

# 공간시계열모형에 대한 베イズ 추론

이성덕<sup>1,a</sup>, 김인규<sup>b</sup>, 김덕기<sup>a</sup>, 정애란<sup>a</sup>

<sup>a</sup>충북대학교 정보통계학과, <sup>b</sup>우송정보대학교 컴퓨터정보통신계열

## 요약

공간시계열모형은 공간의 위치와 시간의 흐름에 따라 동시에 관측되는 분야인 기상, 지질, 천문, 생태, 역학 등에서 넓이 사용되고 있는 매우 복잡한 모형이다. 본 논문은 공간시계열모형에 대한 모수 추정에 있어서 기존의 최대우도추정 방법이 가지는 컴퓨팅의 문제를 해결하기 위하여 모수에 대한 사전정보와 자료의 정보를 모두 이용하는 깁스샘플링과 같은 MCMC 방법으로 모수를 추정하고, 실제 적용사례분석으로 여러 가지 측도를 구해서 추정된 모수에 대한 수렴진단을 수행하였다.

주요용어: 공간시계열자료, 깁스샘플링, Mumps 자료, STARMA, STBL.

## 1. 서론

공간시계열자료(Space Time Series Data)란 공간 위치에서 수집된 시계열의 자료이거나 시간 변화에 대한 공간 자료의 집합이라 볼 수 있다. 이러한 공간시계열자료는 상당히 큰 공간적 변동성(Spatial Variability)과 비정상성(non-stationarity)을 가지고 있으며, 이러한 성질은 기존의 통계적 모형이나 방법들의 효율성에 제약을 가하게 되는데, 베이지안 방법은 이러한 성질에도 불구하고 상당히 안정된 방법으로서 최근 들어 이에 관한 많은 연구들이 활발히 이루어지고 있다. 공간시계열 자료에 대한 연구로 Sansó와 Guenni (1999)이 공간시계열모형을 통한 베이지안(Bayesian) 방법을 이용해 강수량 자료를 분석 및 통계적 추론을 실시하였고, Pfeifer와 Deutsch (1980)는 ARMA(자기회귀이동평균) 모형에 공간정보를 포함시킨 STARMA(공간시계열자기회귀이동평균) 모형을 제시하였다. 본 논문에서는 첫 번째로 여러 공간자료의 유형 중 격자망(Grid)에서의 공간적 의존성을 지닌 공간구조에 대해 살펴보고, 공간 의존성을 가중치행렬로 나타내는 방법을 제시하고, 이러한 공간적 구조를 가진 시계열 모형인 STARMA 모형에 대해 고려했다. 두 번째로 Tierney와 Kadane (1986)에 의해 제안된 MCMC(Markov Chain Monte Carlo) 방법 등을 이용하여 앞서 제시한 공간시계열모형에서 모수들에 대한 완전조건부 분포를 유도하고, MCMC 방법 중 Gelfand와 Smith (1990)에 의해 제안된 깁스샘플링(Gibbs-Sampling) 방법을 적용한다. 세 번째는 실증분석으로 미국의 12개 주의 Mumps(유행성 이하선염 또는 항아루손님) 자료를 이용해서 공간시계열모형의 모수들에 대한 사전분포를 정의하고 깁스샘플링을 이용해 사후분포(Posterior distribution)로부터 표본들을 추출하여 모수를 추정하고 수렴성을 진단한다.

## 2. 가중행렬과 공간시계열모형

### 2.1. 가중행렬(Weight Matrix)

그림 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 집

이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임. (KRF-2007-313-C00115)

<sup>1</sup>교신저자: (361-763) 충북 청주시 개신동 12번지, 충북대학교 정보통계학과, 교수. E-mail: sdlee@chungbuk.ac.kr

