

Statistical Inference for Space Time Series Model with Application to Mumps Data¹⁾

Ae-Ran Jeong²⁾ · Sun-Woo Kim³⁾ · Sung-Duck Lee⁴⁾

Abstract

Space time series data can be viewed either as a set of time series collected simultaneously at a number of spatial locations or as sets of spatial data collected at a number of time points. The major purpose of this article is to formulate a class of space time autoregressive moving average (STARMA) model, to discuss some of the their statistical properties such as model identification approaches, some procedure for estimation and the predictions. For illustration, we apply this STARMA model to the mumps data. The data set of mumps cases consists of the number of cases of mumps reported from twelve states monthly over the years 1969-1988.

Keywords : AIC, Maximum likelihood estimation, Mumps data, Prediction, Space time autoregressive moving average model

1. 서론

공간의 위치와 시간의 변화에 따라 동시에 관찰되어지는 공간시계열 자료 (Space-time series data)는 기상학, 지질학, 천문학, 생태학 등과 더불어 환경공학 및 공간적 역학과 같은 전염병 확산 모형에서까지도 광범위하게 나타나고 있다. 최근에는 공간시계열 자료의 분석에 있어서 단편적이고 전통적인 공간통계학과 시계열분석보다는 공간통계학과 시계열분석의 융합(Convergence)으로 이루어진 공간시계열모형의 개념을 적용시키는데 많은 관심을 가져왔다.

일반적으로 공간 자료는 특정 위치 또는 특정 지역에서의 관측 값이나 측정값으로

1) 본 논문은 2005년 충북대학교 학술연구 지원사업에 의하여 연구되었음.

2) 충북 청주시 개신동 충북대학교 전자계산학과 전산통계전공 이학박사
E-mail : arjeong@freechal.com

3) 충북 청주시 개신동 충북대학교 자연과학대학 정보통계학과 박사과정
E-mail : statg023@hanmail.net

4) 교신저자 : 충북 청주시 개신동 충북대학교 정보통계학과, 기초과학연구소 교수
E-mail : sdlee@chungbuk.ac.kr

구성되며, 특히, 공간 자료는 지리통계자료(Geostatistical data), 격자자료(Lattice data), 공간 점 패턴자료(Spatial point patterns data)의 세 가지 유형으로 분류되며 본 논문에서는 격자자료를 중심으로 연구 한다.

Pfeifer 와 Deutsch(1980)는 시계열 분석에서 사용된 ARMA(Autoregressive Moving Average) 모형에 공간의 정보를 포함시킨 공간 ARMA 모형, 즉 STARMA(Space-time ARMA) 모형을 제시하였다. 이 STARMA 모형은 ARMA 모형을 시간과 공간으로 확장시킨 모형이고, 각 위치에서 이웃하고 있는 계층적인 공간 차수를 사용한다.

Dai와 Billard(2000)는 전염병 데이터를 공간시계열 중선형 모형(Space-time bilinear model)에 대해서 적용하였다. STBL 모형은 공간 이웃 구조 중선형 성질을 가지고 있는 시계열 모형을 사용한 다중 중선형 시계열의 공간적인 형태이다. STBL 모형 실용성과 식별은 Dai와 Billard(1998)에 의해서 제안 되었고 논의되었다. 그들은 STBL 모형에 대해서 최우추정법과 결측치에 대한 모수 추정을 상태 공간(State space)으로 재 표현하고 칼만 필터(Kalman filter)를 이용해서 모수 추정을 하였다.

Kamarianakis와 Prastacos(2002)는 교통량 흐름을 공간시계열 모형에 적용하였다. 교통량 흐름 자료는 연속적인 공간시계열 형태로 되어 있고 시간의 상수 구간 사이에서 특정한 위치에서 수집되는데 여러 가지 위치에서 기본적인 거리를 추정한 가중 행렬을 사용한 STARIMA 모형을 적용하였다. 이들은 정상 공간시계열 모형이 교통량 흐름의 단기 예측에 유용하고 또 다른 교통망을 변화하는 교통량 흐름에 영향을 줄 수 있다는 것을 보였다.

