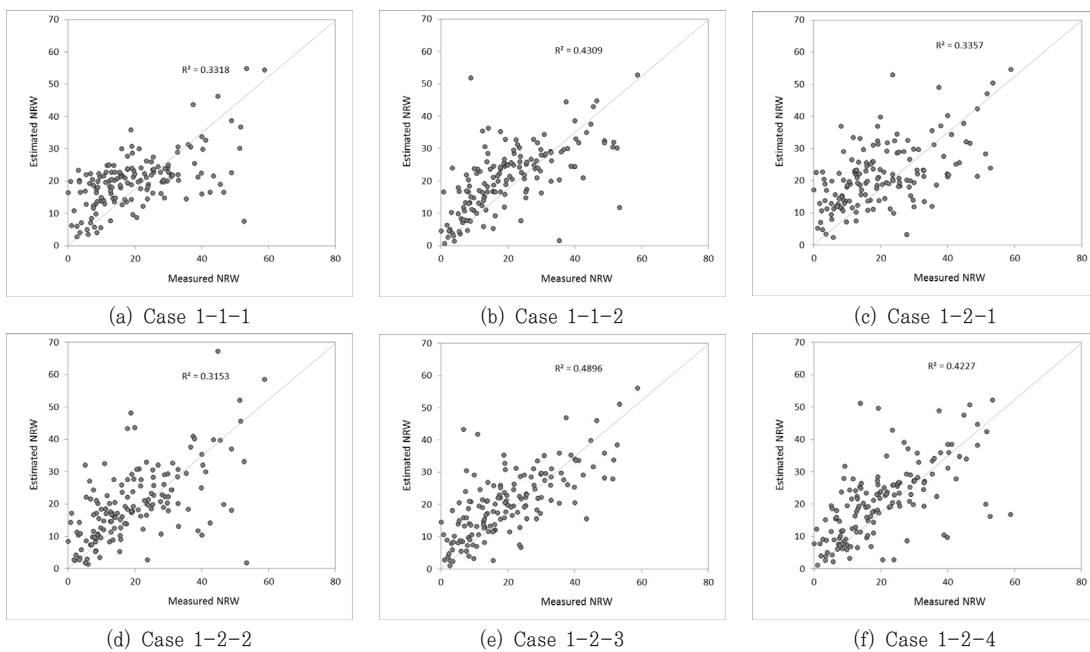


Figure 5. ANN simulation results using original data

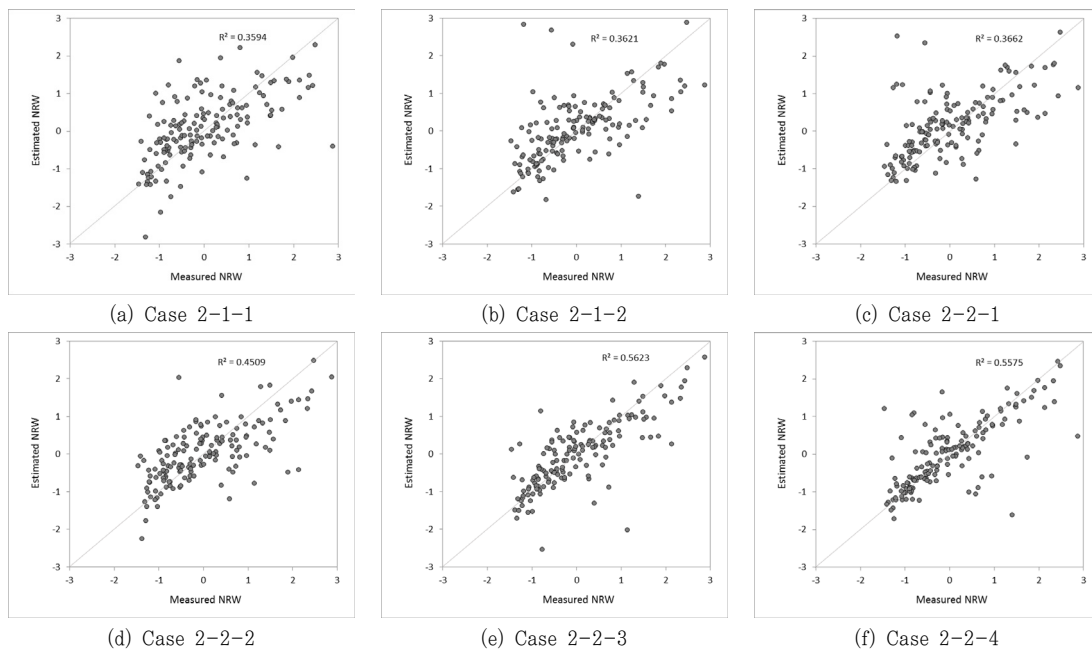


Figure 6. ANN simulation results using 6 principal components data

여 인공신경망을 통하여 무수율을 추정한 결과로써 x축은 실측 무수율을 y축은 인공신경망에 의해 모의된 무수율이다.

〈Figure 7〉의 (a), (b)는 인공신경망 내 은닉층이 1개인 조건에서 은닉층 내 뉴런의 수가 6개와 12개인 조건이다. 결정계수(R^2)의 값이 뉴런이 12개인 조건이 6

개인 조건보다 높은 것으로 나타나 모든 실험조건에서 단일 은닉층 조건에서는 사용된 독립변수 개수의 2배에 해당하는 12개의 뉴런을 사용할 때가 6개 뉴런 조건보다 실측무수율을 잘 모의하는 것으로 나타났다. 