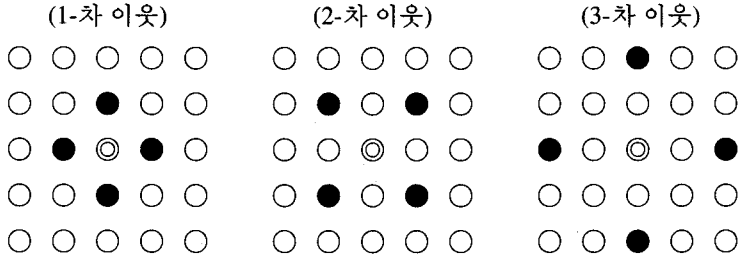


그림 1: 격자자료에서 공간차수의 이웃하고 있는 구조

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

그림 2: 1-차 이웃한 가중치행렬

합(●)이며, 2차 이웃의 경우 1차 이웃보다 먼 동일한 유클리디안 거리를 갖는 집합으로 나타낸다. 3~4차 이웃의 경우도 마찬가지로 나타낼 수 있다.

위  $5 \times 5(n = 25)$  격자자료의 경우 각 위치는 왼쪽 위에서부터 오른쪽 아랫방향으로 가중치행렬(Weighting Matrix)을 그림 2와 같이 결정할 수 있다.

2.2. 공간시계열 자기회귀 이동평균모형(STARMA)

공간시계열 자기회귀 이동평균모형(Space-Time Autoregressive Moving Average Model)은 STARMA( $p_\lambda, q_\eta$ )로 표현할 수 있으며, 모형은 다음과 같다.

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

- $p, q$ : 최대 자기회귀차수, 최대 이동평균차수,
- $\lambda_i, \eta_j$ :  $i$ 번째 자기회귀항의 차수와  $j$ 번째 이동평균항의 차수,
- $\phi_m^i, \theta_n^j$ : 공간차수가  $m, n$ , 시간차수  $i, j$ 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).

### 3. 깁스 표본기법(Gibbs Sampling Method)

베이지안 분석에서 모형이 복잡해지고 적분변수가 많아지는 경우 다차원 적분계산이 필요하게 된다. 깁스 표본기법은 이러한 다차원 적분을 효과적으로 쉽게 해결할 수 있는 기법이며 깁스 표본기법의 적용을 위해서는 각 적분변수의 완전조건부확률분포(full conditional distribution)의 유도가 용이하고 그 분포의 난수 생성이 쉬운 형태여야 한다는 조건이 필요하다.

깁스 표본기법을 3차원 적분의 경우로서 예를 들어 설명하면 다음과 같다. 모수 벡터  $\theta = (\theta_1, \theta_2, \theta_3)$ 에 대한 사후 분포를  $p(\theta_1, \theta_2, \theta_3 | z)$ 라 하고, 초기 값이  $(\theta_1^0, \theta_2^0, \theta_3^0)$ 라고 하자. 깁스 샘플링을 적용하는 방법은 다음 루프를 반복시키는 것이다.

for  $i=1$  to  $n$  {  
 $p(\theta_1 | \theta_2^{i-1}, \theta_3^{i-1}, z)$ 에서  $\theta_1^i$ 를 표본 추출  
 $p(\theta_2 | \theta_1^i, \theta_3^{i-1}, z)$ 에서  $\theta_2^i$ 를 표본 추출  
 $p(\theta_3 | \theta_1^i, \theta_2^i, z)$ 에서  $\theta_3^i$ 를 표본 추출. }

샘플링이 어느 정도 궤도에 오를 때까지 얻은 초기 표본들은 다 버리게 된다. 이런 초기 단계(Burn-in period)가 끝난 후 반복수  $n$ 을 크게 할 때 얻어지는 세 원소  $(\theta_1^i, \theta_2^i, \theta_3^i)$ 는 정규화된 사후 분포  $p(\theta_1, \theta_2, \theta_3 | z) / \int p(\theta_1, \theta_2, \theta_3 | z) d\theta_1 d\theta_2 d\theta_3$ 로부터의 추출로 간주한다. 그리고  $\theta_j^i$  같은 개개의 원소는 주변 사후 분포  $p(\theta_j | z) / \int p(\theta_j | z) d\theta_j$ 로부터의 표본으로 간주 된다.

