

시계열모델을 이용한 하수처리장 유입수 성상 예측

Forecast of Influent Characteristics in Wastewater Treatment Plant with Time Series Model

김병근^{1,*} · 문용택¹ · 김홍석¹ · 김종락²

Byung-Goon Kim^{1,*} · Yong-Taik Moon¹ · Hong-Suck Kim¹ · Jong-Rack Kim²

1 한국수자원공사 수자원연구원

2 캔지아21

(2007년 9월 3일 논문 접수; 2007년 11월 26일 최종 수정논문 채택)

Abstract

The information on the incoming load to wastewater treatment plants is not often available to apply to evaluate effects of control actions on the field plant. In this study, a time series model was developed to forecast influent flow rate, BOD, COD, SS, TN and TP concentrations using field operating data. The developed time series model could predict 1 day ahead forecasting results accurately. The coefficient of determination between measured data and 1 day ahead forecasting results has a range from 0.8898 to 0.9971. So, the correlation is relatively high.

We made forecasting program based on the time series model developed and hope that the program will assist the operators in the stable operation in wastewater treatment plants.

Key words: Time series model, influent forecasting, ARIMA, supervisory management system

주제어: 시계열 모델, 유입수 예측, ARIMA, 하수처리장 관리시스템

1. 서론

하·폐수처리 공정의 안정적인 유출수질을 유지하는 동시에 경제적인 운영을 하기 위해서는 공정의 유입수 및 유출수 자료가 필수적으로 요구된다. 실제 현장의 공정 운전자들은 보통 여러 해 동안의 운전 경험에 의해 축적된 전문 지식을 바탕으로 하·폐수처리 공정을 운전, 관리하고 있다. 하지만 최근 신설

되고 있는 하·폐수처리장의 수는 빠르게 증가하고 있으며, 처리공정 또한 고도화되고 복잡해지고 있는 반면에 잘 훈련된 운전자의 수는 한정적이며, 공정 전문가로 훈련시키는 데 오랜 기간을 필요로 한다. 그러므로 이러한 문제를 해결하기 위해 하·폐수처리 공정의 계측, 제어 및 자동화(Instrumentation, Control and Automation, ICA) 분야에 대한 연구가 지난 십여 년 동안 수행되었다(Krüger, 2004).

이러한 연구의 결과로 하·폐수처리공정에 대한

*Corresponding author Tel: +82-42-870-7551, FAX: +82-42-870-7549, E-mail: bgkim@kwater.or.kr (Kim, B.G.)

다양한 운전 제어전략이 제시되었고, SMAC(Smart Control of Wastewater Systems Krüger, 2004)과 TELEMAC(Telemonitoring and Advanced Telecontrol of high yield wastewater treatment plants Bernard et al., 2004)과 같은 유럽공동체 프로젝트에서 실 규모 하·폐수처리장에 실제 적용되기도 하였다. 제안된 운전 제어전략을 실규모 하·폐수처리공정에 적용하기 위해서는 사전에 적용 가능성을 평가하는 것이 필요하며, 이 과정에서 공정으로 유입되는 부하에 대한 미래 정보가 요구되지만, 이를 획득하는 것은 현실적으로 매우 어렵다(Krüger, 2004). 유입 능도가 제공된다면, 수학적 모델을 이용하여 최적의 운전 전략을 도출하는 데 사용될 수 있으며, 자동 제어기의 설계 및 여러 가지 유입수 주입 전략들을 도출하는 데 사용될 수 있다(Bernard et al., 2004).

본 연구에서는 공정의 유입수 성상 가료를 예측하기 위한 도구로써 시계열모델의 활용을 제안하였다. 일반적으로 하·폐수처리공정의 유입 유량 및 유입 성분의 농도는 일간, 월간 및 연간의 특정 패턴을 가지고 변화하는 특성을 가지고 있으며, 이러한 주기적 변화는 시계열모델로 잘 기술할 수 있다

본 연구의 목적은 하수처리장 유입유량, BOD, COD_{Mn}, SS, TN 및 TP 농도를 대상으로 1일 또는 2일 후의 일일 농도를 예측하는 시계열모델을 도출하고, 이를 바탕으로 하수처리장 운전자가 유입수 성상을 예측하여 사전에 대비할 수 있는 유입수 예측 프로그램을 개발하는 데 있다.

