

推計學的 Modelling에 必要한 合成資料의 發生

柳明辰·金聲得*·黃成一**

土木工學科

〈요 약〉

대기온도, 수온 같은 時系列을 推計學的으로 분석함으로써 자료를 합성할 수 있는 模擬發生모델을 시도하였다. 周期成分은 Fourier급수 형태로 표시하였고, 推計學的的成分은 1차 Markov모델로 나타내었다. 일반적으로 수온자료는 회귀하거나 단기인 것이 보통인데 대기온도와 수온과의 관계를 이용하여도 수온을 발생시킬 수 있을 것이다.

Generation of Synthetic Data for Stochastic Modelling in Water Pollution Control

Myong Jin Yu, Seong Deuk Kim and Sung Il Hwang

Dept. of Civil Engineering

〈Abstract〉

This paper presents a model for synthesising daily average air temperature and water temperature data. Periodic components are described by the Fourier series and stochastic components are produced using a first order Markov model. A cross correlation model can be used to generate water temperatures, providing correlation coefficient is high.

I. 서 론

推計學的 模擬發生技法이 水文學 분야에서 기록연한이 짧은 流量資料를 확충하는 방법으로 많이 연구되어 왔다. 環境工學者의 관심의 대상에도 水質觀測記錄, 廢水放出量, 大氣汚染記錄 같은 연속적인 기록이 많이 있다. 이러한 時系列의 분석은 현상의 이해에 대한 통찰력만 아니라 미래의 추세에 대한 예측에 도움이 될 것이다.

대기온도 및 수온의 예측은 환경공학자뿐만 아니라 기상학자, 호소학자 등이 오랫동안 관심을 가져왔다. 수온은 수질 특성을 나타내는 가장 중요한 변수 중의 하나라써 수온이 증가함에 따라 溶存酸素의 濃度는 감소하는 반면 生物의 新陳代謝는 활발

하여져서 용존산소의 농도를 더욱 감소시킨다. 水處理의 경우에도 수온변화의 영향은 매우 크다. 차츰 발전소의 冷却水熱에 의한 河川水와 海水의 수온 상승과 이에 따른 水質 및 水中生態系에 대한 영향에 관심이 증대되고 있으며 발전량의 증가는 더욱 방출되는 廢熱量을 증가시킬 것이다.

인위적 요인에 의한 수온의 변화에 대한 연구에 앞서 자연적 요인에 의한 수온의 변화에 대한 연구가 선행되어야 할 것이다. 하천 수온은 대기온도와 주로 상관관계가 있고 대기온도 자료를 이용하여 수온을 예측하는 방법이 시도된 바 있으며, (1), (2), (3) 해수의 온도도 이와 유사할 것으로 예상된다. 본 논문에서는 수온과 대기온도 기록을 통계학적으로 분석함으로써 각각의 발생모델을 만들고, 대기온도와 수온의 관계를 이해함으로써 기록연한이 일반적으로

*蔚山工專 土木科

**國民大學 土木工學科

짧은 수온자료의 합리적인 확충방법을 모색하고자 한다.

수온자료로는 울산만에 위치한 한국전력울산화력 발전소에서 냉각용수 취수시 측정된 자료를 이용하였다. 日平均氣溫은 매시간 관측치를 평균한 값으로 하였으며 1972년부터 1977년까지 5년간의 자료를 이용하였다. 대기온도 자료는 중앙관상대 발행 기상월보로부터 1961년부터 1977년까지 16년간의 일평균 대기온도를 이용하였다.

II. 분 석

1. 週期成分 및 長期變動成分

대기온도와 수온의 변화는 지구의 지축에 대한 자전(일간변동)과 태양 주위의 공전(계절성분)에 기인한다고 할 수 있으며, 그 변화는 평균치, Fourier급수와 잔류치의 합으로 표시할 수 있을 것이다.

$$x_i = \bar{x} + \sum_{i=1}^m A_i \sin(i\omega t) + \sum_{i=1}^m B_i \cos(i\omega t) + x_{r,i} \quad (1)$$

여기서 \bar{x} = 시계열의 평균

$x_{r,i}$ = x_i 의 잔류성분

A_i, B_i = i 번째 Fourier계수

m = 고려할 harmonic의 수

그림 1, 그림 2는 3년간의 대기온도 및 수온변화이며 실선은 평균과 4개의 harmonic을 합친 것이다.