Allcroft와 Glasbey(2005)은 STARMA 과정을 태양 복사 에너지에 적용시키고, Gaussian Markov Random Fields(GMRF)와 Partially Orderd Markov Models(POMM) 에 대해서 다루었다.

김선우 등(2005)은 공간자료에 대한 지리적 가중회귀 모형 과 kriging의 비교를 하였다. 공간 상관이 존재하는 지리통계 자료(Geostatistical data)에 대하여 일반적으로 널리 사용되는 kriging 모형과 통계학적 공간자료 분석모형인 지리적 가중회귀 모형을 고려하고, 미지의 위치에 대한 예측력을 비교하였다.

본 논문에서는 STARMA의 실제 적용 사례로서 공간적, 시간적으로 퍼져나가는 전염성이 강한 질병인 Mumps(유행성 이하선염) 자료(1969년 1월부터 1988년 12월 까지 미국 12개 주에 대한 월별 자료)를 이용하여 적절한 공간 시계열 모형을 가정, 모형식별을 하고 구체적으로 공간시계열 모형 중에서 STAR과 STARMA모형을 이용하여 최우추정법을 이용해서 모수추정을 하였고 예측하였다. 모형 선택으로는 AIC(Akaike information criterion), AICc(Corrected Akaike information criterion), 또는 BIC(Bayesian information criterion)의 측도를 이용해서 비교분석 하였다.

2장에서는 공간시계열 모형에 대해 소개하였고, 3장에서는 공간시계열 모형에 대하여 STACF와 STPACF를 이용하여 모형식별을 하였다. 4장에서는 공간시계열 자기회귀 이동평균 모형에 대해서 최우추정법을 이용해서 모수를 추정하였고, 5장에서는 모형 선택 통계량으로 AIC, AICc BIC 통계량을 구해서 모형 선택을 해서 STAR (1_1) 모형과 STARMA $(1_1, 1_1)$ 모형이 잠정적으로 채택 되었으므로, 이 두 모형을 비교 분석하였다. 6장에서는 실제 데이터를 가지고 모형식별, 모수추정, 모형 선택 통계량을 구하였고, 1989년도 12개 주에 대한 월별 자료를 두 모형으로 예측하여 비교하였다.

2. 공간시계열자기회귀 이동평균모형 (STARMA Model)

STARMA 모형은 위치 i ($i=1, 2, \dots, g$)와 시간으로 나타내는 확률변수 $Z_i(t)$ 로 표현되어 진다.

$Z(t)$ 는 시점 t 에서 $g \times 1$ 관측치 벡터이다. 즉, $Z(t) = [Z_1(t), Z_2(t), \dots, Z_g(t)]'$ 여기서, g 은 이차원 공간의 전체위치 수이다. STARMA(p, λ, q, η) 모형은 다음과 같다.

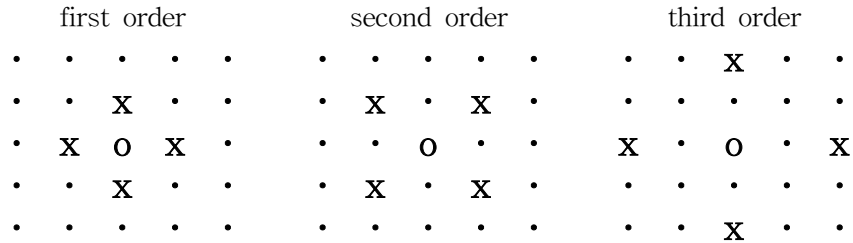
$$Z(t) = \sum_{i=1}^p \sum_{m=0}^{\lambda_i} \phi_m^i W^{(m)} Z(t-i) - \sum_{j=1}^q \sum_{n=0}^{\eta_j} \theta_n^j W^{(n)} e(t-j) + e(t) \quad (2.1)$$

여기서 p 는 최대 자기회귀 차수이고, q 는 최대 이동평균 차수이고, λ_i 는 i 번째 자기회귀 항의 차수이고, η_j 는 j 번째 이동평균항의 차수이고, ϕ_m^i 은 시간차수 i 와 공간차수 m 의 자기회귀 모수이고, θ_n^j 은 시간차수 j 와 공간차수 n 의 이동평균 모수이고, $e(t) = [e_1(t), e_2(t), \dots, e_g(t)]'$ 는 t 시점에서 확률잡음 벡터이다. $W^{(m)}$ 은 공간 차수 m 에 대한 $g \times g$ 가중행렬을 나타낸다.

