〈論 文〉

漢江 下流部의 水質變動에 對한 推計學的 特性(Ⅰ)

——특히 뚝도 및 노량진 地點의 DO, 濁度, 水溫의 變動을 中心으로——

Stochastic Properties of Water Quality Variation in Downstream Part of Han River

李 弘 根*
Hong Keun, Lee

ABSTRACT

The stochastic variations and structures of time series data on water quality were examined by employing the techniques of autocorrelation function, variance spectrum, Fourier series, autoregressive model and ARIMA model. These time series included hourly and daily observations on DO, turbidity, conductivity pH and water temperature. The measurement was made by automatic recording instrument at Noryangjin and Dook-do located in the downstream part of Han River during 1975 and 1976.

Hourly water quality time series varied with the dominant 24-hour periodicity, and the 12-hour periodicity was also observed. An important factor affecting 24-hour periodic variation of DO is believed to be photosynthesis by algae. These phenomena might be attributable to periodic discharges of municipal sewage. Noryangjin site showed the more distinct 12-hour periodicity than Dook-do site did, and tidal effect might be responsible for the difference.

The water quality, as measured by DO and turbidity, was better in the afternoon compared with the quality in the morning. This change can be explained by the periodic variation of DO, temperature and the amount of municipal sewage discharge.

It was also observed that the water temperature at Noryangjin was higher than the temperature at Dook-do. This difference might have been caused by the pollutants that were added to the section between two sites.

The correlation coefficients between some of the variables were fairly high. For example, the coefficient was −0.88 between DO and water temperature, 0.75 between turbidity and river flow, and 0.957 between water temperature and air temperature. The lag time of heat transfer from the air to the water was estimated as 23 days.

The first order auto-regressive model was appropriate for explaining standardized hourly DO time series. The ARIMA model of (1, 0, 0) type provided relatively satisfactory results for daily DO time series after the removal of significant harmonic values.

記 號 目 錄

記 號	定 義	記 號	定 義
X_t	原時系列	ρ_k	지체 k에 對한 自己相關係數
$X_{T(t)}$	X_t의 傾向成分	γ_k	지체 k에 對한 自己相關係數의 推定値
$X_{P(t)}$	X_t의 週期成分, 調和函數	C_k	지체 k에 對한 自己共分數
ε_t, a_t	確率成分 또는 殘差成分	m	지체 k의 最大값

*本學會 編輯委員, 서울大學校 保建大學院 副敎授 工博

N	資料數
$L(\alpha)$	基準 $\alpha(\%)$에 對한 γ_k의 信賴限度
A_j, B_j	j調和函數에 對한 Fourier係數
C_j	j調和函數의 振幅
$C_j^2/2$	j調和函數의 分散
θ_j	j調和函數의 位相角
T_j,	j調和函數의 週期
P_m	全體分散에 對한 m調和函數까지의 說明分散比
f_j	j調和函數에 對한 頻度
$I_{(f_j)}$	頻度 f_j에 對한 時系列 資料의 spectral density
$I_{(f)}$	頻度領域에서 任意의 頻度 f에 對한 母集團의 spectral density
$V_{(f)}$	$I_{(f)}$의 推定値
V_k, U_k	지체 k인 時間領域의 spectral density 및 平滑化된 spectral density의 推定値
m_τ, S_τ	τ時에 對한 X_t의 平均과 標準偏差
m_t, S_t	m_τ와 S_τ에 對한 調和函數
$X_{p,\tau}$	p日째 τ時의 X_t
TB	濁度
WT	水溫
CT	電導率
WD	流量
M	時系列의 平均分散 $\bar{U}$에 對한 週期 T에 해당하는 spectral density U_T의 比
$P_{(t)}$	藻類의 光合成에 依해서 t時에 發生되는 酸素量
P_{max}	光合成에 依한 酸素의 最大發生量
Z_t	標準化 時系列, 離散型 時系列
$\phi_p(B)$	p次 自己回歸模型의 演算子(operator)
ϕ_i	自己回歸模型의 i番째 係數
B	backward shift 演算子
$\bar{Z}$	$Z_t-\bar{Z}$
$\theta_q(B)$	q次 移動平均模型의 演算者
d	隔差數(difference)
V	$1-B$
w_t	置換된 定常時系列
s	季節週期
D	s의 階差數
$\Phi_P(B^s)$	P次 自己回歸模型의 演算子
$\Theta_Q(B^s)$	Q次 移動平均模型의 演算子
$\hat{\phi}_{kk}$	지체 k의 偏自己相關係數
$S_{(\hat{\theta},\hat{\phi})}$	ARIMA模型에 對한 殘差 自乘合
Q	a_t에 對한 χ^2인 統計量
$S_{(r_k)}$	γ_k의 標準偏差
$S_{(\hat{\phi}_{kk})}$	$\hat{\phi}_{kk}$의 標準偏差

1. 序　　論

河川의 水質은 물의 物理的 擧動으로 나타나는 流量의 變化와 密接한 關係를 가지면서 時間에 따라 끊임없이 變化하고 있다. 水質은 水文量과 함께 하나의 推計學的 現象으로 볼 수 있는데, 水質은 水文量보다 人爲的인 生活環境의 影響을 많이 받고 있기 때문에 水質의 變動特性은 한층 複雜한 樣相을 나타내고 있다. 水質 現象은 長期間으로 생각할 때 特定의 傾向을 가지고 있다고 判斷될 수 있으나 一個年 동안에도 季節的 變化의 週期性을 가지고 있으며 1日 동안에도 역시 時間的인 週期成分을 나타내고 있다. 傾向과 週期成分 外에도 說明할 수 없는 確率的 成分이 水質現象의 系列에 包含되어 있기 때문에 連續的으로 測定된 水質의 時系列은 時推計的인 現象으로 간주되고 있다.

1.1 河川水質의 一般的 特性과 研究背景

河川의 水質은 外部로 부터 河川으로 附加되는 汚染物質의 負荷量과 流水狀態의 函數關係로 說明될 수 있는데 水質에 영향을 미치는 條件을 크게 두가지로 分類해 보면 人爲的인 生活環境條件과 自然環境條件이다 生活環境條件에서 惹起되어 河川 水質에 큰 영향을 미치는 要因으로는 都市下水, 工場廢水, 農業排水等의 要因이 있고, 自然環境의 條件인 경우에는 氣象과 流域의 地形的 特性을 附與하는 諸要因이 있다. 生活環境과 自然環境의 條件이 時間的으로 계속 變化하기 때문에 두가지 條件의 變化에 따라 上記 諸要因이 質的으로나 量的으로 變化되고 있고 質的인 狀態를 意味하는 水質 또한 時間的으로 계속하여 變化 또는 變動되고 있다.

河川의 流水에 包含되어 水質特性을 나타내는 不純物은 溶存物質과 浮遊物質로 大別할 수 있다. 이와같은 不純物은 물과 稀釋되고 擴散과 流水의 進路에 依해서 移動하고 있기 때문에 單位 體積當 不純物의 무게로 定義되는 水質은 流水의 狀況, 即 水文量에 依해 決定的 영향을 받게 된다. 不純物이 물속에서 完全한 保存性 物質(conservative material)이라 假定하면 河川의 水質은 水文量의 函數로서만 說明될 수 있다. 그러나 水中의 不純物은 時間의 變化에 따라 化學的이고 生物學的인 反應이 계속 일어나기 때문에 水文量의 變化만으로서 說明될 수 없다. 그러므로 水質의 變化는 水文量의 變化에 많은 영향을 받고 있으며 水質의 變動樣相은 水文量의 變動樣相보다 훨씬 複合的이고 豫測하기 어려운 樣相을 나타내고 있다.

河川水質의 時間的 變動은 上述한 바와같이 河川 水

文量의 變動에 큰 영향을 받고 있으므로 河川水質 水動에 對한 硏究는 水文量의 變動에 對한 硏究와 類似한 點이 많다. 어느 限定된 水系에서 一定時間동안의 水質變化는 다음과 같은 敍述的 平衡方程式으로 記述될 수 있다.

物質의 累積量=∑(物質의 流入과 生成量)
−∑(物質의 流出과 消滅量)…(1.1)

即 物質의 累積量을 限定된 水系內의 全體貯水量으로 나누면 一定期間동안의 濃度變化가 되기 때문에 上式은 水質의 變化를 나타내는 것이고 上式을 適用하기 爲해서는 水文量의 變化를 알아야만 한다. (1.1)式은 여러가지 實驗과 假定下에서 確定的인 力學的 法則의 項으로 水系內의 擧動을 說明할 수 있는 경우에 適用되는 것으로써 確定的 模型에 해당하는 것이다. 反面 不確實한 要因이 많아 (1.1)式을 確定的인 數式으로 說明할 수 없을 때에는 水系內의 水質資料 또는 水質과 關聯되는 資料를 統計的으로 處理하여 模型을 導出하는 경우가 많다. 이러한 模型은 經驗的인 推計學的 模型에 해당되는 것이다.

