시계열 환경변수 분포도 작성 및 불확실성 모델링: 미세먼지(PM₁₀) 농도 분포도 작성 사례연구

박 노 욱*

인하대학교 지리정보공학과, 402-751, 인천광역시 남구 용현동 253

Time-series Mapping and Uncertainty Modeling of Environmental Variables: A Case Study of PM₁₀ Concentration Mapping

No-Wook Park*

Department of Geoinformatic Engineering, Inha University, Incheon 402-751, Korea

Abstract: A multi-Gaussian kriging approach extended to space-time domain is presented for uncertainty modeling as well as time-series mapping of environmental variables. Within a multi-Gaussian framework, normal score transformed environmental variables are first decomposed into deterministic trend and stochastic residual components. After local temporal trend models are constructed, the parameters of the models are estimated and interpolated in space. Space-time correlation structures of stationary residual components are quantified using a product-sum space-time variogram model. The ccdf is modeled at all grid locations using this space-time variogram model and space-time kriging. Finally, e-type estimates and conditional variances are computed from the ccdf models for spatial mapping and uncertainty analysis, respectively. The proposed approach is illustrated through a case of time-series Particulate Matter 10 (PM₁₀) concentration mapping in Incheon Metropolitan city using monthly PM₁₀ concentrations at 13 stations for 3 years. It is shown that the proposed approach would generate reliable time-series PM₁₀ concentration maps with less mean bias and better prediction capability, compared to conventional spatial-only ordinary kriging. It is also demonstrated that the conditional variances and the probability exceeding a certain thresholding value would be useful information sources for interpretation.

Keywords: time-series, space-time kriging, uncertainty modeling, PM₁₀

요 약: 이 논문에서는 환경변수의 시계열 분포도 작성과 불확실성 모델링을 위해 시공간영역으로 확장된 다중 가우시안 크리깅을 제안하였다. 다중 가우시안 틀 안에서, 우선 정규점수변환된 환경변수를 결정론적 경향 성분과 확률론적 잔차 성분으로 분해하였다. 그리고 시간 경향 모델 계수의 내삽을 통해 경향 성분의 시계열 공간 분포도를 작성하였다. 정상성 잔차 성분의 시공간 상관 구조는 곱-합 시공간 베리오그램 모델을 이용하여 정량화하였고, 이 베리오그램 모델과 시공간 크리깅을 이용하여 국소적 누적 확률분포함수를 모델링하였다. 이 국소적 누적 확률분포함수로부터 평균값과 조건부 분산을 계산하여 공간분포도 작성과 불확실성 분석에 각각 이용하였다. 제안 기법의 적용성 평가를 위해 인천광역시에서 3년간 13개 관측소에서 측정된 월 평균 미세먼지(PM₁₀) 농도 자료를 이용한 시계열 분포도 작성 사례 연구를 수행하였다. 사례연구 결과, 제안 기법을 통해 기존 공간 정규 크리깅에 비해 작은 편향과 높은 예측 능력을 가진 시계열 미세먼지(PM₁₀) 농도 분포도 작성이 가능함을 확인할 수 있었다. 또한 조건부 분산과 특정 농도값을 초과할 확률값들은 해석을 위한 유용한 보조 정보를 제공하였다.

주요어: 시계열, 시공간 크리깅, 불확실성 모델링, 미세먼지(PM₁₀)

*Corresponding author: nwpark@inha.ac.kr
Tel: +82-32-860-7607
Fax: +82-32-863-1506

서 론

광상학에서 출발한 지구통계학은 이론의 확장, 컴퓨터 자원의 고급화 등으로 광상학, 석유공학, 지형학, 토양, 대기 및 해양학 등의 기존 분야뿐만 아니라, 생태학, GIS, 원격탐사 및 의료 지리학 등 다양한 분야에서 공간 자료 분석을 위한 유용한 방법론으로 인식되고 있다(Goovaerts, 1997; Deutsch and Journel, 1998). 특히 위치 정보와 함께 속성이 제시되는 지구과학 자료의 특성상 지구통계학은 적용 가능성이 매우 높은 공간 분석 방법론으로 간주되고 있다(Oh, 2000; Park, 2004; Park, 2008). 기존 연구에서는 공간적으로 소수의 타겟으로부터 샘플링되지 않은 미지의 지점에서의 값을 유추하는 일종의 내삽 과정에 크리깅과 시뮬레이션을 적용하여 공간 분포도를 작성해 왔다. 이와 더불어 조건부 누적확률분포함수(conditional cumulative distribution function: ccdf) 모델링과 시뮬레이션의 적용을 통해 국소 및 공간 불확실성 모델링이 가능하다(Goovaerts, 1997; Chilès and Delfiner, 1999).

