

STBL 모형의 모수추정 및 예측방법의 비교

김덕기¹⁾, 이성덕²⁾, 김성수³⁾, 이찬희⁴⁾, 이건명⁵⁾

요약

본 논문은 공간시계열자료가 공간의 위치와 시간의 흐름에 따라 동시에 관측되는 분야인 기상, 지질, 천문, 생태, 역학 등에서 아주 넓이 사용되고 있고 그 수요가 점차 증가하는 이 시기에 복잡한 공간시계열 증선형(STBL) 모형에 대한 모수 추정 방법 중 수치 해석적 방법인 Newton-Raphson 방법과 Kalman-Filter 방법을 비교하고, 두 가지 방법에 의한 예측력을 비교하여 보았다.

주요용어 : 공간시계열 증선형모형(STBL), 예측오차제곱합(SSF), 최대우도추정(MLE)
Kalman-Filter, Newton-Raphson.

1. 서론

공간시계열자료(Space-time series data)는 공간의 위치와 시간의 흐름에 따라 동시에 관측되어지는 자료로서 기상학, 지질학, 천문학, 생태학, 환경공학 및 역학 등의 분야에 폭넓게 나타나고 있다. 일반적으로 공간자료는 특정위치 또는 지역에서의 관측 값이나 측정값으로 구성되며, 공간자료는 지리통계자료, 격자자료, 공간 점 패턴자료의 3가지 유형으로 분류될 수 있으며 본 연구에서는 격자자료(Lattice data)에 대하여 다루었다.

공간시계열자료에 대한 기존 연구를 짧게 살펴보면, Pfeifer 와 Deutsch(1980-1981)는 ARMA 시계열모형에 공간의 정보를 포함시킨 공간시계열

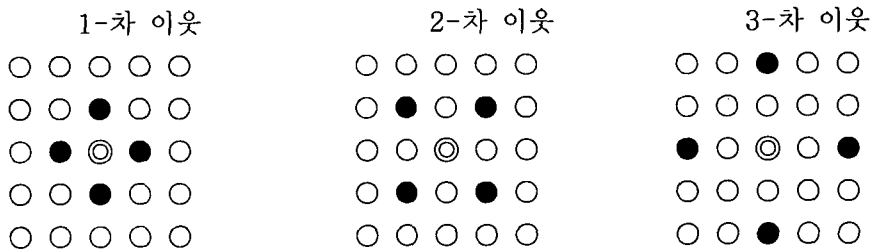
-
- 1) (361-763) 충북 청주시 개신동 12번지, 충북대학교 정보통계학과, 대학원생
E-mail : k1d2k3@hanmail.net
 - 2) (361-763) 충북 청주시 개신동 12번지, 충북대학교 정보통계학과, 교수
E-mail : sdlee@chungbuk.ac.kr
 - 3) (361-763) 충북 청주시 개신동 12번지, 충북대학교 전기공학과, 교수
E-mail : sungkim@chungbuk.ac.kr
 - 4) (361-763) 충북 청주시 개신동 12번지, 충북대학교 미생물학과, 교수
E-mail : chlee@chungbuk.ac.kr
 - 5) (361-763) 충북 청주시 개신동 12번지, 충북대학교 컴퓨터공학과, 교수
E-mail : kmlee@chungbuk.ac.kr

ARMA(이하 STARMA)모형을 제시하였고, Dai와 Billard(1998)는 공간시계열중선형 모형(이하 STBL)에 대한 식별과 추정방법에 대해 연구하였고, 정애란과 이성덕(2005)은 미국 12개 주에 대한 Mumps 자료에 대한 STAR-STARMA 모형에 대한 추정과 예측력을 비교하였다. 본 논문은 복잡한 비선형 모형으로 공간(Space)과 시간(Time)이 결합된 중선형모형인 STBL 모형의 모수추정 방법에 있어서의 어려움을 수치 해석적 방법인 Newton-Raphson방법과 모형을 상태공간모형으로 재구성하여 추정하는 Kalman Filter방법을 비교하여 어떤 방법이 더 적합한 모형을 이끌어내는지 비교하는 것에 그 의의를 두었고, 이를 위해 국내외 실제 공간시계열자료를 이용해 예측력을 비교하여 보았다. 본 논문의 구성은 2장에서 가중행렬과 공간시계열모형을 살펴보고, 3장에서는 모형식별과 두 가지 모수추정방법을 살펴보고, 4장에서는 실증분석을 위해 미국 12개 주의 Mumps 자료와 우리나라 8개도의 유행성이하선염 자료를 이용하였고, 5장에서는 결론과 추후 연구계획으로 구성하였다.