인접한 위치에 따른 가중치 행렬은 <그림 1>과 같이 주어진다. 즉, 가장 가까운 이웃 값들(1차 neighbours)에 대한 가중치는 가장 인접한 유클리디안 거리를 갖는 위치들에 대하여 동일하게 주고 그 다음 가까운 이웃 값들(2차 neighbours)에 대한 가중치는 다음 그림에서와 같이 그 다음 인접한 유클리디안 거리를 갖는 위치들에 대하여 동등하게 가중치를 준다. <그림 2>는 공간차수에서 이웃하고 있는 1, 2, 3차 이웃구조를 나타낸다.

•	•	•	•	•	•	•
•	•	•	3	•	•	•
•	•	2	1	2	•	•
•	3	1	0	1	3	•
•	•	2	1	2	•	•
•	•	•	3	•	•	•

<그림 1> 위치에 대한 가중치



<그림 2> 공간 차수에서 이웃하고 있는 구조

3. 모형식별방법

공간시계열 모형의 식별 단계에서는 먼저 시계열에 대한 정상성(stationarity) 여부를 검토한 다음 공간시계열 자기상관함수(Space-time autocorrelation function: STACF)와 공간시계열 부분 자기상관 함수(Space-time partial autocorrelation function: STPACF)를 나타내서 식별한다. 공간 차수 l 과 시간 차수 s 사이의 STACF는 다음과 같이 나타낸다.

$$\rho_{lk}(s) = \frac{\gamma_{ll}(s)}{\sqrt{\gamma_{ll}(0)\gamma_{kk}(0)}}$$

여기서, $\gamma_{lk}(s)$ 는 시차 s 에 이웃하고 있는 차수 l 번째와 k 번째 사이의 공간시계열 자기공분산함수이다. 공간시계열 자기공분산 함수는 다음과 같다. 여기서, g 는 위치의 차(space lag)이고, s 는 시차(time lag)이다

$$\gamma_{lk}(s) = E_t \left[\frac{[W^{(l)}Z(t)]'[W^{(k)}Z(t+s)]}{g} \right]$$

공간시계열 자기공분산 함수의 추정은 다음과 같다.

$$\hat{\gamma}_{lk}(s) = \frac{1}{g} \frac{\sum_{t=1}^{T-s} [W^{(l)}Z(t)]'[W^{(k)}Z(t+s)]}{T-s}$$

따라서, 공간시계열 자기상관함수(STACF)는 다음과 같이 추정된다.

$$\hat{\rho}_{lk}(s) = \frac{T}{T-s} \frac{\sum_{t=1}^{T-s} [W^{(l)}Z(t)]'Z(t+s)}{\left(\sum_t [W^{(l)}Z(t)]'[W^{(l)}Z(t)] \sum_t [Z(t)'Z(t)] \right)^{1/2}}$$

공간시계열 부분 자기상관함수(STPACF)는 s 와 h 를 조건으로 한 BLP(Best Linear Predictor) $Z(t)$ 는

$$\widehat{Z}(t)_{s,h} = \sum_{j=1}^s \sum_{l=0}^{\lambda} \phi_{jl} W^{(l)} Z(t-j)$$

여기서, $s=1, \dots, p$, $h=0, \dots, \lambda$, $\{\phi_{jl} \mid j=1, \dots, k, \quad l=0, \dots, \lambda\}$ 계수의 집합은 BLP $\widehat{z}(t)_{s,h}$ 으로 찾아낼 수 있다. 공간 시계열 편자기함수 ϕ_{kl} 를 얻기 위하여 STAR 모형에 $[W^{(h)}Z(t-s)]'$ 를 양변에 곱하면 다음과 같다.