### 4. STAR 모형의 베이지안 추정방법

STAR(1<sub>1</sub>) 모형은 다음과 같이 정의한다.

$$z(t) = \phi_1 z(t-1) + \phi_2 W^{(1)} z(t-1) + e(t), \quad t = 1, \dots, T,$$

여기서,  $W^{(1)}$ 은 1차 가중행렬을 나타내며  $e(t)$ 는 평균 0, 분산  $\sigma^2$ 을 따르는 정규분포를 나타낸다. 사후 분포를 유도하기 위하여 먼저 사전분포를 정의하여야 하며, STAR 모형의 모수에 대한 사전 분포는 다음과 같이 정의 하였다. 모수  $\Phi = (\phi_1, \phi_2, \sigma^2)$ 이다.

여기서  $\Phi$ 은 모든 응답단위들에 대한 것이다. 흔히 응답률의 역수에 해당하는  $\phi_i^{-1}$ 을 무응답 보정 상수라고 한다. 응답률을 기초로 한 추가적인 가중치는 추출확률에 따른 가중치와 유사한 개념으로 이해할 수 있다. 만약 각 단위들에 대한 응답률을 알 수 있으면 단위 무응답이 발생해도 비편향 추정이 가능하지만 실제로는 응답률  $\phi_i$ 를 알 수 없기 때문에 이를 추정해야 한다.

$$\begin{aligned} \phi_1 &\sim N(0, \tau_1^2), \\ \phi_2 &\sim N(0, \tau_2^2), \\ \frac{1}{\sigma^2} &\sim \Gamma(\alpha, \beta), \end{aligned}$$

여기서, 비정보 사전 분포를 위하여  $\tau_1^2 = \tau_2^2 = 1000$ ,  $\alpha = \beta = 0.001$ 로 정의하였다.  $z(t)$ 에 대한 우도함수는 다음과 같이 표현된다.

$$\pi(z(t, s) | \Phi) = \left( \frac{1}{2\pi\sigma^2} \right)^{NT} \exp \left[ -\frac{1}{2\sigma^2} \left\{ \sum \sum (z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)} z(t-1, s)) \right\}^2 \right].$$

따라서 우도함수  $\pi(z(t, s)|\Phi)$ 와 사전분포의 결합으로 STAR 모형의 사후분포를 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
\pi(\Phi|z(t, s)) &= \left(\frac{1}{2\pi\sigma^2}\right)^{NT} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum\sum(z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)} z(t-1, s))\right\}^2\right] \\
&\quad \times \frac{1}{\sqrt{2\pi\tau_1^2}} \exp\left(-\frac{\phi_1^2}{2\tau_1^2}\right) \times \frac{1}{\sqrt{2\pi\tau_2^2}} \exp\left(-\frac{\phi_2^2}{2\tau_2^2}\right) \times \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{\sigma^2}\right)^{\alpha-1} \exp\left(-\frac{1}{\beta\sigma^2}\right) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{NT+\alpha-1} \exp\left[-\frac{1}{2\sigma^2}\sum\sum\left\{z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)} z(t-1, s)\right\}^2 - \frac{\phi_1^2}{2\tau_1^2} - \frac{\phi_2^2}{2\tau_2^2} - \frac{1}{\beta\sigma^2}\right] \\
&= \left(\frac{1}{\sigma^2}\right)^{NT+\alpha-1} \exp\left[-\frac{1}{2\sigma^2\tau_1^2\tau_2^2\beta}\left\{\beta\tau_1^2\tau_2^2\sum\sum(z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)} z(t-1, s))^2\right.\right. \\
&\quad \left.\left.+ \beta\sigma^2\tau_2^2\phi_1^2 + \beta\sigma^2\tau_1^2\phi_2^2 + 2\tau_1^2\tau_2^2\right\}\right] \\
&= \left(\frac{1}{\sigma^2}\right)^{NT+\alpha-1} \exp\left[-\frac{1}{2\sigma^2\tau_1^2\tau_2^2\beta}\left\{\beta\tau_1^2\tau_2^2\left(\sum\sum z^2(t, s) + \phi_1^2\sum\sum z^2(t-1, s)\right.\right.\right. \\
&\quad \left.\left.+ \phi_2^2 W^{(1)2} z^2(t-1, s) - 2\phi_1\sum\sum z(t, s)z(t-1, s) - 2\phi_2\sum\sum z(t, s)W^{(1)} z(t-1, s)\right.\right. \\
&\quad \left.\left.+ 2\phi_1\phi_2\sum\sum z(t-1, s)W^{(1)} z(t-1, s) + \beta\sigma^2\tau_2^2\phi_1^2 + \beta\sigma^2\tau_1^2\phi_2^2 + 2\tau_1^2\tau_2^2\right\}\right] \\
&= \left(\frac{1}{\sigma^2}\right)^{NT+\alpha-1} \exp\left[-\frac{1}{2\sigma^2\tau_1^2\tau_2^2\beta}\left\{\phi_1^2\left(\beta\tau_1^2\tau_2^2\sum\sum z^2(t-1, s) + \beta\sigma^2\tau_2^2\right)\right.\right. \\
&\quad \left.\left.- 2\phi_1\beta\tau_1^2\tau_2^2\sum\sum(z(t, s)z(t-1, s) - \phi_2 z(t-1, s)W^{(1)} z(t-1, s))\right\}\right]
\end{aligned}$$