水質變化에 對한 推計學的 分析을 利用한 硏究는 氣象分野와 河川 水文分野에서 硏究되기 시작한 後 1960年代의 後半부터 活發해 지고 있다. 1967年 Thoman[1]은 環境界의 時系列은 4個 範疇 中 하나에 屬하는 것으로 提示하였는데 各 範疇의 分類條件은 時系列 속의 週期의 長短과 明確히 나타나는 傾向成分과 週期成分, 確率的 成分 等의 構成狀態를 檢討한 것이다. 그는 여기에서 週期成分을 찾아내는데 spectral分析을 利用하였고 이 分析技法으로 日別 水溫과 溶存酸素의 變動을 調査하였다. Hahn[2]도 역시 1972年에 spectral 分析으로 日別 水質變動의 週期性을 調査하였다. 1972年에 McMichael과 Hunter[3]는 水溫과 流量에 對한 推計學的 模型開發에서 週期性을 除去하기 爲하여 Fourier級數에 依한 調和分析을 實施하였다.

1970年 Box와 Jenkins[4]가 一般的으로 觀測되는 時系列을 獨立的이 아닌 從屬的인 時系列로 생각하고 母數(parameter)의 數가 적고 간단한 推計學的 模型(一名 Box and Jenkins Model)을 提示한 以來, 經濟, 工學, 自然科學 等에서 이 模型의 普及이 점차 많아지고 있다. Mehta와 Ahlert[5]는 1975年에 流量, 水溫, BOD 및 溶存酸素의 日別資料를 가지고 Box and Jenkins 模型인 ARIMA模型을 利用하여 推計學的 分析을 試圖하였다. Hipel[6,7]等은 1977年 Box-Jenkins Model에 對한 模型 構造와 應用의 改善方法을 爲하여 여러가지 理論과 技法을 提示하고 例를 들어 說明하였다.

水溫에 對한 推計學的 模型의 硏究로는 1970年 Norse[8]와 1971年 Chen[9]이 上流部의 熱汚染 流入點으로 부터 下流部의 水溫을 豫測하기 爲하여 推計學的 模型을 開發하였다. Long[10]은 1976年에 Fourier series와 通信理論을 利用하여 日平均 水溫을 60日까지 推定하는 方法을 發表하였다. 1977年 Song과 Chien[11]은 河川의 日別 水溫에 對해 여러가지 形態의 線型 回歸模型을 適用시켜 大氣와 水溫의 相關性을 檢討하고 日別 大氣溫度의 變動에 依한 反應을 分析하였다. 1978年에 Song과 Leung[12]는 여기에서도 日別 水溫에 對한 變動狀態를 推計的으로 分析하였는데 大氣溫度의 變動으로 因한 水溫의 反應에 對하여 물의 凍結과 貯水의 영향을 分析한 것이다. Kothandaraman[13]은 1971年에 日平均 水質의 季節에 對한 週期 變動과 非週期 變動의 特性을 調査하고 氣象資料로서 水溫을 豫測하는 方法을 開發試圖하였다.

Lettenmaier[14]는 水質資料의 時系列에서 傾向을 나타내는 檢證方法을 提示하였는데 檢證方法을 獨立 時系列에서 展開하여 從屬 時系列로 擴張시켰다. Whitehead와 Young[15]은 河川水質에 對한 推計學的 模型에 'Monte Carlo' simulation을 적용하고 下流部의 水質을 豫測할 수 있는 模型을 導出하였다.

國內에서 自己相關係數와 spectral density의 技法을 利用하여 時系列의 週期性을 分析한 것은 1976年에 鮮于[16]가 地域別로 月降水量 週期性分을 導出해 比較 分析한 것이다. 1977年에는 李[17]가 溶存酸素의 時間別 時系列에 對해 週期性을 分析하고 自己回歸模型에 對한 檢討를 實施하였다.

1.2 河川水質의 時間的 持續性

特定한 地點에서 連續하여 測定한 河川水質의 資料는 하나의 時系列로 나타난다. 河川水質의 時系列 X_t는 1.1)에서 記述한 바와같이 傾向成分 $X_T(t)$와 週期成分 $X_F(t)$ 및 確率成分 ε_t의 合으로 생각할 수 있기 때문에 다음과 같은 式으로 記述할 수 있다.

$$X_t = X_T(t) + X_F(t) + \varepsilon_t \cdots\cdots\cdots\cdots\cdots\cdots (1.2)$$

河川水質의 時系列은 前述한 바와같이 推計的인 現象이기 때문에 時系列은 時間에 따라 變化하고 있는 同時에 時系列內의 自體相關性을 가지고 있다. 即 時系列 中의 各 값은 前 時間에 發生한 값에서 영향을 받아 서로 相關性이 內包된 時間的 持續性이 있는 것이다. (1.2)式 X_t에서 $X_T(t)$와 $X_F(t)$를 除去한 確率成分인 ε_t는 時間的으로 獨立的인 確率成分을 가진 時系列이라기 보다는 從屬的인 殘差成分의 時系列로 볼 수 있고 確定成分의 時系列 $X_T(t)$와 $X_F(t)$도 從屬的 特

性을 갖고 있다.

時系列에서 時間的인 持續性을 決定하는 技法으로 自己相關分析을 實施한다. 自己相關係數는 하나의 時系列에서 지체 k와의 값 間에 相關性을 나타내주는 하나의 尺度이다. 無限한 離散型 時系列 $\{X_t\}_{t=1}^{\infty}$가 定常時系列인 경우 自己相關係數 ρ_k는 다음 式과 같이 定義된다.

$$\rho_k=\frac{\mathrm{Cov}(X_t, X_{t+k})}{\mathrm{Var}X_t} \quad \cdots\cdots\cdots\cdots\cdots\cdots(1.3)$$

時系列의 持續性이 있다고 判斷되는 경우에 持續性의 程度는 測定 時間區間(time interval)이 짧을수록 增加하고 길수록 減少한다. 連續測定된 水質의 時系列에서 時間과 日別로 본 時間區間에 對한 持續性은 언제나 크게 存在하고 있다. 自然環境 中에서 河川水質에 많은 영향을 주는 水文量의 持續性은 河川水質의 持續性과 密接한 連關을 가진다. 1970年 Chow와 Kareliotis[18] 및 1968年 Quimpo[19,20]는 水文量의 時系列에서 變化의 樣相을 찾아내기 爲하여 自己相關係數를 分析하였다.

1.3 河川水質의 自己回歸模型(Autoregressive Model)과 ARIMA 模型

離散型 資料의 水質時系列은 時間的으로 持續性이 있고, 資料의 測定間隔에 따라 時系列의 持續性이 다른 것을 1.2)에서 記述하였다. 그런데 水質의 時系列은 汚染排出量의 增加, 사람 活動의 規則的인 反復 樣相 및 自然環境의 週期性 等으로 因하여 (1.2)式의 現象으로 나타난다고 볼 수 있기 때문에 非定常 時系列(nonstationary time series)이다.

連續的으로 測定된 水質資料의 分析은 水質의 時間的 變動을 보다 合理的으로 說明할 수 있는 模型의 展開에 큰 意義를 두고 있다. 推計學的 模型의 展開는 時系列의 確定成分과 確率成分을 分離시키고 兩成分을 統計學的으로 差異가 없는 範圍內에서 再結合시킬 수 있는 數式을 誘導해 내도록 해야 한다. 時系列의 模型化 過程에서는 時系列을 定常時系列(stationary time series)로 만드는 데서 부터 始作된다. 定常時系列에 對한 模型은 自己回歸模型 또는 이의 變形을 흔히 利用하고, 非定常 時系列은 ARIMA 模型의 開發以後 이의 利用이 漸次 普及되고 있다.

自己回歸模型化 過程에서 先行되어야 할 問題는 原時系列의 定常時系列 與否를 檢討해야 한다. 定常時系列은 各各의 時系列組(Ensemble)가 서로 같은 統計的 特性을 가지고 있는 것을 意味한다. 非定常 時系列로부터 確定成分을 除去시켜 殘差成分의 時系列이 定常化되도록 하는 過程이 自己回歸模型化에 重要한 節次가 된다. 定常時系列로 置換된 時系列은 크게 두가지 型으로 分離할 수 있는데 하나는 時間的으로 持續性이 없는 獨立的 時系列이고 다른 하나는 時間的으로 持續性이 있는 從屬的 時系列이다. 이와같은 두가지 型의 時系列 判定은 自己相關係數나 spectral分析으로 實施하는 경우가 많다. 時間的으로 從屬的인 時系列인 경우는 時系列 自體間에 回歸性이 存在하는 것을 뜻하므로 回歸方程式으로 展開되는 回歸模型을 作成할 수 있다.

ARIMA 模型은 時系列이 非定常 時系列이라 할지라도 模型內에서 定常系列로 置換시키는 特徵을 가지고 있다. 그러므로 模型化 以前에 非定常 時系列을 定常時系列로 置換시킬 必要가 없는 큰 長點을 가지고 있다. 模型內에서 時系列을 定常時系列化하고 自己回歸模型과 移動平均模型(Moving Average Model)을 混合시켜 構成된 것이 Autoregressive-Integrated Moving Average Model인데 이 略字를 ARIMA 模型이라 한다.

1.4 硏究의 目的과 範圍

우리나라 河川의 水質에 對한 硏究는 測定資料의 貧困으로 여러가지 問題點을 內包하고 있다. 個人 또는 關聯 硏究機關에서 河川의 水質에 對한 間歇的 調査硏究가 實施되어 오고 있으나 特定 時點의 斷片的 調査에 根據한 것이기 때문에 一般性이 缺如된 것이 大部分이다. 以上의 問題點으로 因하여 河川水質의 變動狀態가 具體的으로 把握되지 못하고 있고 各種 水質模型化에 利用될 母數의 推定이 어려운 처지에 있다.

