水文諸量의 豫測이나 分析에 있어서 時系列 應用에 對하여(Ⅱ)

崔 榮 博*

2. 물 需要의 將來 豫測

水資源開發이나 地域開發計劃에 있어서, 물需給計劃의 策定에 있어서는 各種用水의 需要가 他要因에 依해서 決定된 다음에 주어지게되며, 供給을 이것에 뒤따라가는 方法으로 計劃을 取하는 경우가 一般的이다.

그런데 이와같은 물需要와 供給을 떼어 놓아 樹立하는 計劃이 果然 妥當性이있 는가에 對하여 생각하여 보기로 한다.

即, 水資源開發에 依해서 供給을 圖謀할 경우 물 需要者는 이것에 所要되는 費用의 負擔能力에 自然히 限度가 있는 것은 當然하다고 보아도 좋다. 따라서 各 需要者가 어느程度까지 負擔할 수 있는가? 或은 各地域에 있어서 妥當한 물의 價格이 얼마인가에 따라 물이 가지는 地域性에서 본 물의 供給能力에 限度가 있는 것이다. 이것은 물이라는 것을 앞으로는 自由財에서 經濟財로 보지 않으면 안될 것을 말하며, 물의需要와 供給사이에는 물 코스트의 問題가 介在하고 이것을 考慮하지 않고는 물 需給을 論할 수 없다는 것이다. 그래서 將來 물 코스트를 어느 程度까지 올릴수 있는가에 따라 물 供給能力에 對한 물 코스트라는 拘束要因은 작게되고 또 물의 廣域性도 높아져가는 것이라고 생각된다.

以上은 地域開發에 있어서 물의 需給밸런스를 생각해가는 境遇 물의 供給可能量은 물 코스트에 크게 支配된다는 것을 論한 것인데 逆으로 需要量에 對하여는 어떠한가? 다시말하면 需要量이라는 것은 그 發生要因만으로서 一義的으로 決定될 것인가의 問題이다. 地域計劃에 있어서 물 需要想定에 있어서는 當該地域의 各種因子(人口, 工業出荷額, 農業生産高)에서 算定하는 手法이 使用되고있다. 그리하여 이에 對한 供給에 對해서는 量的인 可能性은 多少의 配慮를 한다하더래도 물 코스트에 關해서는 거의 檢討를 하지않은 경우가 많다. 그러나 亦是 需要라 하는 것은 코스트에 대우 크게 支配되어 實質的으로 물 코스트와 對比되면 變動하는 것이라고 생각하는 것이 妥當하다.

以上에서 論述한 바와같이 물의 需要, 供給에 對해서는 各自가 單獨으로 決定될 것이 아니고 물 코스트를 媒體로해서 서로關聯되면서 變動하는 것이며 따라서 우리들이 地域開發에 있어서 물 需給計劃을 策定하는데 있어서 물 코스트의 聯關된 check and balance를 行할 必要가있다.

以上觀點에서 단지 供給面만의 檢討에 始終함이 없이 需要面에서의 檢討를 充分히 進行시키며, 경우에 따라서는 물이 各種開發計劃에 對한 制約要因도 된다는 것을 생각하면 水資源面에서 이들의 開發에 對한 Comment가 있어도 좋다고 본다. 여기서는 여러豫測手法中에서 가장 잘 使用되는 回歸分析을 應用한 豫測法은 이를 省略하고 外國에서 새로 登場 應用되는 時系列에 依한 豫測法을 論述하고 例로서 都市用水需要의 將來豫測을 하는 한 手法을 紹介하기로 한다.

* 技術士(建設部門)
理學博士
高麗大學校 理工大 敎授

時系列이라는것은 統計資料가 時間의 經過順序에 따라서 配列된 것을 말한다. 水文資料가 어느時點에 있어서 各地域의 資料가 아니고 어떤 特定의 것에 對한 時系列의 資料로서 나타나 있을 경우에는 그 時系列 모델을 回歸分析의 手法으로서도 決定할 수 있다. 이 경우는 資料가 時系列的이라는 것뿐으로서 回歸分析과 全혀 같은 立場이므로 이 例는 여기서는 省略한다. 時系列分析에서는 係數決定은 반드시 最小自乘法을 使用한다고는 限定하지 않는다. 여기서는 잘 使用되는 中間法을 說明하기로 한다.