〈Figure 7〉의 (c)~(f)는 인공신경망 내 2개의 은닉층이 사용된 조건으로 순서대로 6-6, 6-12, 12-6, 12-12 (은닉층 1 -

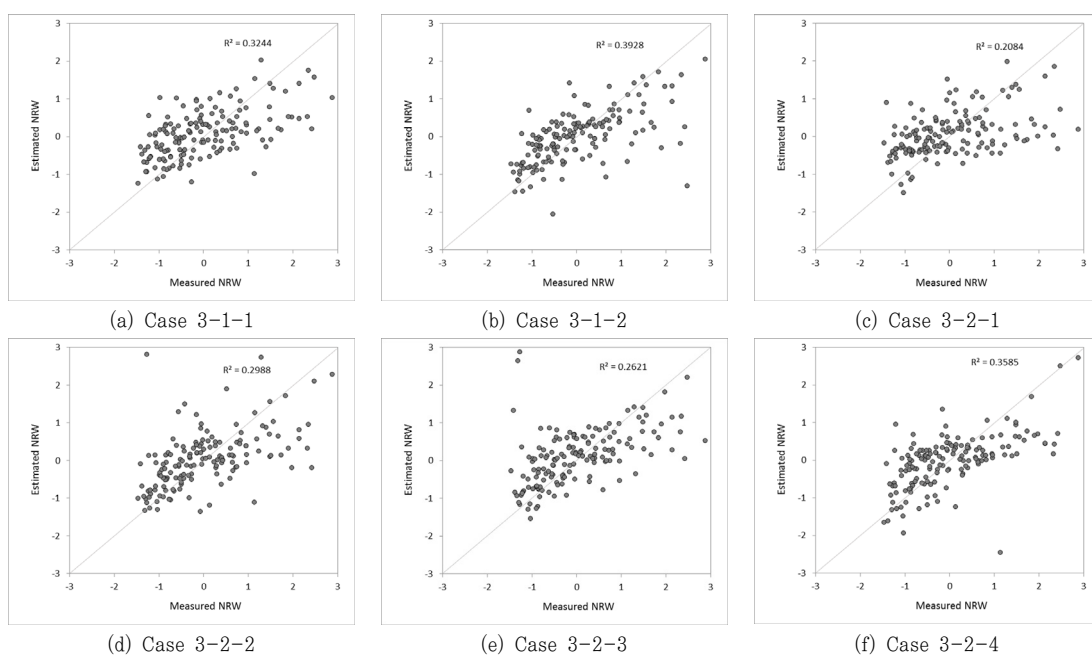


Figure 7. ANN simulation results using 4 principal components data

은닉층 2)의 뉴런이 사용된 무수율 산정결과이다. 2개의 은닉층이 사용된 조건에서는 은닉층 1과 2 모두 12개의 뉴런이 사용된 조건에서 결정계수(R^2)가 0.36으로 가장 높게 나타났는데 이는 원 데이터 및 6개의 주성분을 이용하는 조건과 다른 결과를 나타내었다. 또한 4개의 주성분을 활용하는 조건에서는 은닉층 2개를 사용하는 조건보다 은닉층 1개를 사용하는 조건에서 실측 무수율을 더욱 잘 나타내는 것으로 나타났다.