지구통계학은 기존 결정론적 내삽 기법과 마찬가지로 주제도 작성에 기본적으로 이용될 수 있지만, 차별화된 모델링 이슈들이 존재한다. 우선 자료 통합의 관점에서 공간적으로 산재되어 있는 소수의 자료가 가지는 한계를 부가자료를 이용함으로써 내삽의 불확실성을 완화시킬 수 있다. 다변량 크리깅 기법을 적용함으로써 다양한 형태의 부가자료의 통합이 가능하며, 특히 최근 이용이 급증하고 있는 원격탐사 자료나 GIS 주제도의 결합이 가능하여 단일 자료만을 이용할 때에 비해 높은 정확도를 가지는 신뢰성 높은 주제도 생성이 가능하다(오석훈, 2005; Park et al., 2009). 두 번째로는 자료 통합시 다양한 스케일 자료의 이용이 가능하다는 것이다. 포인트 기반의 자료와 특정 영역을 차지하는 부가자료의 통합시 상호 스케일 차이를 베리오그램 모델링과 크리깅 적용에 직접적으로 반영할 수 있는 장점이 있다. 이러한 다중 스케일 자료의 통합 연구로 Kyriakidis(2004)는 이론적 배경을 제시하였으며, Goovaerts(2010)는 중금속 오염도 분포도 작성과 질병 자료 분포도 작성에 적용하였다. 국내에서는 박노욱(2010a)과 오석훈과 한성미(2010)가 각각 중저해상도 원격탐사 자료 기반의 주제도 및 물리탐사 자료에 적용하였다. 마지막 분야

로는 시공간 지구통계학 분야로, 기존 공간적 자기상관성만을 고려하는 지구통계학적 분석에 시간 차원을 추가하여 시공간 자기상관성을 직접적으로 분석에 이용하는 경우이다. 지구통계학은 시간적인 변이가 적은 지질학적 타겟을 대상으로 이론이 발전되었지만, 다양한 시공간 자료에 적용이 가능하다. 공간 개념과 서로 물리적인 의미가 다른 시간 개념을 모델에 포함시키는 것은 단순히 차원 하나가 늘어나는 문제로 간주할 수 없으며 상호 관계성을 정량화할 필요가 있는데, 이러한 시공간 지구통계학 모델에 대한 리뷰는 Kyriakidis and Journel(1999)을 참조할 수 있다. 응용 사례로는 대기 오염 물질 분석(Kyriakidis and Journel, 2001; De Iaco et al., 2002), 강수 및 기온 자료의 시계열 분석(Kyriakidis et al., 2004; Fernandez-Cortes et al., 2006, Spadavecchia and Williams, 2009) 등을 들 수 있다. 처음 2가지 이슈는 국내에서도 사례연구가 수행되었지만, 3번째 시공간 지구통계학적 분석의 경우 국내에서는 최근 박노욱과 장동호(2011)에 의해 시계열 시간별 기온 분포도 작성 연구가 수행되었을 뿐 아직 활성화되지 못한 상황이다.

그동안 시공간 지구통계학 분석과 관련된 연구는 대부분 시계열 분포도 작성에 치중되어 왔다. 그러나 속성값의 추정과 더불어 값 예측에 수반되는 불확실성 분석도 가능하다. 기존 시공간 지구통계학 연구는 일부 시뮬레이션 연구(Kyriakidis and Journel, 2001; Kyriakidis et al., 2004)를 제외하고는 이러한 시공간 영역에서의 국소적 불확실성 모델링 연구는 수행되지 않았다.

