

수질자료의 추이분석기법에 관한 연구

Trend Analysis Techniques for Water Quality Data

○ 백 중 철* · 허 준 행** · 조 원 철***

1. 서론

수질 시계열 자료는 비정규분포, 계절성(seasonality), 유량과의 상관성, 결측자료, 검출한계이하의 자료, 계열상관 등의 특성을 포함하므로 분석이 난해하다. 본 연구에서는 이러한 계절성, 결측자료 그리고 검출한계 이하의 자료를 포함하는 수질자료의 추이 분석에 적합한 비매개변수 검정기법(nonparametric test) 즉, 계절적 Kendall 검정기법 그리고 추이의 크기를 추정하기 위한 계절적 Kendall 변화율 추정기법을 소개하고자 한다.

수질자료의 특성을 파악하기 위해서 지난 20여년간 시계열 수질자료를 분석할 수 있는 방법을 개발하려는 연구가 많이 이루어 졌다. Steele 등(1974)은 독립적인 통계치(statistics) 즉 연평균 자료를 이용하여 수질변동의 추이를 분석한 바 있다. Lettenmaier(1976)은 종속성이 있는 수질 시계열 자료의 선형 추이(linear trend)와 축차 추이(step trend)를 분석하기 위해서 각각 비매개변수 Spearman rho 검정법 그리고 Mann-Whitney 검정법을 이용하는 기법을 제안하였다. Hirsch 등(1982)은 계절적(seasonal) 변동 특성을 갖는 수질 시계열 자료에서의 추이를 검색하기 위한 Kendall 비매개변수 검정(nonparametric test) 기법을 제안하였으며, 나자료(裸資料, raw data) 대신에 추이분석을 위해서 유량-농도 관계로부터 잔차(residuals)를 분석함으로써 유량의 변동이 농도에 미치는 영향을 고려할 수 방법을 제안하였다. Hirsch와 Slack(1984)는 수질 구성물사이의 계열상관(serial correlation)을 고려할 수 있는 일반화된 비매개변수 계절적 Kendall 검정법을 제안하였다. 그 후 Hirsch 등(1982) 그리고 Hirsch와 Slack(1984)이 제안한 기법을 수질시계열자료에 적용하여 추이를 분석한 연구가 많이 이루어 졌다(Alexander와 Smith, 1988 ; Hipel 등, 1988 ; Berryman 등, 1988 ; Walker, 1991 ; Stanley, 1993). Lettenmaier 등(1991)은 비매개변수 계절적 Kendall 검정법과 다변량 검정법을 이용하여 403개 관측점에서의 수집한 수질 시계열 자료를 분석한 바 있다.

본 연구에서는 계절성 및 결측자료를 포함하는 수질 시계열자료의 추이분석을 위한 비매개변수 검정기법 및 변화량 추정기법을 소개한 후, AR 및 ARMA 모형 등에 의해서 발생시킨 주기성 및 비주기성 추이를 포함한 자료에 대해서 선형회귀분석과 함께 비매개변수 검정기법을 적용하여 그 결과를 비교분석하고, 적용성을 파악하였다.

2. 기본이론

일반적으로 대부분의 월별 수질자료는 계절의 변화에 따른 주기성 즉, 계절성(seasonality)을 가지고 변화하므로, 이러한 자료에 대해서는 주기성을 포함하는 추이가 있는지를 분석해야 하며, 만약 추이가 존재한다면 그 변화율을 추정하게 된다.

2.1 계절적 Kendall 검정기법(Seasonal Kendall Test)

Mann(1945)은 시간에 대한 무작위성을 분석하기 위해서, 상관관계에 대한 Kendall 검정기법 즉, Mann-Kendall 검정기법(Kendall, 1975)의 특수한 형태인 비매개변수 검정기법을 처음으로 제안하였다.

