

공간시계열 모형의 칼만필터 추정과 예측

이성덕^{1,a}, 한은희^a, 김덕기^a

^a충북대학교 정보통계학과

요약

공간적, 시간적으로 퍼져나가는 전염성이 강한 질병인 수두자료를 이용하여 공간 시계열 자료를 분석하는데 있어 일반적으로 알려진 ARIMA 모형에 적합하여 분석을 행하면 공간적인 정보를 반영하지 못하기 때문에 기존에 시간만을 고려한 시계열 분석방법에 공간통계의 공간적 정보를 반영한 공간시계열 모형을 고려한다. 공간시계열 모형에서 공간의 위치 및 영향은 시계열 모형에 공간적 정보로써 가중치행렬을 더함으로써 처리 가능해진다. 가중치행렬은 지리적으로 인접한 지역일수록 공간의존도가 높다는 것을 반영한 것이며 공간시계열 모형의 연구에서 가중치행렬은 인접한 지역들은 동일한 영향을 줄 것이라 가정하였다. 따라서 본 논문에서는 공간시계열 모형인 STARMA 모형과 STBL 모형에 대한 식별방법, 통계적 추론 및 예측력 비교에 대해 연구하였고 특히, 모수추정의 알고리즘 비교와 공간시계열 모형의 예측력 비교를 통해 Kalman-Filter 방법의 우수성을 보이고자 한다.

주요용어: STARMA, STBL, 가중치행렬, STACF, STPACF, Newton-Raphson, Kalman-Filter, SSF.

1. 서론

공간시계열 자료는 시계열 자료(Time series data)와 공간 위치의 변화에 따라 관찰되는 공간자료(Space data)의 결합이다. 다시 말해 공간시계열 자료는 특정 위치에서 관측되어진 자료가 과거의 시간과 주변 공간의 영향을 동시에 받아 나타나는 자료이다. 공간시계열 자료는 기상학(Metrology), 지질학(Geology), 천문학(Astronomy), 생태학(Ecology) 및 역학(疫學, Epidemiology)과 같은 전염병 확산 모형에까지도 광범위하게 나타나고 있다. 공간시계열 자료에 기존의 전통적인 ARIMA 모형을 적용할 경우 인근지역의 공간적 정보가 배제되어 각 지역별로 여러 모형과 추정해야할 많은 모수가 고려되어 모수절약의 원칙에도 맞지 않으며 추정과 예측의 오차가 커지고 공간적인 의존성을 반영하지 못한 단순한 자료분석이 된다. 이러한 단점을 보완하기 위해 시계열 모형에 공간통계를 결합한 공간시계열 모형에 대한 추정 및 예측방법에 대한 연구가 활발히 이루어지고 있다 (이성덕 등, 2007). 하지만 공간시계열모형은 모형이 복잡하여 모형식별과 모수추정 및 예측의 과정이 쉽지않다. 공간시계열 모형은 시간의 변화와 더불어 공간의 위치도 영향을 주기 때문에 시계열 모형에 가중행렬을 공간정보로 포함시킨다.

본 논문에서는 Pfeifer와 Deutsch (1981)가 제시한 공간시계열 모형으로 기존의 자기회귀이동평균(ARMA) 모형에 공간정보를 포함시킨 공간시계열자기회귀이동평균(STARMA) 모형과 Billard와 Dai (1998)가 제시한 비선형 시계열 모형인 중선형(BL) 모형에 공간정보로써 가중치행렬을 결합한 공간시계열중선형(STBL) 모형에 대한 통계적 추론에 있어서 모형의 복잡성과 비선형성때문에 수치해석적 방법을 고려해야 하는데 Billard와 Dai (2003)가 기존의 연구에서 제한한 알고리즘 중 추정의 정도와

¹ 이 논문은 2009학년도 충북대학교 학술연구지원 사업의 연구비 지원에 의하여 연구되었음.

^a 교신저자: (361-763) 충북 청주시 흥덕구 충북대학교 정보통계학과, 자연과학연구원, 교수.