2. 가중행렬과 공간시계열모형

2.1. 가중행렬[Weight Matrix]

그림 2.1은 공간차수에서 격자자료(lattice-data)의 1차~3차 이웃(neighbor)하고 있는 구조의 예를 나타낸 그림이다. 1차 이웃의 경우 한 위치(◎)에서 가장 가까우며 동일한 유클리디안 거리를 갖는 집합(●)이며, 2차 이웃의 경우 1차 이웃보다 먼 동일한 유클리디안 거리를 갖는 집합으로 나타낸다. 3~4차 이웃의 경우도 마찬가지로 방법으로 나타낼 수 있다.



[그림 2.1] 공간차수의 격자자료에서 이웃하고 있는 구조

위 5×5 ($n=25$) 격자자료의 경우 각 위치는 왼쪽 위에서부터 오른쪽 아랫방향으로 가중행렬(Weighting Matrix)을 그림 2.2와 같이 결정할 수 있다.

$$W^{(1)} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \dots \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \dots \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

[그림 2.2] 1-차 이웃한 가중치행렬

2.2. 공간시계열 중선형모형

공간시계열 중선형모형(Space-Time Bilinear Model)은 $STBL(p_{\lambda_i}, q_{\eta_j}, r_{\xi_i}, s_{\mu_j})$ 로 표현할 수 있으며, 모형은 다음과 같다.

$$Z(t) = \sum_{i=1}^p \sum_{m=0}^{\lambda_i} \phi_m^i W^{(m)} Z(t-i) + \sum_{j=1}^q \sum_{n=0}^{\eta_j} \theta_n^j W^{(n)} e(t-j) \tag{2.1}$$

$$+ \sum_{i=1}^r \sum_{j=1}^s \sum_{m=0}^{\xi_i} \sum_{n=0}^{\mu_j} \beta_{mn}^{ij} [W^{(m)} Z(t-i)] \# [W^{(n)} e(t-j)] + e(t)$$

p, q : 최대 자기회귀차수, 최대 이동평균차수

λ_i, η_j : i 번째 자기회귀항의 차수와 j 번째 이동평균항의 차수

ϕ_m^i, θ_n^j : 공간차수가 m, n , 시간차수 μ_j 인 자기회귀모수, 이동평균모수

$W^{(m)}$: 공간차수가 m 인 $n \times n$ 가중치행렬(weighting matrix)

$z(t) = [z_1(t), \dots, z_n(t)]^T$: $n \times 1$ 확률벡터과정(random vector process)

$e(t) = [e_1(t), \dots, e_n(t)]^T$: $n \times 1$ 확률잡음벡터(random noise vector)

r, s : 중선형항의 최대자기회귀차수, 최대이동평균차수

ξ_i, μ_j : 중선형항의 i, j 번째 시간차수에서 자기회귀항의 공간차수, 이동평균항의 공간차수

β_{mn}^{ij} : 자기회귀와 이동평균의 i, j 번째 시간차수에서 m, n 번째 공간차수를 갖는 중선형모수

여기서, $(r, s, \xi, \mu) = (0, 0, 0, 0)$ 이면 공간시계열 자기회귀 이동평균모형(Space-Time Autoregressive Moving Average Model)인 다음의 STARMA(p_{λ}, q_{η}) 모형이 된다.

$$z(t) = \sum_{i=1}^p \sum_{m=0}^{\lambda_i} \phi_m^i W^{(m)} z(t-i) + \sum_{j=1}^q \sum_{n=0}^{\eta_j} \theta_n^j W^{(n)} e(t-j) + e(t) \tag{2.2}$$

3. 모형식별과 모수추정

3.1. 모형식별(Model Identification)

(1) 공간시계열 자기회귀이동평균모형(STARMA)

STARMA모형의 식별은 [표 3-1]과 같이 Z_t 의 STACF, STPACF를 구하여 시간차수와 공간차수를 식별해 낸다.

[표 3-1] STARMA 모형의 식별

모형	공간시계열자기상관함수 (STACF)	공간시계열부분자가상관함수 (STPACF)
STAR($p_{\lambda_1, \dots, \lambda_p}$)	Tails off	Cuts off after p-time lag and λ_p spatial lag
STMA($q_{\eta_1, \dots, \eta_q}$)	Cuts off after q-time lag and η_q spatial lag	Tails off
STARMA($p_{\lambda_p}, q_{\eta_q}$)	Tails off	Tails off

(2) 공간시계열 중선형모형(STBL)

STBL모형의 식별은 우선 모형을 STARMA모형으로 가정하여 Z_t 의 STACF, STPACF를 구하여 [표3-1]과 같이 시간차수와 공간차수를 식별하고, STARMA모형 \hat{Z}_t 을 추정한 후 잔차의 제곱 $\epsilon_t^2 = (Z_t - \hat{Z}_t)^2$ 을 새로운 데이터로 취하여 이 데이터의 STACF, STPACF를 다시 구하여 중선형 항의 시간차수와 공간차수를 식별해 낸다.