STAR 모형식은 식 (3.1)과 같다.

$$Z(t) = \sum_{j=1}^k \sum_{l=0}^{\lambda} \phi_{jl} W^{(l)} Z(t-j) + e(t) \quad (3.1)$$

$$\begin{aligned} & Z'(t-s) W^{(h)'} Z(t) \\ &= \sum_{j=1}^s \sum_{l=0}^{\lambda} \phi_{jl} Z'(t-s) W^{(h)'} W^{(l)} Z(t-j) + Z'(t-s) W^{(h)'} e(t) \end{aligned} \quad (3.2)$$

식 (3.2)을 기대값을 취하고 양변을 N으로 나누면 다음과 같은 자기공분산함수 식을 얻는다.

$$\gamma_{h0}(s) = \sum_{j=1}^k \sum_{l=0}^{\lambda} \phi_{jl} \gamma_{hl}(s-j), \quad s=1, 2, \dots, k, \quad h=0, 1, \dots, \lambda.$$

여기서, $E[Z(t-s)'e(t)] = 0$, $s > 0$ 이다.

Yule-Walker 방정식으로부터 공간시계열 부분자기상관함수 ϕ_{kl} 를 구할 수 있다.

4. 모수 추정

STARMA 모형의 최우추정방법은 앞에서 정의한 식(2.1)에 대하여 $e(t)$ 는 다음과 같이 정의된다.

$$e(t) = z(t) - y(t)\Theta$$

여기서,

$$\begin{aligned} y(t) &= [w^{(0)}z(t-1), \dots, w^{(\lambda_1)}z(t-1), \dots, w^{(0)}z(t-p), \dots, w^{(\lambda_p)}z(t-p), \\ & \quad w^{(0)}e(t-1), \dots, w^{(\eta_1)}e(t-1), \dots, w^{(0)}e(t-q), \dots, w^{(\eta_q)}e(t-q)] \\ \Theta &= (\phi_0^1, \dots, \phi_{\lambda_1}^1, \dots, \phi_0^p, \dots, \phi_{\lambda_p}^p, \theta_0^1, \dots, \theta_{\eta_1}^1, \dots, \theta_0^p, \dots, \theta_{\eta_q}^p)^T \text{ 이다.} \end{aligned}$$

$\{e(t)\}$ 는 정규분포라 가정한다. $t=1, \dots, T$ 이다. 그러면 우도함수는 다음과 같다.

$$L(\phi, \theta, \sigma^2) = (2\pi)^{-\frac{Tg}{2}} \sum_{t=1}^T e^{-\frac{1}{2} [e(t)^T \Sigma^{-1} e(t)]}$$

또한 로그우도함수를 나타내면 다음과 같다.

$$\ln L(\phi, \theta, \sigma^2) = -\frac{Tg}{2} \ln(2\pi) - \frac{Tg}{2} \ln \left| \Sigma \right| - 2^{-1} \sum_{t=1}^T [z(t) - y(t)\theta]^T \Sigma^{-1} [z(t) - y(t)\theta]$$

여기서, $Q(\theta)$ 는 다음과 같이 주어진다.

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

모수 θ 와 Σ 에 대해서 $L \equiv L(\phi, \theta, \sigma^2)$ 를 최대화 하기 위해서는, θ 을 고정하고, Σ 을 최대화 시키면 (Anderson, 1984) 다음을 얻는다.

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

Σ 를 고정하고, L을 최대화 시키는 것은 $Q(\theta)$ 를 최소화 시키는 방법과 같다.

$\{e(t)\}$ 는 비선형이기 때문에 명확하게 최대우도 함수를 구할 수 없다. 그래서 $Q(\theta)$ 을 최소화 시키는 Newton-Raphson 방법과 같은 수치적인 방법을 사용한다. $Q(\theta)$ 을 최소화 시키는 Newton-Raphson 반복적인 방정식 다음과 같다.