Email: sdlee@cbnu.ac.kr

예측력이 우수한 Kalman-Filter 방법에 초점을 맞춰 공간시계열 모형에 대한 통계적추론방법을 연구하고 실증분석으로 한국의 수두(Chicken Pox)자료를 이용해 Newton-Raphson과 Kalman-Filter 방법의 추정 정도와 공간시계열 모형에 따른 예측력을 비교분석한다.

본 논문의 구성은 2장에서 Kalman-Filter 이론에 대해 소개하고 3장에서는 공간자료의 정의와 유형과 공간시계열 모형 및 가중행렬에 대해 소개하였다. 4장에서는 공간시계열 모형에 대한 통계적 추론에 대해 소개하고 5장에서는 실증분석을 위해 전염성이 강한 질병인 수두(Chicken Pox)자료를 이용해 모형식별, 모수추정 및 예측력을 비교하였다.

2. Kalman Filter 추정방법

상태공간 모형의 추정방법에는 Filtering, Smoothing, Prediction의 세 방법이 있다. Filtering은 모든 과거 자료들에 기초하여 현재 시간의 상태를 추정하는 것이며 Smoothing은 모든 과거 자료들과 현재 자료들에 기초하여 과거 상태의 추정이며 Prediction은 현재 및 과거 자료들에 기초하여 미래 상태의 추정을 하는 것이다. Filtering은 t 시점의 상태를 추정하기 위해 $t-1$ 시점까지의 자료를 필요로 하고 t 시점의 자료를 획득하면 이를 토대로 $t+1$ 시점의 상태를 추정하는 연속적인 성질을 갖고 있다. Filtering 기법의 일종인 Kalman-Filter 기법 또한 각 추정시점마다 최적예측을 추구하는 연속적 최적상태의 성질을 가지고 있으며 시간 영역에서 모형을 형성하므로 수학적으로 다루기 쉬운 많은 장점이 있어서 널리 활용되고 있다. Kalman-Filter 모형은 측정되지 않은 상태변수를 추정해 내고 추정된 잡음(noise)의 영향을 최소화하는 것이며 기본 전제는 관측값에도 오차가 있다는 것이다. 오차가 포함된 관측값 정보를 이용하여 추정값을 갱신하고 최적추정(optimal estimation)을 한다. Kalman-Filter에서는 추정값 행렬과 관측값 행렬간의 선형결합에 의해 최적 추정행렬을 구할 수 있고, Kalman-Filter 모형은 크게 세 부분으로 구분할 수 있는데 시스템방정식, 관측방정식, Kalman-Filtering의 세부분이다.

$$\text{시스템 방정식: } X(t) = \phi(t-1)X(t-1) + w(t),$$

$$\text{관측 방정식: } Z(t) = H(t)X(t) + v(t),$$

$$\text{상태추정식: } \hat{x}(t|t-1) = \phi(t-1)\hat{x}(t-1|t-1),$$

$$\text{오차공분산: } P(t|t-1) = \phi(t-1)P(t-1|t-1)\phi(t-1)^T + Q(t-1),$$

$$\text{갱신방정식: } \hat{x}(t|t) = [I - K(t)H(t)]\hat{x}(t|t-1) + K(t)z(t|t-1),$$

$$P(t|t) = [I - K(t)H(t)]P(t|t-1),$$

$$\text{Kalman Gain: } K(t) = P(t|t-1)H^T(t)[H(t)P(t|t-1)H^T(t) + R(t)]^{-1}.$$

2.1. 공간시계열 모형: STARMA(Space Time Autoregressive Moving Average) 모형

공간시계열자기회귀이동평균(STARMA) 모형을 STARMA($p_{\lambda_i}, q_{\eta_j}$)라 표기하며, 모형은 다음과 같다.