目的하는 資料의 時系列變化를(1)式으로 表示한다.

$$y_i'=a+bt_i \cdots\cdots\cdots\cdots (1)$$

이資料의 値 y와 推計値 y'와의 偏差를 d로할 때 (1)式이 成立한다.

$$\begin{aligned}\Sigma d_i&=\Sigma(y_i)-y_i')\\&=\Sigma(y_i-a-bt_i) \cdots\cdots\cdots\cdots (2)\end{aligned}$$

이 a,b 決定에 있어서는 資料를 2等分해서 그 各各에 對한 Σdi의 値를 0로 놓으면 다음式이 얻어진다.

$$\sum_{i=1}^{n}(y_i-a-bt_i)+\sum_{i=n_1+1}^{n}(y_i-a-bti)$$

여기서 $n_1=\frac{n}{2}$, 但 n가 寄數의 경우는 n_1쪽에 한 資料가 餘分으로 보태여 진다.

$$\left.\begin{aligned}&\sum_{i=1}^{n}(y_i-a-bt_i)=0\\&\sum_{i=n_1+1}^{n}(y_i-a-bt_i)=0\end{aligned}\right\} \cdots\cdots\cdots (3)$$

여기서 부터

$$\left.\begin{aligned}\Sigma y_{i1}&=n_1a+b\Sigma t_{i1}\\\Sigma y_{i2}&=n_2a+b\Sigma t_{i2}\end{aligned}\right\} \cdots\cdots\cdots (4)$$

로되고 (4)式은 y_{i1}, y_{i2}, t_{i1}, t_{i2} 모두 旣定數이므로 a, b에 對한 聯立方程式이 成立한다. 이 方法은 最小自乘法 보다도 매우 計算勞力이 적어서 좋다.

時系列分析에서는 자주 1次式이 아닌 線形이 傾向線으로서 使用된다. 그 多大數가 多項式으로서 指數式이나 指數式에서의 變形인 logistic 曲線도 成長現象을 수반하는 事象의 豫測에 자주 使用된다. 多項式의 경우 係數決定은 電子計算機에 依한 回歸分析을 사용하면 問題가 없는데 計算을 自身이 行할 경우는 直交多項式法을 利用하는 것이 便利하다. 다음(5)式과 같은 2次式에 對한 解法을 表示하면 다음과 같다.

$$y=\alpha+\beta x_1+\gamma x_2 \cdots\cdots\cdots\cdots (5)$$

여기서

$x_1=a_1+t$, $x_2=a_2+b_2t+t^2$ 으로서 表示되는 것과 같은 函數(t^2 의 係數가 1이 되도록 한다) 로한다.

一般의 中間法과 같이 다음式

$$\Sigma x_1=0 \qquad \Sigma x_2=0 \qquad \Sigma x_1\cdot x_2=0$$

를 滿足시킬 때, (α,β,γ)는 直交系를 이룬다하고 (5)式을 直交多項式이라고 부른다.

(5) 式의 α, β, γ를 求하기 爲하여 最小 自乘法을 適用해서

$$\left.\begin{aligned}\Sigma y&=n\alpha+\beta\Sigma x_1+\gamma\Sigma x_2\\\Sigma x_1\cdot y&=\alpha\Sigma x_1+\beta\Sigma x_2+\gamma\Sigma x_1x_2\\\Sigma x_2\cdot y&=\alpha\Sigma x_2+\beta\Sigma x_1\,x_2+\gamma\Sigma {x_2}^2\end{aligned}\right\} \cdots\cdots (6)$$

여기서 直交系의 性質을 利用해서 다음 (7)式을 얻는다.