이외에 시계열은 자연적인 기후변화나 인위적인 관리계통의 변경에 의하여 장기적인 성형이 포함될 수 있다. 예로써 발전소 건설에 따른 냉각수의 방출은 하류의 수온을 상승시킬 것이다. 일반적으로 이용할 수 있는 대기온도 또는 수온기록은 장기변동 성분(trend component)을 감지할 만큼 길지는 못하다. 장기변동성분은 調和分解(Fourier analysis) 전에 제거되어야 할 것이며 장기회 합요가 모으면

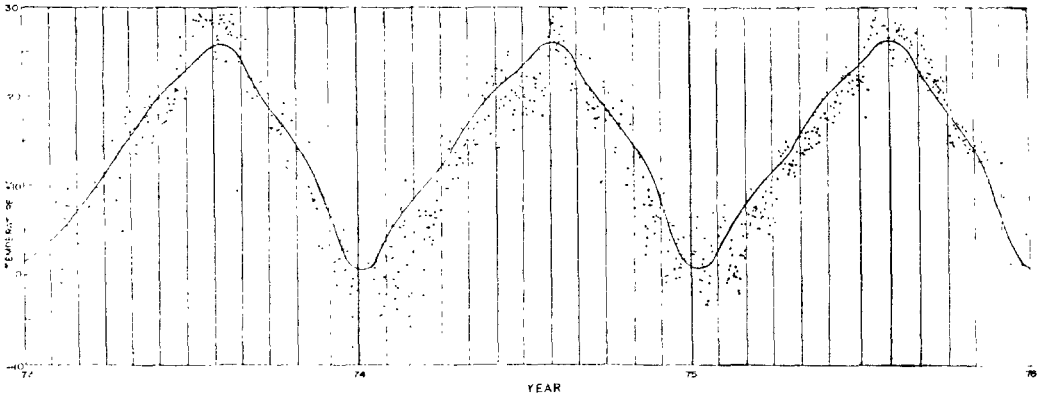


그림 1. 대기온도 기록

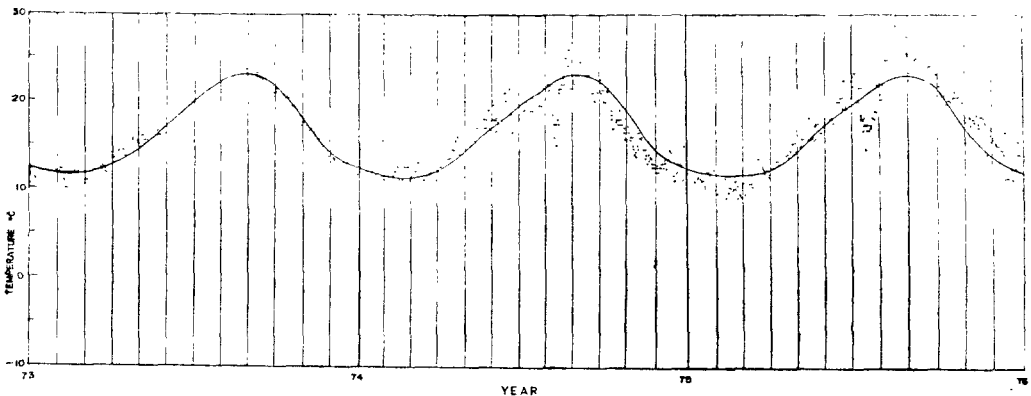


그림 2. 수온 기록

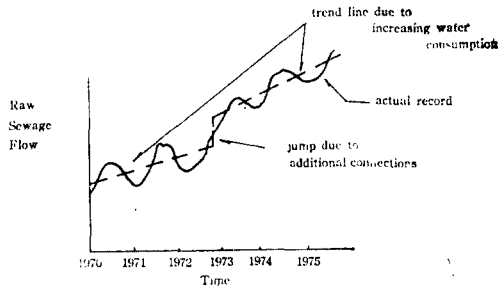


그림 3. 추세와 급전을 보이는 하수량의 변화

통계적 방법에 의하여 急轉(jump)과 趨勢(trend)를 구할 수 있을 것이다. 그림 3은 下水량의 변화에 있어 인구증가와 소비수량의 증가에 따른 증가추세와 배수구역의 확장에 따른 급전을 보인다.