4) 무수율 모의 정확도 평가

주성분 분석과 인공신경망을 이용하여 대상지역의 무수율을 추정하기 위한 가장 적합한 인공신경망 조건을 선정하였다. 사용된 소블록 자료에서 실측 무수율 데이터와 전체 모의결과를 절대평균오차(MAE, Mean Absolute Error), 평균제곱오차(MSE, Mean Squared Error), 편향성 분석(BIAS), G 값(Goodness of prediction measure)을 산출하여 정확도를 판별하였다. <Table 4>는 모든 모의조건에 대한 정확도 평가결과이다.

MAE와 MSE는 값이 작을수록 추정된 값이 정확한

것을 의미하는데 원데이터를 이용한 조건에서는 은닉층이 두 개이고, 은닉층 1이 12개, 은닉층 2가 6개의 뉴런이 사용된 조건(Case 1-2-3)에서 실측무수율과 가장 정확한 모의결과가 나타났다. 또한, G값에서도 100에 근접한 결과는 Case 1-2-3번으로 나타났다. 반면 자료의 편향성을 나타내는 BIAS에서는 두 개의 은닉층에서 12개의 뉴런이 사용한 조건이 편향성이 가장 적은 것으로 나타났다.

6개의 주성분 인자가 사용된 결과에서는 은닉층이 두 개이고, 은닉층 1과 은닉층 2가 각각 12개의 뉴런이 사용된 조건 (Case 2-2-4)에서 실측무수율과 가장 정확한 모의결과가 나타났다. BIAS는 소수점 둘째 자리까지 0값이 나타났기 때문에 모의결과가 편향되지 않았다고 할 수 있다. MAE, MSE, BIAS, G 값까지 4개 항목 모두에서 가장 실측값과 정확한 추정이 가능한 모형으로 선정되었다. 또한 Case 1, 3번과 비교 시에도 가장 정확도가 높은 모형으로 선정되었다.

4개의 주성분 인자가 사용된 결과에서는 Case 1, 2번과 달리 은닉층이 한 개이고, 12개의 뉴런이 사용된

Table 4. Results of accuracy assessment by ANN simulation

Classification	Hidden layer	Number of neurons	Modeling case	MAE	MSE	BIAS	G value (%)	Optimal model
6 Original data	Single layer	6	1-1-1	8,294	119,7	0,79	31,3	
		12	1-1-2	7,172	102,9	-0,23	40,9	
	Multiple layer	12(6:6)	1-2-1	8,841	122,3	-1,02	29,8	
		18(6:12)	1-2-2	8,181	148,0	0,97	15,1	
		18 (12:6)	1-2-3	7,106	92,4	0,27	47,0	O
	24(12:12)	1-2-4	7,217	112,8	0,08	35,3		
6 PCA data	Single layer	6	2-1-1	0,622	0,699	-0,07	26,5	
		12	2-1-2	0,544	0,687	-0,04	27,8	
	Multiple layer	12(6:6)	2-2-1	0,575	0,674	-0,11	29,1	
		18(6:12)	2-2-2	0,556	0,538	-0,02	43,4	
		18(12:6)	2-2-3	0,454	0,478	0,05	49,8	
	24(12:12)	2-2-4	0,401	0,448	-0,00	52,9	O	
4 PCA data	Single layer	6	3-1-1	0,650	0,652	-0,06	31,4	
		12	3-1-2	0,547	0,587	0,03	38,3	O
	Multiple layer	12(6:6)	3-2-1	0,683	0,775	-0,04	18,5	
		18(6:12)	3-2-2	0,619	0,717	-0,02	24,6	
		18(12:6)	3-2-3	0,610	0,766	-0,09	19,5	
	24(12:12)	3-2-4	0,609	0,629	0,05	33,9		

조건 (Case 3-1-2)에서 실측무수율과 가장 정확한 모의결과가 나타났다. G값이 38.3으로 다른 모의조건에 비하여 정확도가 비교적 낮게 나타났고, 두 개의 은닉층이 사용된 모의결과에서도 정확도가 낮은 것으로 나타났다.