2. 재료 및 방법

2.1. 모델개발에 사용된 유입수 성상

모델 개발에 사용된 자료는 J 하수처리장의 2005년 1월에서 2006년 8월 사이에 측정된 유입 유량, BOD, COD_{Mn}, SS, TN, TP 농도 그리고 강수량 자료이다. 하수처리장으로 유입되는 하수관로는 합류식으로 구성되어 있고, 하수처리 공정은 산화구 공정으로 운전되고 있다. Table 1은 모델 개발에 사용된 유입수 성상을 보여주고 있다. 모델 개발을 위해 2005년 1월부터 2006년 5월까지의 자료를 사용하였고, 모델 검증을 위해 2006년 6월부터 8월까지의 3개월간의 자료를 사용하였다.

Table 1. Influent characteristics of J wastewater treatment plant

	Range	Average
Flowrate (m ³ /day)	473-3,212	1,112
BOD (mg/L)	6-266	87
COD _{Mn} (mg/L)	12-151	55
SS (mg/L)	8-327	71
TN (mg/L)	4-51	21
TP (mg/L)	0.3-7.0	2.6

2.2. 시계열모델 개발을 위한 접근법

하수처리장으로 유입되는 각각의 관로에서 유량과 오염물질의 농도를 측정하는 것이 가능하다면 관망해석을 포함한 수리학적 모델이나 블랙박스형태의 신경회로망 모델과 같은 비선형모델을 사용하여 하수처리장 유입수 성상을 예측하는 것이 가능할 것이다. 하지만 현재 운영 중인 국내 대부분의 하수처리장으로부터 얻을 수 있는 자료는 일간 측정되는 유입 유량 및 성분 농도, 그리고 강수량뿐이므로, 모델의 입력 변수로 사용 가능한 항목은 과거 유입수 자료와 강수량이다. 그러므로 유입수 성상 예측모델로 사용가능한 것은 시계열모델 혹은 선형회귀모델을 고려해 볼 수 있다. 본 연구에서는 단기예측에 좋은 성능을 보여주는 ARIMA 시계열모델을 도출하고자 하였으며, 예측 성능을 향상시키기 위해 유입수와 밀접한 관계를 가지는 강수량을 모델에 포함하였다.

시계열 자료(time series data)는 시간에 따라 변화하는 현상을 기록한 자료이며, 시계열 모델은 과거의 자료를 가지고 미래의 값을 예측한다. 시계열 분석에 사용되는 자료는 평균과 분산이 시간에 따라 변하지 않고 일정한 값을 가져야 하는 정상성(stationarity) 조건을 만족하여야 하며, 비정상 시계열자료는 차분 혹은 변수변환을 사용하여 정상시계열로 만들어야 한다. ARIMA(autoregressive integrated moving average model) 시계열모델을 도출하는 시계열 분석과정을 Fig. 1에 나타내었다(Box and Jenkins, 1976). 모델의 형태를 결정하는 식별단계(model identification), 식별된 모델의 모수들을 추정하는 추정단계(model estimation), 최종 모델의 타당성을 검토하는 진단단계(diagnostic checking), 결정된 모델을 사용하여 미래의 값을 예측하는 예측단계(forecast model)의 순서로 예측모델을 도출하게 된다.

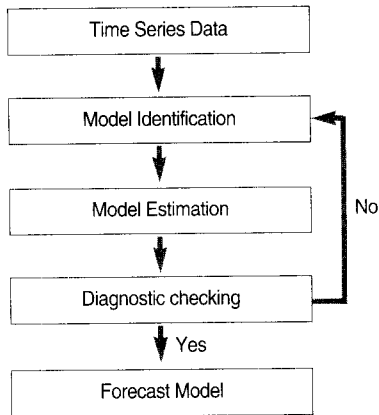


Fig. 1. Overall procedure for development of time series model.