Mann의 연구에 의하면 무작위(randomness)의 歸無假說(null hypothesis) H_0 는 자료(x_1, \dots, x_n)가 n 개의 독립적이고 동등하게 분포된 확률변수(random variables)의 표본임을 의미한다. 양측기각역검정(two tailed test)의 對立假說 H_1 은 x_k 와 x_j 의 분포가 $k \neq j$ 인 모든 $k, j \leq n$ 에 대해서 동일하지 않다는 것이다.

한편 Hirsch 등(1982)은 계절성(seasonality)의 존재에 대해서 민감하지 않은 계절적 Kendall 검정기법을 제안하였다. 계절적 Kendall 검정법은 Mann-Kendall 검정법의 일반화된 형태이다(Gilbert, 1987). 이 검정법에서 귀무가설(null hypothesis) H_0' 은 대립가설(alternative hypothesis) H_0 의 완화된 형태이다. X 를 하부표본

* 연세대학교 대학원 토목공학과 박사과정

**연세대학교 공과대학 토목공학과 조교수

***연세대학교 공과대학 토목공학과 교수

X_1 부터 X_{12} 까지로 구성된 전체표본이라하고 하부표본 X_i 는 i 월로부터 n_i 개의 연간 자료를 포함한다고 하면, 계절성 Kendall 검정법에 대한 귀무가설 H_0' 는 X 가 독립확률변수(x_{ij})의 표본이며 X_i 가 독립적이고 동일하게 분포된 확률변수 $i = 1, \dots, 12$ 의 하부표본이라는 것이다. 대립가설은 하나 이상의 월에 대해서 하부표본은 동일하게 분포하지 않는다는 것이다. 추정치 S_i 는 다음과 같이 정의된다.

$$S_i = \sum_{k=1}^{n_i-1} \sum_{j=k+1}^{n_i} \text{sgn}(x_{ij} - x_{ik}) \quad (1)$$

H_0' 하에서 하부표본 X_i 는 Mann 검정법에서의 귀무가설 H_0 를 만족시킨다. Hirsch 등(1982)에 의하면 S_i 에 대한 평균과 분산은 다음과 같이 정의되며, S_i 는 $n_i \rightarrow \infty$ 에 대해서 정규분포이다.

$$E[S_i] = 0 \quad (2)$$

$$\text{Var}[S_i] = \frac{n_i(n_i - 1)(2n_i + 5) - \sum_{t_i} t_i(t_i - 1)(2t_i + 5)}{18} \quad (3)$$

한편 $S' = \sum_{i=1}^{12} S_i$ 에 대한 기대치 및 분산은 다음과 같다.

$$E[S'] = \sum_{i=1}^{12} E[S_i] = 0 \quad (4)$$

$$\text{Var}[S'] = \sum_{i=1}^{12} \text{var}[S_i] + \sum_{i=1}^{12} \sum_{j=1, j \neq i}^{12} \text{cov}(S_i, S_j), \quad i \neq j \quad (5)$$

여기서 S_i 와 S_j ($i \neq j$)은 독립확률변수의 함수이므로 $\text{cov}(S_i, S_j) = 0$ 이다. 따라서 식 (5)의 $\text{Var}[S']$ 는 식 (6)과 같이 표현되며, 정규분포화에 의한 표준정규변량 Z' 은 식 (7)과 같이 정의된다.

$$\text{Var}[S'] = \sum_{i=1}^{12} \text{var}[S_i] = \sum_{i=1}^{12} \frac{n_i(n_i - 1)(2n_i - 5)}{18} \quad (6)$$

$$Z' = \begin{cases} \frac{S' - 1}{(\text{Var}(S'))^{1/2}} & \text{if } S' > 0 \\ 0 & \text{if } S' = 0 \\ \frac{S' + 1}{(\text{Var}(S'))^{1/2}} & \text{if } S' < 0 \end{cases} \quad (7)$$

증가(upward) 또는 감소(downward) 추이의 대립가설 H_1 대한 무(無) 추이의 귀무가설 H_0 를 검정하는데 있어서, 양측기각역검정(two-tailed test)의 경우, 만약 $|Z| \leq z_{\alpha/2}$ 이면 H_0 는 수용된다. 여기서 $F_N(z_{\alpha/2}) = \alpha/2$ 이며, F_N 은 표준정규가분포함수이고, α 는 검정에 대한 유의수준(significance level)이다. 양의 S' 값은 시간에 따라서 값이 증가하는 증가추이(upward trend)를 그리고 음의 S' 값은 감소추이(downward trend)를 의미한다.