V. 결론

본 연구에서는 인공신경망과 주성분 분석을 이용하여 무수율 추정연구를 수행하였다. 이를 위하여 상수관망 주요 인자 자료를 수집하고, Z-score 방법을 통한 통계데이터 변환 후 무수율 추정 모형을 구축하였다. 모의조건별로 무수율 산정모형을 모의한 후 정확도 분석을 통하여 최적 인공신경망 모형을 선정하였다. 연구를 통한 결론은 다음과 같다.

선정된 상수관망 주요인자 중 데이터 변환을 거치지 않은 원 데이터를 이용하여 인공신경망을 적용하는 것보다 주성분 분석을 통해 산출된 주성분 인자를 사용하여 인공신경망에 적용하는 것이 무수율 추정 정확도를 높이는 것으로 나타났다. 특히 주성분 분석을 통하여 산출된 모든 인자를 이용하는 것이 무수율 예측 정확도를 향상시키는 것으로 나타났다.

연구에서 사용된 인공신경망은 은닉층이 1개와 2개인 경우로 구분하여 시뮬레이션을 수행하였다. 6개의 주요인자, 주성분을 사용하는 조건에서는 은닉층이 1개인 조건에 비하여 2개인 조건에서 평균적으로 무수율 추정 정확도가 높은 것으로 나타났으나 4개의 주성분을 이용하여 무수율을 추정하는 경우에는 은닉층이 1개인 조건에서 정확도가 높은 것으로 나타났다. 사용되는 인자, 모의 조건에 따라서 은닉층 수에 따른 정확도 추정 확률이 변하기 때문에 대상지역, 모의조건을 고려한 은닉층 수가 결정되어야 할 것으로 판단된다.

본 연구에서는 무수율 추정을 위하여 6개의 물리적, 운영적 인자를 사용하였고, 이에 따라 인공신경망 은닉층의 뉴런수를 사용된 독립변수의 1, 2배에 해당하는 조건으로 시뮬레이션을 하였다. 사용된 독립변수와 동

일한 수의 뉴런보다는 2배에 해당하는 뉴런에서 무수율 예측정확도가 증가하는 것으로 나타났다.

기존 연구를 통하여 선정된 급수전당 공급량, 급수전당 관로연장, 관로노후도, 누수발생횟수, 평균관경, 급수에너지비 인자를 본 연구에 적용하였고, 결정계수 (R^2)가 0.5 이상의 추정 정확도를 확보하였다. 연구 대상지역인 인천시의 경우 관 노후도가 높고, 블록시스템 구축이 이루어지지 않은 곳이 신규 개발도시에 비하여 많은 편이다. 블록구축이 잘 이루어져 있고, 상수관망 운영이 잘 이루어지고 있는 지역을 선정하여 주성분 분석과 인공신경망 모형을 적용할 경우 무수율 예측 정확도가 상승할 것으로 보인다.

연구를 통하여 무수율 추정방안으로 주성분 분석과 인공신경망을 이용한 예측방법론을 제시하였고, 향후 국내 무수율 추정연구에 적용할 수 있다. 또한, 상수관망 운영관리를 위한 관망 내 누수, 파손위험에 대한 원인분석 및 관련 관망운영사업 우선순위를 결정하는 데에도 도움이 될 것으로 기대된다. 특히 이 연구는 상수관망 내에 가장 큰 손실을 나타내는 지표인 누수량을 상수관망의 주요인자를 통하여 가늠할 수 있다는 점에서 상수관로 파손에 따른 싱크홀, 단수 피해로부터의 재난위기를 사전에 방지하기 위한 감시시스템에 기여할 수 있을 것으로 판단된다.