위의 식에서 표현된 사후 분포로부터 각 모수에 대한 조건부 분포를 유도할 수 있으며  $\phi_1$ 에 대한 조건부 분포는 정규분포를 하고 평균이 다음과 같이 표현된다.

$$\begin{aligned}
&\frac{\sum\sum(z(t, s)z(t-1, s) - \phi_2 z(t-1, s)W^{(1)} z(t-1, s))\tau_1^2\tau_2^2\beta}{\beta\tau_1^2\tau_2^2\sum\sum z^2(t-1, s) + \beta\sigma^2\tau_2^2} \\
&= \frac{\sum\sum(z(t, s)z(t-1, s) - \phi_2 z(t-1, s)W^{(1)} z(t-1, s))\tau_1^2}{\tau_1^2\sum\sum z^2(t-1, s) + \sigma^2}
\end{aligned}$$

분산은 다음과 같다.

$$\frac{\sigma^2\tau_1^2}{\tau_1^2\sum\sum z^2(t-1, s) + \sigma^2}$$

비슷한 방법으로  $\phi_2$ 에 대한 조건부 분포는 정규분포를 하고 평균은 다음과 같이 표현된다.

$$\frac{\sum\sum(z(t, s)W^{(1)} z(t-1, s) - \phi_1 z(t-1, s)W^{(1)} z(t-1, s))\tau_1^2\tau_2^2\beta}{\beta\tau_1^2\tau_2^2\sum\sum W^{(1)2} z^2(t-1, s) + \beta\sigma^2\tau_1^2}$$

$$= \frac{\sum \sum (z(t, s)W^{(1)}z(t-1, s) - \phi_1 z(t-1, s)W^{(1)}z(t-1, s)) \tau_2^2}{\tau_2^2 \sum \sum W^{(1)}z^2(t-1, s) + \sigma^2}$$

분산은 다음과 같다.

$$\frac{\sigma^2 \tau_2^2}{\tau_2^2 \sum \sum W^{(1)}z^2(t-1, s) + \sigma^2}$$

$\pi(1/\sigma^2 | z(t, s))$ 의 분포는 감마 분포를 하고 다음과 같이 나타내어진다.

$$\begin{aligned} \pi\left(\frac{1}{\sigma^2} | \cdot\right) &= \left(\frac{1}{\sigma^2}\right)^{NT+\alpha-1} \exp\left[-\frac{1}{2\sigma^2} \sum \sum (z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)}z(t-1, s))^2 - \frac{\phi_1^2}{2\tau_1^2} - \frac{\phi_2^2}{2\tau_2^2} - \frac{1}{\beta\sigma^2}\right] \\ &\propto \left(\frac{1}{\sigma^2}\right)^{NT+\alpha-1} \exp\left[-\frac{1}{2\sigma^2\beta} \left\{\beta \sum \sum (z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)}z(t-1, s))^2 + 2\right\}\right]. \end{aligned}$$

그래서  $1/\sigma^2$ 의 완전 조건부 분포는 다음과 같다.