3.2. STBL 모형의 모수추정

(1) Newton-Raphson 방법을 이용한 최대우도추정(MLE)

2.2.의 STBL모형에 대한 모수의 조건부최대우도추정(conditional maximum likelihood estimation)은 다음과 같은 절차에 따라 구할 수 있다.

$$e(t) = z(t) - y(t)\theta, \quad t = m, \dots, T, \tag{3.1}$$

여기서, y_t 는 다음과 같다.

$$y(t) = [W^{(0)}z(t-1), \dots, W^{(\lambda_1)}z(t-1), \dots, W^{(0)}z(t-p), \dots, W^{(\lambda_p)}z(t-p)],$$

$$\begin{aligned} & W^{(\eta_1)} e(t-1), \dots, W^{(\eta_1)} e(t-1), \dots, W^{(0)} e(t-q), \dots, W^{(\eta_q)} e(t-q), \\ & W^{(0)} z(t-1) \# W^{(0)} e(t-1), \dots, W^{(\xi_1)} z(t-1) \# W^{(\mu_1)} e(t-1), \dots, \\ & W^{(0)} z(t-1) \# W^{(0)} e(t-s), \dots, W^{(\xi_1)} z(t-1) \# W^{(\mu_s)} e(t-s), \dots, \\ & W^{(0)} z(t-r) \# W^{(0)} e(t-1), \dots, W^{(\xi_r)} z(t-r) \# W^{(\mu_1)} e(t-1), \dots, \\ & W^{(0)} z(t-r) \# W^{(0)} e(t-s), \dots, W^{(\xi_r)} z(t-r) \# W^{(\mu_s)} e(t-s) \end{aligned}$$

여기서, θ 는 다음과 같다.

$$\begin{aligned} \theta = & (\phi_0^1, \dots, \phi_{\lambda_1}^1, \dots, \phi_0^p, \dots, \phi_{\lambda_p}^p, \theta_0^1, \dots, \theta_{\eta_1}^1, \dots, \theta_0^p, \dots, \theta_{\eta_p}^p, \beta_{00}^{11}, \dots, \beta_{\xi_1 \mu_1}^{11}, \dots, \beta_{00}^{1s}, \dots, \beta_{\xi_1 \mu_s}^{1s}, \\ & \dots, \beta_{00}^{r1}, \dots, \beta_{\xi_r \mu_1}^{r1}, \dots, \beta_{00}^{rs}, \dots, \beta_{\xi_r \mu_s}^{rs})^T \end{aligned}$$

여기서, $e(t)$ 는 다음과 같이 가정한다.

$$e(t) = [e_1(t), \dots, e_n(t)]^T, t = 1, \dots, m-1 \sim iid N(0, \Sigma)$$

e_0 의 조건부결합밀도함수는 다음과 같다.

$$L(\phi, \theta, \beta, \sigma^2 | e_0) = -2^{-1} n (T-m+1) \log |\Sigma| - 2^{-1} \sum_{t=m}^T e(t)^T \Sigma^{-1} e(t)$$

$e(t)$ 로부터 $Z(t)$ 로의 변환 시 Jacobian은 단위행렬이고 우도함수는 다음과 같다.

$$L(\phi, \theta, \beta, \sigma^2 | e_0) = -2^{-1} n (T-m+1) \log |\Sigma| - 2^{-1} \sum_{t=m}^T [z(t) - y(t)\theta]^T \Sigma^{-1} [z(t) - y(t)\theta]$$

여기서, $Q(\theta)$ 는 다음과 같이 나타낼 수 있다.

$$Q(\theta) = \sum_{t=m}^T e(t)^T e(t) = \sum_{i=1}^n \sum_{t=m}^T [z_i(t) - y_i(t)\theta]^2$$

여기서, θ, Σ 에 대해서 $L \equiv L(\phi, \theta, \beta, \sigma^2 | e_0)$ 를 최대화하기 위해, θ 을 고정하고 Σ 을 최대화시키면 다음을 얻는다.

$$\hat{\Sigma} = [n(T-m+1)]^{-1} \sum_{t=m}^T [z(t) - y(t)\theta][z(t) - y(t)\theta]^T$$

또한, Σ 을 고정하고 θ 를 최대화시키는 것은 앞의 $Q(\theta)$ 을 최소화하는 것과 같아진다.