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

여기서, $\mathbf{z}(t) = [z_1(t), \dots, z_n(t)]^T$ 인 $n \times 1$ 확률벡터과정(Random vector process)이다.

p 는 자기회귀차수, q 는 이동평균차수

λ_i 는 i 번째 자기회귀항의 공간차수, η_j 는 j 번째 이동평균항의 공간차수

ϕ_m^i 는 시간차수 i , 공간차수 m 인 자기회귀모수

θ_n^j 는 시간차수 j , 공간차수 n 인 이동평균모수

$\mathbf{W}^{(m)}$ 는 공간차수가 m 에 대한 $n \times n$ 가중치행렬

$e(t) = \{e_1(t), e_2(t), \dots, e_n(t)\}^T$ 는 t 시점의 확률잡음벡터

$$e(t) \sim \text{Normal}(0, \sigma^2),$$

$$E[e(t)e(t+j)^T] = \begin{cases} \mathbf{G}, & j = 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$E[\mathbf{z}(t)e(t+j)^T] = 0, \quad \text{if } j > 0.$$

2.2. 공간시계열 모형: STBL(Space Time Bilinear) 모형

공간시계열중선형(STBL) 모형을 $\text{STBL}(p_{\lambda_i}, q_{\eta_j}, r_{\xi_i}, s_{\mu_j})$ 라 표기하며, 모형은 다음과 같다.

$$\begin{aligned} z(t) = & \sum_{i=1}^p \sum_{m=0}^{\lambda_i} \phi_m^i [\mathbf{W}^{(m)} z(t-i)] + \sum_{j=1}^q \sum_{n=0}^{\eta_j} \theta_n^j [\mathbf{W}^{(n)} e(t-j)] \\ & + \sum_{i=1}^r \sum_{j=1}^s \sum_{m=0}^{\xi_i} \sum_{n=0}^{\mu_j} \beta_{mn}^{ij} [\mathbf{W}^{(m)} z(t-i)] \# [\mathbf{W}^{(n)} e(t-j)] + e(t) \end{aligned} \quad (2.2)$$

여기서,

r 은 중선형 항의 자기회귀차수,

s 는 중선형 항의 이동평균차수,

ξ_i 는 i 번째 중선형 항의 자기회귀차수,

μ_j 는 j 번째 중선형 항의 이동평균차수,

β_{mn}^{ij} 는 중선형 항의 i, j 번째 자기회귀와 이동평균 시간차수에 대한 m, n 번째 공간차수,

$\#$ 은 행렬의 원소간의 곱을 나타낸다.

예로 $A = (a_{ij})$ 와 $B = (b_{ij})$ 이면 $c_{ij} = a_{ij}b_{ij}$ 나머지 기호는 앞의 STARMA 모형의 정의와 같다.

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

위 5×5 ($n = 25$) 격자자료의 경우 각 위치는 왼쪽 위에서부터 오른쪽 아랫방향으로 가중행렬을 그림 2와 같이 결정할 수 있다.

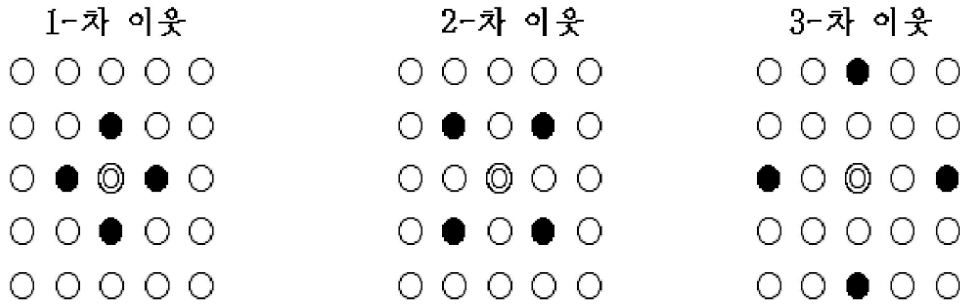


그림 1: 공간차수의 격자자료에서 이웃하고 있는 구조

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

그림 2: 1차 이웃한 가중치행렬

3. Kalman Filter 방법에 의한 모수추정

3.1. 상태공간 모형(Space state model)