本 硏究의 目的은 노량진 및 뚝도地點의 DO, 濁度 水溫等의 特定水質에 對하여, (1) 水質의 週期性 發見 (2) 水質變動의 特性把握, (3) 時別 DO에 對한 自己回歸模型의 推論, (4) 日別 DO에 對한 ARIMA 模型의 適用評價에 對한 것이다.

本 硏究에 利用된 水質資料는, 漢江下流部인 노량진과 뚝도에서 自動水質測定機로 連續測定된 5個 項目의 時系列 資料인데 測定項目은 溶存酸素(DO), 電導率(conductivity), 濁度(turbidity), 水溫(water temperature), 水素이온濃度(pH)이다. 連續時系列 資料를 求하기 어렵고 資料의 信憑性 問題로 因하여 長期間의 時系列 資料를 利用할 수 없는 與件에 있기 때문에, 比較的 많은 情報를 얻을 수 있도록 時間 區間이 1時間과 1日인 離散型 時系列을 擇하여 分析에 利用하였다.

水質의 週期性과 變動特性에 對한 分析은 推計學的인 것으로 調和分析, Spectral分析, 自己相關係數의 技法을 利用하였고 水質變動의 模型構造分析은 自己回歸

模型과 ARIMA模型인 推計的 模型에 局限시켰다.

이와같은 研究結果는, 河川의 水質監視 및 上水의 取水計劃과 下水의 排水計劃 等 河川의 水質管理에 利用될 수 있고, 水質管理에 必要한 模型 構造의 選定에 活用될 것으로 期待한다.

1.5 研究의 接近方向

本 研究의 基本的인 接近方向은, (1) 時別 및 日別 水質에 대한 週期와 推計的 性質에 對한 特性을 究明하고, (2) 時別 및 日別 DO의 推計的 模型構造의 推論에 根據를 두고 있다.

時系列 資料의 選定은 自動水質測定所에서 連續 測定된 資料로 부터 1時間 間隔의 時別 時系列을 取한 것이고 또한 1日間의 平均値를 求하여 日別 時系列을 取한 것이다.

選定된 時系列을 Fourier級數, 自己相關係數, spectral density의 分析技法을 사용하여 時別 週期分析, 時系列의 自體相關性 또는 從屬性 等의 變動特性을 究明하면서 上下流의 特性의 差異點을 檢討하였다.

時別 時系列에 對한 推計學的 回歸模型의 設定方向은 一次的으로 時系列을 定常的인 時系列로 置換시키는 조작을 試圖한 것인데 이 過程에서 週期成分을 除去하고 時系列의 平均이 0이고 標準偏差가 1인 標準化 時系列로 置換토록 한 것이다. 이와같이 置換된 時系列에 對하여 correlogram과 spectral density의 分析을 通해 回歸模型 또는 Markov 模型의 構造로 推論토록 하였다.

日別 DO時系列에 對한 推計學的 模型은 ARIMA模型을 適用시키고 原時系列에 日差를 나타내는 階差(difference) d를 주어 時系列을 置換시키며 置換된 時系列의 自己相關係數와 偏相關係數의 分析을 通해 模型의 構造를 判別하였다. 階差數를 주어 時系列을 置換시키는 것은 定常時系列化의 한 方法이며, 自己相關係數와 自己偏相關係數의 分析은 置換된 時系列의 定常化 與否의 斷折과 模型의 判別을 爲한 것이다.

2. 分析技法과 水質模型의 理論

2.1 自己相關係數

時系列에 對한 持續性의 程度인 系列相關의 分析은 水文量의 時系列의 경우에서 널리 利用되고 있는데 水質의 時系列인 경우에서도 널리 利用되고 있다.

系列相關係數 또는 自己相關係數 ρ_k는 離散型 時系列의 경우, k만큼 지체된 時系列의 自體 變數 사이에 存在하는 相關係數로서 定義된다.

無限히 계속되는 定常時系列의 離散型 時系列에 對한 ρ_k는 (1.3)式으로 表示되는데 $\mathrm{Cov}(X_t, X_{t+k})$와 $\mathrm{Var}\ X_t$는 各各 母集團의 共分散과 分散이다. 資料數가 N인 時系列 資料에서 $\mathrm{Cov}(X_t, X_{t+k})$의 推定値 C_k와 ρ_k의 推定値 r_k는 다음 式과 같이 求할 수 있다.

$$C_k=\frac{1}{N-k}\sum_{t=1}^{N-k}(X_t-\bar{X}_t)(X_{t+k}-\bar{X}_{t+k}) \quad \cdots(2.1)$$

$$r_k=\frac{\sum_{t=1}^{N-k}(X_t-\bar{X}_t)(X_{t+k}-\bar{X}_{t+k})}{(N-K)\ S_t\cdot S_{t+k}} \quad \cdots\cdots\cdots\cdots(2.2)$$

여기서 $\bar{X}_t$와 $\bar{X}_{t+k}$는 X_t와 X_{t+k}의 平均値이고 S_t와 S_{t+k}는 X_t와 X_{t+k}의 分散의 平方根인 標準偏差이다. (2.1)式과 (2.2)式에서는 이들 값은 다음 式에서 求한다.

$$\left.\begin{aligned}\bar{X}_t&=\frac{1}{N-k}\sum_{t=1}^{N-k}X_t\\ \bar{X}_{t+k}&=\frac{1}{N-k}\sum_{t=1}^{N-k}X_{t+k}\end{aligned}\right\} \quad \cdots\cdots\cdots\cdots(2.3)$$

$$\left.\begin{aligned}S_t^2&=\frac{1}{N-k}\sum_{t=1}^{N-k}(X_t-\bar{X}_t)^2\\ S^2_{t+k}&=\frac{1}{N-k}\sum_{t=1}^{N-k}(X_{t+k}-\bar{X}_{t+k})^2\end{aligned}\right\} \quad \cdots\cdots\cdots(2.4)$$

그런데 資料의 數가 많고 定常時系列인 경우에는 $\bar{X}_t$와 $\bar{X}_{t+k}$는 母平均 μ의 推定値인 (2.5)式의 $\bar{X}$와 同一한 값에 接近하고, S_t^2과 S^2_{t+k}은 母分散 σ_x^2의 推定値인 (2.6)式의 S_x^2과 同一한 값에 接近한다.

$$\bar{X}=\frac{1}{N}\sum_{t=1}^{N}X_t \quad \cdots\cdots\cdots\cdots(2.5)$$

$$S_x^2=\frac{1}{N-1}\sum_{t=1}^{N}(X_t-\bar{X})^2 \quad \cdots\cdots\cdots\cdots(2.6)$$

이와같은 경우 (2.1)式과 (2.2)式은 다음의 (2.7)式과 (2.8)式으로 고쳐 쓸 수 있다.

$$C_k=\frac{1}{N-k}\sum_{t=1}^{N-k}(X_t-X)(X_{t+k}-X) \quad \cdots\cdots\cdots(2.7)$$

$$r_k=\frac{\sum_{t=1}^{N-k}(X_t-X)(X_{t+k}-X)}{(N-k)S_x^2} \quad \cdots\cdots\cdots\cdots(2.8)$$

r_k의 값을 (2.2)式과 (2.8)式에서 比較하여 볼 때 k의 값이 적은 경우에 兩式에서 求한 r_k의 값의 差異가 無視될 정도이나 k가 큰 값인 경우에는 (2.8)式에서 計算된 r_k의 값에 比하여 (2.2)式에서 r_k의 값이 더 큰 bias를 가지고 있고 또한 計算도 (2.2)式이 複雜하다. 이와같은 理由로 本 研究에서는 共分散 C_k와 自己相關係數 r_k를 計算하는데 (2.7)式과 (2.8)式을 使用하였다. k의 값은 $0, 1, 2\cdots\cdots m$까지 取할 때 最大의 값 m은 時系列의 資料數 N보다 훨씬 작은 값을 擇하는 것이 보통이다. Box와 Jenkins[4]는 N은 적어도 50以上

이 必要하며 m은 $N/4$보다 적은 값을 擇하도록 提案하였다.

時系列을 分析하는데 自己相關函數 r_k는 다음과 같은 事項 等에 利用되고 있다.

(1) 1次 自己相關係數 r_1은 時系列의 分析에 가장 重要한 값으로 생각할 수 있다.

(2) correlogram의 形態에 따라 時系列 模型 選擇의 判斷根據가 될 수 있다.

(3) r_k는 週期性의 判斷에 利用되고 spectral density의 作成에 利用된다.

時系列 X_t가 獨立的인 時系列의 與否를 判斷하기 위하여 信賴限度 α에서 $r_k=0$인 有意性의 檢定을 實施한다. Anderson[21]은 有意基準 α에 對한 信賴限度(confidence limit) $L(\alpha)$을 다음 式과 같이 提示하였다.

$$L(\alpha)=\frac{-1\pm n\alpha\sqrt{N-k-2}}{N-k-1} \quad \cdots\cdots(2.9)$$

여기서 n_α는 α에 對한 正規分布의 標準偏差로 α가 95%일 때 n_α는 1.96으로 주어진다.