이 논문에서는 시공간 지구통계학적 분석에서 국소적 불확실성 모델링이 가능하도록 기존 공간 영역에서의 다중 가우시안 크리깅을 확장한 시공간 다중 가우시안 크리깅을 제안하고자 한다. 시공간 다중 가우시안 틀 안에서, ccdf 모델링에 필요한 평균과 분산 추정시 시공간 자기상관성 정보를 설명할 수 있도록 확장하였다. 이렇게 모델링된 ccdf로부터 속성 값과 조건부 분산 혹은 특정 임계치를 초과할 확률 계산이 가능하여 불확실성 분석이 가능하게 된다. 인천광역시 도시대기측정망에서 시계열로 측정된 미세먼지 PM_{10} 자료를 이용한 사례연구를 통해 제안기법의 적용 과정을 예시하였고, 주요 이슈에 대해 논의하였다.

시공간 지구통계학

공간 시계열 모델

지상의 소수 측정소에서 시계열로 측정된 환경변수는 개별 위치별로 서로 차이가 나는 공간 시계열 자료로 간주할 수 있다. 연구지역에서 이산 시간 영역 T 와 연속적인 공간 영역 D 에서 측정된 공간 시계열 자료값은 식 (1)과 같이 공간적으로 상관된 시계열 모델의 집합체인 확률함수 모델로 나타낼 수 있다.

$$\{Z(u, t), t \in T, u \in D\} \quad (1)$$

특정 공간 위치 u 와 특정 시간 t 에서의 불확실성은 식 (2)의 cdf의 형태로 모델링된다.

$$F_z(u, t; z | \text{info}) = \text{Prob}\{Z(u, t) \leq z | \text{info}\} \quad (2)$$

여기서 (info)는 u 및 t 주변에 포함되는 시공간 자료값을 나타낸다.

이 연구에서는 식 (2)의 cdf 모델링을 위해 모수적 접근인 다중 가우시안 크리깅을 적용하고자 한다. 다중 가우시안 방법은 cdf를 정규 분포로 가정하여, 정규 분포의 주요 모수인 평균과 분산을 크리깅을 통해 얻고 최종적으로 cdf를 모델링하게 된다(Goovaerts, 1999). 이러한 다중 가우시안 접근법을 적용하기 위해, 원 자료에 대해 정규점수변환(normal score transform, Deutsch and Journel, 1998)을 적용하여 평균 0, 분산 1인 표준정규분포를 따르는 정규점수변환된 공간 시계열을 얻었다. 이후 다중 가우시안 확률 분포하에 식 (2)의 cdf는 식 (3)으로 표현된다.

$$F_z(u, t; z | \text{info}) = G\left(\frac{z - E\{Z(u, t) | \text{info}\}}{\sqrt{\text{Var}\{Z(u, t) | \text{info}\}}}\right) \quad (3)$$

여기서 $G(\cdot)$ 는 정규 가우시안 cdf(cumulative distribution function)를 나타내며, $E\{Z(u, t) | \text{info}\}$ 와 $\text{Var}\{Z(u, t) | \text{info}\}$ 는 주어진 자료 (info)로부터 얻어지는 기댓값(평균)과 분산을 각각 나타낸다. 이 다중 가우시안 확률 분포 모델에서, 기댓값과 분산은 단순 크리깅의 추정치와 분산으로 표현될 수 있다.

식 (3)과 같이 특정 공간 위치 u 와 특정 시간 t 에서의 불확실성을 모델링하는 확률변수 $Z(u, t)$ 는 일반적으로 결정론적 경향 성분($m(u, t)$)과 잔차 성분($R(u, t)$)으로 분해될 수 있다(Kyriakidis and Journel, 1999).

$$Z(u, t) = m(u, t) + R(u, t) \quad (4)$$

여기서 경향 성분 $m(u, t)$ 은 경향성을 모델링하는 결정론적 평균 함수($E\{Z(u, t)\} = m(u, t)$)를 나타내며, 계절적 혹은 주기적 변동 요인과 시간적 추이에 따른 증가 및 감소를 나타내는 경향성을 모두 포함하는 성분이다. 반면 잔차 성분 $R(u, t)$ 은 평균 0인 2차 정상성 확률 함수($E\{R(u, t)\} = 0$)로 간주된다. 식 (4)의 성분 분해는 원 자료의 스케일에서도 가능하고, 이 연구에서는 다중 가우시안 접근법을 적용하였기 때문에 실제로 식 (4)는 모두 정규점수변환된 값에 대해 분해를 수행하게 된다. 앞으로 자료값에 아래 첨자 G 가 붙은 변수는 모두 정규점수변환된 것을 의미한다.