상태 공간 모형은 두 개의 방정식, 즉 시점 t 에서의 체계의 상태를 표현하는 상태 방정식과 관측되지 않은 상태 벡터와 관측 오차의 함수인 관측 방정식으로 구성된다. 이렇게 상태-공간 모형으로 표현된 모형은 칼만필터(Kalman-Filter)를 통하여 모형을 반복적으로 계산함으로써 추정오차를 줄이는 모형을 선택할 수 있다 (Harvey, 1989).

※STARMA(1₁, 1₁)의 상태공간 모형

$$y(t) = Zb(t),$$

$$b(t) = T_t b(t-1) + Fe(t)$$

여기서, $y(t) = z(t)$ 이고, $Z = [I_n, O_n]$, $b(t) = \begin{bmatrix} z(t) \\ e(t) \end{bmatrix}$, $F(t) = \begin{bmatrix} I_n \\ O_n \end{bmatrix}$, $T_t = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$, $a = \phi_0 + \phi_1 W^{(1)}$, $b = \theta_0 + \theta_1 W^{(1)}$ 이다.

※STBL(1₁, 1₁, 1₁, 1₁)의 상태공간 모형

$$y(t) = Zb(t),$$

$$b(t) = T_t b(t-1) + Fe(t)$$

여기서, $y(t) = z(t)$ 이고, $Z = [I_n, O_n]$, $b(t) = \begin{bmatrix} z(t) \\ e(t) \end{bmatrix}$, $F(t) = \begin{bmatrix} I_n \\ O_n \end{bmatrix}$, $T_t[y_t] = \begin{bmatrix} cca & b+c(t) \\ 0 & 0 \end{bmatrix}$, $a = \phi_0 + \phi_1 W^{(1)}$, $b = \theta_0 + \theta_1 W^{(1)}$, $c(t) = \beta_{00} \text{diag}[z(t-1)] + \beta_{01} \text{diag}[z(t-1)]W^{(1)} + \beta_{10} \text{diag}[W^{(1)}z(t-1)] + \beta_{11} \text{diag}[W^{(1)}z(t-1)]W^{(1)}$.

3.2. Kalman Filter 모수추정

Kalman (1960)에 의해 도입된 칼만 필터 방법은 상태공간 모형을 추정하는 도구로서, 상태공간 모형에서 $u(t), v(t)$ 의 정규성이 만족될 때, 시점 t 에서 이용 가능한 정보에 기초하여 시점 t 의 비관측 변수들인 상태공간 벡터 $b(t)$ 의 조건부 평균과 분산을 반복과정을 통해 계산해 내는 기법이다. 시점 t 까지의 이용 가능한 정보를 바탕으로 $b(t)$ 을 추정하는 것이 기본 필터(basic Filter)이다. 칼만 필터 방정식은 새로운 관측치가 사용 가능하게 됨과 동시에 추정치가 Update 되는 것을 가능하게 해주는 방정식의 집합이라고 할 수 있다. $\beta(t|t-1)$ 에 의한 관찰치 $y(1), y(2), \dots, y(t-1)$ 가 주어졌을 때 상태 벡터인 $b(t)$ 을 t 시점에서의 실현치에 의한 모수 추정치로 정의하자. 정의에 의하여 행렬 $\beta(t|t-1)$ 에 대한 평균 잔차 제곱합 행렬(Mean squared error matrix)은 다음과 같이 표현될 수 있다.

$$\sigma^2 C(t|t-1) = E[\beta(t|t-1) - b(t)][\beta(t|t-1) - b(t)]^T$$

t 번째의 실현치가 나타나는 시간에 우리는 $\beta(t|t-1)$ 과 $C(t|t-1)$ 을 얻는데 이들은 $(t-1)$ 번째의 실현치에 의해 계산될 수 있다. t 번째의 실현치가 나타나는 $C(t|t)$ 을 따라 추정치 $\beta(t|t)$ 을 계산할 수 있으며 이 과정은 다음과 같은 예측 방정식으로 표현될 수 있다.