2. 調和分解(Fourier analysis)

조화분해에 있어 Fourier계수는 다음과 같이 구할 수 있다.

$$A_i = \frac{2}{n} \sum_{t=1}^n x_t \sin(i\omega t) \quad (2)$$

$$B_i = \frac{2}{n} \sum_{t=1}^n x_t \cos(i\omega t) \quad (3)$$

여기서 n =자료의 수

t =시간(일)

ω =첫번째 harmonic의 각속도

i 번째 harmonic을 하나의 cosine함수로 표시하면 (1)식은

$$x_t = \bar{x} + \sum_{i=1}^m C_i \cos(i\omega t - \phi_i) + x_{r,t} \quad (4)$$

여기서 $C_i = \sqrt{A_i^2 + B_i^2}$

$\phi_i = \arctan(A_i/B_i)$

i 번째 harmonic에 의하여 설명되는 分散(variance)는

$$i < \frac{n}{2} \text{ 일때 } \sigma_i^2 = \frac{C_i^2}{2} \quad (5)$$

$$i = \frac{n}{2} \text{ 일때 } \sigma_i^2 = C_i^2 \quad (6)$$

조화분해는 대부분의 경우 존재할 것으로 기대되는 harmonic에 맞추어 행하는 것이 타당하다고 생각된다. 존재할 harmonic의 주기는 spectral analysis 또는 自己相關分析(autocorrelation analysis)에 의하여 찾을 수 있는데 Quimpo는 수문자료 분석에서 두 방법을 비교하여 자기상관분석이 이 경우에 값이 덜 소모되고 잔류성분을 분기하는데까지 영

장될 수 있다는 利點을 말했다. (6) 이러한 관점에서 본 논문에서 자기상관분석은 이용하려 한다.

相關圖(correlogram)은 차수(lag) k 에 대하여 系列相關係數(serial correlation coefficient)를 표시한 것으로 k 에 대한 계열상관 계수는 다음과 같다

$$r_k = \frac{1}{n-k} \frac{\sum_{t=1}^{n-k} x_t \cdot x_{t+k}}{S^2} - \left(\frac{1}{n-k} \right)^2 \frac{\sum_{t=1}^{n-k} x_t \cdot \sum_{t=1}^{n-k} x_{t+k}}{S^2} \quad (7)$$

8544개의 대기온도 자료와 1827개의 수문 자료에 대하여 $k=1095$ 까지의 계열상관계수를 구하였다. 그림 4(a)는 대기온도에 대한 상관도이고 수문의 경우에는 유사한 상관도를 얻었다. 그림에서 365의 배수의 차수마다 규칙적으로 peak가 있으므로 일년주기성분이 있다는 것을 알 수 있다. 따라서 기본주기가 365일인 Fourier series로 표시하면 確定論的成分(deterministic component)은

$$y_t = \bar{x} + \sum_{i=1}^m \left(A_i \cos \frac{2\pi i t}{365} - B_i \sin \frac{2\pi i t}{365} \right) \quad (8)$$

설명에 필요한 subharmonic의 수를 결정하기 위하여 i 를 하나씩 증가시켜서 얻은 殘留數列(remainder series)는

$$z_t = x_t - y_t \quad (9)$$

이 잔류수열에 대하여 계열상관계수 r_k 를 차례로 계산하여 상관도를 그리면 그림 4와 같게 된다. 그림 4에서 4개의 harmonic을 제거한 후에는 상관도의 규칙적 변화는 거의 식별할 수 없고 따라서 Fourier series 표시에 4개의 harmonic으로 충분하다 할 수 있다.