여기서, $Q(\theta)$ 을 최소화시키기 위해 Newton-Raphson의 반복 알고리즘을 이용한다.

$$\theta^{(i+1)} = \theta^{(i)} - H(\theta)^{-1} G(\theta^{(i)})$$

여기서, $\theta^{(i)}$: i -번째 반복에서 얻어진 추정모수벡터

gradient vector : $G(\theta) = \frac{dQ}{d\theta}$, Hessian matrix : $H(\theta) = \frac{d^2Q}{d\theta d\theta^T}$

$$\frac{dQ}{d\theta} = 2 \sum_{t=m}^T \frac{de(t)^T}{d\theta} e(t), \quad \frac{d^2Q}{d\theta d\theta^T} = 2 \sum_{t=m}^T [I_R \otimes e(t)^T] \frac{d^2e(t)}{d\theta d\theta^T} + 2 \sum_{t=m}^T \frac{de(t)^T}{d\theta} \frac{de(t)}{d\theta^T}$$

여기서, $e(t)$ 에 대한 도함수는 반복방정식을 사용하여 구할 수 있다.(Dai, Billard 2000).

(2) Kalman-Filter 방법을 이용한 모수추정

상태공간모형은 Kalman과 Bucy에 의해 도입되었다. 일반적인 상태공간 형태는 조 건부적인 Gaussian 모형을 적용하여 다음과 같은 형태를 따른다.

관측방정식(Observation Equation) : $y(t) = Z_t[y(t)]b(t) + u(t), t = 1, 2, \dots$

여기서, $y(t)$ 는 시점 t 에서의 관측 값인 $n \times 1$ 벡터, $Z(t)$ 는 $n \times q$ 행렬, $b(t)$ 는 $q \times 1$ 상태 벡터.

$$u(t) \sim \text{independently } MN(0, \sigma^2 Q(t))$$

상태방정식(State Equation) : $b(t) = T_t[y(t)]b(t-1) + F_t[y(t)]v(t), t = 1, 2, \dots$

여기서, F_t 는 $q \times p$ 행렬, T_t 는 전이행렬(transition matrix).

$$v(t) \sim \text{independently } MN(0, \sigma^2 R(t))$$

예로써 STBL($1_1, 1_1, 1_1, 1_1$)모형의 상태 공간 표현은 다음과 같다.

$$\begin{aligned} z(t) = & \phi_0 z(t-1) + \phi_1 W^{(1)} z(t-1) + \theta_0 e(t-1) + \theta_1 W^{(1)} e(t-1) \\ & + \beta_{00} z(t-1) \# e(t-1) + \beta_{01} z(t-1) \# [W^{(1)} e(t-1)] \\ & + \beta_{10} [W^{(1)} z(t-1)] \# e(t-1) + \beta_{11} [W^{(1)} z(t-1)] \# [W^{(1)} e(t-1)] + e(t), \end{aligned}$$

$t = 1, 2, \dots, T$

상태-공간표현 : $y(t) = Zb(t), \quad b(t) = T[y(t)]b(t-1) + Fe(t), \quad t = 1, 2, \dots, T,$

($y(t) = z(t)$)

여기서, $Z = [I_n \ 0_n]$, $b(t) = \begin{bmatrix} z(t) \\ e(t) \end{bmatrix}$, $F = \begin{bmatrix} I_n \\ I_n \end{bmatrix}$, $T[y(t)] = \begin{bmatrix} a & b+c(t) \\ 0 & 0 \end{bmatrix}$

$$a = \phi_0 + \phi_1 W^{(1)}, \quad b = \theta_0 + \theta_1 W^{(1)},$$

$$c(t) = \beta_{00} \text{diag}[z(t-1)] + \beta_{01} \text{diag}[z(t-1)] W^{(1)} + \beta_{10} \text{diag}[W^{(1)} z(t-1)] + \beta_{11} \text{diag}[W^{(1)} z(t-1)] W^{(1)}.$$

상태벡터 $b(t)$ 의 추정량은 $\beta(t|t-1)$ 에 의한 $y(1), y(2), \dots, y(t-1)$ 의 관찰 값으로 주어지고, t -번째 시점의 $\beta(t|t-1)$ 와 $C(t|t-1)$ 는 $t-1$ 번째 시점의 기원으로부터 계산되고, t -시점동안 $\beta(t|t)$ 와 $C(t|t)$ 추정치를 계산할 수 있다. 이러한 추정치의 갱신방정식(updated equation)은 다음과 같다.

$$\beta(t|t) = \beta(t|t-1) + C(t|t-1)Z^T H^{-1}(t)v(t), \quad C(t|t) = C(t|t-1) - C(t|t-1)Z^T H^{-1}(t)C(t|t-1)$$

여기서, $v(t) = y(t) - Z\beta(t|t-1)$, $H(t) = ZC(t|t-1)Z^T$, v_t 는 한 단계 후의 예측오차이고, $\sigma^2 H(t)$ 는 v_t 의 공분산 행렬이다.