2.2 계절적 Kendall 변화율 추정량(Seasonal Kendall Slope Estimator)

계절적 Kendall 변화율 추정법은 추이의 크기를 추정하는 기법으로서, 선형 추이의 변화율을 추정하는 불편의(unbiased) 기법이며, 자료가 심하게 왜곡된 경우에도 상당히 높은 정확도를 가지고 있다(1987, Gilbert).

계절적 Kendall 변화율 추정법은 계절성을 고려하기 위해서 Thell(1950)과 Sen(1968)이 제안한 방법을 일반화 한 것이다. 추이를 나타내는 시계열 자료를 분석하기 위해서는 추이의 크기(magnitude)를 추정하여야 한다. 이 크기는 변화율(시간당 변화량)로서 나타내어진다.

계절적 Kendall 변화율 추정량 B는 먼저 모든 i 개의 계절(season)에 대해서 다음의 식(8)과 같이 d_{ijk} 값을 계산한다. 여기서 $x_{ij}, x_{ik} (i = 1, 2, \dots, 12; 1 \leq k < j \leq n_i)$ 에 대해서 계산된 d_{ijk} 값들 중 중간값(median)이 계절적 Kendall 변화율 추정량이다.

$$d_{ijk} = \frac{x_{ij} - x_{ik}}{j - k} \quad (8)$$

추정량 B는 계절적 Kendall 검정 통계량 S'과 관계가 있다. 즉, S'는 양(+)의 d_{ijk} 값들의 개수에서 음(-)의 d_{ijk} 값들의 개수를 뺀 값이고 B는 d_{ijk} 값중 중간값이므로, 만약 $S' > 0$ 이면 $B \geq 0$ 그리고 $S' < 0$ 이면 $B \leq 0$ 이 된다.

3 기법의 적용

앞절에서 설명한 계절적 Kendall 검정기법과 계절적 Kendall 변화율 추정법의 적용성을 파악하기 위해서, 본 연구에서는 먼저 확률모형에 의해서 발생시킨 시계열 자료에 대해서 적용하여 두가지 기법을 검증한 후, 실측된 수질 시계열 자료에 대해서 적용해 보았다. 확률모형에 의한 시계열자료와 실제 관측된 수질 시계열자료에 대해서 기술하면 다음과 같다.

3.1 확률모형에 의해서 발생시킨 자료

본 연구에서는 식(9.1) 및 식(9.2)와 같은 두가지 추이가 없는 확률모형(Hirsch 등, 1982)에 의해서 발생시킨 시계열자료에 대해서 식(12)를 이용하여 선형 추이를 부가함으로써 추이를 갖는 시계열자료를 발생시켰다.

(1) 정규 ARMA(1,1) 모형(NARMA)

$$x_{ij} = 0.75[x_{ij}]_L + 0.97 \cdot \epsilon_{ij} - 0.57[\epsilon_{ij}]_L \quad (9.1)$$

(2) 주기성을 갖는 대수정규 AR 모형(LNARS)

$$x_{ij} = (0.5) \cdot \exp[0.22[x_{ij}]_L + 0.80 \cdot \epsilon_{ij} - 0.35] - 0.71 + \sin\left(\frac{\pi}{3} + \frac{\pi}{6} i\right) \quad (9.2)$$

시계열자료는 $i=1, 2, \dots, 12; j=1, 2, \dots, n; n=5, 10$ 에 대해서 발생시켰다. 여기서, $[x_{ij}]_L$ 은 $i = 2, 3, \dots, 12$ 에 대해서는 $i-1, j$ 를 그리고 $i = 1$ 에 대해서는 $12, j-1$ 를 나타낸다. 변수 ϵ_{ij} 는 평균이 '0'이고 분산이 1인 정규확률변수(normal random variable)이다.