이제 시계열은 確定論的成分(deterministic component)와 推計學的的成分(stochastic component)을 다음과 같이 표시할 수 있다.

$$x_t = \underbrace{\bar{x} + \sum_{i=1}^4 C_i \cos(i\omega t - \phi_i)}_{\text{확정론적 성분}} + \underbrace{x_{r,t}}_{\text{추계학적 성분}} \quad (10)$$

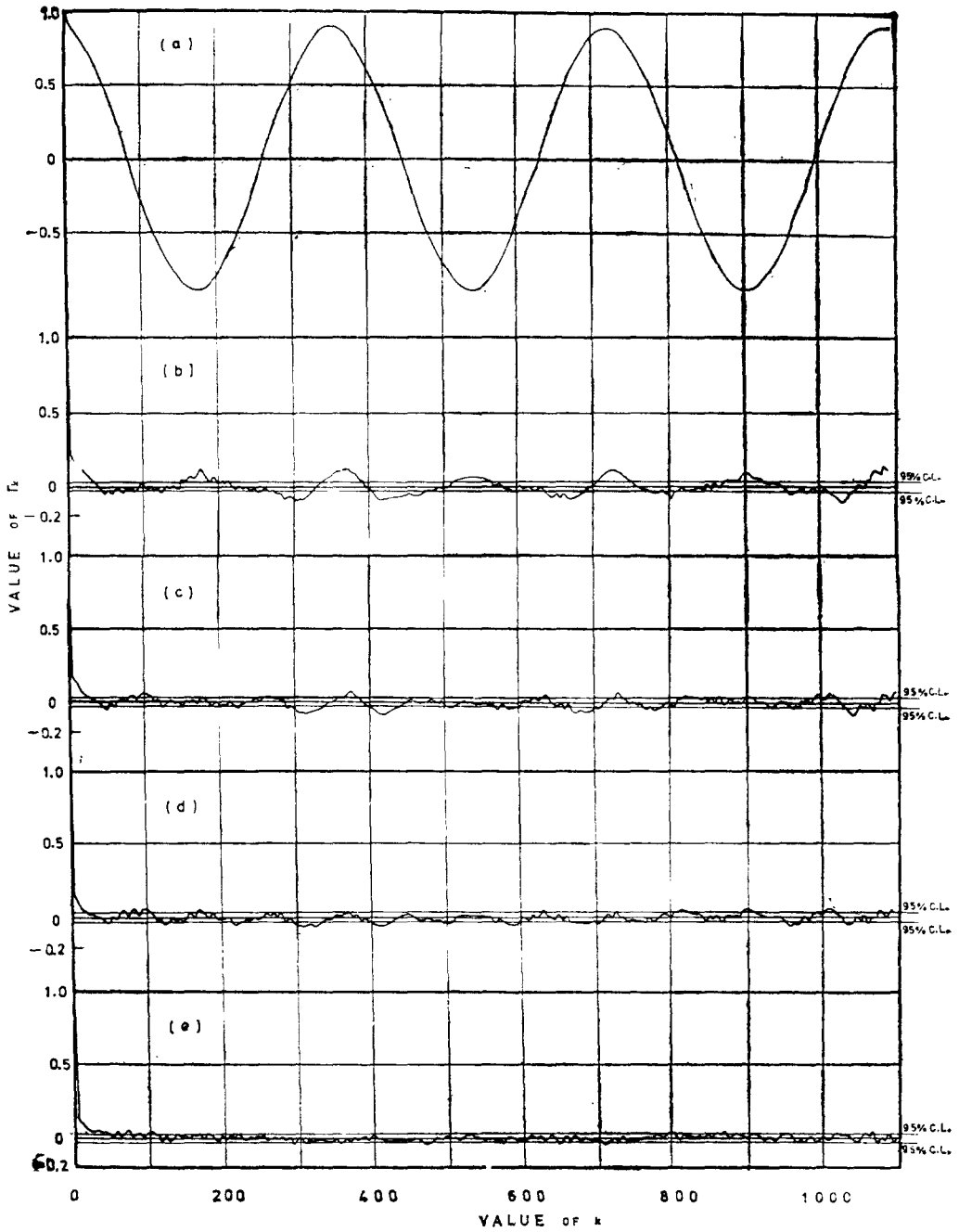


그림 4. 대기온도 자료의 상관도 : (a) 원시계열 : (b) 365일 주기 제거 후 : (c) 365일 주기와 하나의 subharmonic 제거 후 : (d) 365일 주기와 두개의 subharmonic 제거 후 : (e) 365일 주기와 세개의 subharmonic 제거 후