$$\Gamma\left[NT + \alpha, \frac{1}{2} \sum \sum \left\{z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)}z(t-1, s)\right\}^2 + \beta\right]$$

즉,

$$\frac{\sum \sum \left\{z(t, s) - \phi_1 z(t-1, s) - \phi_2 W^{(1)}z(t-1, s)\right\}^2 + 2\beta}{\sigma^2} \sim \chi_{(NT+\alpha)}^2.$$

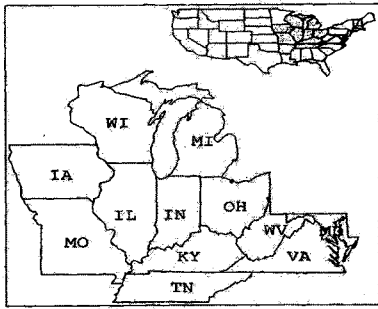
## 5. 실제 응용

### 5.1. STAR 모형의 베이지안 방법

STAR 모형을 베이지안 방법으로 완전 조건부 분포를 유도하여 MCMC 방법을 이용해서 모수들의 결합 분포로부터 모수를 추출한다. MCMC 방법을 이용한 알고리즘은 Geman과 Geman (1984), Tanner와 Wong (1987), Gelfand와 Smith (1990)에 의해서 소개되었다. 실제분석을 위하여 베이지안 분석을 위한 상용 소프트웨어인 WinBUGS(Bayesian Inference Using Gibbs Sampling Based on Windows)를 사용하여 모수를 추정하였으며, 추정된 모수에 대한 수렴진단(Convergence diagnostic)을 수행하였다. 이 프로그램은 사후분포로부터 모수를 추출하는 것뿐만 아니라 각 모수(Parameter)에 대한 평균, 분산 등의 추정치를 구해주고, 여러 가지 그래프들을 제공해 준다(<http://www.mrc-bsu.cam.ac.uk/bugs>에서 프로그램을 얻을 수 있다). 사후분포로부터 모수들을 생성하기 위하여 깃스 샘플링을 하게 된다. 자료는 12개 주에 대한 1968-1988년도 사이에 발생한 월 별 Mumps 자료를 사용하였다.

그림 3a은 본 논문의 관심영역인 12개 주를 나타내며, 그림 3b는 1차-이웃한 가중행렬을 나타낸 것이고, 표 1은 각 주에서 첫 번째 이웃하고 있는 집합을 나타낸다.

Mumps 자료를 사용한 STAR 모형에서 세 개의 초기치를 이용하여 각 초기치에서 사후분포 표본을 10,000개씩 추출하고 그 표본들 중에서 초기 생성자료 3,000개(Burn-in period)를 제거하여 사후분포로부터 전체 21,000개의 사후분포 표본들을 통하여 모수를 추정 하였다. 추정된 모수 값의 수렴상태를 보기 위하여 본 논문에서는 앞서 언급했듯이 세 가지의 다양한 초기치를 사용하였으며, 따라서 세 개의 초기치를 통한 사후분포 표본을 통하여 모수 추정 값을 얻을 수 있었으며, 세 초기치를 통하여 얻어진 Gelman-Rubin Diagnostic은 그림 5는 STAR 모형의 세 모수의 추정 값은 수렴하는데 특별한 문제가 없다는 것을 보여준다. 모수의 추정 값과 추출된 표본을 이용하여 각 모수들의 평균(Mean), 표준편차(Standard deviation), 중위수(Median), 분위수(Percentile) 등의 추정 값들은 표 2와 같다.