$$\begin{aligned}\beta(t|t) &= \beta(t|t-1) + C(t|t-1)Z^T H^{-1}(t)v(t) \\ C(t|t) &= C(t|t-1) - C(t|t-1)Z^T H^{-1}(t)ZC(t|t-1)\end{aligned}$$

여기서, $v(t) = y(t) - Z\beta(t|t-1)$ 이고, $H(t) = ZC(t|t-1)Z^T$ 이다. $v(t)$ 는 1단계의 예측 추정치의 오차이며 $\sigma^2 H(t)$ 는 $v(t)$ 에 대한 분산공분산 행렬이다. 이 경우 우리는 다음의 예측 방정식에 의해 주어진 $C(t+1|t)$ 을 따라 1단계 예측 추정치인 $\beta(t+1|t)$ 을 계산함으로써 t 번째의 실현치를 완성시킬 수 있다.

$$\begin{aligned}\beta(t+1|t) &= T[y(t+1)]\beta(t|t) \\ C(t+1|t) &= T[y(t+1)]C(t|t)T[y(t+1)]^T + FQ(t+1)F^T\end{aligned}$$

상태벡터 $b(t)$ 가 정상적이라는 가정 하에서 칼만 필터의 초기치로 사용할 추정치가 주어져야 하는데 이 값은 $\beta(1|0) = E[b(0)]$, $C(1|0) = FQ(0)F^T = \sigma^{-2}\text{var}[b(0)]$ 으로 각각 부여할 수 있다.

4. 사례연구

4.1. 자료설명 및 정상화

수두(Chichen Pox)는 홍역, B형 간염 등과 더불어 우리나라에서 제 2군 전염병으로 분류하고 있으며, 수두 바이러스에 의해 감염되거나 다른 오염된 물질이 코나 입으로 들어가서 감염된다.

본 논문에서 사용된 수두는 한국질병관리본부에서 2005년 8월부터 2010년 2월까지 전염성 감시체계로 전산 보고된 15개 시도의 월별 자료이다. 사용된 수두 자료는 연속성 자료가 아닌 빈도 자료로 포아송 분포를 따르는 변수이므로, 원래의 자료를 분산안정화변환과 1차로 계절차분을 통해 계절성을 제거한 후 제거 후 추세성도 보여 추세제거 후 정상시계열이 되었고, 이 자료를 정규분포로 근사시키기 위해 각 시도별로 평균과 표준편차를 구해 자료를 표준화하였다. 여기서 2005년 8월부터 2009년 8월 자료까지의 월별자료만을 이용하여 모형을 추정하였고 나머지 자료는 예측력을 비교하는데 사용하였다. 많은 지역에서 발생한 수두 자료를 정상성 가정을 만족시키기 위해 분산안정화 변환을 하고 원래 자료의 계절성을 조정하기 위해 12차분을 실시하고 추세제거를 위해 1차 차분을 실시하였다. 분산안정화 변환은 $h(z) = \sqrt{z}$ 의 변환을 취하였으며, 자료를 표준화하여 근사적으로 정규분포에 따르도록 변환하였다.