표 1. 조화분해 결과

	대 기 온 도				수 온			
	$A_i(^{\circ}C)$	$B_i(^{\circ}C)$	$C_i(^{\circ}C)$	$\phi_i(^{\circ})$	$A_i(^{\circ}C)$	$B_i(^{\circ}C)$	$C_i(^{\circ}C)$	$\phi_i(^{\circ})$
1차 harmonic	-5.2	-10.8	12.0	25.8	-4.2	-4.0	5.8	46.4
2차 harmonic	0.6	-0.9	1.1	-31.1	0.7	-0.5	0.8	-52.1
3차 harmonic	-0.5	0.2	0.6	-70.3	0.4	0.3	0.5	57.1
4차 harmonic	0.4	-0.6	0.7	-34.8	-0.1	0.1	0.2	-33.4

표 2. Harmonic 증가와 평균 분산 표준편차

	대 기 온 도			수 온		
	평균($^{\circ}C$)	분산($^{\circ}C^2$)	표준편차($^{\circ}C$)	평균($^{\circ}C$)	분산($^{\circ}C^2$)	표준편차($^{\circ}C$)
원시 계열	13.2	81.4	9.0	16.5	20.7	4.6
1차 harmonic 제거후	0.0	8.7	2.9	0.0	3.6	1.9
2차 harmonic 제거후	0.0	8.1	2.8	0.0	3.3	1.8
3차 harmonic 제거후	0.0	7.9	2.8	0.0	3.2	1.8
4차 harmonic 제거후	0.0	7.6	2.7	0.0	3.2	1.8

4개의 harmonic을 계산하였을 때의 결과는 표1, 표2와 같고, 365일 주기인 첫번째 harmonic의 amplitude가 가장 크고 전체 분산의 대기온도의 경우 90%, 수온의 경우 83%를 타내고 있다.

3. 推計學的的成分(stochastic component)의 분석

잔류성분 시계열 $x_{r,t}$ 는 상관도가 주기성분을 보이는 중요성분(significant component)을 그 이상 나타내지 않을 때까지 subharmonic을 제거함으로써 구할 수 있다. 중요성(significance)를 시험하기 위한 신뢰한계(confidence limits)는 다음과 같이 계산된다.

$$C.L.(\rho_1) = \frac{-1 \pm Z_{\alpha}(n-2)^{\frac{1}{2}}}{n-1} = \frac{-1 \pm 1.64(n-2)^{\frac{1}{2}}}{n-1} \quad (11)$$

여기서 Z_{α} 는 α sinificance level에 해당하는 normal deviate이며 α 가 5%일때 1.64이다. 대기온도 자료에서 4번 subharmonic을 제거한 후에는 그림 4(e)와 같이 95% 신뢰구간(confidence limit)안에 들었으며, 수온자료에서도 이와 유사한 결과를 보였다. 이제 4번의 subharmonic을 제거한 후의 잔류성분에 대하여 추계학적 분석을 행한다.

Kottogoda 는 수문자료의 잔류성분의 해석에서 1

차 Markov model에 의하여 만족한 결과를 얻었고 환경문제에도 적당하리라 생각된다.⁽¹⁶⁾

1차 Markov model의 일반적인 식형은^{(10), (12)}

$$x_{r,t} = a_0 + a_1 x_{r,t-1} + \delta_t \quad (12)$$

또는 $x_{r,t} = m_t + r_1(x_{r,t-1} - m_t) + \delta_t$

여기서 a_0, a_1 = 自己 回歸係數(autoregressive coefficient)

m_t = 평균

잔류성분은 Chi 검정 또는 Kolmogorov-Smirnov 검정에 의하여 표시할 수 있는 分布가 조사되어야 하며, 대기온도 및 수온자료의 $x_{r,t}$ 열은 Kolmogorov-Smirnov 검정에 의하여 正規分布(normal distribution)를 나타내었다. 標準化(standardization)된 자료에서는 평균 즉 $m_t = 0$ 임으로

$$z_{r,t} = r_1(z_{r,t-1}) + \delta_t \quad (13)$$

여기서 $z_{r,t}$ = 표준화된 자료

즉 $m_t = 0$ 이면 $a_0 = 0, a_1 = r_1$ 이다.