2.2 Fourier 級數의 分析

河川水質과 其他 環境에 對한 時系列의 數式(1.2)에서 傾向成分 $X_T(t)$와 週期成分 $X_F(t)$는 時系列을 構成하는 重要한 確定成分(deterministic component)이다. 長期間의 資料가 없을 때에는 $X_T(t)$의 成分을 찾아내기 어렵고 回歸分析으로 公式化한다 하여도 統計的인 意味를 부여하는데 無理한 點이 많다. 그러므로 本硏究에서는 $X_T(t)$를 無視하고 $X_F(t)$만을 생각하여 確定成分으로 하였다. 이 경우 (1.2)式의 X_t는 다음과 같이 생각할 수 있다.

$$X_t=X_F(t)+\varepsilon(t) \quad \cdots\cdots(2.10)$$

時系列의 資料數가 N인 경우 週期成分 $X_F(t)$는 다음 式과 같이 Fourier 級數의 展開式으로 나타낼 수 있다.

$$X_F(t)=\bar{X}+\sum_{j=1}^{q}\left(A_j\cos\frac{2\pi_j}{N}t+B_j\sin\frac{2\pi_j}{N}t\right)+A_N/2\cos\left[\frac{2\pi_N}{N}t\right] \quad \cdots\cdots(2.11)$$

여기에서 $\bar{X}$는 時系列 X_t의 平均이고, A_j와 B_j는 j調和函數의 係數로서 最小自乘法으로 誘導된 (2.12)式을 가지고 推定한다. N이 奇數인 경우 $A_N/2=0$이고 $q=(N-1)/2$이며 N이 偶數인 경우에는 $q=(N-1)/2$이다.

$$\left.\begin{aligned}A_j&=\frac{2}{N}\sum_{t=1}^{q}X_t\cos\frac{2\pi_j}{N}t\\B_j&=\frac{2}{N}\sum_{t=1}^{q}X_t\sin\frac{2\pi_j}{N}t\end{aligned}\right\} \quad \cdots\cdots(2.12)$$

$X_F(t)$에 對한 (2.11)式을 cosine의 項으로 代置시키면 (2.13)式과 같이 된다.

$$X_F(t)=X+\sum_{j=1}^{q}C_j\cos\left(\frac{2\pi_j}{N}t-\theta_j\right) \quad \cdots\cdots(2.13)$$

여기서

$$C_j=(A_j^2+B_j^2)^{1/2} \quad \cdots\cdots(2.14)$$

$$\theta_j=\tan^{-1}\frac{B_j}{A_j} \quad \cdots\cdots(2.15)$$

C_j와 θ_j는 各各 振幅과 位相角을 나타내는데 週期分析에 必要한 경우가 많다.

(2.11)式 또는 (2.13)式은 j次 週期에 해당하는 sine과 cosine의 合成曲線 또는 cosine의 曲線이므로 Fourier 級數로 展開된 調和函數이다. $C_j^2/2$은 j次 調和函數의 分散이고 이를 $H(f_j)$로 놓으면 (2.14)式의 關係로부터 $H(f_j)$는 다음과 같게 된다.

$$H(f_j)=\frac{1}{2}(A_j^2+B_j^2) \quad \cdots\cdots(2.16)$$

여기서 f_j는 $\frac{2}{N}$와 같으며 j次 調和函數의 頻度이고 0에서 0.5까지의 範圍에 있다. f_j에 對한 $H(f_j)$圖表는 一種의 periodogram이다. 이는 時系列의 週期成分을 찾아내고 確率成分을 檢討하는데 많이 利用되고 있다.

X_t의 全體分散을 σ_x^2라고 하고 殘差成分 ε_t가 時間的으로 獨立된 確率成分이라면 (2.10)式으로 부터 다음과 같은 關係가 成立한다.

$$\sigma_x^2=\sigma_F^2+\sigma_\varepsilon^2 \quad \cdots\cdots(2.17)$$

여기서 σ_F^2와σ_ε^2는 $X_F(t)$와 ε_t의 分散이다. 그런데 σ_x^2에 대한 σ_F^2의 比는 X_t와 $X_F(t)$의 多變數 回歸係數(multiple correlation coefficient)의 自乘과 同一한 값이다. 또한 σ_F^2은 $C_j^2/2$의 合으로 構成되었기 때문에 (2.17)式은 다음과 같은 式이 成立된다.

$$\sigma_x^2=\sum_{j=1}^{q}\frac{C_j^2}{2}+\sigma_\varepsilon^2 \quad \cdots\cdots(2.18)$$

C_j^2은 σ_x^2을 構成하는 分散에 해당되어 j次 調和函數가 X_t에 기여하는 尺度 또는 X_t를 說明하는 分散이 된다. 따라서 $\frac{C_j^2}{2}$은 하나의 說明分散(explained variience)이 된다. 따라서 1次에서 m次까지 合成된 m個의 調和函數로 說明되는 說明分散比 P_m은 다음 式과 같이 表示된다.

$$P_m=\sum_{j=1}^{m}\frac{C_j^2/2}{\sigma_x^2} \quad \cdots\cdots(2.19)$$

週期成分을 決定하는 데는 包含될 調和函數의 適正數가 重要하다. 調和函數의 數가 너무 많이 包含되었을 경우에는 資料의 測定誤差分이 調和函數의 各 係數 推定値에 包含되기 때문이고 數가 너무 적을 경우에는 週期成分의 部만을 代表하기 때문이다.

Fourier級數의 分析에 對한 長點으로는, (1) 調和函數의 說明分散을 이용하여 週期成分을 찾아낼 수 있고 (2) 決定된 週期成分의 調和函數를 直接推計學的 模型追求에 利用할 수 있으며, (3) 調和函數의 振幅과 位相을 分析하여 時系列 自體變動과 時系列 相互間의 變動分析에 利用할 수 있다는 點 等이다.

2.3 分散 Spectrum

時系列에 對한 Fourier 分析에서는 여러가지 固定되어 있는 週期成分만을 찾아낼 수 있다. 分散 spectrum은 固定된 週期뿐만 아니라 0에서 0.5까지의 任意의 頻度에 對해서 週期成分을 發見할 수 있다. 이와 같은 利點으로 因하여 分散 spectrum 分析은 時系列의 特性分析에 效率的으로 利用되고 있다. $H(f_j)$의 關係式인 (2.16)의 兩邊에 時系列 X_t의 資料數인 N을 乘하여 左邊을 $I(f_j)$라 하면 다음 式과 같이 되는데 一般的으로 f_j에 對한 $I(f_j)$를 periodogram이라 한다.

$$I(f_j)=\frac{N}{2}(A_j^2+B_j^2) \quad \cdots\cdots(2.20)$$

N이 無限히 接近하는 경우에 $I(f_j)$는 $X_t(t)$의 j次 調和函數에 대한 自乘의 合이고 다음 關係式이 成立된다.

$$\sum_{t=1}^{N}(X_t-X)^2=\sum_{j=1}^{q} I(f_j) \quad \cdots\cdots(2.21)$$

(2.20)式에 $I(f_j)$를 頻度 f의 函數로 놓으면 (2.22)式과 같이 놓을 수 있다.

$$I(f)=\frac{N}{2}(A_f^2+B_f^2) \quad \cdots\cdots(2.22)$$

$I(f)$는 資料에서 얻을 수 있는 spectral density이다. Wiener-Khinchine 方程式을 利用하면[22,23] 時系列의 連續 共分散函數 $C(\tau)$를 가지고 任意의 頻度 f에 對한 母集團의 spectral density $I(f)$를 求할 수 있다 $I(f)$와 $C(\tau)$는 다음式과 같은 Fourier 變換의 雙으로 表示된다.

$$I(f)=\int_{-\infty}^{+\infty}C(\tau)e^{-i_2\pi f\tau}$$

$$d\tau=\int_{-\infty}^{+\infty} C(\tau)\cos 2\pi_f\tau\, e\tau \cdots\cdots(2.23)$$

$$C(\tau)=\int_{-\infty}^{+\infty}I(f)e^{-i_2\pi f\tau}$$

$$df=\int_{-\infty}^{+\infty}I(f)\cos 2\pi_f\tau\, df \cdots\cdots(2.24)$$

Fourier의 雙으로 이루어진 上記 두 式에서 自己共分散을 알면 spectral density을 求할 수 있고 spectral density를 알면 自己共分散을 求할 수 있게 된다. $\tau=0$인 경우 $C(0)=\sigma_x^2$이므로 (2.24)式으로부터

$$\sigma_x^2=\int_{-\infty}^{\infty}I(f)df \quad \cdots\cdots(2.25)$$

가 成立된다. 따라서 spectral density의 全體 合은 時系列의 分散 σ_x^2과 같다. 時系列의 標準化된 경우에는 $\sigma_x^2=1$ 이므로 spectral density의 全體合은 1이다.