식 (4)에서 경향 성분과 잔차 성분으로의 분해는 실제 경향 성분을 나타내는 자료가 없기 때문에 주관적인 모델링 방식에 따라 성분들이 달라질 수 있다(Diggle and Ribeiro, 2007). 예를 들어 경향 성분 추출을 위해 시계열 자료 분석에 사용하는 이동 평균방법을 사용하거나, 자료 특성을 반영할 수 있는 특정 기본 함수 성분에 대한 최소자승 모델링을 이용할 수 있다. 이렇게 서로 다른 경향 성분 추출에 따라 잔차 성분들은 달라지게 되며, 결과적으로 잔차 성분은 경향 성분으로 설명되지 않는 자료의 성분을 궁극적으로 나타내게 된다.

경향 성분 모델링

연구지역에 존재하는 N 개의 특정 측정소($\alpha=1, 2, \dots, N$)에서의 정규점수변환된 시계열을 $z_G(u_\alpha, t_i)$ 로 나타내었을 때, α 번째 관측소에서의 경향 성분($m_G(u_\alpha, t_i)$)은 다음과 같이 시간에 대한 기본 성분 함수들의 합으로 표현할 수 있다(Kyriakidis and Journel, 1999).

$$m_G(u_\alpha, t_i) = \sum_{k=0}^K b_k(u_\alpha) f_k(t_i), \quad i=1, 2, \dots, T_\alpha \quad (5)$$

여기서 f_k 는 경향 성분을 모델링하는 기본 성분 함수를 나타내며, 일반적으로 $f_0=1$ 로 간주한다. 만약 고려하고 있는 자료에 주기성과 경향성이 모두 포함된다면, k 는 절편, 선형 기울기, 주기함수의 진폭 및 위상 등 총 4개가 된다.

이 연구에서는 기본 성분 함수를 이용한 경향 성분 모델링을 수행하는데 있어서, 주기성, 경향성의 존재 유무와 관계없이 적용 가능하도록 Kyriakidis et al.(2004)에 의해 제안된 연구 지역내 공간 평균

시계열을 기본 성분 함수로 사용하였다.

$$m_G(\mathbf{u}_\alpha, t_i) = b_0(\mathbf{u}_\alpha) + b_1(\mathbf{u}_\alpha) \left[\frac{1}{N} \sum_{\alpha=1}^N z_G(\mathbf{u}_\alpha, t_i) \right], \quad i=1, 2, \dots, T_\alpha \quad (6)$$

각 측정소별로 시계열 관측값을 연구 지역내의 공간 평균 시계열 함수 식 (6)에 대해 선형회귀분석을 수행하면, 관측소별로 식 (6)의 절편과 기울기를 각각 나타내는 $b_0(\mathbf{u}_\alpha)$ 와 $b_1(\mathbf{u}_\alpha)$ 를 얻게 된다. 이 계수값들은 연구 지역 전체의 공간 평균 시계열 함수와 측정소별 관측 시계열 값 사이의 유사도를 나타내는데, 절편과 기울기가 각각 0과 1에 근접할수록 연구지역 전체의 공간 평균 시계열과 유사함을 의미한다. 일단 측정소별로 이 경향 성분을 모델링하는 두 계수들을 얻게 되면, 이 계수들을 전체 대상 지역으로 내삽한 후에 식 (6)과 함께 결합함으로써 고려하고 있는 시간대별로 경향 성분 분포도를 얻게 된다.

잔차 성분 및 ccdf 모델링

각 측정소별로 경향 성분을 모델링하면, 측정소에서의 잔차는 식 (7)로 표현된다.

$$r_G(\mathbf{u}_\alpha, t_i) = z_G(\mathbf{u}_\alpha, t_i) - m_G(\mathbf{u}_\alpha, t_i) \quad (7)$$

경향성이 제거된 잔차 성분을 2차 정상성 확률함수로 간주하기 때문에, 측정소별 잔차 성분에 대한 공간 내삽을 위해서는 잔차 성분의 시공간 자기상관성을 정량화하는 베리오그램 모델링을 수행해야 한다.