$$\left.\begin{aligned}\Sigma x_1&=0\\\Sigma x_2&=0\\\Sigma x_1\cdot x_2&=0\end{aligned}\right\} \cdots\cdots\cdots\cdots (7)$$

따라서 (7)式은

$$\left.\begin{aligned}\Sigma y&=n\cdot\alpha\\\Sigma x_1\cdot y&=\beta\Sigma {x_1}^2\\\Sigma x_2\cdot y&=\gamma\Sigma {x_2}^2\end{aligned}\right\} \cdots\cdots\cdots\cdots (8)$$

로되고

$$\left.\begin{aligned}\alpha&=\frac{1}{n}\Sigma y\\\beta&=\frac{\Sigma x_1\cdot y}{\Sigma {x_1}^2}\\\gamma&=\frac{\Sigma x_2\cdot y}{\Sigma {x_2}^2}\end{aligned}\right\} \cdots\cdots\cdots\cdots (9)$$

로된다.

여기서

α는 直線傾向을 表示하는 切片의 推定値가 되고 β는 x_1, γ는 x_2 만의 函數로서 計算이된다.

이 直交多項式方法은 多項式의 高次數를 順次附加해갈 때 새로운 係數η에 對해서 새로운 x_n만큼이 必要로하게 되므로서 便利하다.

指數曲線에서 表示되는 時系列에 對하여는 一

般으로 다음(10)式形으로 表示되는 函數의 값을 求하는 것이 된다.

$$y=a\cdot b^{t} \cdots\cdots(10)$$

이 指數曲線의 適合度는 期間 t 와 $t+1$에 있어서 y 의 伸長量 Δy 가 y의 값에 對해서 어느程度 y의 값을 表示 하는 가에 따라 決定된다 式으로 表示하면 (11)式과 같다.

$$\left.\begin{aligned}&\Delta y=y_{t+1}-y_{t}\\&\frac{\Delta y}{y}=\frac{ab^{t+1}-a\cdot b^{t}}{a\cdot b^{t}}=b-1\end{aligned}\right\} \cdots\cdots(11)$$

即, 指數曲線으로서의 適用의 良否는 伸長率 $\Delta y/yx$, 恒常一定値를 얻는가의 與否에 따라 定해진다. 係數 a, b의 決定은 前節에서 取扱한 對數를 取해서 最小自乘法을 適用해도 좋고 (11)式의 關係에서 (12)式과 같이 表示함으로써 b가 求하여진다.

$$\Sigma\frac{\Delta y_{i}}{y_{i}}=n(b-1) \cdots\cdots(12)$$

또 a의 값은 (10)式의 兩邊의 對數를 取하여

$$\Sigma\log y=n\log a+\log b\cdot\Sigma t \cdots\cdots(13)$$

을 얻는데, 여기서 $\Sigma t=0$로 할 수 있음으로 (14)式과 같이 求하여진다.

$$\log a=\frac{\Sigma\log y}{n} \cdots\cdots(14)$$

이 方法은 一般으로 最小自乘法보다도 훨씬 쉬우므로 計算例는 省略한다.

다음에 指數式의 變形으로서 (15)式과 같은 函數가 자주 利用된다.

$$y=K-a\cdot b^{t} \cdots\cdots(15)$$

이 函數는 經濟現象等에서 처음 急成長하다가 그後 漸進하는 경우라든가 그 反對의 경우에 잘 利用된다. 이경우의 函數의 決定도 (11)式에서 (14)式을 유도한 경우와 같이 求하여진다. 即

$$\Delta y=y_{t+1}-y_{t}=a\cdot b^{t}(1-b) \cdots\cdots(16)$$

(16)式의 右邊 第2項에 對해서 變形을 시킨後 代入하면

$$\Delta y=K(1-b)-y(1-b) \cdots\cdots(17)$$

(17)式에서 다음式이 求하여진다.

$$\left.\begin{aligned}&\sum_{i=1}^{n_1}\Delta y_{i}=n_{1}\cdot K(1-b)-(1-b)\sum_{i=1}^{n_1}y_{i}\\&\sum_{i=n_1+1}^{n}\Delta y_{i}=(n_{1}+1)K(1-b)-(1-b)\sum_{i=n_1+1}^{n}y_{i}\end{aligned}\right\}(18)$$

여기서

$$n_{1}=n/2$$

여기서 K, b의 값이 求하여진다. a값을 求하는 것은 (13)式에서 (14)式을 求하는 경우와 같다.