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

여기서, $\theta^{(i)}$ 는 i 번을 반복해서 얻은 모수 추정치의 벡터이다. 기울기(gradient) 벡터 $G(\theta)$ 와 헤시안(Hessian) 행렬 $H(\theta)$ 는 다음과 같다.

$$G(\theta) = \frac{dQ}{d\theta}, \quad H(\theta) = \frac{d^2Q}{d\theta d\theta^T}$$

여기서, $\frac{dQ}{d\theta} = 2 \sum_{t=1}^T \frac{de(t)^T}{d\theta} e(t)$ 이다.

$$\frac{d^2Q}{d\theta d\theta^T} = 2 \sum_{t=1}^T [I_R \otimes e(t)^T] \frac{d^2e(t)}{d\theta d\theta^T} + 2 \sum_{t=1}^T \frac{de(t)^T d\theta}{d\theta d\theta^T}$$

5. 모형 선택을 위한 통계량

모형 선택을 위한 통계량으로 AIC(Akaike's information criteria), AICc, BIC 통계량으로 일반적인 식은 각각 다음과 같다.

$$AIC = -2 \ln(\hat{\sigma}_a^2) + 2M,$$

$$AICc = n \log(\hat{\sigma}_a^2) + \frac{1 + M/n}{1 - (M+2)/n},$$

$$BIC = n \ln \hat{\sigma}_a^2 - (n - M) \ln(1 - \frac{M}{n}) + M \ln n + M \ln [((\frac{\hat{\sigma}_a^2}{\hat{\sigma}_z^2}) - 1) / M]$$

여기서, n 은 관측치들의 개수, M 은 모수들의 개수이고, $\hat{\sigma}_a^2$ 는 σ_a^2 의 최우추정량이고, $\hat{\sigma}_z^2$ 는 시계열의 표본분산이다.

한편 공간시계열 모형인 STARMA 모형의 AICc 통계량은 다음과 같이 나타낸다.

$$AICc = Tg \log(\hat{\sigma}_a^2) + \frac{Tg + k(\hat{\phi}) + K(\hat{\theta})}{Tg - K(\hat{\phi}) - k(\hat{\theta}) - 2}$$

BIC 통계량은 다음과 같이 나타낸다.

$$BIC = Tg \log(\hat{\sigma}_a^2) + \frac{S(\hat{\beta})}{\hat{\sigma}_z^2} + (K(\hat{\phi}) + K(\hat{\theta})) \log(Tg)$$

6. 사례 연구

6.1 자료설명

Mumps(유행성 이하선염, 항아리 손님)는 공간적, 시간적으로 퍼져나가는 전염성이 강한 질병이고, 이러한 전염병 자료는 의사를 비롯한 여러 과학자들에게 처음부터 연구대상이 되어 왔다. Mumps는 예방 백신을 하지 않으면 청소년과 노년층 사이에서 주기적으로 아주 크게 발생한다는 것을 알 수 있다.

질병대책센터(CDC : Centers for Disease Control and Prevention)에서는 여러 가지 전염병에 대한 기록과 대책을 세우고 있으며, 특히 질병대책센터는 각 전염병에 대한 주별 자료와 연간 자료를 발간하며, 주별 자료는 병으로 인한 사망을 기록하고, 연간 자료는 전염성 질병에 대한 전체적 고찰을 다루고 있다.

본 논문에서는 질병대책센터에서 수집한 미국의 12개 주(Maryland, Virginia, West Virginia, Tennessee, Kentucky, Indiana, Ohio, Michigan, Illinois, Wisconsin, Iowa,

Missouri)에 대해서 1969년 1월부터 1988년 12월까지의 월별로 발생한 Mumps 자료를 사용했다. 자료의 정상성(Stationarity) 가정을 만족시키기 위해서는 분산 안정화 변환(Variance stabilizing transformation)을 하고 원래의 자료를 계절성(Seasonality)을 조정하기 위하여 12차분 구조로 변환 하였다. 분산 안정화 변환을 위하여 $h(z) = \sqrt{z}$ 와 같은 변환을 취하였다

<그림 3> 12개 주의 Mumps 관심대상 지역 지역

각 주에서 이웃(Neighbor)은 보통 지형적인 공통경계(Common border)를 중심으로 나누어져 있다. <그림 3> 안에 작은 지도는 미국 전체를 나타내며 그 중 본 연구의 흥미영역인 12개의 주에 대하여 큰 지도로 나타내었다. <표 1>에서 12개 주에 대한 가중치 행렬(Weight matrix), $W^{(l)}$ 은 차원이 12인 12×12 행렬이고, 그 원소 $W^{(l)}$ 은 l 번째 이웃하는 주의 개수의 역비례로 구할 수 있다. <표 2>는 각 주에서 첫 번째 이웃하고 있는 집합을 표시한 것이다. 따라서 12개 주에 첫 번째 이웃의 가중 행렬은 다음과 같다.