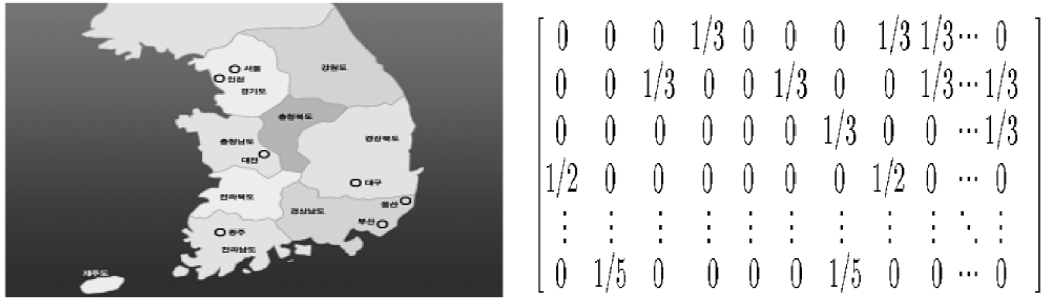


그림 3: 15개 시도의 연구관심영역과 1-차 이웃한 가중행렬

	STACF					STPACF			
	L=0	L=1	L=2	L=3		L=0	L=1	L=2	L=3
S=1	-0.431932	-0.195851	-0.162713	-0.1294	S=1	-0.431932	-0.06836	-0.054349	-0.002527
S=2	-0.065055	0.0203896	-0.008167	0.0023386	S=2	-0.309546	0.016053	-0.101703	0.016409
S=3	0.0380413	-0.025024	0.0144831	0.0028603	S=3	-0.183516	-0.059725	-0.062991	0.0165308
S=4	-0.043755	0.0499711	0.0782831	0.057547	S=4	-0.180616	0.094259	0.0960854	0.1007261
S=5	-0.030319	-0.083021	-0.10831	-0.083808	S=5	-0.184816	-0.054638	-0.031326	0.0500982
S=6	0.0147286	-0.045397	0.0079034	-0.041312	S=6	-0.175247	-0.223574	-0.015782	-0.064066
S=7	0.0104787	0.1326113	0.0508015	0.0749356	S=7	-0.159808	0.003444	-0.000774	-0.04494
S=8	0.0669889	-0.037354	-0.00676	-0.008897	S=8	-0.051067	-0.076531	0.0063042	-0.060608
S=9	-0.074167	0.0184921	0.0133809	0.0218759	S=9	-0.098826	0.0967991	0.0655417	0.0253178
S=10	0.0265269	-0.05	-0.034566	-0.021763	S=10	-0.046733	-0.084385	0.0489268	0.0365926
S=11	0.1665678	0.1199761	0.0816353	0.0417522	S=11	0.1948197	0.0768962	0.0242867	-0.054297
S=12	-0.311495	-0.15367	-0.109697	-0.061295	S=12	-0.1203	0.0571636	0.070377	0.0326884

그림 4: STBL 모형의 STACF, STPACF

각 지역에서 이웃(Neighbor)은 보통 지형적인 공통경계를 중심으로 나누어져 있다. 관심영역인 15개 시도에 대한 가중행렬(Weight Matrix) $W^{(l)}$ 은 공간에서 이웃하고 있는 순서에 관계가 있고, 본 자료에서 $W^{(l)}$ 은 l -번째 이웃하는 차원이 15인 15×15 행렬이다. 첫 번째 이웃하고 있는 지역수의 역비례로 구할 수 있는데 앞에서 언급한 것처럼 연계되어 있는 지역에 각각 동일한 가중치를 주는 것으로 15개 시도지역에 대한 연구관심영역과 가중행렬은 그림 3과 같다.

공간시계열 모형의 식별은 시간차수와 공간차수를 선택하는 문제이며, 공간시계열 자기상관함수(STACF)와 공간시계열 편자기상관함수(STPACF)를 구하여 감소패턴과 절단형태를 고려하여 시간차수와 공간 차수를 결정하는데 본 연구에 활용된 자료의 경우 두 함수 모두에서 지수적으로 감소하는 경향을 보이는 것을 그림 4를 통해서 알 수 있다. 따라서 STACF와 STPACF의 명확한 패턴을 보이지 않으므로 모형선택에 어려움이 있어 공간과 시간차수를 적절히 고려하여 경쟁 모형들의 AIC값을 비교해 공간시계열 자기회귀 모형은 STARMA(1₁, 1₁)로 공간시계열 중선형 모형은 STBL(1₁, 1₁, 1₁, 1₁) 모형으로 최종 선택하였다.