經濟現象, 自然現象의 增加現象에 對해서 가장 잘 사용되는 것에 logistic 曲線이 있다.

logistic曲線의 一般式은 다음(19) 式과 같다.

$$y=\frac{k}{1+e^{g(t)}} \cdots\cdots(19)$$

單純한 logistic 曲線에서는 $g(t)=a_{0}+a_{1}t$ 로서 表示된다. 따라서 $e^{a_0}=m$, $a_{1}=-a$로 놓으면

$$y=\frac{k}{1+me^{-at}} \cdots\cdots(19')$$

이 曲線은 成長當初는 t의 增加와 함께 增加速度도 急增하고 어느點에서 最大伸長을 表示하고 以後는 t의增加에 따라 增加比率을 減하고 最後에는 어느極限値에 接近하는 曲線을 나타낸다.

(19')式을 微分하면

$$\frac{dy}{dt}=ay-\frac{a}{k}y^{2} \cdots\cdots(20)$$

그리고 다시 微分하면

$$\frac{d^{2}y}{dt^{2}}=a^{2}y\left(1-\frac{y}{k}\right)\left(1-\frac{2y}{k}\right) \cdots\cdots(21)$$

로된다.

(19')式은 $a>0$ 이면 $t=-\infty$일 때 y에 가까워지고 $t\to+\infty$에서는 $y=k$ 로되고 또 $y=0$ $y=k$ 以外의 값에 對해서는 (20)式은 0이 되지않음으로 $y=k$ $y=k$ 의 두極限値 사이에는 極大및 極小値가 없고 이 曲線의 變化는 單調增加를 나타내는 것이 된다.

(21)式 에서 이 曲線은 $y=k/2$일때 增加의 加速度는 0 이되고, 即 이 時點에서는 하나의 變曲點을 가지게 된다. 다음에 函數形을 求하기 爲하여 (19')式의 兩邊의 逆數를 取하면

$$\frac{1}{y}=\frac{1}{k}+\frac{m}{k}e^{-at} \cdots\cdots(22)$$

여기서 $Y=\frac{1}{y}, K=\frac{1}{k}, A=\frac{m}{k}$, $B=e^{-a}$ 로 놓으면 $Y=K+A\cdot B^{t}$ 가되고 (15)式의 修正指數曲線과 같은 形이 된다.

Hoteling은 그 計算方法으로서 따로 $dy=\Delta y$, $dt=\Delta t$로 해서 (20)에서 (23)式을 유도하여

$$\frac{\Delta y}{\Delta t}=ay-\frac{a}{k}y^2 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (23)$$

$\Delta t=1$ 로해서 $\Delta y/y$ 를 求하면 (24)式을얻는다.

$$\frac{\Delta y}{y}=a-\frac{a}{k}y \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (24)$$

여기서 $\Delta y/y=\alpha$, $a=\beta$, $-a/k=\gamma$ 로 하면 (24)式은 (25)式과 같은 直線式을 나타낸다.

$$\alpha=\beta+\gamma y \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (25)$$

따라서 觀測資料에 依해서 α, y를 計算해 두면 係數 β, γ를 求하는 것이 쉽다.

上記와 같은 關係에서 正規方程式은 다음과 같이 해서 求한다.

$$\left.\begin{array}{l}\Sigma\alpha=n\beta+\gamma\Sigma y\\ \Sigma\Delta y=\beta\cdot\Sigma y+\gamma\Sigma y^2\end{array}\right\}\cdots\cdots\cdots\cdots\cdots\cdots (26)$$

計算例 : 어느都市의 年平均 1日生活用水의 供給實績의 系年別 값은 다음表와 같다. 이것을 logistic 曲線으로서 計算하라.