한 단계 이후 추정치 $\beta(t+1|t)$ 와 $C(t+1|t)$ 는 다음의 예측방정식(prediction equation)에 의해 계산된다.

$$\beta(t+1|t) = T[y(t+1)]\beta(t|t), \quad C(t+1|t) = T[y(t+1)]C(t|t)T[y(t+1)]^T + FQ(t+1)F^T$$

상태벡터 $b(t)$ 가 정상성인 가정 하에서, 칼만 필터에서 첫 번째 시점의 초기 추정 값은 각각 $\beta(1|0) = E[b(0)]$, $C(1|0) = FQ(0)F^T = \sigma^{-2} \text{var}[b(0)]$ 이다.

Harvey(1981)는 로그우도함수가 예측오차 v_t , $t=1, \dots, T$ 의 결합분포로 분해될 수 있음을 보였고, $\{z(t-1), z(t-2), \dots\}$, $t=1, \dots, T$ 에 대한 $z(t)$ 의 조건부 분포는 정규분포이므로 로그우도함수는 다음과 같다.

$$\begin{aligned} L[\phi, \theta, \beta, \sigma^2; z(1), \dots, z(T)] &= \sum_{t=2}^T \log f[z(t)|z(t-1)] + \log f[z(1)] \\ &= -2^{-1}nT \log \sigma^2 - 2^{-1} \sum_{t=1}^T \log |H(t)| \end{aligned}$$

$$-2^{-1}\sigma^2 \sum_{t=1}^T v(t)^T H^{-1}(t)v(t)$$

모수 $\phi, \theta, \beta, \sigma^2$ 에 대한 $-2L[\phi, \theta, \beta, \sigma^2; z(1), \dots, z(T)]$ 의 최소화가 최대우도추정치이며,

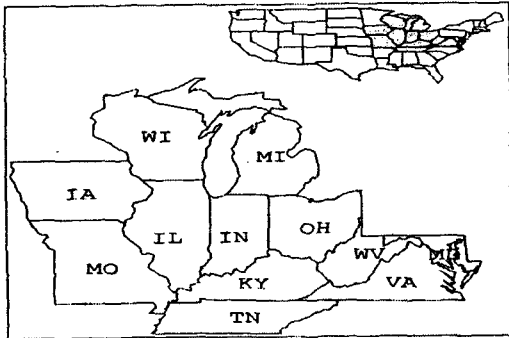
결국 $\hat{\sigma}^2(\phi, \theta, \beta) = SS(\phi, \theta, \beta)/(nT)$ 이 $-2L[\phi, \theta, \beta, \sigma^2; z(1), \dots, z(T)]$ 안의 σ^2 을 대신할 수 있으므로 모수 ϕ, θ, β 에 대한 $-2L_c[\phi, \theta, \beta; z(1), \dots, z(T)]$ 을 최소화 시키는 문제로 축소될 수 있다.

4. 사례연구

Mumps(유행성 이하선염)는 공간적, 시간적으로 퍼져나가는 전염성이 강한 질병이고, 이러한 전염병 자료는 의사를 비롯한 여러 과학자들에게 처음부터 연구대상이 되어왔다. 본 논문에서는 질병대책센터에서 수집한 미국의 12개 주(Maryland, Virginia,

West Virginia, Tennessee, Kentucky, Indiana, Ohio, Michigan, Illinois, Wisconsin, Iowa, Missouri)에 대해서 1969년 1월부터 1990년 12월까지의 월별 Mumps 자료를 사용했고, 국내자료로는 한국질병관리본부에서 2001년 1월부터 2006년 4월까지의 전염병 감시체제로 전산 보고된 16개 시도의 월별 Mumps 자료이다. 여기서 한국의 Mumps 자료는 8개도로 재분류하였다. 사용된 Mumps 자료는 연속성 자료가 아닌 빈도 자료로 Poisson분포를 따르는 변수이므로, 원래의 자료를 분산안정화변환(여기서, \sqrt{Z})과 12차 계절차분을 취한 후 정상시계열이 되었고, 이 자료를 정규분포로 근사시키기 위해 평균과 표준편차를 구해 자료를 표준화 하였다.