이상과 같이 발생시킨 시린 시계열 자료에 대해서 식 (10)을 이용하여 선형추이를 부가함으로써, 추이를 갖는 시계열 자료를 발생시켰다.

$$v_{ij} = x_{ij} + \beta\left(\frac{i}{12} + j\right) + p \quad (10)$$

각각의 시계열자료에 대해서 3가지의 β 값을 적용하였다. 즉, $n = 5$ 에 대해서는 0.0(추이 없음)과 0.4를 그리고 $n = 10$ 에 대해서는 0.0과 0.2를 이용하였다. p는 발생시킨 수질시계열자료가 음의 값을 가지지 않도록 더해준 값으로서 본 연구에서는 3을 이용하였다. 이상과 같이 두 개의 확률모형에 의해서 발생시킨 시계열자료에 대한 왜곡도계수와 Lag 1 계열상관계수를 계산하여 정리하면 표 1과 같다.

표 1 발생시킨 자료의 통계치

자 료 내 용	NARMA ($\beta=0.0$)	NARMA ($\beta=0.2$)	NARMA ($\beta=0.0$)	NARMA ($\beta=0.4$)	LNARS ($\beta=0.0$)	LNARS ($\beta=0.2$)	LNARS ($\beta=0.0$)	LNARS ($\beta=0.4$)
자료길이 (월)	120	120	60	60	120	120	60	60
왜 곡 도 계 수	0.113	-0.122	-0.131	0.286	0.920	0.458	0.464	0.311
Lag1 계열상관계수	0.102	0.340	0.083	0.306	0.625	0.717	0.701	0.774

NARMA 모형과 LNARS 모형에 의해서 발생시킨 시계열자료를 도시하면 그림 1 ~ 그림 4와 같다.

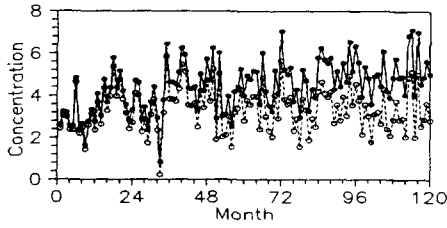


그림 1 NARMA 모형에 의해서 발생시킨 자료
(●: B=0.4, ○: B=0.0)

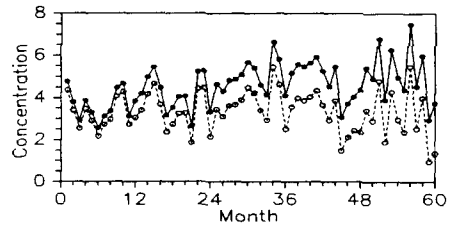


그림 2 NARMA 모형에 의해서 발생시킨 자료
(●: B=0.4, ○: B=0.0)

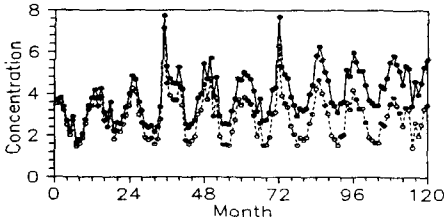


그림 3 LNARS 모형에 의해서 발생시킨 자료
(●: B=0.2, ○: B=0.0)

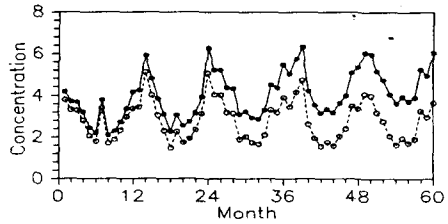


그림 4 LNARS 모형에 의해서 발생시킨 자료
(●: B=0.4, ○: B=0.0)