$z_{r,t}$ 가 정규분포이면 δ_t 도 정규분포이고 자료의 변환에 의한 定常時系列의 가정으로부터 δ_t 는 일정한 분산을 가지고 $z_{r,t-1}$ 에 좌우되지 않는다. $z_{r,t}$ 의 분산은 δ_t 의 分散(variance)과 다음과 같은 관계를 가진다.

$$\sigma_z^2 = \sigma_{\delta}^2(1 - r_1^2) \quad (14)$$

여기서 σ_z^2 = random 성분 δ_t 의 분산

σ_{δ}^2 = $z_{r,t}$ 의 분산

만약 t_u 가 평균이 0이고 분산이 1인 정규분포의 亂數(random variable)라면 $t_u \sigma_z = t_u \sigma_z \sqrt{1-r_1^2}$ 은 평균이 0이고, 분산이 $\sigma_z^2(1-r_1^2)$ 과 같은 정규분포의 random變量이다. 따라서 (13)식은 다음과 같이 변형시킬 수 있다.

$$z_{r,t} = r_1 z_{r,t-1} + t_u \sigma_z \sqrt{1-r_1^2} \quad (15)$$

여기서 $\sigma_z=1$ 이다. 여기서 다시 標準化 變換式을 대입하면,

$$r_1 z_{r,t-1} + t_u \sqrt{1-r_1^2} = z_{r,t} = \frac{x_{r,t} - \bar{x}_{r,t}}{\sigma_x} \quad (16)$$

따라서

$$x_{r,t} = \bar{x}_{r,t} + \sigma_x (r_1 z_{r,t-1} + t_u \sqrt{1-r_1^2}) \quad (17)$$

잔류성분의 평균 $\bar{x}_{r,t}$ 는 보통 0이므로 (17)식은 다시

$$x_{r,t} = r_1 x_{r,t-1} + e_t \sqrt{1-r_1^2} \quad (18)$$

여기서 e_t 는 평균이 0이고 분산이 $x_{r,t}$ 와 같은 정규분포의 random變量이다.

Markov모델에 의한 모의발생이 정당하려면 e_t 列이 완전히 random하여야 하는데, 그림 5 e_t 列의 상관도에 의하여 완전히 random하다는 것을 알 수 있다. 따라서 이 방법에 의한 모의발생이 타당하다고 생각되고 이제 時系列의 발생을 위한 완성된

식은

$$x_t = \bar{x} + \sum_{i=1}^m C_i \cos(i\omega t - \phi_i) + r_1 x_{r,t-1} + e_t \sqrt{1-r_1^2} \quad (19)$$

(19)식에 의한 x_t 의 발생을 위하여 크기 $x_{r,t-1}$ 의 값은 0으로 한다.

대기온도에 있어 原資料와 合成된 資料의 전체와 잔류성분의 통계적 성질을 비교하면 표3과 같고, 만족할 만한 모의발생 은도자료가 얻어짐이 판명되었다.

표 3. 대기온도의 원자료와 합성자료의 통계적 비교

	원 자료		합성 자료	
	원자료	잔류성분	합성 자료	잔류성분
평균(°C)	13.2	0.0	12.6	-0.7
분산(°C ²)	81.4	7.6	81.7	7.9
표준편차(°C)	9.0	2.7	9.0	2.8

4. 대기온도와 수온의 관계에 의한 모의발생

수온 자료는 (19)식에 의하여 발생시킬 수 있겠으나, 다른 방법으로 상관관계모델(cross correlation model)을 이용할 수 있을 것이다. 즉 내기온