공간과 시간의 이격 간격 (h_u, h_t) 에서의 실험적 시공간 베리오그램은 식 (8)로 표현할 수 있다.

$$\begin{aligned} \gamma_{st}(h_u, h_t) \\ = \frac{1}{2n(h_u, h_t)} \sum_{\alpha=1}^{n(h_u, h_t)} [z_G(\mathbf{u}_\alpha, t_\alpha) - z_G(\mathbf{u}_\alpha + h_u, t_\alpha + h_t)] \end{aligned} \quad (8)$$

여기서 $n(h_u, h_t)$ 는 이격 간격 (h_u, h_t) 에 포함되는 공간 자료 쌍의 개수를 나타낸다.

식 (8)을 통해 계산된 실험적 시공간 베리오그램을 계산한 후에, 모든 시간 및 공간의 이격 간격에서의 베리오그램을 정의하는 시공간 베리오그램을 모델링한다. 시공간 지구통계학에서는 경향 성분 분해 및 모델링과 더불어 시공간 베리오그램 모델링이 주요 연구 이슈가 되어왔는데(Kyriakidis and Journel, 1999), 이 연구에서는 곱-합 (product-sum) 모델(De Cesare

et al., 2001)을 적용하였다. 이 모델은 실험적 시공간 베리오그램을 얻은 후에, 공간 이격 간격과 시간 이격 간격을 각각 0으로 지정함으로써 얻게 되는 시간 베리오그램($\gamma_t = \gamma_{st}(h_u=0, h_t)$)과 공간 베리오그램($\gamma_s = \gamma_{st}(h_u=h_t, h_t=0)$)만을 모델링한 후에, 시공간 문턱값(sill)을 함께 이용하여 전체 시공간 베리오그램 모델을 얻게 된다. 기존 연구에서 사용된 메트릭(metric) 모델(Dimitrakopoulos and Luo, 1994)은 공간과 시간 향에 대해 이방성 모델을 적용하여 시공간 베리오그램을 모델링하는데, 시간 및 공간 베리오그램 모델의 형태가 같아야 하는 제약이 있다. 그래서 기존 공간 베리오그램 모델링 과정과 유사한 곱-합 베리오그램 모델을 적용하였다.

곱-합 시공간 베리오그램 모델 $\gamma_{st}(h_u, h_t)$ 은 다음과 같이 개별적인 공간 베리오그램 모델 γ_s 와 시간 베리오그램 γ_t 의 곱과 합의 결합으로 표현된다(De Cesare et al., 2001).

$$\begin{aligned} \gamma_{st}(h_u, h_t) = & (k_1 C_s(0) + k_3) \gamma_t(h_t) + (k_1 C_t(0) + k_2) \gamma_s(h_u) \\ & - k_1 \gamma_s(h_u) \gamma_t(h_t) \end{aligned} \quad (9)$$

여기서 $C_s(0)$ 와 $C_t(0)$ 는 공간 베리오그램과 시간 베리오그램의 문턱값을 각각 나타내며, 식 (9)의 정의에 필요한 파라미터 k_1 , k_2 및 k_3 은 식 (10)으로 정의된다.

$$\begin{aligned} k_1 &= [C_s(0) + C_t(0) - C_{st}(0, 0)] / C_s(0) C_t(0) > 0 \\ k_2 &= [C_{st}(0, 0) - C_s(0)] / C_s(0) \geq 0 \\ k_3 &= [C_{st}(0, 0) - C_t(0)] / C_t(0) \geq 0 \end{aligned} \quad (10)$$

여기서 $C_{st}(0, 0)$ 는 시공간 베리오그램의 문턱값을 나타낸다.