離散型 時系列에서 spectral density의 推定値는

$$I(f)=2[C_0+2\sum_{k=1}^{\infty}C_k\cos 2\pi fk],\quad 0\leqq f<0.5$$

$\cdots\cdots(2.26)$

과 같다. 이 式을 利用하여 資料數가 N인 離散型 時系列의 spectral density를 求하는 경우에는, 自己共分散의 次數인 k값의 最大値 m을 適切히 決定해야 한다. k값이 1에서 m까지인 경우, (2.26)式의 $I(f)$는 다음式으로 表示되는 $V(f)$값으로 推定된다.

$$V(f)=2[C_0+2\sum_{k=1}^{m}C_k\cos 2\pi fk] \quad \cdots\cdots(2.27)$$

여기에서 自己共分散 C_k는 (2.7)式에서 計算되는 自己共分散의 推定値이다. 時系列 X_t가 獨立的인 確率成分만으로 構成되었다면 $C_0=\sigma_x^2$이고 $k>0$인 모든 C_k값은 0이기 때문에 $I(f)$ 또는 $V(f)$는 (2.26)式과 (2.27)式에서 알 수 있는 바와같이 $2\sigma_x^2$이 된다. 即 時系列上의 어느 區間에서나 spectral density가 一定할 때 時系列은 持續性이 없는 獨立時系列로 생각할 수 있다. (2.27)式을 頻度領域(Frequency Domain)에서 時間領域(Time Domain)으로 바꾸어 놓으면

$$V_k=\frac{P}{m}\left(C_0+2\sum_{i=1}^{m-1}C_i\cos\frac{1k\pi}{m}+C_m\cos k\pi\right)$$

$\cdots\cdots(2.28)$

이 된다.[24] 이 式에서 $k=0, 1, 2\cdots\cdots m$이며 P값은 k가 o와 m인 경우에 $\frac{1}{2}$이고 그 外의 경우에는 모두 1이다.

지체 k와 頻度 f_k의 關係式은

$$f_k=\frac{k}{2m\Delta t} \quad \cdots\cdots(2.29)$$

인데 spectral density를 f_k의 函數로서 表示할 때 適用된다. 여기에서 Δt는 離散型 時系列 資料의 等 間隔單位時間間隔(time interval)이다. Spectral density V_k를 平滑하게 하기 爲하여 Blackman과 Tukey[25]가 開發한 두가지 方法의 平滑法 中의 하나인 "hanning"으로 통칭되는 式은

$$\left.\begin{aligned} U_0&=\frac{1}{2}(V_0+V_1)\\ U_k&=\frac{1}{4}V_{k-1}+\frac{1}{2}V_k+\frac{1}{4}V_{k+1}\\ U_m&=\frac{1}{2}(V_{m-1}+V_m)\end{aligned}\right\} \quad \cdots\cdots(2.30)$$

과 같다.

spectral density의 有意性 檢定은 平均 spectral density에 對한 推定된 density의 比인 R[24]을 가지고

實施한다. 信賴限度 95%의 基準에 해당되는 R은 다음과 같다.

$$R=10^{\frac{2.4}{\sqrt{b-1}}} \quad \cdots\cdots(2.31)$$

여기서 b는 다음과 같은 關係를 가진다.

$$b=2\left(\frac{M}{m}-\frac{1}{3}\right) \quad \cdots\cdots(2.32)$$

때로는 spectral density의 信賴限度를 spectral density의 平均값으로 하는 경우가 있는데 f의 連續函數인 (2.27)式의 경우는 $\sqrt{2\sigma_x^2}$으로, (2.30)式의 경우는

$$\bar{U}=\frac{1}{m+1}\sum_{k=0}^{m}U_k \quad \cdots\cdots(2.33)$$

을 限度로 한다.

2.4 時別 時系列의 模型 設定

水中生物의 生存과 有機物質의 酸化에 決定的 役割을 하는 溶存酸素(DO)는 1日동안에도 時別 變動이 뚜렷한 경우가 많다. DO에 큰 영향을 끼치는 水溫과 汚染負荷量 等의 時別 變動은 週期性을 가지는 것이 一般的이다. 그러므로 DO는 時別로 各各 다른 統計的 特性을 가지게 되며 非定常性(nonstationarity)인 週期性의 日別 樣相을 가지게 된다.

時系列의 推計的 模型에 對한 設定에서 優先되어야 할 事項은 時系列을 定常時系列로 置換시키는 것이다. 時系列의 n次 定常性은 各 時點에 對한 n次 모멘트(moment)가 서로 同一한 경우이다. 本 硏究의 模型設定에서 취급되는 DO의 時別 時系列 $\{X_t\}_{t=1}^{N}$는 다음 두 式을 滿足시킬 수 있는 2次 定常性으로 置換토록 하였다.

$$E[Z_t]=\mu=\text{constant} \quad \cdots\cdots(2.34)$$

$$E[Z_tZ_{t+k}]=\rho_k\sigma^2+\mu^2=\text{constant} \quad \cdots\cdots(2.35)$$

여기에서 Z_t는 DO의 時系列 資料中 時點 t (1, 2, 3, ……N)에서 測定된 原時系列 X_t의 定常化된 置換값이고 k는 지체(Lag)이며, ρ_k와 σ^2은 各各 置換된 母集團의 k次 自己相關係數와 分散이다. 上記 두 式에서 (2.34)式을 滿足하는 경우는 1次 定常性이고 (2.34)式과 (2.35)式을 同時에 滿足시키는 경우는 2次 定常性이다.

DO의 測定値가 季節과 1日동안의 時別 週期性이 同時에 存在한다는 前提下에 本 硏究에서는 다음과 같은 假定을 하여 特定月에 對한 時別 DO의 模型設定을 試圖하였다. (1) 特定月인 1個月 內에서 日平均 DO의 時系列은 確率成分으로만 構成되었고, (2) 1個月內의 DO時系列은 統計的 特性이 서로 다른 24個의 組로 構成되었다.

即, 1個月 內에서 日平均으로 構成된 DO時系列은 傾向成分과 週期成分을 無視한 確率成分으로 생각한 것이며, 時間間隔이 1時間인 特別 時系列은 24時間의 週期性이 있기 때문에 統計的 特性이 서로 다른 24個의 組로 構成되었다고 본 것이다.

本 硏究에서는 以上과 같은 特定月의 非定常 時系列을 平均이 0이고 標準偏差가 1이 되도록 다음과 같은 方法으로 定常化시킨 것이다.

1) 時別 平均에 依한 定常 時系列化

原時系列 X_t에 對해서 τ時(1, 2, ……, 24)의 平均値를 m_τ라 하고 τ時의 標準偏差를 S_τ라고 하면 다음 式으로 表示되는 Z_t는 平均이 0이고 標準偏差가 1인 標準化된 時系列이다.

$$Z_t=\frac{X_t-m_\tau}{S_\tau} \quad \tau=1, 2, 3, \cdots\cdots, 24 \cdots\cdots(2.36)$$

여기서 τ와 t와의 關係는 $t=\tau+24(n-1)$이고 n은 觀測順位日을 表示한다. (2.36)式은 $E(Z_t)=0$이기 때문에 (2.34)式을 滿足하고 있다. 또한 $E[Z_t^2]=1$이고 $E[Z_tZ_{t+k}]=\rho_k$가 되어 (2.35)式을 滿足한다. 그러므로 (2.36)式으로 表示되는 Z_t는 2次 定常時系列이다.

2) 時別平均의 調和函數에 依한 定常時系列化

時別 平均 m_τ와 標準偏差 S_τ는 (2.11)式의 Fourier 級數로 m_τ와 S_τ에 接近하는 重要週期成分인 m_t와 S_t를 찾아 낼 수 있다. m_t와 S_t는 $\tau=0, 1, 2, 3, \cdots\cdots$24인 24個의 資料에서 導出해 낸 調和函數이기 때문에 各 週期에 따라 同一한 樣相으로 反復된다. m_t와 S_t를 利用하여 X_t를 다음과 같이 變換시켜

$$Y_t=\frac{X_t-m_t}{S_t} \quad \cdots\cdots(2.37)$$

라 하면 m_t와 S_t가 m_τ와 S_τ의 重要週期만에 限한 것이어서 Y_t의 平均値와 標準偏差는 0와 1이 되지 않을 可能性이 많다. 그러므로 完全히 標準化되었다고 볼 수 없다. Y_t의 平均値와 標準偏差를 各各 $\bar{Y}$와 S_y라 하면 (2.37)式은 다음 式과 같이 標準化된다.

$$Z_t=\frac{Y_t-\bar{Y}}{S_y}=\frac{X_t-\bar{Y}S_t-m_t}{S_tS_y} \quad \cdots\cdots(2.38)$$

(2.38)式은 $E(Z_t)=0$이고 $E[Z_t^2]=1$이 되어 (2.34)式과 (2.34)式과 (2.35)式을 滿足하고 있기 때문에 標準化된 2次 定常時系列이다.

上記한 (2.36)式과 (2.38)式으로 表示된 Z_t가 時間 t의 變化에 따라 從屬性이 없는 獨立的 確率成分이면

$$X_t=m_\tau+S_\tau Z_t \quad \cdots\cdots(2.39)$$

$$X_t=m_t+YS_t+S_tS_yZ_t \quad \cdots\cdots(2.40)$$

로 X_t를 模型化 시킬 수 있다.