1) 모델의 식별 (model identification)

모델의 형태를 식별하기 위해서는 시도표, 자기상관함수 (autocorrelation function: ACF), 편자기상관함수 (partial autocorrelation function: PACF) 등을 이용한다. 자료의 그림으로부터 파악할 수 있는 것은 추세, 계절성, 분산의 변환, 이상값 등의 비정상적인 현상들이며, 만약 비정상적인 요소들이 발견되면 정상 시계열로 변환해야 한다. 시계열의 분산이 시간에 따라 변한다면 차분작업을 수행한다. 이때 d차 차분한 ARMA(p, d) 모델을 ARIMA(p, d, q) 모델 (autoregressive integrated moving average model) 이라고 한다. ARMA 모델의 차수인 p와 q를 결정하여야 하며, 이는 ACF와 PACF의 그래프 모양으로부터 결정할 수 있다. ACF는 MA(q) 모델의 모델구조 q를 선택하는데 이용되고, PACF는 AR(q) 모델의 모델구조 p를 선택하는데 이용된다.

2) 모델의 추정 (model estimation)

일반적으로 사용되는 식별된 ARMA(p, q) 모델의 모수추정 방법으로 적률추정법 (method of moment estimation), 조건부 최소제곱추정법 (conditional least square estimation method), 비조건부 최소제곱법 (unconditional least square estimation method), 최대우도추정법 (maximum likelihood estimation method) 등이 사용된다. 본 연구에서는 모수 추정을 위해 Microsoft Excel 프로그램의 해찾기 기능인 Generalized Reduced Gradient method (Lasdon and Waren, 1978)를 사용하

였다. Generalized Reduced Gradient method를 사용한 모수 추정 과정에서 최적의 모수 값을 도출하기 위해 모델에 의해 예측한 유입수 값과 실제 측정된 유입수 값의 오차를 최소화하도록 목적함수를 구성하였다.

3) 모델의 진단 (diagnostic checking)

통계적방법을 사용한 모수추정 결과 도출된 모델의 차수 및 모수값에 대한 타당성을 평가하기 위해 일반적으로 잔차분석과 과적합진단 방법을 사용한다. 하지만 본 연구에서는 최적화 기법을 사용하여 모수를 추정하였기 때문에 모델 개발에 사용되지 않은 자료를 도출된 모델에 적용하여 모델의 타당성을 검증하였다.

4) 강수량을 고려한 시계열모델

일반적으로 강우는 하수처리장 유입수 성장 변화에 큰 영향을 미치는 것으로 인식되고 있다. 그러므로 유입수 성장 예측을 위한 시계열모델에 강수량을 고려하고자 하였으며, 강우 후 하수처리공정의 유입수 성장 변화에 영향을 미치는 시간은 일정하지 않고, 보통 0~1일 정도 소요되는 것으로 나타났다. 따라서 예측모델에 고려되는 강수량은 당일, 전일의 값만을 사용하였으며, 다음과 같은 선형 회귀 방정식으로 예측 모델을 구성하였다.

$$Z_{t+1} = c + \alpha_0 Z_t + \alpha_1 Z_{t-1} + \dots + \alpha_p Z_{t-p} + \beta_0^* (R_{t+1} + R_t)$$

여기서, Z_t 는 시간 t에서 유입수 성장 시계열 자료이며, R_t 는 시간 t에서 강수량, c, α_0 , α_1 , ..., α_p , β_0 는 시계열 모델의 계수(모수)들, p는 유입수 성장을 예측하기 위해 사용된 최대 과거 자료의 수이다.