(a) 12개주의 Mumps 관심 대상지역

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

(b) 1차 이웃한 가중치 행렬

그림 3: 12개주의 Mumps 관심 대상지역과 1차 가중치 행렬

표 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 2: STAR 모형의 MCMC를 이용한 사후분포에 대한 모의실험 결과

Parameter	Mean	SD	MC error	2.5%	Median	97.5%	Start	Sample
$\phi_1$	0.6887	0.01433	9.828E-5	0.6603	0.6887	0.7167	3001	21000
$\phi_2$	0.1229	0.02206	1.457E-4	0.0798	0.1228	0.1667	3001	21000
$\sigma^2$	0.4795	0.01315	8.986E-5	0.4541	0.4794	0.5055	3001	21000

### 5.2. 모수에 대한 수렴진단

다음의 그림 4와 5은 깃스 샘플링 알고리즘을 이용하여 생성된 21,000개의 모수들을 가지고 작성한 것이다. 그림 4는 WinBUGS 프로그램을 사용해서 추정된 모수들이 얼마나 잘 혼합(Well-mixing)되어 추출되었는지를 보여준다. 만약 그림에서 반복(Iteration)하는 동안 값이 변하지 않고 일정하게 유지되는 부분이 자주 존재한다면 이 표본은 잘 혼합되어 추출되었다고 볼 수 없다. 따라서 세 가지 모수에 대한 그림을 살펴보면 어느 반복지점에서 값이 변하지 않고 일정하게 유지되는 부분을 찾아볼 수 없으므로 생성된 모수들이 잘 혼합되어 추출되었다고 할 수 있다. 그림 5는 추정된 모수들의 커널 밀도함수를 그래프로 표현한 것이다.

모수에 대한 수렴진단 방법으로는 Gelman과 Rubin (1992), Raftery와 Lewis (1992), Ritter와 Tanner (1992) 등의 방법들이 있는데, 최근에 가장 관심을 모으고 있는 Gelman과 Rubin (1992) 방법을 이용하