References

- Alegre, H., W. Hirner, J.M. Baptista, and R. Parena. 2000. *Performance Indicators for Water Supply Services; IWA*. Publishing: London, UK.
- Choi, Gye Woon, Yung Gyu Chang, and Sueng Woo Lee. 2006. Effect of Estimation Method of Demand Water on the Analysis of Water Distribution System. *Korea Water Resources Association Conference*. 1425-1430.
- Gonalves, F. V., H. M. Ramos, and L. F. Reis. 2010. Hybrid Energy System Evaluation in Water Supply System Energy Production: Neural Network Approach. *International Journal of Energy and Environment*. 1(1): 21-30.

- Haykin, S. 1994. *Neural Networks: A Comprehensive Foundation*. NY: Macmillan.
- Heaton, J. T. 2005. *Introduction to Neural Networks with Java*. UK: Heaton Research, Inc.
- Jang, Dong Woo. 2017. Estimation of Non-Revenue Water Ratio Using PCA and ANN in Water Distribution Systems. Ph.D. Dissertation. Incheon National University.
- Jang, Dong Woo and Gye Woon Choi. 2017. Estimation of Non-Revenue Water Ratio for Sustainable Management Using Artificial Neural Network and Z-Score in Incheon, Republic of Korea. *Sustainability*. 9(11): 1-13.
- Jang, Dong Woo, Hyo Seon Park, and Gye Woon Choi. 2018. Estimation of Leakage Ratio Using Principal Component Analysis and Artificial Neural Network in Water Distribution Systems. *Sustainability*. 10(3): 1-15.
- Jo, Hyoung Geun. 2017. Study on Influence Factors of Non-revenue Water for Sustainable Management of Water Distribution Networks. Ph.D. Dissertation. Incheon National University.
- Jo, Hyoung Geun, Gye Woon Choi, and Dong Woo Jang. 2016. Development of the Non-revenue Water Analysis Equation through the Statistical Analysis of Main Parameter in Waterworks System in Incheon City. *Crisisonomy*. 12(11): 63-75.
- Jun, Hwan Don, Seok Hyeon Kim, Moo Jong Park, Joong Hoon Kim, and Hwan Goo Lee. 2008. Development of a Method to Reduce Damages by Pipe Failures. *Journal of the Korean Society of Hazard Mitigation*. 8(6): 31-36.
- Kim, Dong Hee, Dong Woo Ryu, Young Min Choi, and Woo Jin Lee. 2010. Application of Kriging and Inverse Distance Weighting Method for the Estimation of Geo-Layer of Songdo Area in Incheon. *Journal of Korean Geotechnical Society*. 26(1): 5-19.
- Kim, Dong Sik. 2007. Prediction of Soft Clay Settlement Using Artificial Neural Network. Ph.D. Dissertation. University of Suwon.
- Kim, Seok Hyeon, Hwan Don Jun, Chun Woo Baek, and Joong Hoon Kim. 2007. Research for Improvement of Water Distribution System Reliability. *KSCE 2007 Conference & Civil Expo*. 1397-1400.
- Kim, Seok Woo. 2004. Development of Optimal Decision-Making System for Rehabilitation of Water Distribution Systems Divided by Small Division. Master's Thesis. Korea University.
- Koo, Doo Jin. 2006. Management of Water Distribution Pipe Networks Using BFS Algorithms. Master's Thesis. Chonnam National University.
- Kreyszig, E. 1979. *Advanced Engineering Mathematics (Fourth Edition)*. John Wiley & Sons Inc.
- Lambert, A. O., T. G. Brown, M. Takizawa, and D. Weimer. 1999. A Review of Performance Indicators for Real Losses from Water Supply Systems. *J. Water SRT Aqua*. 48(6): 227-237.
- Lee, Mee Ran. 2010. A Study on Deterioration Evaluation Model of Water Main Using Integrated PCA and ANN. Master's Thesis. University of Seoul.
- Moon, H. and P. J. Phillips. 2001. Computational and Performance Aspects of PCA-Based Face Recognition Algorithms. *Perception-London*. 30(3): 303-322.
- National Institute of Environmental. 2011. *A Study on Early Forecasting for Algal Blooms Using Artificial Neural Networks (II)*. Technical Report.
- Pal, S. K. and S. Mitra. 1992. Multilayer Perceptron, Fuzzy Sets, and Classification. *IEEE Transactions on Neural Networks*. 3(5): 683-697.
- Park, Han Yung. 2008. Suggestions for Scientific Management of Pipe Network. *KSCE Journal of Civil Engineering*. 56(2): 77-82.
- Park, In Chan, Ki Won Kwon, Won Cheol Cho, and Kwan Hee Cho. 2006. Study on the Decision Priority of Rehabilitation for Water Distribution Network Based on Prediction of Pipe Deterioration. *Korea Water Resources Association Conference*. 1391-1394.
- Park, Su Wan, Tae Young Kim, Ki Young Lim, and Hwan Don Jun. 2011. Fuzzy Techniques to Establish Improvement Priorities of Water Pipes. *Journal of Korea Water Resources Association*. 44(11): 903-913.
- Park, Yeong Sik. 2014. A Study on Long Term Replacement and Maintenance Plan for Multi-region Water Pipelines Considering Economics. Master's Thesis. Seoul National University.