본 논문의 관심영역과 가중치행렬(weight matrix)은 미국의 경우 그림 4.1과 그림 4.2이며 [표 4.1]은 각 주의 첫 번째 이웃하고 있는 집합을 나타내며, 우리나라의 경우는 그림 4.3과 그림 4.4이며 [표 4.2]는 각 도의 첫 번째 이웃하고 있는 집합을 나타낸다.



$$W^{(1)} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

그림 4.2 1-차 이웃한 가중행렬

그림 4.1 12개주의 Mumps 관심 대상지역

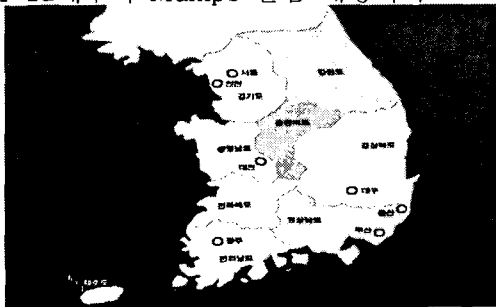


그림 4.3 8개도의 Mumps 대상지역

$$W^{(1)} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

그림 4.4 1-차 이웃한 가중행렬

[표 4.1] 각 주에서 첫 번째 이웃하고 있는 주의 집합

Site	state	Neighbor States
1	Maryland, MD	VA, WV
2	Virginia, VA	MD, WV, TN, KY
3	West Virginia, WV	MD, VA, KY, OH
4	Tennessee, TN	VA, KY, MO
5	Kentucky, KY	VA, WV, TN, IN, OH, IL, MO
6	Indiana, IN	KY, OH, MI, IL
7	Ohio, OH	WV, KY, IN, MI
8	Michigan, MI	IN, OH, WI
9	Illinois, IL	KY, IN, WI, IA, MO
10	Wisconsin, WI	MI, IL, IA
11	Iowa, IA	IL, WI, MO
12	Missouri, MO	TN, KY, IL, WI

[표 4.2] 각 주에서 첫 번째 이웃하고 있는 주의 집합

Site	Province	Neighbor States
1	KG	KW, CN, CB
2	KW	KG, CB, KB
3	CN	KG, CB, JB
4	CB	KG, KW, CN, JB, KB
5	JB	CN, CB, KB, KN, JN
6	JN	JB, KN
7	KB	KW, CB, JB, KN
8	KN	KB, JB, JN

4.1. 모형식별

비선형 시계열 모형인 STBL 모형식별은 시계열 정상성(stationary)을 만족하도록 변환되고 표준화된 자료를 이용해 우선 모형을 선형시계열 모형인 STARMA 모형으로 간주한 후 모형식별을 위하여 공간시계열자기상관함수(STACF)와 공간시계열편자기상관함(STPACF)를 구한다. 그 결과가 [표 4.3], [표4.4]에 있으며, 앞의 표 3.1을 참고로 하여 공간-시간차수를 식별하고, AIC, AICc등의 통계량을 이용해서 최종모형을 식별한 결과 STARMA(1₀, 1₁)으로 식별되었다.

식별된 STARMA모형을 앞의 2.2.절의 비선형 모형인 STBL 모형 식에서 추정된 선형모형식인 STARMA(1₀, 1₁)를 뺀 잔차의 제곱으로 취한 자료를 새로운 자료로 취하여 $Z(t)^2$ 의 STACF와 STPACF를 다시 구하여 증선형 항의 공간-시간차수를 식별

한 결과 최종 식별모형은 $STBL(1_1, 1_1, 1_1, 1_1)$ 로 결정되었다.

[표 4.3] $Z(t)$ 의 STACF(공간시계열자기상관함수)

	L=0	L=1	L=2	L=3
S=1	0.713150	0.360819	0.209175	0.157405
S=2	0.591886	0.309771	0.172458	0.137593
S=3	0.499260	0.259676	0.139606	0.111409
S=4	0.379767	0.220269	0.115650	0.082931
S=5	0.298184	0.168880	0.091436	0.051672
S=6	0.234337	0.143793	0.073273	0.023808
S=7	0.151589	0.119658	0.052311	0.007977
S=8	0.101101	0.101949	0.046367	-0.009243
S=9	0.021934	0.078328	0.033267	-0.029496
S=10	-0.049063	0.047273	0.020915	-0.033300
⋮	⋮	⋮	⋮	⋮

[표 4.4] $Z(t)$ 의 STPACF(공간시계열편자기상관함수)