<표 1> 가중행렬

$$W^{(1)} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/4 & 0 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 \\ 0 & 1/7 & 1/7 & 1/7 & 0 & 1/7 & 1/7 & 0 & 1/7 & 0 & 0 & 1/7 \\ 0 & 0 & 0 & 0 & 1/4 & 0 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\ \dots & \dots & \dots & & & & & & & & & \end{pmatrix}$$

<표 2> 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

6.2 모형식별

모형식별은 두 단계를 비교해서 나타낸다. 첫 번째는 기본모형의 선형성분을 결합해서 모형의 차수를 결정하는 방법이다. 두 번째 방법은 관측 값 $z(t)$ 에 대해서 공간 차수 l 와 시간차수 s 에 대해서 공간시계열 자기 상관함수(Space-time autocorrelation function:STACF)와 공간시계열 부분 자기상관함수(Space-time partial autocorrelation function:STPACF)를 이용해서 모형식별을 한다. Mumps 자료를 분석하기 위해 적합한 STARMA 모형을 식별하고자 한다. STAR(p_λ)모형은 공간시계열 자기상관함수는 시간과 공간 차수에서 지수적으로 감소하고 공간시계열 부분 자기상관함수 값은 시간 차수 p 차 이후에 공간 차수 λ 차 이후에 절단하는 것을 볼 수 있다. STMA(q_m)모형은 STACF는 시간 차수 q 차 이후에 공간 차수 m 차 이후에 절단되고, STPACF는 시간과 공간 차수에서 지수적으로 감소한다. STARMA(p_λ, q_m)모형은 STACF와 STPACF는 시간과 공간 차수에서 지수적으로 감소하는 것을 볼 수 있다.

본 논문에서는 이러한 내용을 종합적으로 볼 때 공간시계열 모형 중에서 STAR(1_1)과 STARMA($1_1, 1_1$) 모형으로 식별할 수 있다.

6.3 모수추정

STAR(1_1) 모형, STARMA($1_1, 1_1$) 모형에 대해서 모수들을 추정하기 위해서 <표 3>에 Newton-Raphson 반복법을 이용한 최대우도 추정치를 나타내었다. <표 3>에서 두 가지 모형에 의한 모수 추정 값을 구하였다. 모수의 추정 값을 구하기 위해서 우도함수방법의 경우에는 SAS/IML 모듈을 사용하여서 모수를 추정하였다. , .

<표 4>는 두 가지 모형에 대해 최우추정방법으로 적합 시킨 후 계산된 AIC, AICc, BIC 통계량을 표시하였다. Newton-Raphson 반복법을 이용한 최대우도 추정방법에서

STARMA 모형이 가장 작은 AIC와 AICc, BIC 값을 가지는 것을 볼 수 있다.

<표 3> 공간시계열(Space-time) 모형에 대한 모수추정치

방법	모형	모수	추정치	표준오차	T 값
MLE	STAR	ϕ_1	0.6946208	0.0004567	32.50443
		ϕ_2	0.1133934	0.001084	3.4440892
		σ^2	0.4784042		
	STARMA	ϕ_1	0.818722	0.0005692	34.317521
		ϕ_2	0.0297903	0.0013883	0.7995346
		θ_1	-0.237609	0.0014736	-6.1897
		θ_2	0.1066078	0.003933	1.699131
		σ^2	0.465837		

<표 4> 공간시계열 모형의 선택 통계량

방법	모형	통계량	추정치
MLE	STAR	AIC	-1851.994
		AICc	668.01551
		BIC	-1834.498
	STARMA	AIC	-1915.077
		AICc	605.95346
		BIC	-1885.917

6.4 예측(Prediction)

Mumps 자료를 가지고 STAR 모형과 STARMA 모형에 적합 시키고 최우추정법과 칼만 필터를 이용해서 모수를 추정하고 각 12개 주에 대해서 1969-1989년도 월 별 자료 중 1989년도 월 별 자료를 제외한 1969-1988년도 월 별 자료를 이용하여 모형을 적합하고 그 나머지 12개 주에 대한 1989년도 월 별 자료를 예측하였다.