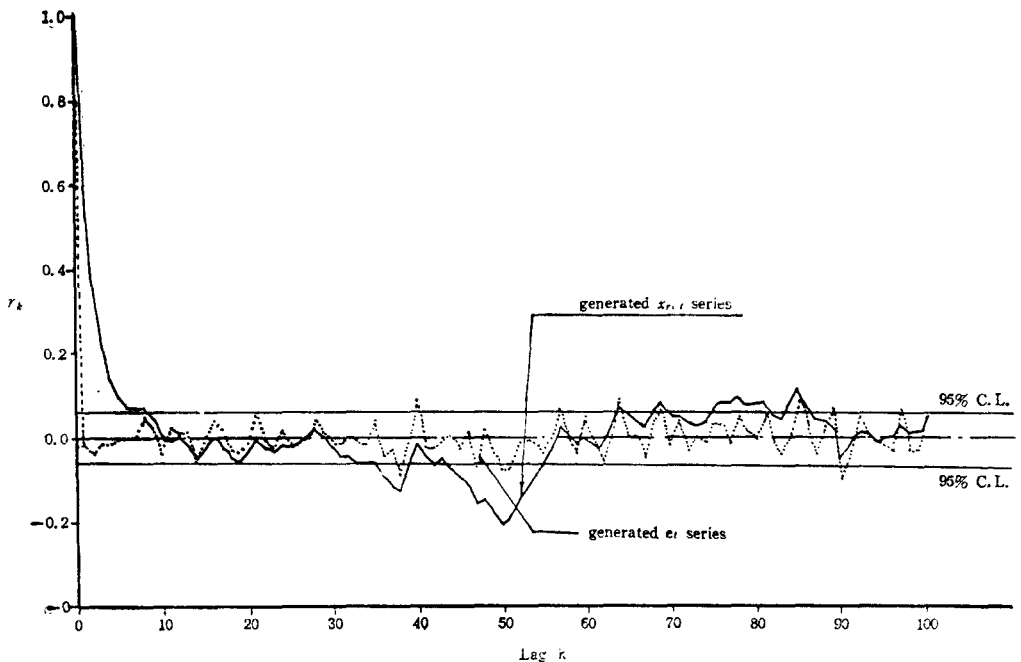


그림 5. 합성자료의 $x_{r,t}$ 와 e_t 列의 상관도

도와 수온의 상관계수(cross correlation coefficient)가 높으면 비교적 장기간의 기록이 있는 대기 온도 자료의 통계학적 특성을 이용하여 수온자료를 확충할 수 있을 것이다. 수온의 確定論的 成分은 수온자료의 調和分解로써 구하고, 偶然變動成分(random component)에 대기온도의 殘留値와 수온의 殘留値와의 相關關係를 도입하여 단기기록에서 불충분하게 분석될 가능성이 있는 우연변동성분을 장기기록의 통계적 특성으로 보충하려는 시도이다. 결과적으로 다음과 같은 모의 발생식을 만들 수 있다. (3)

$$x_t = \bar{x} + \sum_{i=1}^m C_i \cos(i\omega t - \phi_i) + r_1 x_{t-1} + (1-r_1^2)^{\frac{1}{2}} \{be_i' + f_i(1-\rho_c^2)^{\frac{1}{2}}\} \quad (20)$$

여기서 e_i' = 대기온도의 잔류치열

ρ_c = e_i 열과 e_i' 열과의 상관계수

e_i = 수온의 잔류치열

b = e_i' 에 대한 e_i 의 回歸直線의 기울기

f_i = 평균이 0이고 표준편차가 e_i 와 같은 無作為列(random series)

본 논문에서 이용한 자료에서는 대기온도와 수온의 잔류치열의 상관관계가 0.2로 대단히 낮았다. 이는 海水의 온도는 대기온도의 변화뿐 아니라 潮流, 바람이 일으킨 흐름 등의 영향을 받을 뿐 아니라 여기서는 冷却水를 取水하는 지점이 사용후 放流하는 지점과 접근하여 있어 경우에 따라서는 廢熱水의 영향도 있었기 때문일 것이다. 따라서 해수온도의 변화는 이러한 요소들을 다 포함하여 回歸分析을 하여야만 예측이 가능할 것이다. 그러나 대기온도 이외에는 연속된 기록이 없어 시도하지 못하였다. 하천수의 경우에는 대기온도의 영향의에는 별로 없을 것이고 (20)식의 사용이 가능할 것이다.