3.2 실측된 수질 시계열자료

본 연구에서는 실제 관측된 수질자료로서 양평수질관측소에서 관측된 9년(1985~1993)간의 수질 시계열 자료를 선정하여 적용하였다. 대상 수질항목으로는 용존산소, BOD, COD, pH의 4가지 항목을 선정하여 추이를 분석하였다. 실측 시계열자료에 대한 왜곡도계수와 Lag 1 계열상관계수를 계산하여 정리하면 표 2와 같다. 양평지점에서 관측된 4가지 수질 시계열 자료를 도시하면 그림 5, 그림 6과 같다. 그림에서와 같이 용존산소의 경우는 계절성(주기성)을 크게 나타내고 있으며, COD와 BOD는 비교적 적은 주기성을 나타내고, pH의 경우는 주기성을 나타내지 않는 것으로 나타났다.

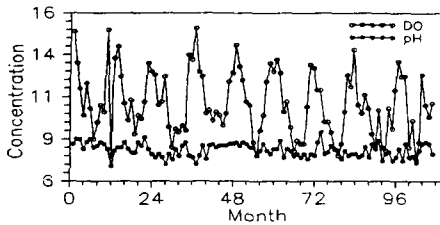


그림 5 양평지점의 DO와 pH의 시계열자료

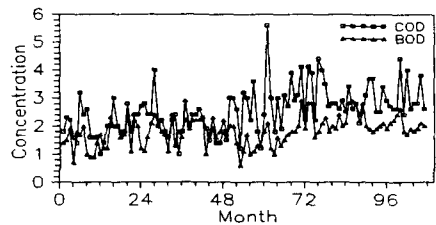


그림 6 양평지점의 COD와 BOD의 시계열자료

4 비교고찰

확률모형에 의해서 발생시킨 시계열자료와 실제 관측된 수질 시계열자료에 대해서 계절적 Kendall 검정기법과 계절적 Kendall 변화율 추정기법을 적용한 결과를 요약하면 표 3, 표 4와 같다. 표 3을 분석해 보면 $\beta=0.0$ 인 경우, NARMA 모형과 LNARS 모형에 의한 자료에 대해서 모두 $|Z| \leq z_{\alpha/2}$ 확률이 약 0.8정도이므로 귀무가설

표 2 양평지점 수질자료에 대한 통계치

내용 자료	자료의 길이 (월)	왜곡도계수	tlag1 계열 상관계수
DO	107	0.305	0.520
pH	107	-0.049	0.348
COD	107	0.718	0.320
BOD	107	-0.530	0.432

H_0 가 수용된다. 즉 추이가 없음을 나타내고 있다. 한편, $\beta=0.2$ 또는 0.4 인 경우 $|Z| \leq z_{\alpha/2}$ 확률이 0.0 으로서 귀무가설 H_0 는 기각되므로 추이가 존재함을 나타내고 있다. 다음으로 표 4를 분석해 보면, 먼저 NARMA 모형에 의해서 발생시킨 자료에 대한 계산 결과의 경우, $n=120$ 인 경우 결과치가 상당히 양호한 반면에 $n=60$ 의 경우 오차가 어느 정도 차이를 보이고 있으나 선형회귀분석에 의한 결과와 비교해 보면 오차는 약 절반 정도인 것으로 나타났다. 그리고 LNARS 모형에 의해서 발생시킨 자료에 대한 계산 결과의 경우, 계절적 Kendall 변화량 추정법에 의한 결과는 상당히 양호한 결과를 나타내고 있으며, 선형회귀분석에 의한 계산결과와 비교해 보면 오차가 약 6~18배 정도로 적은 것으로 나타났다.