- Shilehwa, C. M. 2013. Factors Influencing Water Supply's Non Revenue Water: A Case of Webuye Water Supply Scheme. Master's Thesis. University of Nairobi, Nairobi, Kenya.
- Shinde, V. R., N. Hirayama, A. Mugita, and S. Itoh. 2013. Revising the Existing Performance Indicator System for Small Water Supply Utilities in Japan. *Urban Water J.* 10(6): 377-393.
- Surendran, S., T. T. Tanyimboh, and M. Tabeshb. 2005. Peaking Demand Factor-based Reliability Analysis of Water Distribution Systems. *Advances in Engineering Software.* 36: 789-796.
- Waterworks Headquarters Incheon Metropolitan City. 2015. *Technical Diagnostics Report for Re-establish Basic Plan of Waterworks Maintenance in Incheon Water Distribution Network.*
- Waterworks Headquarters Incheon Metropolitan City. 2015. *Waterworks Status; Incheon Metropolitan City.*
- Wikipedia. 2017. Eigenvalue, https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors.
- Wikipedia. 2017. Non-RevenueWater. https://en.wikipedia.org/wiki/Non-revenue_water#cite_note-11.
- Wikipedia. 2017. Z-score. https://en.wikipedia.org/wiki/Standard_score.
- Winarni, W. 2009. Infrastructure Leakage Index (ILI) as Water Losses Indicator. *Civ. Eng. Dimens.* 11: 126-134.
- Woo, Hyoung Min. 2008. Development Plans for Distribution Network Improvement in Waterworks System. *Water for Future. Korea Water Resources Association.* 41(11): 16-23.
- Wyatt, A. S. 2010. *Non-Revenue Water: Financial Model for Optimal Management in Developing Countries.* Press: Amman, Jordan.
- Xie, Z., J. Hall, I. P. McCarthy, M. Skitmore, and L. Shen. 2016. Standardization Efforts: The Relationship between Knowledge Dimensions, Search Processes and Innovation Outcomes. *Technovation.* 48(49): 69-78.
- Korean References Translated from the English*
- 구두진. 2006. BFS알고리즘을 이용한 상수도 배·급수 관망관리. 전남대학교 석사학위논문.
- 국립환경과학원. 2011. 인공신경망 기법을 이용한 조류발생 조기예측기법 연구(Ⅱ). 기술보고서.
- 김동식. 2007. 인공신경망을 이용한 연약지반의 침하 예측. 수원대학교 박사학위논문.
- 김동휘, 류동우, 최영민, 이우진. 2010. 인천 송도지역 지층분포 추정을 위한 크리깅과 역거리가중치법의 적용. 한국지반공학회논문집. 26(1): 5-19.
- 김석우. 2004. 상수관망의 최적 개량 의사결정 시스템의 개발: 구역별 우선순위 결정. 고려대학교 석사학위논문.
- 김석현, 전환돈, 백천우, 김중훈. 2007. 상수관망의 신뢰도 향상 방안에 관한 연구. 대한토목학회 학술대회. 1397-1400.
- 박수완, 김태영, 임기영, 전환돈. 2011. 상수관로 개량 우선순위 수립을 위한 퍼지 기법. 한국수자원학회논문집. 44(11): 903-913.
- 박영식. 2014. 경제성을 고려한 광역상수도 관로의 장기 교체 및 보수계획에 관한 연구. 서울시립대학교 석사학위논문.
- 박인찬, 권기원, 조원철, 조관희. 2006. 상수관로 노후도 평가를 통한 개량 우선순위 결정에 관한 연구. 한국수자원학회 학술발표회 논문집. 1391-1394.
- 박한영. 2008. 과학적 상수도관망 관리를 위한 제언. 대한토목학회지. 56(2):77-82.
- 우형민. 2008. 상수도 배·급수 관망개선을 위한 기술개발 방안. 물과 미래. 한국수자원학회. 41(11): 16-23.
- 이미란. 2010. 주성분 분석과 인공신경망을 이용한 상수관로 노후도 추정 모델 연구. 서울시립대학교 석사학위논문.
- 인천광역시상수도사업본부. 2015. 인천 수도정비 기본계획 재수립 상수도관망기술진단 보고서.
- 장동우. 2017. 주성분 분석과 인공신경망을 이용한 상수관망 내 무수율 추정. 인천대학교 박사학위논문.
- 전환돈, 박무중, 김중훈, 이환구. 2008. 상수관 파괴에 의한 피해 경감기법의 개발. 한국방재학회논문집. 8(6): 31-36.
- 조형근, 최계운, 장동우. 2016. 인천시 상수관망 주요인자의 통계 분석을 통한 무수율 산정 공식개발, *Crisisonomy.* 12(11): 63-75.
- 조형근. 2017. 상수관망의 지속가능한 관리를 위한 무수율 영향인자에 관한연구. 인천대학교 박사학위논문.
- 최계운, 장연규, 이승우. 2006. 배수관망해석에 수요량 적용방법이 미치는 영향. 한국수자원학회학술발표회. 1425-1430.