3. 결과 및 고찰

3.1. 유입수 예측 시계열모델 개발

유입 유량, BOD, COD, SS, TN, TP 농도의 모델 구조 동정을 위한 ARIMA 분석을 수행하였다. Fig. 2는 각 유입수 성장의 시계열자료에 대한 ACF와 PACF 그래프를 보여준다. ACF는 현재 시점의 자료 값이 영향을 미치는 최대 미래의 시점을 의미하고, PACF는 현재 시점과 미래 시점의 자료값의 직접적

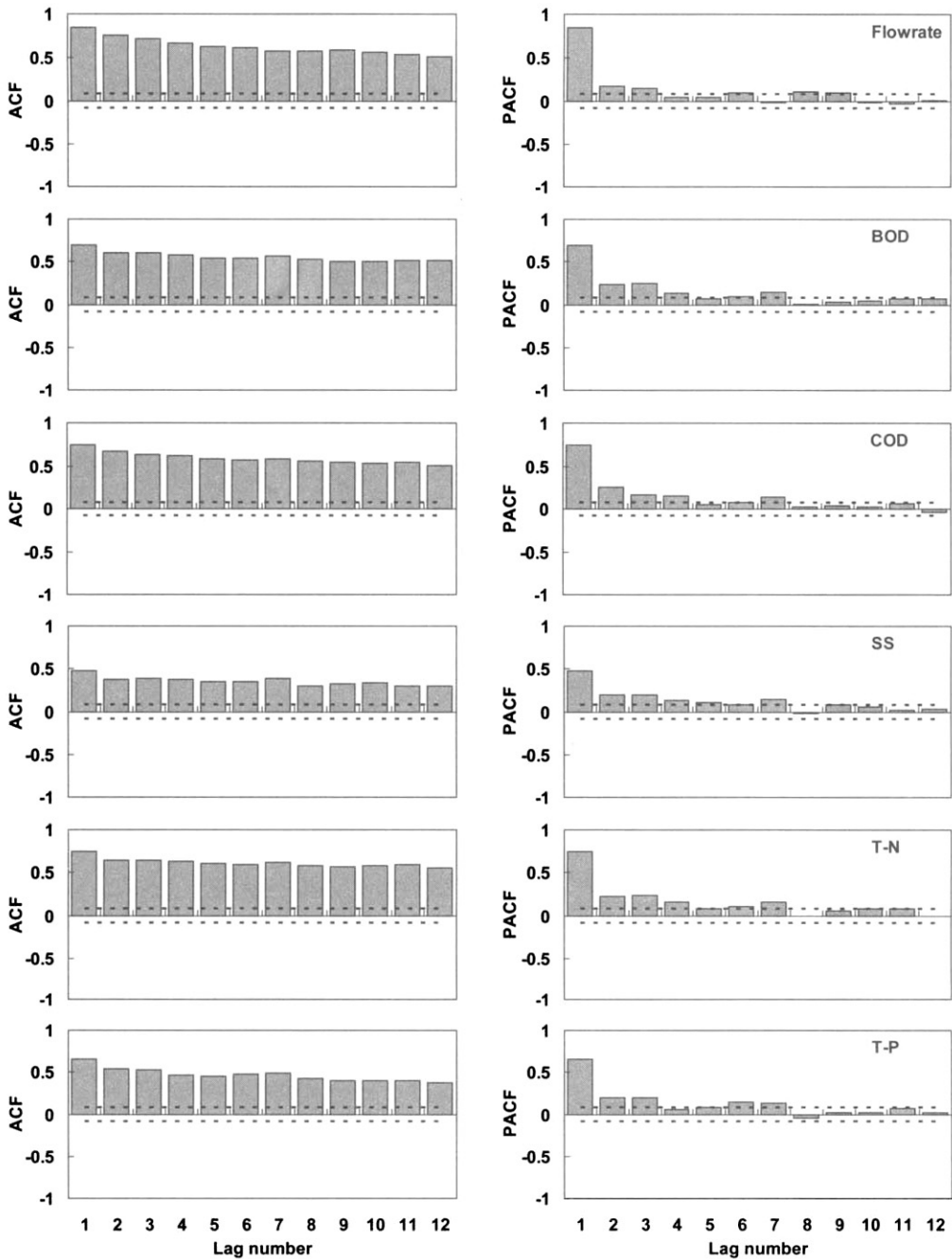


Fig. 2. ACF and PACF for flowrate, BOD, COD, SS, TN and TP concentrations of influent.