	L=0	L=1	L=2	L=3
S=1	0.713150	0.134451	0.0008301	0.0135536
S=2	0.159082	-0.064428	-0.0485200	-0.0006140
S=3	0.060623	-0.059372	-0.0218540	-0.0249290
S=4	-0.062317	0.015633	0.0266167	-0.0396760
S=5	-0.014361	-0.051407	0.0171428	-0.0516380
S=6	0.006623	0.043054	0.0074412	-0.0327150
S=7	-0.062842	0.026888	-0.0151890	0.0238461
S=8	-0.015841	0.040039	0.0284608	-0.0198520
S=9	-0.097198	0.013107	0.0144077	-0.0238510
S=10	-0.078355	-0.004166	0.0214072	0.0398651
⋮	⋮	⋮	⋮	⋮

4.2. 모수추정

앞의 미국 12개 주 Mumps 자료의 최종 식별된 모형인 $STBL(1_1, 1_1, 1_1, 1_1)$ 에 대한 모수추정은 Kalman Filter 방법과 Newton-Raphson 방법의 최우추정치(MLE)를 구하였고, 추정치중 통계적으로 유의하지 않은 계수를 제거한 후의 최종 모형은 Kalman의 경우 $STBL(1_0, 1_1, 0_1, 1_0)$ 이고, Newton-Raphson의 경우 $STBL(1_1, 1_1, 0_1, 1_0)$ 로써 모수를 추정한 결과는 [표 4.5]와 같다. 그리고 국내 Mumps 자료의 경우 같은 절차와 방법으로 구하여진 최종 모형은 $STBL(1_1, 1_1, 1_1, 1_1)$ 으로 [표 4.6]와 같다. [표 4.5]~[표 4.6]에서 알 수 있듯이 kalman filter 방법에 의한 모수추

정치를 초기 값으로 하여 구한 Newton-Raphson MLE가 더 작은 MSE와 AIC를 보이므로 더 적합한 추정방법이라 할 수 있겠다.

[표 4.5] (미국 Mumps 자료) 최종모형의 모수추정 결과

계수	Kalman-Filter	Newton-Raphson
ϕ_0	0.8270	0.8129
ϕ_1		0.0461
θ_0	-0.2687	-0.2537
θ_1	0.1711	0.0970
β_{01}	-0.0730	-0.0872
β_{10}	0.0221	0.0258
MSE	0.4668	0.4460
AIC	-2403.69	-2555.66

[표 4.6] (한국 Mumps 자료) 최종모형의 모수추정 결과

계수	Kalman-Filter	Newton-Raphson
ϕ_0	0.5670	0.5199
ϕ_1	0.3173	0.3266
θ_0	-0.2425	-0.1897
θ_1	-0.2380	-0.2837
β_{00}	-0.0280	-0.0181
β_{01}	-0.2739	-0.1848
β_{10}	0.1514	0.1021
β_{11}	0.0378	0.0429
MSE	0.7825	0.7820
AIC	-87.98	-102.09

4.3. 예측(Prediction)

위의 최종 모형은 1968년 1월~1990년 12월까지의 미국 12개 주의 mumps(유행성이하선염)월별자료를 이용해 추정되었고, 12개 주에 대한 1991년 1월~12월까지를 예측력 비교에 사용 하였다. 예측력 비교는 아래의 예측 오차 제곱합(Sum of Square Forecast Error)을 이용 하였고, 그 결과는 [표 4.7]과 같다.

$$SSF = \sum_{i=1}^{12} \sum_{j=1}^{12} \{(OBS_{ij} - Fore_{ij})\}^2, i = \text{주}, j = \text{월}$$

[표 4.7] 월별 Mumps 자료에 대한 관측값과 예측값의 비교

Site-month	관측 값	Kalman Prediction	Newton prediction
MD-1	14	15	3
MD-2	14	10	5
MD-3	5	27	16
MD-4	6	21	12
MD-5	8	16	9
MD-6	13	11	6
MD-7	5	3	1
MD-8	2	12	9
MD-9	2	5	3
⋮	⋮	⋮	⋮
MO-10	0	0	0
MO-11	0	0	0
MO-12	0	0	0
SSF		15120	11927

한국 Mumps 자료의 최종모형은 2001년 1월 ~ 2006년 5월 자료까지의 8개도의 Mumps(유행성이하선염)월별자료를 이용해 추정되었고, 8개도에 대한 2006년 6월 ~ 2006년 9월까지를 예측력 비교에 사용하였다. 예측력 비교는 [표 4.8]과 같다.

$$SSF = \sum_{i=1}^8 \sum_{j=6}^9 \{(OBS_{ij} - Fore_{ij})\}^2, i = \text{도}, j = \text{월}$$