反面 Z_t가 時間의 變化에 따라 持續性이 있는 從屬的인 時系列인 때에는 (2.41)에 記述된 自己回歸模型

을 適用시켜 Z_t를 模型化하는 경우가 많다. 自己回歸模型의 構造는 時系列의 correlogram의 特性에 따라 달라진다. 自己相關係數가 $\rho_k=\rho_1{}^k$의 關係가 成立되면 時系列 Z_t의 推計的 模型은 Markov模型 또는 1次 自己回歸模型으로

$$Z_t=\rho_1 Z_{t-1}+\varepsilon_t \quad \cdots\cdots(2.41)$$

와 같이 表示된다. 여기서 ε_t는 時間 t에 對한 獨立的인 確率成分이다. 一般的인 p次의 自己回歸模型은 (2.5)에서 記述될 (2.42)式과 같다.

2.5 ARIMA模型

ARIMA模型은 Autoregressive Integrated Moving Average 模型의 略字로 Box-Jenkins 模型으로도 命名되고 있는데 1970年 Box와 Jenkins[4]가 提案한 以後 經濟, 水資源, 環境分野 等에 많이 利用되고 있다. 本 硏究에서는 河川의 日別 DO의 時系列에 對해 ARIMA 模型의 適用性을 檢討하고 漢江 下流部에 對한 DO 模型을 提示하였다.

推計學的 模型은 定常時系列模型(stationary model과 非定常 時系列模型(nonstationary model)으로 區分할 수 있다. 一般的으로 推計學的 模型의 開發은 時系列을 定常時系列로 假定하고 出發하며 非定常系列인 경우에도 時系列을 標準化하든가 또는 時系列 中의 非定常性 成分을 除去하여 定常時系列을 만든 後 模型을 推論하는 것이 보통이다. 그런데 ARIMA 模型에서는 非定常時系列이라 할지라도 模型自體 內에서 時系列을 定常化시키기 때문에 ARIMA 模型의 效用性은 크다고 본다.

ARIMA模型의 理論은 自己回歸模型(Autoregressive model: AR model) 移動平均模型(Moving Average model: MA model)의 理論에서 派生된 것이다. 等間隔의 時間으로 測定된 離散型 時系列 $Z_t(t=1,2,3,\cdots N)$가 定常時系列인 경우, AR(p) 模型은

$$\phi_p(B)\tilde{Z}_t=a_t \quad \cdots\cdots(2.42)$$

으로 表示되는데, 여기서

$\phi_p(B)=1-\phi_1B-\phi_2B^2\cdots\cdots\phi_pB^p$: p次에 對한 AR模型의 演算子(operator)이고 $\phi_i(i=1,2\cdots p)$는 AR模型의 母數(parameter).

B : backward shift operator로 $B\tilde{Z}_t=\tilde{Z}_{t-1}$인 關係가 있고, $\tilde{Z}_t=Z_t-\bar{Z}$로 $\bar{Z}$는 Z_t의 平均값.

a_t : NID$(0,\sigma_a)$ 即 標準化된 獨立的 時系列分布의 殘差成分으로 平均이 0이고 分散이 $\sigma_a{}^2$임.

MA(q)模型은

$$\tilde{Z}_t=\theta_q(B)a_t \quad \cdots\cdots(2.43)$$

으로 表示된다. 여기서

$\theta_q(B)=(1-\theta_1B-\theta_2B^2\cdots\cdots\theta_qB^q)$: q次에 對한 MA模型의 演算子이며 $\theta_i=(i=1,2,3,\cdots,q)$는 MA模型의 母數.

a_t는 白色雜音(white noise)의 時系列이기 때문에 線型 filter過程의 關係를 가진 (2.42)式과 (2.43)式으로부터

$$\phi_p(B)\tilde{Z}_t=\theta_q(B)a_t \quad \cdots\cdots(2.44)$$

인 混合型 自己回歸模型과 移動平均 模型인 ARMA (p,q)模型(Autoregressive-Moving Average Model)을 얻을 수 있다.

Z_t가 非定常時系列인 때에는 時系列 Z_t에서 定常時系列이 되도록

$$w_t=(1-B)^dZ_t=V^dZ_t \quad \cdots\cdots(2.45)$$

를 導出한다. 여기서는 d는 階差數(difference)이며 一般的으로 $d=0$, 1 또는 2가 된다. $d>1$인 경우는 언제나 $V^dZ_t=V^dZ_t$의 關係가 成立되어 (2.45)式에서 Z_t 代身 Z_t를 사용한다. 定常時系列化된 w_t와 (2.44)式의 關係에서

$$\phi_p(B)\,V^dZ_t=\theta_q(B)a_t \quad \cdots\cdots(2.46)$$

인 $ARIMA(p,d,q)$ 模型을 얻을 수 있다. 이 模型은 一名(p,d,q) 模型이다. 이 式은 $ARMA$模型에 V^d項을 追加함으로써 非定常時系列인 Z_t를 定常時系列로 變換시키는 過程이 包含된 것이므로 ARMA模型의 擴張으로 생각할 수 있다.

時系列 Z_t에 季節的인 週期性分이 있는 경우에

$$\Phi_p(B^S)\,V_S{}^DZ^t=\Theta_Q(B^S)\alpha_t \quad \cdots\cdots(2.47)$$

인 關係가 成立한다. 여기서

s; 季節週期이며

$V_S{}^D=(1-B^S)^D$; 季節週期 s의 D階差로 季節週期를 가진 時系列을 定常時系列로 만들어 주는 演算子이며 $D=0$, 1 또는 2로 함이 보통.

$\Phi_p(B^S)=1-\Phi_1B^S-\Phi B^{2S}-\cdots\cdots-\Phi_pB^{PS}$: P次인 AR模型의 演算子.

$\Theta_Q(B^S)=1-\Theta_1B^S-\Theta_2B^{2S}-\cdots\cdots-\Theta_QB^{Qs}$: Q次인 MA模型의 演算子.

(2.47)式의 殘差 α_t는 季節成分이 除去된 것일지라도 時系列間에 從屬性이 있을 可能性이 많으므로 非定常時系列로 생각할 수 있다. 이와같은 경우 (2.46)式의 Z_t를 a_t로 代置시키고 代置된 (2.46)式에 (2.47)式을 곱하면 다음과 같은 季節成分이 있는 時系列의 ARIMA 模型으로 誘導된다.

$$\phi_p(B)\Phi_P(B^S)\,V^dV_S{}^D\,Z_t=\theta_q(B)\Theta_Q(B^S)a_t \quad \cdots\cdots(2.48)$$

이와 같은 ARIMA 模型은 $(p,d,q)\times(P,D\,Q)_s$模

型으로도 불려진다.

$D=0$, $d=0$이면 上式은 $(p,q)\times(P,Q)_s$型의 ARMA 模型이다.

(p,d,q)模型과 $(p,d,q)\times(P,D,Q)_s$ 模型은 d 또는 D로 階差化시켜 模型自體內에서 時系列을 定常化하고 있는 만치, 模型外部에서 階差化된 定常時系列을 ARMA 模型에 入力시켜 나타나는 結果는 (2.48)式의 ARIMA 模型의 경우와 類似한 結果를 얻는다. 이러한 경우에 ARIMA 模型의 一般式은 다음과 같이 쓸 수도 있다.

$$\phi_p(B)Y_t=\theta_q(B)a_t,\quad d>0,\ D>0 \cdots\cdots\cdots(2.49)$$

$$Y_t=(1-B)^d(1-B^s)^D Z_t \cdots\cdots\cdots\cdots\cdots\cdots(2.50)$$

여기서는 Z_t는 原時系列인데 $d=0$와 $D=0$에서는 Z_t는 $\tilde{Z}=Z_t-\bar{Z}$로 代置시켜야 한다. Y_t는 非定常時系列 Z_t를 定常化시킨 時系列이다. (2.49)式은, 季節의 週期 s의 階差數인 D가 存在할때, ARIMA 模型의 (p,d,D,q)型이 되며, $D=0$일때는 (2.46)式과 同一한 (p,d,q)型이 된다.

本 硏究에서 取扱되는 日別 DO時系列은 觀測數가 1年分인 366日이므로 $D>0$인 季節模型에 해당하는 $(p,d,q)\times(P,D,Q)_s$ 또는 (p,d,D,q)型의 模型을 利用하는 데 制限이 있다. 日別 DO時系列에도 流量 또는 其他 河川水文量과 마찬가지로 366日의 週期가 있을 可能性은 많으므로 日別 DO時系列 資料數 $N=366$으로서는 $S=366$이 되어 (P,d,D,q)型의 模型을 適用시킬 수 없다. 그러므로 時系列의 重要 週期成分이 있을 경우에는 調和分析 等을 實施하여 週期를 除去한 後, (p,d,q) 模型을 사용하는 것이 바람직하다. Mehta[5]等은 流量, 水溫, BOD, DO等의 日別時系列에 對하여 年度別로 ARIMA模型의 (p, d, q)型을 提示하였는데 原時系列에서 週期成分이 있을 때는 週期成分을 除去하고 除去된 時系列을 가지고 模型을 檢討 確認했다. 1972年 Hipel等[6][7]은 ARIMA模型의 構造와 應用面을 分析 檢討하였는데 되도록 ARIMA模型에서는 週期成分 s를 除去한 (p,d,q)型을 時系列에 適用시키는 것이 때로는 必要함을 强調하였다. 本 硏究에서도 上記와 같은 理由로 原時系列에서 有意한 週期成分을 除去한 後 ARIMA模型인 (p,d,q)型을 適用시켰다. 一般的으로 模型을 評價하는 過程은 (1) 模型의 判別(identification), (2) 判別된 模型의 母數 推定, (3) 模型의 檢定으로 생각할 수 있는데 本 硏究에서의 ARIMA模型에 對한 評價過程은 다음과 같다.