한편, 표 1에서 나타난 통계치를 참고해 보면, 8개의 자료 모두에 대해서 정규분포인 것과는 상관없이 추정결과는 상당히 양호하게 나타났다. 즉 비매개변수 검정방법은 주기성(계절성)을 갖는 비정규분포 자료를 양호하게 처리할 수 있다는 것을 보여주고 있다. 그리고 계절상관계수가 큰 경우에는 계절적 Kendall검정방법(1982)에서 '0'로 처리한 공분산을 고려하여야 하지만, 이러한 경우는 40년 이상의 긴 시계열자료에 대해서만 유용하다(Hirsch와 Slack, 1984). 따라서 본 연구에서는 10년 이하의 자료를 사용하였으므로 공분산을 고려하지 않았으며, 그에 따른 계산결과는 양호한 것으로 나타났다.

표 3 발생 자료에 대한 계절적 Kendall 검정방법의 계산결과, () = β

발생자료원	자료수	Seasonal Kendall	$p(Z \leq z_{\alpha/2})$	발생자료원	자료수	Seasonal Kendall	$p(Z \leq z_{\alpha/2})$
NARMA(0.0)	120	0.232	0.816	LNARS(0.0)	120	-0.284	0.776
NARMA(0.2)	120	5.448	0.000	LNARS(0.2)	120	9.424	0.000
NARMA(0.0)	60	0.212	0.832	LNARS(0.0)	60	0.354	0.724
NARMA(0.4)	60	3.606	0.000	LNARS(0.4)	60	6.010	0.000

표 4 발생 자료에 대한 계절적 Kendall 변화율 추정량 계산결과

발생자료원	α	LL	변화율 (unit /y)	UL	LR	발생자료원	α	LL	변화율 (unit /y)	UL	LR
NARMA ($\beta=0.0$; $n=120$)	.010	-.078	.004	.089	.009	LNARS ($\beta=0.0$; $n=120$)	.010	-.027	-.002	.016	-.012
	.050	-.051	.004	.061			.050	-.022	-.002	.013	
	.100	-.045	.004	.049			.100	-.019	-.002	.010	
	.200	-.034	.004	.030			.200	-.014	-.002	.007	
NARMA ($\beta=0.2$; $n=120$)	.010	.122	.204	.289	.208	LNARS ($\beta=0.2$; $n=120$)	.010	.173	.198	.216	.187
	.050	.149	.204	.261			.050	.178	.198	.213	
	.100	.155	.204	.249			.100	.181	.198	.210	
	.200	.166	.204	.230			.200	.186	.198	.207	
NARMA ($\beta=0.0$; $n=60$)	.010	-.244	.031	.235	-.077	LNARS ($\beta=0.0$; $n=60$)	.010	-.071	.003	.072	-.048
	.050	-.162	.031	.192			.050	-.052	.003	.056	
	.100	-.146	.031	.186			.100	-.035	.003	.053	
	.200	-.123	.031	.142			.200	-.027	.003	.038	
NARMA ($\beta=0.4$; $n=60$)	.010	.156	.431	.635	.314	LNARS ($\beta=0.4$; $n=60$)	.010	.329	.403	.472	.345
	.050	.238	.431	.592			.050	.348	.403	.456	
	.100	.254	.431	.586			.100	.365	.403	.453	
	.200	.277	.431	.542			.200	.373	.403	.438	

단, α : Size of significance level. LL : Lower Limit, UL : Upper Limit, LR : 선형회귀분석 결과치

실측자료에 대해서 적용해 본 결과를 요약하면 표 5 및 표 6과 같다. 표 5에서와 같이 4가지 수질항목 모두에 대해서 $p(|Z| \leq z_{\alpha/2}) = 0.0 \sim 0.009$ 로서 귀무가설 H_0 는 기각되므로 추이가 존재하는 것으로 나타났다. 표 6을 분석해 보면 pH와 BOD는 각각 미소하게 감소하고 증가하는 추이를 보이고 있으며, DO와 COD는 각각 비교적 크게 감소하고 증가하는 추이를 보이고 있다. 즉, 양평지점은 1985년~1993년 사이에 DO -0.18mg/l/년 감소, BOD 0.07 mg/l/년 증가 그리고 COD 0.15 mg/l/년 증가하므로 수질이 악화되고 있는 것으로 나타났다. 한편, 일반적으로 하천수질은 수질항목 및 위치에 따라서 하천유량과 밀접한 관계가 있다(Smith 등, 1982). 따라서 유량을 고려하여 변형시킨 농도를 이용하여 수질의 추이를 분석하는 것이 실제 오염물질의 양적인 추이를 분석하는데 적합할 것이며, 이에 대한 연구는 추후 실시하기로 한다.