[표 4.8] 월별 Mumps 자료에 대한 관측값과 예측값의 비교

Site-month	관측 값	Kalman Prediction	Newton prediction
경기-6	184	6	9
경기-7	120	5	6
경기-8	92	6	7
경기-9	37	0	0
강원-6	10	0	0
강원-7	14	0	0
강원-8	6	1	1
강원-9	2	0	0
⋮	⋮	⋮	⋮
경남-6	37	10	7
경남-7	24	3	1
경남-8	18	0	0
경남-9	8	1	0
SSF		56958.0	55712.0

5. 결론

비선형 공간시계열 모형인 STBL 모형에서 모수추정의 경우 Kalman-Filter와 수치해석 방법인 Newton-Raphson방법을 비교한 결과는 평균제곱오차 $MSE(=\hat{\sigma}^2)$ 와 AIC통계량을 비교해볼 때 Kalman-Filter보다 Newton-Raphson 방법이 더 적합한 것으로 나타났다. 하지만 복잡한 비선형 모형인 STBL의 경우 Kalman-Filter는 임의의 초기 값에 영향을 덜 받고 수렴속도가 빠른 장점이 있으며, Newton-Raphson 방법은 초기 값을 임의로 부여할 경우 수렴성이 떨어지거나 수렴 속도가 느린 단점이 있다. 이 경우 Kalman-Filter 방법에 의한 추정치를 초기 값으로 하여 Newton-Raphson방법으로 MLE를 추정하는 것이 바람직하며 추정의 정도가 더 좋아진다. 예측방법의 경우 Kalman-Prediction의 예측오차제곱합(SSF)보다 칼만필터 추정치를 초기 값으로 이용한 Newton-Raphson 방법에 의한 조건부기대치(conditional expectation)를 이용한 예측력이 더 좋게 나왔다. 즉, Kalman Filter 추정치를 초기 값으로 한 Newton-Raphson 추정 시 수렴성과 수렴속도를 향상시킬 수 있었으며, 실제 자료에서 MSE와 AIC등의 결과를 보아도 탁월함을 알 수 있었다. 또한 Kalman Prediction 방법보다 Newton-Raphson Prediction 방법이 더 우수한 예측력을 보임을 알 수 있었다. 추후 연구계획으로는 본 연구에서 가중행렬(weight matrix)을 이웃(neighbor)한 지역에 동일한 가중치를 부여하였는데 공간(Site)의 사전정보에 의해 이웃한지역의 가중치를 달리 부여하거나 2~3차 가중행렬의 경우 거리가 멀어짐에 따라 일정한 패턴으로 적은 가중치를 부여하는 방법에 대한 연구를 진행할 예정이다.

Acknowledgement

This work was supported by the Chungbuk BIT Research-oriented Consortium(CBITRC)

References

- [1] 김선우, 정애란, 이성덕(2005) 공간자료를 이용한 세미베리오그램 추정, 응용통계연구, 18(2), 271-280.
- [2] 손건태, 백지선(1997), 공간자기회귀모형의 식별, 응용통계연구, 10(1), 121-136.
- [3] Alcroft, D.J. and Glasbey, C.A.(2005). STARMA process Applied to Solar Radiation, Biomathematics and Statistics Scotland, 1-24.

- [4] Ali.M.M.(1979) Analysis of Stationary Spatial-temporal Processes: Estimation and Prediction. *Biometrika*, 66, 513-518
- [5] Bennett, R.J.(1971) *Spatial Time Series Analysis-Forecasting-Control*. Pion, London.
- [6] Billard, L., and Yuqing. Dai(2003). Maximum Likelihood Estimation in Space Time Bilinear model, *Journal of Time Series Analysis*, 25-44.
- [7] Cliff, A.D. and J.K. Ord(1975). Space Time Modeling with an Application to Regional Forecasting, *Transaction of the institute of British Geographers*, 64, 119-128.
- [8] Dai, Y. and Billard, L.(1998) A Space-Time Bilinear Model And Its Identification, *Journal of Time Series Analysis*, 19, 657-679.
- [9] Harvey, A.C.(1989). *Forecasting, Structural Time Series Models and The Kalman Filter*, Cambridge University Press: Cambridge.
- [10] Kalman, R. E.(1960). A New Approach to Linear Filtering and Prediction Problems, *Transaction ASME Journal of Basic Engineering Series D*, 83, 95-108.
- [11] Li, W.K.(1984) On the Autocorrelation Structure and Identification of Some Bilinear Time Series. *Journal of Time Series Analysis* 5, 173-181.
- [12] Pfeifer, P.E. and S.J. Deutsch(1980b). Identification and Interpretation of First order space-time ARMA Models, *Technometrics*, 22, 397-408.