STARMA(1₁, 1₁) 모형은 $\Theta = (\phi_0, \phi_1, \theta_0, \theta_1)$ 인 4개의 모수를 가진다. 또한 STBL(1₁, 1₁, 1₁, 1₁) 모형은 $\Theta = (\phi_0, \phi_1, \theta_0, \theta_1, \beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$ 인 8개의 모수를 가진다. 모수추정을 위해 수치해석적 알고리즘

표 1: 공간시계열 모형의 모수추정결과

모형 모수	STARMA (뉴턴랩슨)	STARMA (칼만필터)	STBL (뉴턴랩슨)	STBL (칼만필터)
ϕ_0	0.5526	0.5527	0.6322	0.6321
ϕ_1	0.1839	0.1842	0.3246	0.3247
θ_0	-0.7956	-0.7959	-0.9047	-0.9047
θ_1	0.1007	0.1000	-0.1518	-0.1518
β_{00}			-0.0690	-0.0909
β_{01}			-0.9087	-0.9380
β_{10}			0.1000	0.1075
β_{11}			0.1874	0.1800
MSE	0.552	0.114	1.451	0.110
AIC	-318.9	-1166.4	202.9	-1176.7

표 2: 공간시계열 모형의 예측비교

Province	Sum of Squared Forecasts Errors	
	STARMA	STBL
서울	1044	896
부산	1217	1122
⋮	⋮	⋮
경북	407	314
경남	382	585
Total	20018	17966

으로 기존에 많이 활용되고 있는 Newton-Raphson 방법과 본 연구에서 초점을 두고 있는 Kalman-Filter 방법을 통해 추정한 결과는 표 1과 같다.

STARMA 모형에서 뉴턴랩슨 방법과 칼만필터 방법에 의한 모수추정 결과가 비슷하게 추정된 것을 알 수 있다. 또한 STBL 모형에서의 모수추정 결과도 뉴턴랩슨 방법이나 칼만필터 방법 역시 비슷한 결과로 추정됨을 알 수 있다. 다만 MSE값과 AIC통계량이 뉴턴랩슨방법보다 칼만필터 방법이 더 작게 추정되어 모수 추정방법으로 더 우수함을 알 수 있었다.

그리고 공간시계열 모형의 예측력 비교에 있어서는 표 2에서 알 수 있듯이 선형 모형인 STARMA 모형의 예측결과 예측오차제곱합(SSF)은 20018이고, 비선형 모형인 STBL 모형의 예측결과 예측오차제곱합(SSF)은 17966로써 공간시계열 모형의 비교에서 비선형 공간시계열 모형인 STBL 모형의 예측력이 더 좋게 나타났음을 알 수 있다. 따라서 본 연구에 사용된 수두자료의 예측에 있어서 STBL 모형이 더 적합하다고 볼 수 있다. 이는 수두 자료가 일정 시점에서 갑작스럽게 변동성이 커지는 현상을 잘 반영하는 중선형(BL) 모형에 공간 가중치를 반영한 공간시계열중선형(STBL) 모형이 보다 적합함을 나타낸 결과라 하겠다.

5. 결론

시계열 모형과 공간통계의 융합인 공간시계열 모형은 기존의 각 지역에 대한 개별 모형들을 고려하여 단순히 시간만을 고려한 단변량 시계열분석으로 ARIMA 모형과 BL 모형을 이용하였는데 이는 추정에 있어 매우 많은 모수를 갖게 되어 모수절약의 원칙에도 어긋나며 추정과 예측에 있어 정확도가 떨어지는 단점이 있었다.