인 상관관계를 의미한다. 이러한 관계를 바탕으로 두 함수의 형태로부터 예측 모델 형태를 동정할 수 있다. ACF 함수는 증감하는 형태를 가지고, PCF는 초기 시차(lag)에서 절단하는 형태를 보여주므로, 모두

자료의 모델은 AR(p) 형태로 이루어짐을 알 수 있다. 각 유입수 성상별로 동정된 시계열 예측 모델의 구조와 모델 식을 Table 2에 나타내었다.

일간 유입수 자료 중 유입 유량과 TP의 경우,

Table 2. Types of time series model developed through ARIMA analysis

	Model	Type of model equation
flow rate TP	AR(3)	$Z_{t+1} = c + \alpha_0 Z_t + \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \beta_0 (R_{t+1} + R_t)$
BOD COD SS T-N	AR(4)	$Z_{t+1} = c + \alpha_0 Z_t + \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \alpha_3 Z_{t-3} + \beta_0 (R_{t+1} + R_t)$

* R_t = rainfall rate (mm/hr)

Table 3. Time series model for forecasting influent characteristics in J wastewater treatment plant

	Model	Equation
Flow rate	AR(3)	$Z_{t+1} = 147.74 + 0.65Z_t + 0.13Z_{t-1} + 0.03Z_{t-2} + 6.19(R_t + R_{t-1})$
BOD	AR(4)	$Z_{t+1} = 5.62 + 0.75Z_t - 0.18Z_{t-1} + 0.25Z_{t-2} - 0.025Z_{t-3} + 0.017(R_t + R_{t-1})$
COD _{Mn}	AR(4)	$Z_{t+1} = 5.68 + 0.74Z_t + 0.04Z_{t-1} + 0.029Z_{t-2} - 0.035Z_{t-3} + 0.01(R_t + R_{t-1})$
SS	AR(4)	$Z_{t+1} = 7.90 + 0.61Z_t + 0.26Z_{t-1} - 0.055Z_{t-2} - 0.015Z_{t-3} + 0.02(R_t + R_{t-1})$
TN	AR(4)	$Z_{t+1} = 8.94 + 0.50Z_t - 0.10Z_{t-1} + 0.156Z_{t-2} - 0.023Z_{t-3} + 0.003(R_t + R_{t-1})$
TP	AR(3)	$Z_{t+1} = 0.539 + 0.61Z_t + 0.093Z_{t-1} + 0.008Z_{t-2} + 0.0003(R_t + R_{t-1})$

* R_t = rainfall rate (mm/hr)

PACF 그래프에서 시차 1, 2, 3에서 유의한 값을 보였기 때문에 AR(3) 모델로 선정하였고, BOD, COD, SS, T-N은 시차 1, 2, 3, 4에서 유의한 값을 보였기 때문에 AR(4) 모델로 선정하였다. 또한 강수량이 유입수 변동에 미치는 영향을 고려하기 위해 강수량에 대한 항을 모델 식에 추가하였다.

모델에 포함된 계수들의 값을 추정하기 위해 Excel의 해 찾기 기능을 사용하였으며, 2005년 1월부터 2006년 5월까지의 유입수 자료와 강수량을 사용하였다. 이로부터 구한 예측 모델식을 Table 3에 나타내었다.