식 (9)와 같이 잔차 성분의 시공간 베리오그램 모델링이 완료되면, 다중 가우시안 확률 분포 모델의 가정하에 ccdf 모델링에 필요한 기댓값과 분산은 크리깅을 통해 얻게 된다. 잔차 성분의 단순 크리깅은 식 (11)로 표현되며, 이 때 얻어진 크리깅 추정값에 식 (6)으로부터 얻어진 경향 성분값을 더하여 ccdf의 평균으로, 단순 크리깅의 분산값을 분산값으로 사용하여 식 (3)에 적용함으로써 연구지역의 모든 위치에서의 ccdf 모델링이 가능해진다. 결국 시공간 다중 가우시안 크리깅과 공간 다중 가우시안 크리깅의 차이는 정규분포의 파라미터를 추정하는 단계에서 시공간 베리오그램 모델을 이용하여 시공간 자료를 크리깅 과정에 포함하느냐로 좌우된다.

$$z_G^*(\mathbf{u}, t) = \sum_{\alpha=1}^{n(\mathbf{u}, t)} \lambda_{\alpha}(\mathbf{u}, t) z(\mathbf{u}_{\alpha}, t_{\alpha})$$

$$\sigma_G^*(\mathbf{u}, t) = C_{st}(0, 0) - \sum_{\alpha=1}^{n(\mathbf{u}, t)} \lambda_{\alpha}(\mathbf{u}, t) C_{st}(\mathbf{u}_{\alpha} - \mathbf{u}, t_{\alpha} - t) \quad (11)$$

여기서 $C_{st}(\mathbf{u}_{\alpha} - \mathbf{u}, t_{\alpha} - t)$ 는 추정 위치와 주변 이웃값의 시공간 공분산값을 나타낸다.

속성값 추정 및 불확실성 모델링

식 (11)을 통해 다중 가우시안 틀 안에서 연구지역의 모든 위치에서 ccdf를 결정하는 파라미터를 결정하게 되면, 정규점수변환된 값을 원 자료의 스케일로 변환하는 역정규점수변환을 적용하면 모든 위치에서 원 자료의 스케일에서의 ccdf를 얻게 되고 이로부터 다양한 통계값 계산이 가능하다.

우선 각 위치에서의 속성값 $z^*(\mathbf{u}, t)$ 은 역정규점수변환된 ccdf의 기댓값과 경향 성분을 통해 얻을 수 있다(Park et al., 2009).

$$z^*(\mathbf{u}, t) = \frac{1}{100} \sum_{k=1}^{100} z_p(\mathbf{u}, t) \quad \text{with } p=k/100-0.5/100 \quad (12)$$

여기서 $z_p(\mathbf{u}, t)$ 는 z 의 ccdf $F(\mathbf{u}, t | (\text{info}))$ 의 p 백분위수를 나타낸다.

또한 연구지역의 모든 위치에서 ccdf를 알고 있기 때문에, 이로부터 불확실성 분석을 위한 통계치를 계산할 수 있다. 이 연구에서는 불확실성의 척도로 ccdf의 평균으로부터의 편진 정도를 나타내는 조건부 분산을 이용하였다. 만약 크리깅 분산값만을 이용할 경우, 매월 대상 지역내의 측정소 위치가 모두 동일하기 때문에 전반적인 양상이 유사하게 나타날 것이다. 이는 크리깅 분산값 계산에 주변 자료값이 전혀 반영되지 않고, 공분산 값만이 반영되기 때문이다. 반면 시공간 다중 가우시안 크리깅의 ccdf로부터 계산할 수 있는 조건부 분산값은 실제 자료값이 분산값에 반영되기 때문에 추정값의 불확실성 분석에 직접적으로 활용될 수 있다. 조건부 분산이 상대적으로 클 경우, ccdf가 평균을 중심으로 넓게 퍼져 있음을 의미하여 값 예측의 불확실성이 큼을 의미한다. 조건부 분산 $\text{Var}(\mathbf{u}, t)$ 값은 다음 식 (13)과 같이 자료 범위를 K 로 이산화시켜 계산하였다(Goovaerts, 1997).

$$\text{Var}(\mathbf{u}, t) =$$

$$\sum_{k=1}^{K+1} [\bar{z}_k - z_E^*(\mathbf{u}, t)]^2 [F_z(\mathbf{u}, t; z_k | (\text{info})) - F_z(\mathbf{u}, t; z_{k-1} | (\text{info}))] \quad (13)$$

$$z_E^*(\mathbf{u}, t) = \sum_{k=1}^{K+1} \bar{z}_k \cdot [F_z(\mathbf{u}, t; z_k | (\text{info})) - F_z(\mathbf{u}, t; z_{k-1} | (\text{info}))]$$

여기서 $\bar{z}_k$ 는 (z_{k-1}, z_k) 의 평균값을 나타낸다.