STAR(1, 1) 모형과 STARMA(1, 1, 1) 모형으로 적합하고, 식 (6.1)과 같은 예측 오차 제곱합(Sum of Square Prediction Error)기준을 이용하여 비교한다.

$$SSF = \sum_{i=1}^{12} \{(\text{관측값} - \text{예측값})\}^2, \quad i = \text{해당 월} \quad (6.1)$$

STAR(1₁) 모형식은 식 (6.2)와 같이 나타낸다.

$$Z(t) = \phi_0 Z(t-1) + \phi_1 W^{(1)} Z(t-1) + e(t), \quad t = 1, \dots, T \quad (6.2)$$

STARMA (1₁, 1₁) 모형식은 식(6.3)과 같이 나타낸다.

$$Z(t) = \phi_0 Z(t-1) + \phi_1 w^{(1)} Z(t-1) + \theta_0 e(t-1) + \theta_1 W^{(1)} e(t-1) + e(t) \quad (6.3)$$

1989년도에 대해서 12개 주에 대해서 월 별 자료를 실제값과 예측값을 <표 5>에 나타내었다.

STAR 모형에 대한 예측 오차 제곱 합은 SSF=190910이고, STARMA 모형에 대한 예측 오차 제곱 합은 SSF=173969이다. 이러한 결과로부터 STARMA 모형이 더 적합하다고 볼 수 있다.

<표 5> 월 별 자료에 대한 관측값과 예측값

월 \ 모형	STAR 모형		STARMA 모형	
	관측값	예측값	관측값	예측값
1월	116	75	116	116
2월	163	235	163	195
3월	274	2	274	30
4월	201	385	201	280
5월	285	330	285	170
6월	304	385	304	150
7월	150	1	150	3
8월	102	65	102	90
9월	93	9	93	88
10월	105	227	105	295
11월	95	180	95	125
12월	98	220	98	205
SSF	190910		173969	

7. 결 론

공간시계열 자료의 정의는 “공간 위치에서 수집된 시계열의 자료이거나 시간 변화에서 수집된 공간 자료의 집합으로서, 본 연구는 시계열 ARMA 모형에 공간개념을 확장시킨 STARMA 모형에 대하여 모형식별, 모수추정 및 예측문제를 다루었다. 1968년부터 1989년까지의 미국 중북부 12주의 Mumps 자료를 가지고 STACF와 STPACF를 이용하여 모형식별 한 결과 STAR(1₁) 과 STARMA(1₁, 1₁) 모형이 잠재적으로 선택되었고, 모수 추정 및 모형 선택을 위한 통계량에 의해 STARMA 모형이 예측오차제곱합 기준에서 더욱 우수한 예측 값을 제공함을 알 수 있었다.

참 고 문 헌

1. 김선우 · 정애란 · 이성덕(2005), 공간자료를 이용한 세미베리오그램 추정, *한국통계학회 응용통계연구*, 제18권 2호, 271-280.
2. Anderson, T. W. (1984). *An Introuction to Multivariate Statistical Analysis(2nd ed.)*. John Wiley & Sons, NY.
3. Allcroft, D.J. and Glasbey, C. A. (2005). STARMA Process Applied to Solar Radiation, *Biomathematics and Statistics Scotland*, 1-24.
4. Billard, L.,and Dai, Yuqing. (1998), Maximum Likelihood Estimation in Space Time Bilinear Models, *Journal of Time Series Analysis*, 25-44.
5. Billard, L.,and Dai, Yuqing.(2000), Modeling Spatial-Temporal Epidemics, *Department of Statistic*, 1-34.
6. Kamarianakis, I, and Prastacos, P.(2002), Space-Time Modeling of Traffic Flow, Methods of Spatial Analysis-Spatial Time Series Analysis. ERSA Proceedings Radiation," *Biomathematics and Statistics Scotland*, 1-24.
7. Pfeifer, P .E. and Deutsch, S. J. (1980), A Three-Stage Iterative Procedure for Space-Time Modeling, *Technometrics*, 22, 25-47.

[2006년 4월 접수, 2006년 5월 채택]