3.2. 유입수 예측 시계열모델 검증

유입 유량과 각 성분의 농도별로 각각의 예측 시계열모델을 개발하였으며, 모델 개발에 사용되지 않은

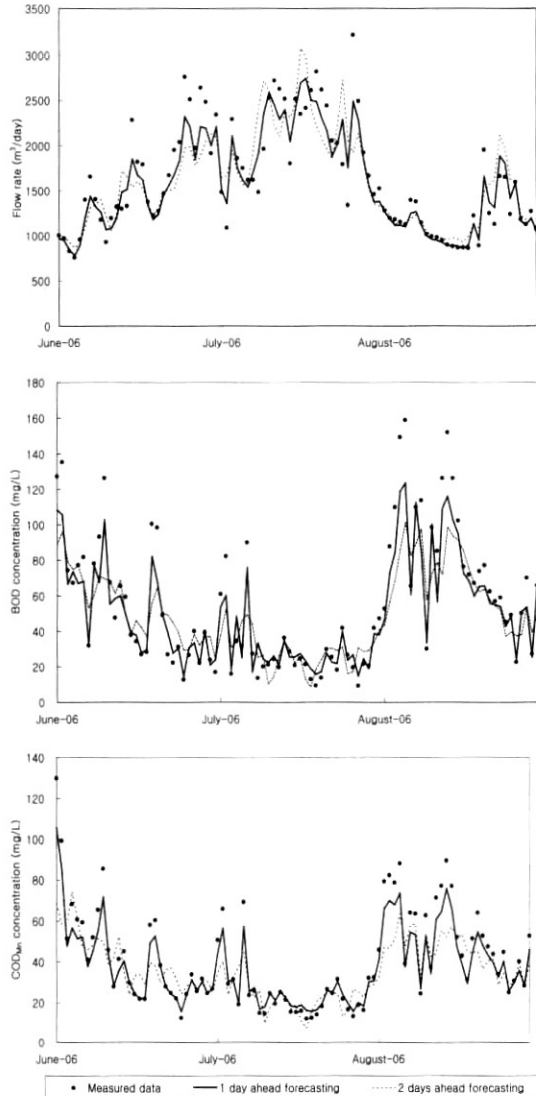


Fig. 3. Comparison of measured data, 1 day ahead forecasting results and 2 day ahead forecasting in flow rate, BOD, COD_{Mn} concentrations.

2006년 6월부터 8월까지의 3개월간의 자료를 사용하여 모델 검증을 수행하였다.

시뮬레이션을 하는 날짜를 현재로 정했을 경우, 개발된 예측모델을 이용하여 과거에 측정된 자료와 현재 측정된 자료를 사용하여 1일 후 유입 유량 및 성분 농도를 예측하였고, 과거 측정자료, 현재 측정자료, 1일 후 예측값을 사용하여 2일 후 유입 유량 및 성분 농도를 예측하였다. 실측된 유입수 정상 자료와