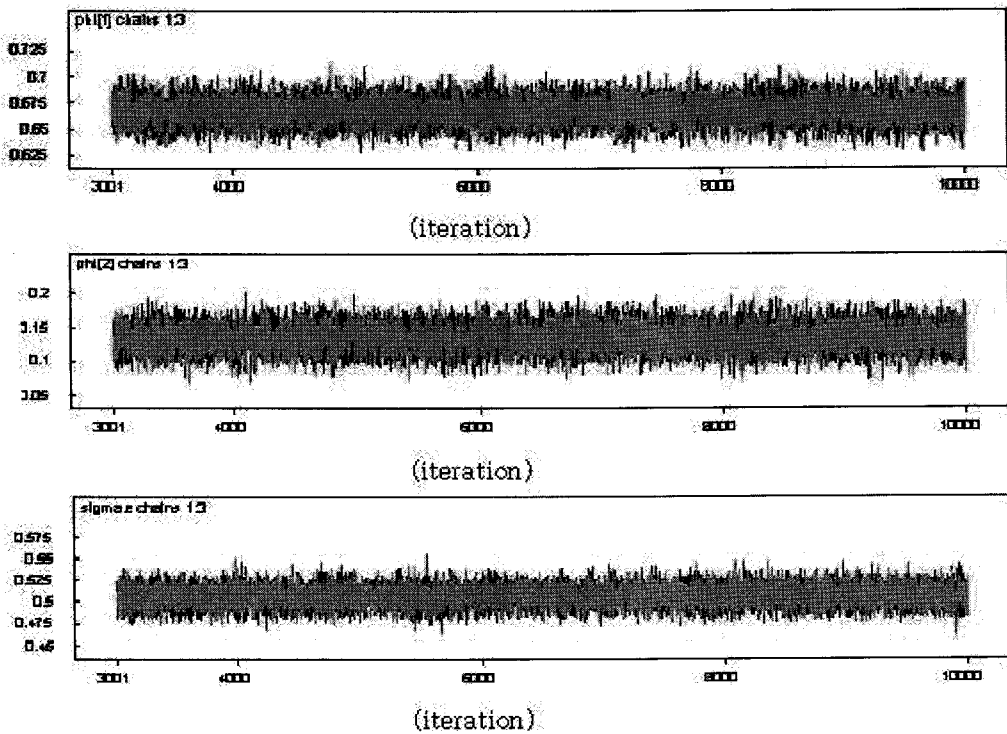


그림 4: STAR 모형의 MCMC를 이용한 사후분포에 대한 모의실험 결과

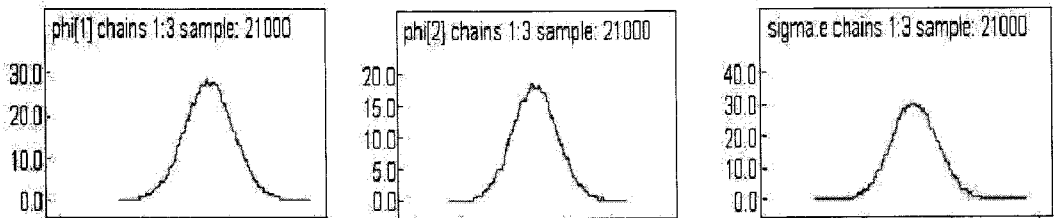


그림 5: STAR 모형 안에 포함된 세 모수의 Smoothed Kernel Density Plot

여 모수에 대한 수렴진단을 한다. 모수의 수렴진단의 방법은 깃스 샘플링 알고리즘에 의해 생성된 각 모수들이 수렴하는가의 여부를 통해 수행될 수 있다.

Gelman과 Rubin (1992)의 방법은 2개 이상의 병렬연쇄(Parallel chains)를 가지고 수렴진단을 하는 것으로써 각 연쇄(chain)는 실제 사후분포에 대하여 과대산포(Over-dispersion)를 갖는 각기 다른 초기 값을 이용하여 수행한다. 만약 연쇄가 무한대로 진행됨에 따라 이 측도의 값이 1에 가까워지면 두 연쇄에 의해 생성된 모수들이 동일한 분포로 수렴한다는 것을 나타낸다.

본 연구에서 STAR 모형에 3개의 연쇄를 적용하고, 각각 21,000번의 반복을 시행하여 Gelman과 Rubin (1992) 방법을 적용하면 그림 6과 같은 결과가 나온다. 이 결과들을 보면 각 모수들의 진단 값들이 1에 가깝게 나타나고 있으므로 생성된 모수들이 잘 수렴하고 있다는 것을 알 수 있다.

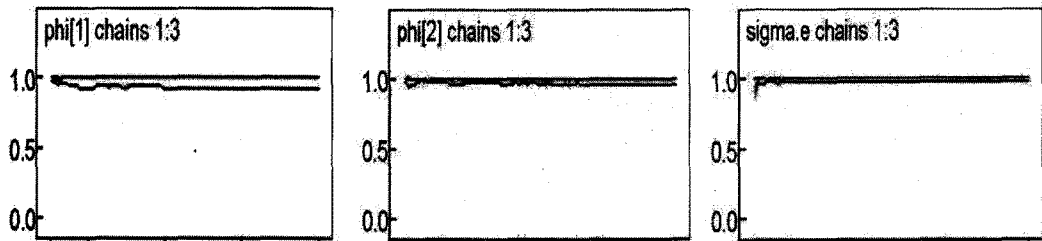


그림 6: STAR 모형 안에 포함된 세 모수에 대한 Gelman과 Rubin의 수렴진단결과

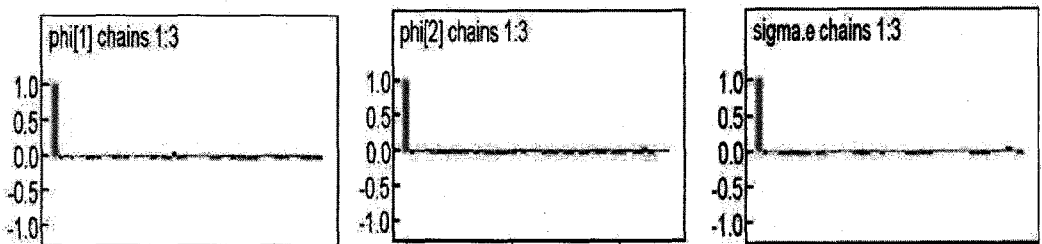


그림 7: STAR 모형 안에 포함된 세 모수에 대한 자기상관 그래프

그림 7은 STAR 모형 안에 포함된 세 모수에 대한 자기상관(autocorrelation) 그래프를 나타낸다. 각 모수들에 대해서 자기상관이 없는 걸로 나타나기 때문에 모수들이 잘 추출 되었다고 볼 수 있다.