III. 결 론

두가지의 系列模擬發生(sequential simulation)이 시도되었는데 정확하게 數學的으로 模型化하는 것이 어려운 경우 유용한 방법이라 생각된다. 이러한 기법들은 水文에서 뿐만 아니라 環境問題에서도 유용한 수단이 될 것이다. 예로 大氣汚染測定資料, 下水量 및 廢水強度의 變化, 水質測定記錄의 분석에 이용하면 長期間의 變動性向과 季節에 따른 變化 등을 이해하는데 도움이 될 것이다.

推計學的 模擬發生은 平均的인 確定論的 成分뿐 아니라, 系列相關成分과 偶然變動成分까지 고려함으

로써 確定論的 模型(deterministic model)의 단점을 보충할 수 있을 것이다.

相關關係를 이용한 모델은 相關係數가 낮아 시도하지 못하였다. 河川水溫과 대기온도의 상관관계는 海水의 경우보다 높으리라 예상되는데 수온은 수질 특성을 나타내는 중요한 자료인데도 하천에서 연속 기록이 희귀하였다. 海水의 온도의 경우 대기온도의 변화 뿐 아니라 風速, 潮流 등 다른 변수를 추가하여 연구하면 좀더 나은 모델이 개발될 수 있을 것이다.

謝 辭

본 연구는 產學協同財團의 지원으로 이루어졌으며, 이에 지면을 통하여 감사를 드립니다.

참 고 문 헌

1. Kothandaraman, V., Analysis of Water Temperature Variations in Large River, Journal of the Sanitary Engineering Division, ASCE, SAI, pp.19~31 (1971)
2. Song, C.C.S. Pabst, A.F. & Bowers, C. E., Stochastic Analysis of Air and Water Temperatures, Journal of the Environmental Engineering Division, ASCE, EE6, pp.785~880, (1973)
3. Green, N.M.D., Statistical Techniques in the Field of Water Pollution Control, Proceedings of Symposium, University of Newcastle upon Tyne, England (1973)
4. Rich, L.G., Environmental Systems Engineering, pp.214~253, McGraw-Hill, New York (1973)
5. Quimpo, R.G., Autocorrelation and Spectral Analysis in Hydrology, Journal of the Hydraulics Division, ASCE, HY2, pp.363~373 (1968)
6. Anderson, R.L., Distribution of Serial Correlation Coefficient, The Annals of Mathematical Statistics, vol. 13, pp.1~13 (1942)
7. Kisiel, C.C., Time Series Analysis of Hydrologic Data, Advances in Hydroscience

- vol. 5, pp. 1~117, Academic Press, New York, (1969)
8. Roesner, L. A. & Yevjevich, V. M., Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff, Hydrology Paper 15, Colorado State University, Fort Collins, Colorado (1966)
 9. Quimpo, R. G., Stochastic Model of Daily River Flow Sequences, Hydrology Paper 18, Colorado State University, Fort Collins, Colorado (1967)
 10. Adamowski, K. & Smith, A. F., Stochastic Generation of Rainfall, Journal of the Hydraulics Division, ASCE, HY11, pp. 1935~1945, (1972)
 11. Fiering, M. B., Streamflow Synthesis, Macmillan, London (1967)
 12. 李舜鐸, 河川流量의 模擬發生에 관한 推計學的 研究, 博士學位論文, 高麗大學校大學院 (1974)
 13. Yevjevich, V. M., Stochastic Processes in Hydrology, Water Resources Publications, Fort Collins, Colorado (1972)
 14. 金聲得·陳致慶, 釜山地方의 水溫變化에 관한 연구, 蔚山工專 論文集 제2집 (1977)
 15. 尹龍男, 水文學, pp. 353~384, 淸文閣 (1976)
 16. Kottegoda, N. T., Statistical Methods of River Flow Synthesis for Water Resources Assessment, Proc. Inst. Civ. Eng. Supplement 18 (1970)