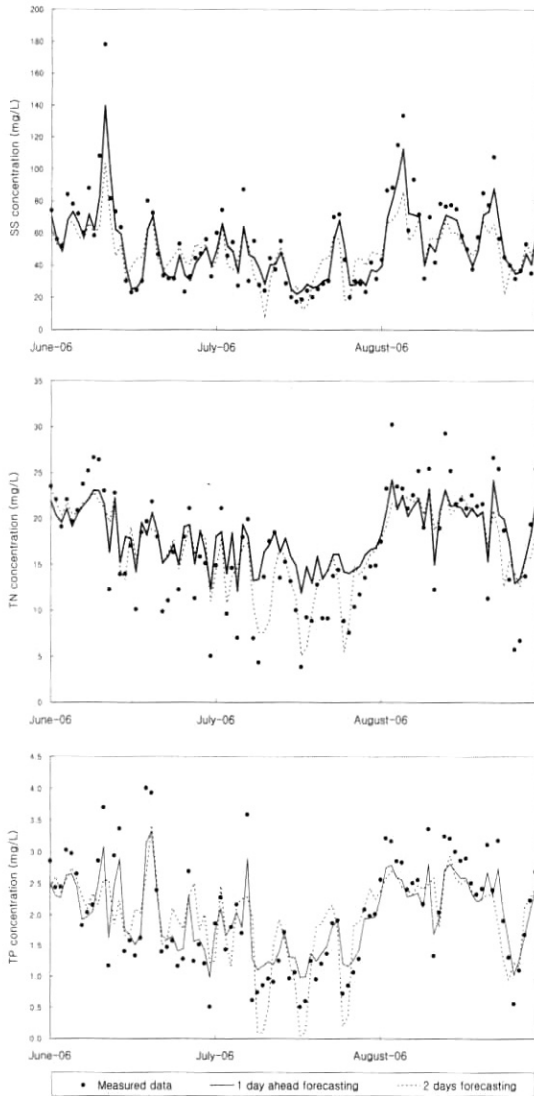


Fig. 4. Comparison of measured data, 1 day ahead forecasting results and 2 day ahead forecasting results in SS, TN, TP concentrations.

함께 시계열모델에 의해 예측된 자료를 비교하여, 모델의 적합성을 검토하고자 하였다.

예측모델을 이용한 유입수 성상 1일 및 2일 후 예측 값과 실측 값을 Fig. 3와 Fig. 4에 각각 나타내었다.

Fig. 3와 Fig. 4에서 볼 수 있듯이 1일 후 유입수 성상 예측결과와 2일 후 유입수 성상 예측결과는 실제 유입수 성상의 거동을 잘 모사하고 있음을 알 수 있으나, 자료의 값이 급변하는 구간에서는 시계열자료

Table 4. Coefficients of determination between measured data and forecasting results.

	Coefficients of determination (R^2)	
	1day ahead forecasting	2days ahead forecasting
Flowrate (m^3/day)	0.8898	0.6171
BCD (mg/L)	0.9508	0.7799
COD_{Mn} (mg/L)	0.9971	0.6923
SS (mg/L)	0.9287	0.7205
TN (mg/L)	0.9423	0.6810
TP (mg/L)	0.9849	0.5443

의 거동을 잘 모사하지 못하는 경향을 보여주고 있다. 이는 시계열 예측모델의 과거 자료의 변동을 기반으로 예측하는 모델이므로, 모델에서 고려되지 않는 자료, 즉 유입수 성상 및 강수량이 아닌 자료의 영향으로 인한 급격한 변화를 예측하는 것이 불가능하기 때문이다.

그리고 1일 후 유입수 성상 예측결과가 2일 후 유입수 성상 예측결과보다 실제 유입수 성상의 거동을 더 잘 모사하는 것으로 나타났으며, 이는 Table 4에 정당한 실제 유입수 성상과 예측결과와의 상관성을 나타낸 결정계수(coefficient of determination, R^2)에서도 잘 나타난다.

실제 유입수 성상과 1일 후 유입수 예측결과와의 결정계수는 0.8898~0.9971로 상관관계가 매우 높은 것으로 나타났으며, 실제 유입수 성상과 2일 후 유입수 예측결과와의 결정계수는 0.5443~0.7799로 상관관계가 상대적으로 낮은 값으로 나타나 1일 후 유입수 예측결과가 보다 실제 유입수 성상의 거동에 가까운 것으로 나타났다.

2일 후 유입수 성상의 예측 결과에 대한 결정계수가 상대적으로 낮다는 것은 J 하수처리장 유입수의 일별 변동이 상대적으로 크다는 것을 나타낸다. 이러한 유입수 성상의 변화가 하수처리장 유출 수질의 악화 및 공정의 처리 효율에 영향을 미칠 경우, 유입수 성상 변화를 감소시킬 수 있는 전처리 단위공정의 도입을 고려해 볼 수 있을 것이다.

결론

본 연구에서는 하수처리장의 유입수 성상을 예측

하기 위한 여러 모델들을 검토한 결과, 대상 하수처리장으로부터 얻을 수 있는 자료가 한정되어 있어, 복잡한 형태의 모델 개발이 어려운 것으로 판단되었다. 이에 하수처리장 유입수 성상의 과거 기록 자료만으로 예측 가능한 ARIMA 시계열 모델을 고려하였으며, 모델 예측성능 향상을 위해 강수량을 추가한 선형회귀 시계열모델을 개발하였다. J 하수처리장 유입수 자료를 사용하여 모델을 개발하였으며, 운전자가 쉽게 모델 개발 및 계수 보정을 할 수 있는 소프트웨어를 Excel 프로그램으로 개발하였다.