## 6. 결론

최근 Metropolis-Hastings 알고리즘과 Gibbs Sampling과 같은 Markov Chain Monte Carlo(MCMC)의 확률적 적분 기법의 발달과 더불어 복잡한 유형의 시계열 모형을 분석하는데 베이지안 방법이 많이 사용되고 있다. 본 논문에서는 베이지안 방법을 공간 시계열 모형인 STARMA 모형에 적용하여, 모수들에 대한 완전 조건부 분포(Full conditional distribution)를 유도하였으며, 실제 분석을 위하여 베이지안 분석을 위한 상용 소프트웨어인 WinBUGS(Bayesian Inference Using Gibbs Sampling based on Windows)를 사용하여 모수를 추정하였다. 그리고 추정된 모수에 대한 수렴진단(convergence diagnostics)을 WinBUGS에 내장 되어있는 여러 가지 측도를 통하여 수행하였다.

위의 결과들로부터 대용량의 자료이거나 복잡한 유형의 모형을 분석하는데 있어서 기존의 최대우도추정법이 가지고 있는 컴퓨팅의 문제를 해결하기 위하여 모수(Parameter)에 대한 사전정보(Prior information)와 자료의 정보(Data information)를 모두 이용하는 베이지안 기법이 적합하게 이용될 수 있다는 것을 보였고 STARMA 모형과 같은 선형모형뿐만 아니라, 비선형 공간시계열 모형인 STBL(Space Time Bilinear) 모형에서도 이와 같은 베이지안 방법으로 모수를 추정하고 예측하는 것이 가능할 것이라고 사료된다. 위의 결과들로부터 대용량의 자료이거나 복잡한 유형의 모형을 분석하는데 있어서 기존의 최대우도추정법이 가지고 있는 컴퓨팅의 문제를 해결하기 위하여 모수(Parameter)에 대한 사전정보(Prior information)와 자료의 정보(Data information)를 모두 이용하는 베이지안 기법이 적합하게 이용될 수 있다는 것을 보였고 STARMA 모형과 같은 선형모형뿐만 아니라, 비선형 공간시계열 모형인 STBL(Space Time Bilinear) 모형에서도 이와 같은 베이지안 방법으로 모수를 추정하고 예측하는 것이 가능할 것이라고 사료된다.



## 참고 문헌

- Gelfand, A. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–472
- Pfeifer, P. E. and Deutsch, S. J. (1980). Identification and interpretation of first order space-time ARMA models, *Technometrics*, **22**, 397–408.
- Raftery, A. L. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo, *Statistical Science*, **7**, 493–497.
- Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddly Gibbs sampler, *Journal of the American Statistical Association*, **85**, 398–409.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Sansó, B. and Guenni, L. (1999). Venezuelan rainfall data analysed by using a Bayesian space-time model, *Applied Statistics*, **48**, 345–362.
- Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528–550.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**, 82–86.

2008년 11월 접수; 2008년 12월 채택

# Bayes Inference for the Spatial Time Series Model

Sungduck Lee<sup>1,a</sup>, Inkyu kim<sup>b</sup>, Dukki Kim<sup>a</sup>, Aeran Chung<sup>a</sup>

<sup>a</sup>Dept. of Information & Statistics, Chungbuk Univ.

<sup>b</sup>Div. of Computer Information & Communication of Information, Woosong Info. Col.

---

## Abstract

Spatial time series data can be viewed either as a set of time series collected simultaneously at a number of spatial locations. In this paper, We estimate the parameters of spatial time autoregressive moving average (STARMA) process by method of Gibbs sampling. Finally, We apply this method to a set of U.S. Mumps data over a 12 states region.

Keywords: Space time series data, Gibbs sampling, Mumps data, STARMA, STBL.

---

---

This work was supported by Government (Ministry of Education & Human resources) (KRF-2007-313-C00115.)

<sup>1</sup> Corresponding author: Professor, Department of Information & Statistics, Chungbuk National University, Cheongju, Chungbuk 361-763. E-mail: sdlee@cbnu.ac.kr