1) 模型의 判別(identification)

原時系列에 對해 自己相關函數(autocorrelation function, ACF), 調和分析 및 spectral density 分析을 通해 週期性을 檢討하고 뚜렷한 週期性이 存在하는 경우에는 原時系列에서 週期成分을 除去한 殘差成分의 時系列을 導出하여 ARIMA模型의 (p,d,q)型을 判別하는 人力資料인 Z_t로 한다.

Z_t에 對한 ACF를 檢討하여 時系列 Z_t의 定常性을 調査한다. ACF의 推定値가 지체 k에 따라 빨리 消滅되지 않으면 時系列 Z_t는 非定常性을 意味하는 것이기 때문에 (2.50)式에 d의 階差化(differencing)를 實施하여 Z_t를 定常時系列인 Y_t가 되도록 한다. 原時系列 Z_t가 定常時系列이면 $d=0$이며 $Y_t=Z_t-\bar{Z}$이다.

Y_t가 完全히 確率成分이면 Y_t에 對한 γ_k는 NID(0, $1/N$)의 分布가 된다. 이 같은 結果를 基準하고 γ_k를 檢討하면 Y_t가 時間的으로 獨立된 確率成分의 與否가 밝혀진다. 時系列 Y_t가 獨立的 確率成分이 아닌 경우에는 다음단계로 Y_t時系列의 定常性을 檢討하게 되는데 檢討技法으로는 ACF와 PACF(Partial Autocorelation Function)를 利用한다. ACF의 推定値는 (2.8)式에서 求하는데 Y_t에 對한 ACF를 檢討하여 模型에서 必要한 AR과 MR의 母數의 數를 決定한다. AR型으로 表示될 수 있는 時系列의 ACF는 p의 次數에 따라 固有한 特性이 있는데 이것은 Yule Walker Equation[4]으로 부터 알아 볼 수 있다. 即 AR(1)이면 ACF는 지체 k의 增加에 따라 指數曲線으로 급격히 減少하며, AR(2)이면 ACF는 週期性을 가진 形態를 나타내며 지체 k의 增加에 따라 AR(1)의 경우보다 다소 완만하게 減少되며 上下로 움직인다.

Y_t가 MA(q)인 時系列에서는 γ_k가 q지체까지 存續되다. q以後에는 消滅된다. 即 이와같은 結果는 MA(q)型의 模型으로 부터 誘導된 γ_k와 係數 θ와의 理論的 關係式에서 나타나는 現象이다. 例를 들면 γ_k가 $k=1$에서 斷切되면 Y_t의 模型은 MA(1)이며, $k=2$에서 斷切되면 MA(2)型의 模型으로 判別할 수 있는 것이다. 一般的으로 MA(q)型에서 $k>q$인 경우 γ_k는 消滅되며 標準偏差는

$$S(\gamma_k)\approx\frac{1}{\sqrt{n}}\{1+2\sum_{v=1}^{q}\gamma^2\}^{1/2},\quad k>q \cdots\cdots(2.51)$$

이 된다.[4,26] 여기서 n은 時系列 資料數 N에서 階差 d를 除한 $N-d$이다.

PACF의 誘導는 ACF와 AR(p)模型의 係數 $p_j(j=1,\cdots\cdots,p)$와의 係數를 나타내는 Yule-Walker 方程式을 利用한 것이다. k次 AR模型에서 j項의 AR係數를 ϕ_{kj}라 하면 最終項인 k項의 AR係數는 $\hat{\phi}_{kk}$이다. 이 $\hat{\phi}_{kk}$를 偏自己相關係數라고 하는데

$$\hat{\phi}_{kk}=\begin{cases}\gamma_1: & k=1\end{cases}$$

$$\left\{\frac{r_k-\sum_{j=1}^{k-1}\phi_{k-1,j}}{1-\sum_{j=1}^{k-1}\phi_{k-1,j}r_j}\right.:k=2,3,\cdots\cdots k\cdots(2.52)$$

로 推定된다. 여기서 ϕ_{kj}는

$$\phi_{kj}=\phi_{k-i,j}-\phi_{kk}\phi_{k-1,k-j}\cdots\cdots(2.53)$$

이고 $\hat{\phi}_{kk}$는 지체 k의 函數로서 PACF이다. AR(p)模型의 最終係數는 $\hat{\phi}_{kk}$이기 때문에 $k=p$가 成立되어 上式에서 알 수 있는 바와 같이 $k\leq p$에서 $\hat{\phi}_{kk}$는 0이 아니며 $k>p$에서는 有意하게 0와 差가 없고 지체 k 以後 $\hat{\phi}_{kk}$는 NID$(0, 1/n)$에 가깝게 된다. $\hat{\phi}_{kk}$가 점점 k의 增加에 따라 消滅되면 Y_t時系列은 MA型에 適合하다. ϕ_{kk}는 지체 k를 20~40範圍에서 取하는 것이 적당하며 全資料數 N의 $\frac{1}{4}$보다 적게 取하는 것이 바람직하다.

ϕ_{kk}의 標準偏差는 $S(\phi_{kk})\approx\frac{1}{\sqrt{n}}$이다.

2) 模型의 母數 推定

여러가지 試圖에 依하여 模型이 ARIMA模型으로 判別되면 다음 단계는 模型의 係數推定이다. (p, d, q)型의 ARIMA模型에 對한 係數推定方法은 實測 時系列인 Z_t와 模型으로 부터 求해진 $\tilde{Z}_t$의 殘差의 自乘合을 最小化시키도록 하는 것이다. (p, d, q)模型에 包含된 全體係數를 代表하는 記號로서 $\hat{\phi}$와 $\hat{\theta}$를 使用하여 殘差의 自乘合을 式으로 表示하면

$$S(\hat{\phi},\hat{\theta})=\sum_{t=1}^{N}(Z_t-\tilde{Z}_t)^2\cdots\cdots(2.54)$$

과 같게 되고 이 式에서 $S(\hat{\phi},\hat{\theta})$가 最小化되도록 $\hat{\phi}$와 $\hat{\theta}$를 推定하는 것이다.

係數의 推定過程에서 第1단계로 $\hat{\phi}$와 $\hat{\theta}$를 假定하여 判別된 (p, d, q) 模型에 假定한 係數를 代入하고 $S(\hat{\phi}, \hat{\theta})$를 求한다. 다시 $S(\hat{\phi}, \hat{\theta})$가 最小化되도록 $\hat{\phi}$와 $\hat{\theta}$를 接近시켜 係數의 推定値를 確定시킨다.

3) 模型의 檢定

模型의 檢定은 模型에서 推定된 時系列의 現象이 어느정도 觀測時系列의 現象과 一置할 수 있는가를 判斷하는 妥當性 檢討이다. 흔히 $Z_t-\tilde{Z}_t=\hat{a}_t$인 殘差가 時間的으로 獨立的인 確率成分인가를 判斷하는 檢定方法으로써 自己相關係數 $r_k(\hat{a}_t)$를 分析하는 技法을 사용하고 있다. 模型이 妥當하면 $\hat{a}_t$는 平均이 0이며 $r_k(\hat{a}_t)$가 0에 接近하는 獨立的인 時系列이 될 것이다.

殘差 $\hat{a}_t$에 대한 獨立的 時系列의 信賴性 檢證을 爲해 다음과 같은 $x^2(k-p-q)$ 分布인 統計量 Q를 사용한다.

$$Q=n\sum_{k=1}^{m}r_k^2(\hat{a}_t)\cdots\cdots(2.55)$$

여기서 $n=N-d$

$k=r_k(a_t)$의 지체.

3. 分析資料

水質은 時間的으로 持續性이 있기 때문에 全般的인 水質의 狀態와 傾向을 分析하기 爲해서는 連續測定 資料가 必要하다. 國內外를 莫論하고 河川의 水質測定은 여러가지 與件上 連續測定한 경우가 적고 大部分 時間的으로 斷片的이고 不規則한 不連續測定인 경우가 많다.

漢江 下流部에는 서울의 뚝도, 보광동, 노량진, 영등포와 仁川市 加陽洞에 位置한 各 水源池에 自動水質測定機가 設置되어 있다. 測定項目은 溶存酸素, 電導率, 濁度, 水素이온濃度, 水溫과 全有機炭素(Total Organic Carbon, TOC)인데 實際的으로 TOC의 測定은 中斷된 狀態이다. 自動水質測定機는 測定機의 運營과 管理에 따라서 資料의 缺測과 신빙성이 問題되는 경우가 많은데 現在 設置된 自動水質測定機의 경우도 여러가지 問題點을 內包하고 있다.

河川水質의 경우와 同時에 分析하고 評價해야 될 事項은 水質에 直接 또는 間接으로 영향이 큰 河川流量, 大氣溫度 等이 있다.

分析에 利用된 流量 資料는 노량진 水位標의 日平均水位[27]로 부터 水位流量曲線[31]에서 換算된 것이며 日平均 大氣溫度[32]는 서울地方에 對한 것이다.