또한 식 (3)의 z 에 특정 임계치값을 대입함으로써 특정 임계치값을 초과하지 않을 확률이나, 초과할 확률을 쉽게 계산할 수 있다.

예측 능력 검증

공간 크리깅이나 시공간 크리깅 모두 샘플 위치에 서는 자료값을 그대로 재생산하기 때문에, 검증을 위해 별도의 자료가 필요하다. 이 연구에서는 공간적으로 자료의 수가 적기 때문에 leave-one-out 기반 교차검증을 이용하였다. 임의로 특정 샘플 위치의 자료를 제거한 후에, 나머지 시공간 자료값들을 이용하여 값이 제거된 위치의 값을 추정하는 방식을 모든 샘플 위치에서 반복함으로써 모든 샘플 위치에서의 예측값을 얻게 된다. 이 예측값과 원래 참값을 비교함으로써 추정 오차를 계산할 수 있는데, 이 연구에서는 평균 오차(mean error: ME), 평균 절대 오차(mean absolute error: MAE) 및 평균 제곱 오차의 제곱근(root mean square error: RMSE)을 식 (14)에 의해 계산하였다. ME를 통해 과추정 혹은 저추정 양상 즉 편향 정도를 파악할 수 있으며, MAE와 RMSE로부터 오차의 크기를 판단할 수 있다.

$$\text{ME} = \frac{1}{N} \sum_{\alpha=1}^N z^*(\mathbf{u}_{\alpha}, t) - z(\mathbf{u}_{\alpha}, t)$$

$$\text{MAE} = \frac{1}{N} \sum_{\alpha=1}^N |z^*(\mathbf{u}_{\alpha}, t) - z(\mathbf{u}_{\alpha}, t)| \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{\alpha=1}^N (z^*(\mathbf{u}_{\alpha}, t) - z(\mathbf{u}_{\alpha}, t))^2}$$

사례연구

제안 기법의 적용성 평가를 위해 인천광역시에 설

Fig. 1. Schematic diagram of a space-time multi-Gaussian kriging approach adopted in this study.

치된 도시대기측정망에서 시계열로 측정된 미세먼지 PM_{10} 자료를 이용한 사례연구를 수행하였다. 여기서 주의할 것은 이 사례연구의 목적은 인천광역시의 미세먼지 농도 분포도 작성을 통한 지역적 특성을 규명하기보다는 제안 기법의 적용 과정을 예시하는데 있다.

Fig. 1은 이 연구에서 제안한 자료 분석 절차를 나타내고 있다. 전체 시계열 자료에 대해 우선 정규점수변환을 수행하였으며, 이 변환된 자료에 대해 경향성분을 모델링하였다. 측정소에서의 경향 성분을 추출한 후에 내삽을 통해 연구지역 전체에 걸친 시간별 경향 성분을 유추하였다. 그리고 측정소에서의 정규점수변환된 자료값에서 경향 성분을 빼줌으로써 잔차 성분을 추출하였다. 이렇게 추출된 잔차 성분을 확률 변수로 가정하여 실험적 시공간 베리오그램을 작성한 후에 곱합 베리오그램 모델을 이용하여 시공간 베리오그램을 모델링하였다. 시공간 베리오그램 모델에 기반을 두어 단순 크리깅을 수행하여 평균과 분산값을 얻음으로써, 앞서 추출한 경향 성분과 결합하여 연구지역의 모든 위치에서 ccdf를 얻게 된다. 이 ccdf로부터 미세먼지 추정치를 얻게 되고, 부가적으로 조건부 분산과 특정 임계치 값을 초과할 확률을 계산함으로써 불확실성 분석을 수행하였다.

Fig. 2. Locations and names of observation stations with administrative boundaries.