年　　度	t	Y_t	ΔY_t	$R=\Delta Y_t/Y_t$	Y_t^2
1958	0	0.05000	0.00229	0.0458	0.00250
59	1	0.05229	0.01275	0.02438	0.00273
60	2	0.06504	0.01688	0.2595	0.00423
61	3	0.08192	0.02138	0.2610	0.00671
62	4	0.10330	0.03268	0.3164	0.01067
63	5	0.13598	0.06658	0.4896	0.01849
64	6	0.2256	0.08936	0.4412	0.04103
65	7	0.29192	0.12675	0.4342	0.04491
66	8	0.41867	0.16112	0.3848	0.17528
67	9	0.57979	0.18776	0.3238	0.33615
68	10	0.76755	0.18839	0.2454	0.58913
69	11	0.95594	-	-	-
	$n=11$	$\Sigma Y_t=2.74902$	$\Sigma\Delta Y_t=0.90594$	$\Sigma R=3.1255$	$\Sigma Y_t^2=1.233185$

(單位 Y_t : 100 000 ton)

(26) 式에 表中의 相當하는 값을 代入해서 β, γ를 求한다. 그런데 이 計算例에서는 $t\to+\infty$일때 y(다시 말하면 k)의 값을 42萬 ton으로 先決한 까닭에서의 값은 假値로서 求해지며 以後의 解析에는 直接關係하지 않는다. 이 때의 β는 $\beta=0.22678$ 로서 求하여진다.

다음에 m 값은 變曲點을 주는 年次와 이 때의 값은 求하여지나 주어진 表의 系列中에 變曲點이 包含되어 있지 않으므로 다음과 같은 方法에 依해서 求한다. 即 表中의 系列中에서 $t=6$, $t=8$, $t=10$일 때의 y의 값을 抽出하여 다음式에 依해서 m값을 平均값으로서 算定 하였다.

$$m=\left(\frac{k}{y}-1\right)e^{\beta t}$$

上式은 一般形으로서 이 結果 求한 값은

$t=6$ 일때　　$m=76.9649$

$t=8$ 〃　　$m=55.4641$

$t=10$ 〃　　$m=43.2295$

이므로 그 平均을 取하면 $m=58.5530$ 이다.

따라서 必要한 係數는 求하여 졌으므로 求하려하는 式은 다음과 같다.

$$y=\frac{4.2}{1+58.5530e^{-0.2268t}}$$

이와같은 計算은 回歸分析을 利用해서 係數를 求할 수 있는데 一般으로는 計算量이 많으므로 電子計算機를 利用하는 것이 希望된다. 時系列豫測을 할 경우에 자주 發生하는 問題로서 資料의 時系列的配列에 매우凹凸變動이 있는 것이있다.

即, 작은 變動을 除去해서 傾向의 大勢를 본다든가 또 變動中에서 特定周期를 가지는 것을 重點的으로 取扱하는 것이 必要하게 된다. 이方法의 하나로서 移動平均法이 있다. 即,

$$\overline{y}_{t+(n-1)/2}=\frac{y_t+y_{t+1}+\cdots\cdots+y_{t+n-1}}{2} \quad (27)$$

(27)式은 單純移動平均의 定義式이다.

例를 들면 $y(t)$가 (41)式

$$y(t)=m(t)+\sum_{i=1}^{s}C_i\cos(\mu_i t+\varphi_i)+\varepsilon(t) \cdots (28)$$

과 같이 表現되는 變動을 가진 函數라 하면 $m(t)$는 圓滑한 曲線으로 되는 것 같이 資料數 n을 適當히 取한 移動平均을 구하면 $m(t)$의 變化를 적게해서 他項을 작게 하는 것이 可能하다.

移動平均法에도 加重平均을 使用해도 支障은 없다.