표 5 양평지점의 수질자료에 대한 계절적 Kendall 검정기법의 계산결과

수질항목	자료수	Seasonal Kendall	$p(Z \leq z_{\alpha/2})$	수질항목	자료수	Seasonal Kendall	$p(Z \leq z_{\alpha/2})$
DO	107	-3.741	.000	COD	107	5.133	.000
pH	107	-2.628	.009	BOD	107	4.119	.000

표 6 양평지점의 수질자료에 대한 계절적 Kendall 변화율 추정량 계산결과

수질항목	α	LL	변화율 (unit /y)	UL	수질항목	α	LL	변화율 (unit /y)	UL
DO	.010	-.300	-.182	-.050	COD	.010	.080	.154	.239
	.050	-.266	-.182	-.067		.050	.100	.154	.224
	.100	-.250	-.182	-.076		.100	.100	.154	.200
	.200	-.233	-.182	-.100		.200	.117	.154	.200
pH	.010	-.100	-.041	.000	BOD	.010	.025	.071	.125
	.050	-.075	-.041	.000		.050	.036	.071	.100
	.100	-.074	-.041	-.015		.100	.040	.071	.100
	.200	-.067	-.041	-.025		.200	.050	.071	.100

5 결론

계절성(주기성)을 갖는 수질 시계열자료를 분석하기 위해서 계절적 Kendall 검정기법과 계절적 Kendall 변화량 추정기법을 적용해 본 결과 다음과 같은 결론을 얻었다.

1. 계절적 Kendall 검정기법을 적용해 본 결과 수질 시계열자료에서의 추이의 존재 여부를 정확하게 파악할 수 있는 것으로 나타났다.

2. 계절적 Kendall 변화량 추정기법은 단순 추이(monotonic trend)에 대해서 양호한 추정 결과를 보이고 있으며, 선형회귀분석에 비해 상당히 높은 정확도를 가지는 것으로 나타났다.

한편, 수질은 유량과 밀접한 관계가 있으므로 추후 연구과제로써는 수량을 고려한 수질의 추이를 분석하고자 한다.

6 참고문헌

- Berryman, D., B. Bobee, D. Cluis, and J. Haemmerli, Nonparametric tests for trend detection in water quality time series, *Water Resources Bulletin* 24(3) 545-556, 1988
- Gilbert, R. O., *Statistical methods for environmental pollution monitoring*, VAN NOSTRAND REINHOLD, New York, 1987.
- Hipel, K. W., A. I. Mcleod, and R. R. Weiler, Data analysis of water quality time series in lake Erie, *Water Resources Bulletin* 24(3) 533-544, 1988
- Hirsch, R. M., and J. R. Slack, A nonparametric trend test for seasonal data with serial dependence, *Water Resour. Res.*, 20(6), 727-732, 1984.
- Hirsch, R. M., and J. R. Slack, and R. A. Smith, Techniques of trend analysis for monthly water quality data, *Water Resour. Res.*, 18(1), 107-121, 1982.
- Lettenmaier, D. P., E. R. Hooper, C. Wagoner, and K. B. Faris, Trends in stream quality in the continental United States, 1978-1987, *Water Resour. Res.*, 27(3), 327-339, 1991.
- Smith, R. A., R. M. Hirsch, and J. R. Slack, A study of trends in total phosphorus measurements at NASQAN station, U.S.G.S., *Water Supply Paper.*, 2190, 1982.
- Walker, W. W., Water quality trends at inflows to everglades national park, *Water Resources Bulletin* 27(1) 59-72, 1991