데이터 정규화 및 주성분 분석 기반의 인공지능망을 이용한 상수관망 내 무수율의 추정

국문초록 상수관망 내 무수율은 관로의 파손, 운영 손실, 물리적 요소 등에 의해 발생하는 수도공급량에 대한 손실 비율을 나타낸다. 국내에서 시행하는 관망정비사업, 유수율 제고 사업에 있어서 무수율은 시·도 및 소블록에 대한 비교 지표로써 무수율에 영향을 주는 인자의 발굴 및 무수율 추정기법에 대한 연구는 수도공급시설의 경제성과 연관되어 점차 중요해지고 있다. 본 연구에서는 무수율 추정하기 위한 방법으로 통계분석 기법 중 주성분 분석(PCA), 인공지능망(ANN)을 이용하여 무수율을 추정하였다. 연구를 위하여 상수관망 주요 영향인자에 대한 데이터를 수집하였고, Z-score방법을 통하여 데이터를 정규화 한 후 PCA-ANN 모형에 적용하여 무수율을 추정하였다. 무수율 모의 결과와 실측 무수율을 비교하기 위하여 정확도 평가를 수행하였다. 연구결과 무수율 예측에 있어 PCA-ANN기법이 원데이터를 이용하여 ANN을 단독으로 모의하는 조건보다 정확도가 높은 것으로 나타났다. 또한 ANN 모형 구축시 은닉층 내의 뉴런수에 따라 추정결과가 상이하며 본 연구에서 사용된 6개의 독립 변수에 대하여, 12개의 뉴런을 이용한 조건에서 무수율 예측정확도가 높은 것으로 나타났다.

주제어 : 인공지능망, 무수율, 주성분 분석, 데이터 정규화, 상수관망

Profiles **Dong Woo Jang** : He received his Master degree from Incheon National University, Korea in 2011. He received his Ph.D. in the Department of Civil & Environmental Engineering at Incheon National University in 2017. His interesting subject and area of research are hydraulics, hydrology and water distribution system(nightray@paran.com).