예측 시계열모델은 유입 유량과 TP는 AR(3) 모델로 개발되었고, BOD, COD, SS, T-N 농도는 AR(4) 모델로 개발되었다. 개발된 시계열모델은 유입유량, BOD, COD_{Mn}, SS, TN, TP 농도를 향후 1일, 2일간 자료의 거동을 비교적 정확히 예측하였으며, 실제 유입수 성상과 1일 후 유입수 예측결과와의 결정계수는 0.8898~0.9971로 상관관계가 높은 것으로 나타났다. 하지만 자료의 급변구간에서는 예측 결과의 신뢰성이 낮으며, 이를 해결하기 위해서는 예측 성능을 높일 수 있는 추가적인 항목을 모델에 고려하거나 유입수 성상 변동을 감소시킬 수 있는 전처리 단위공정의 추가 등을 고려해 볼 수 있다.

본 연구에서 개발된 유입수 예측모델은 실제 하수처리장 운전자에게 1, 2일 후의 유입수 성상을 예측할 수 있게 함으로써 하수처리장의 안정적인 운영을 위한 보조적인 자료로 활용될 수 있을 것으로 생각된다.

참고문헌

1. Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis, Forecasting and Control. 2nd Edn., Holden Day, San Francisco, California.
2. Carstensen, J., Nielsen, M.K. and Strandb K.H. (1998). Prediction of hydraulic load for urban storm control of a municipal WWT plant. *Wat. Sci. Tech.* 37(12), 363-370.
3. Eriksson, L., Hagbert, P., Johansson, E., Rannar, S., Whelehan, O., Astrom, A. and Lindgren, T. (2001). Multivariate process monitoring of a newsprint mill. Application to modelling and predicting COD load resulting from de-inking of recycled paper. *J. Chemometrics*, 15, 337-352.
4. Lasdon, L.S. and Waren A.D. (1978) Generalized Reduced Gradient Software for Linearly and Nonlinearly Constrained Problems, in Design and Implementation of Optimization Software, H.J. Greenberg (ed.), Sijthoff and Noordhoff.
5. Naghdy, G. and Helliwell, P. (1989). Process improvement by computer-aided load smoothing in activated sludge treatment. *Wat. Sci. Tech.*, 21, 1225-1237.
6. Krüger (2004). SMAC SMARt Control of wastewater systems Deliverable No. 18. Summary of Implementations, Report EVK1-CT-2000-00056, Krüger, Denmark.
7. Bernard O., Dantec B.L., Chahuat B., Steyer J.-P., Lardon L., Lambert S., Ratini P., Lema J., Ruiz G., Rodriguez J., Vanrolleghem P., Zaher U., Pauw D.D., Neve K.D., Lievens K., Dochain D., Schoefs O., Farina R., Alcaraz-Gonzalez V., Gonzalez-Alvarez V., Lemaire P., Martinez J., Duclaud O. and Lavigne J. (2004). Telemac: an integrated system to remote monitor and control anaerobic wastewater treatment plants through the internet. Proc. of 10th IWA World Congress on Anaerobic Digestion (AD10), Montreal, Canada, 1, 318-323.
8. El-Din, A.G., and Smith D.W. (2002). A neural network model to predict the wastewater inflow incorporating rainfall events. *Wat. Res.*, 36, 1115-1126.
9. Kim, J.R., Ko, J.H. Im, J.H., Lee, S.H., Kim, S.H. Kim, C.W. and Park, T.J. (2006). Forecasting influent flow rate and composition with occasional data for supervisory management system by time series model. *Wat. Sci. Tech.*, 53(4-5), 185-192.