본 연구에서는 공간적 정보를 공간 가중치행렬을 이용하여 모형에 반영한 공간시계열 모형을 연

구하였고 이 분야에서 모수추정의 경우 모형이 매우 복잡하여 수치해석적인 방법을 이용할 것을 제안해 왔는데 그 중 Newton-Raphson 방법과 Kalman-Filter 방법을 이용하여 모수추정의 수월성을 비교하였다. 뉴턴랩슨 방법의 경우는 초기치 선정문제와 수렴성의 문제가 있는 것으로 알려져 있으며, 이러한 단점을 극복하기 위해 칼만필터 방법을 연구하였고, 실증분석 결과에서도 확인하였듯이 모수추정 및 예측에 있어 칼만필터 방법이 매우 탁월함을 알 수 있었으며, 공간시계열 모형간의 비교에서는 STARMA 모형에 비해 확률적 변동성을 잘 반영해주는 STBL 모형의 예측 정확도가 더 높게 나타났다. 결론적으로 질병관리본부에서 전산보고 되고 있는 여러 질병자료의 경우 대체로 공간과 시간을 동시에 고려해야 하는 경우가 대부분이며 이 때 공간시계열 모형을 이용한 모형 설정과 모수추정 및 예측에 있어서는 칼만필터 방법의 활용을 통해 실제 현장에서 의미 있게 사용되어 질 수 있겠다.

본 연구에서는 전염성이 강한 질병인 수두 자료를 다루었으나 최근 문제가 되고 있는 AI나 신종플루, 구제역 등의 전염병 확산 모형에 대해 연구할 것이며 현재로써는 자료가 충분하지 않아 신뢰성 있는 모형을 수립하기 어렵고 추후의 연구 과제로 남긴다. 또한 자료의 형태와 특성에 따라 여러 비선형 시계열 모형들을 고려하여 공간정보를 포함시켜 공간시계열 모형에 대한 폭넓은 연구가 필요할 것이며 이는 다음의 연구과제로 남긴다.

참고 문헌

- 이성덕, 이용준, 박용석, 주재선, 이건명 (2007). 공간시계열 자료에 대한 STARMA 모형과 STBL 모형의 예측력 비교, <응용통계연구>, **20**, 91-102.
- Billard, L. and Dai, Y. (1998). A space-time bilinear model and its identification, *Journal of Time Series Analysis*, **19**, 657-679.
- Billard, L. and Dai, Y. (2003). Maximum likelihood estimation in space time bilinear model, *Journal of Time Series Analysis*, **24**, 25-44.
- Harvey, A. C. (1989). *Forecasting, Stuctural Time Series Models and The Kalman Filter*, Cambridge University, Cambridge.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Transaction ASME Journal of Basic Engineering Series D*, **83**, 95-108.
- Pfeifer, P. E. and Deutsch, S. J. (1981). Space-time ARMA modeling with contemporaneously correlated innovations, *Technometrics*, **23**, 401-409.

Kalman-Filter Estimation and Prediction for a Spatial Time Series Model

SungDuck Lee^{1,a}, EunHee Han^a, DuckKi Kim^a

^aDepartment of Information and Statistics, Chungbuk National University

Abstract

A spatial time series model was used for analyzing the method of spatial time series (not the ARIMA model that is popular for analyzing spatial time series) by using chicken pox data which is a highly contagious disease and grid data due to ARIMA not reflecting the spatial processes. Time series model contains a weighting matrix, because that spatial time series model influences the time variation as well as the spatial location. The weighting matrix reflects that the more geographically contiguous region has the higher spatial dependence. It is hypothesized that the weighting matrix gives neighboring areas the same influence in the study of the spatial time series model. Therefore, we try to present the conclusion with a weighting matrix in a way that gives the same weight to existing neighboring areas in the study of the suitability of the STARMA model, spatial time series model and STBL model, in the comparative study of the predictive power for statistical inference, and the results. Furthermore, through the Kalman-Filter method we try to show the superiority of the Kalman-Filter method through a parameter assumption and the processes of prediction.

Keywords: STARMA, STBL, weight matrix, STACF, STPACF, Newton-Raphson, Kalman-Filter, SSF.

This paper was supported by the Chungbuk National University in 2009.

¹ Professor, Department of Information and Statistics, Chungbuk National University, Chungbuk 361-763, Korea.
E-mail: sdlee@cbnu.ac.kr