3.1 測定地點의 狀況

漢江 下流部에는 上述한 바와 같이 5個 地點에 自動水質測定機가 設置되어 있으나 計器의 눈금調整, 조작技術의 未備 等으로 資料의 신빙성이 희박하며 또한 그나마 資料의 蒐集도 어려운 與件에 있다.

노량진과 뚝도 水源池에 可動되는 測定機는 1974年에 우리나라에서 最初로 設置되었다. 노량진과 뚝도의 2個 地點을 本 硏究에서 測定地點으로 擇한 것은 (1) 短期間이지만 連續測定된 時系列 資料를 入手할 수 있었고, (2) 오래전 부터 水文量의 分析을 爲한 重要 對象地點이며 水位計가 設置되어 있어 流量資料를 求할 수 있고, (3) 位置가 서울의 河川水質管理에 必要한 地點이기 때문이다.

2個測定地點은 Fig. 3.1에서 보는바와 같고 測定地點의 位置狀況과 測定機의 設置場所를 檢討해 보면 다음과 같다.

1) 노량진 : 流量과 水位測定에 긴 歷史를 가진 이 地點은 河川의 흐름이 緩慢하고 水位差는 8m程度이다.

約 9.5km의 上流에 位置한 뚝도地點에서 이 地點까지의 汚染源은 中浪川, 淸溪川을 爲如하여 右岸의 各種 下水排出과 左岸의 盤浦川의 排水等이다. 測定機의 設置場所는 노량진 淨水場이고 採水點 인근인 人道橋에는 自己水位標가 設置되어 있다. 採水點은 漢江人道橋 左岸의 上水取水口이다.

2) 뚝도 : 永東橋 下流의 水源池 一帶의 流水狀態는 比較的 均一하고 3km 上流部에서 汚染된 炭川의 排水가 이 地點의 水質을 不良하게 하고 있다. 뚝도淨水場에 自動水質測定機가 設置되어 있고 採水點은 水源池 取水口이다. 이 地點의 直上流部에 自己水位表가 設置되어 있다.

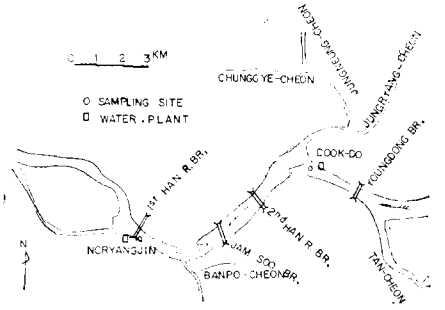

Fig. 3.1 Sampling Sites in Downstream Part of Han River

3.2 時系列의 分析資料選定

本 硏究에서 利用된 測定項目은 測定이 中斷된 TOC를 除外한 5個項目의 水質에 對한 것과 同期間에 測定된 流量과 大氣溫度의 2個 項目이다.

分析된 離散型 時系列의 單位時間 間隔은 時와 日로 하여 時別時系列과 日別 時系列로 分離 使用하였다. 分離된 各 時系列의 分析內容은 다음과 같다.

(1) 時別 時系列 : 時別 週期分析, 水質의 變動特性 分析, 日間水質變化, DO의 時別 推計的 模型設定

(2) 日別 時系列 : 項目間의 相關分析, 日別 水質變動의 週期性, DO의 ARIMA 模型 評價

ARIMA 模型의 判別과 評價分析에 利用된 各 時系列의 測定期間은 表 3.1과 表 3.2에 記述된 바와같이 1976年의 1個年과 1975年의 1個月이다.

短期間의 資料分析으로 水質分析에 나타나는 各種 母數의 決定은 長期間의 水質現象의 說明에 바람직한 것은 아니지만 單位時間 間隔을 時와 日로 하였기 때문에 짧은 時間의 變動特性과 推計的 模型 判別은 可能한 것으로 본다.

3.3 資料의 整理

노량진과 뚝도에 設置된 自動水質測定機로 부터 觀測된 時系列은 다른 地點과 比較하여 그 信賴程度가 良好한 便이고 連續缺測이나 간헐적 缺測이 적다. 그러나 두 地點에서도 1974年 測定機의 設置後 初期에는 技術과 運營上의 問題로 連續測定이 이루어지지 않다가 1976年을 前後하여 缺測이 적은 連續時系列을 測定하게 되었다.

本 硏究에서 分析을 爲한 時別 時系列 資料의 選擇은 表 3.1에 提示되었다. 두 地點의 測定結果를 1時間間隔으로 座標化하고 連續的인 缺測期間이 없고 測定結果에 異常이 적은 時別 時系列을 抽出해낸 것이다. 選擇된 時別 時系列에서 몇개의 缺測이 있거나 異常記錄이 있을 경우는 다음과 같이 補完 修正하였다.

$$X_{p,\tau}=X_{p,\tau-1}+(m_\tau-m_{\tau-1}) \quad \cdots\cdots\cdots\cdots\cdots\cdots (3.1)$$

여기서 $p=1,2,3,\cdots\cdots n$인데 p는 測定期間이 n日인 p番째 日을 나타내고 $\tau=1,2,3,\cdots\cdots 24$는 1日동안의 時를 表示하는 記號이다. m_τ와 $m_{\tau-1}$은 日別 時系列의 總資料數 $N=24n$ 中 τ와 $\tau-1$時의 平均値이다.

노량진과 뚝도에서 1年동안 測定된 時別 資料로 부터 1日 平均値를 算出하여 分析을 爲한 日別 時系列로 取하였다. 10時間以上 完全 缺測되었거나 資料의 異常이 있을 때에는 노량진과 뚝도의 水質資料에서 같은 項目끼리 두 地點의 水質間에 1次 回歸方程式 $y=ax+b$의 關係式을 利用하여 時系列의 資料를 補完 修正하였다.

時別 時系列과 日別 時系列의 補完과 修正을 거친 各 時系列의 統計的 特性은 表 3.1과 表 3.2와 같다. 이들 表에서 測定項目의 略字로 表示한 記號는 必要에 따라 測定項目을 代身하여 사용하기 爲한 것이다.

〈表 3.1〉 時別 時系列의 統計的 特性(1976年)

項目	測定位置	測定期間 (月, 日)	時系列 分類記號	單位	最大	最低	平均	標準偏差	資料數
溶存酸素 (DO)	노량진	3.15~ 3.25	N·DO—3	mg/l	12.4	9.9	11.20	0.6317	264
		5.20~ 5.29	N·DO—5	mg/l	6.4	3.0	4.83	0.9106	240
		10.11~10.17	N·DO—10	mg/l	7.3	5.5	6.36	0.3402	168
		6. 1~ 6.30	D·DO—75						
	뚝 도	3.22~ 3.31	D·DO—3	mg/l	11.0	9.0	9.58	0.4569	240
		10.21~10.31	D·DO—10	mg/l	10.4	7.6	8.79	0.6652	264
電導率 (CT)	노량진	2. 7~ 2.14	N·CT—2	μʊ/cm	154	101	131.08	10.5873	192
		6. 3~ 6. 9	N·CT—6	μʊ/cm	168	139	154.49	5.9512	168
		12. 1~12.18	N·CT—12	μʊ/cm	151	80	109.70	17.6988	432
濁度 (TB)	노량진	3. 3~ 3.11	N·TB—3	mg/l	34	13	25.31	4.8943	216
		7. 1~ 7. 8	N·TB—7	mg/l	25	13	17.59	3.0415	192
		12. 1~12.11	N·TB—12	mg/l	23	10	15.58	3.0415	264
	뚝 도	11. 6~11.12	D·TB—11	mg/l	15.0	4.0	8.11	2.2198	168
水素이온濃度 (pH)	노량진	12. 4~12.13	N·pH—12	—	7.4	6.5	7.08	0.1619	240
	뚝 도	1.11~ 1.15	D·pH—1	—	7.1	5.9	6.62	0.2988	120
水溫 (WT)	노량진	2. 1~ 2.27	N·WT—2	°C	5.2	1.6	3.48	0.7400	648
		6. 1~ 6.30	N·WT—6	°C	26.4	20.7	23.05	1.2289	720
		10. 1~10.20	N·WT—10	°C	20.7	17.7	18.94	0.7398	480
	뚝 도	2. 1~ 2.27	D·WT—2	°C	4.7	2.2	3.10	0.6404	648
		5.12~ 5.28	D·WT—5	°C	19.5	14.0	15.79	0.4504	408

〈表 3.2〉 日別 時系列의 統計的 特性(1976年)

項目	測定位置	單位	最大	最小	平均	標準偏差
溶存酸素	노량진	mg/l	12.3	3.1	7.96	2.4259
	뚝 도	mg/l	12.3	5.0	8.53	1.7682
電導率	노량진	μʊ/cm	200	75	127.18	21.3568
	뚝 도	μʊ/cm	230	34	94.30	43.7545
濁度	노량진	mg/l	175.0	8.0	23.17	18.7428
	뚝 도	mg/l	170.0	2.5	9.69	17.0555
水素이온濃度	노량진	—	8.3	6.6	7.3	0.1674
	뚝 도	—	8.0	5.5	7.16	0.3362
水溫	노량진	°C	27.6	0.0	13.82	8.7688
	뚝 도	°C	27.4	0.9	13.45	7.2699
大氣溫度	서 울	°C	27.1	−14.7	11.48	10.0152
流量	노량진	m³/sec	12,000	130	494.79	882.3586