연구 지역 및 사용 자료

대상 지역인 인천광역시에는 공업단지, 인천항, 수도권 매립지, 인천국제공항 등의 대규모 오염 발생원과 신도시 개발사업 등으로 대기환경 여건이 양호하지 않다고 알려져 있다(신문기 외, 2007). 원래 인천광역시에는 15개의 도시대기 측정소가 있는데, 강화군에 설치된 송해 측정소와 영종도에 설치된 운서 측정소는 인천 도심 대기오염도와의 연계성 부족으로 제외하여, 최종적으로 13개 측정소 자료를 이용하였다(Fig. 2). 각 측정소에서는 아황산가스, 일산화탄소, 이산화질소, 오존, 미세먼지 PM_{10} , 미세먼지 $PM_{2.5}$ 등의 물질을 측정하는데, 이 연구에서는 시범적으로 미세먼지 PM_{10} 을 대상으로 하였다. 미세먼지 PM_{10} 은 대기중 고체상태의 입자와 액적상태의 입자의 혼합물로 공기역학 직경이 $10\mu m$ 이하인 물질을 의미한다(인천광역시 보건환경연구원, 2011). 인천광역시 보건환경연구원 환경정보공개시스템에서 제공하는 2008년 2월부터 2010년 1월까지 3년간 측정된 총 36개월의 월평균 확정 자료를 이용하였는데, 2009년 6월은 모든 측정소의 자료가 부재하여 5월과 7월 자료를 선형 내삽하여 사용하였다. 사례연구에서 고려한 13개 관측소는 상업지역 3개소, 주거지역 7개소, 녹지지역 1개소 및 산업단지 1개소에 위치하고 있다(인

Fig. 3. (a) Space-averaged time profile of normal score transformed monthly PM₁₀ concentration values. Intercept and slope values of local temporal trend models at observation stations are shown in (b) and (c), respectively.

천광역시 보건환경연구원, 2011). 부가적으로 대상 지역은 표고 50 m 이하가 대부분을 차지하는 저지대이며 완전한 경사를 가지므로 지형 효과는 두드러지지 않는다(신문기 외, 2007).

경향 성분 모델링

이 연구에서는 정규점수변환된 값의 공간 평균 시계열을 기본 성분함수로 간주하여 각 측정소별로 선형회귀분석을 수행하여 경향성분을 추출하였다. Fig. 3a는 연구지역에서 매달 13개 측정소에서 관측된 미세먼지의 정규점수변환 평균값을 시계열로 나타내고 있다. 전반적인 경향은 매년 4월을 기점으로 여름철로 갈수록 농도가 약해지다가 다시 가을철에서 증가하는 양상을 나타내고 있는데, 겨울철의 양상은 매년

조금씩 차이를 나타내고 있다. Fig. 3b와 c는 13개 측정소의 시계열과 공간 평균 시계열 사이의 선형회귀분석의 결과로 제시되는 절편과 기울기를 나타내고 있다. 각 측정소별로 계수들의 값이 차이가 남을 확인할 수 있으며, 이러한 차이는 관측소별 경향값의 차이를 유발하게 되고, 결과적으로 시공간 다중 가우시안 크리깅의 분석 대상인 잔차 성분의 차이를 유발하게 된다.

측정소별로 경향 성분계수들이 결정되면, 이 계수들을 연구지역 전체로 내삽을 수행하게 된다. 일반적으로 선형회귀분석 결과인 절편과 기울기는 동일 분석과정을 통해 얻어졌으므로 이 두 계수의 상관관계를 반영하면서 내삽을 수행하여야 한다. 그러나 이 연구에서는 정규점수변환 수행으로 두 계수 사이의

Fig. 4. Spatial distribution of trend components for April, August, October, and December in 2009.

선형 상관계수의 크기가 0.2로 높게 나타나지 않아서, 개별적으로 *radial basis* 내삽 기법을 적용하였다. 이렇게 내삽을 통해 시간 경향 모델 계수의 공간 분포도를 얻게 되면, Fig. 3a의 시간경향성분과 결합하여 매월 경향 성분 분포도를 얻을 수 있다.

Fig. 4는 역정규점수변환을 통해 얻어진 원 미세먼지 농도 스케일의 2009년 4월, 8월, 10월 및 12월의 경향 성분 분포도를 나타내고 있다. 4월과 12월의 경향값이 가장 높게 나타나고 있으며, 상대적으로 8월이 가장 낮은 값들이 분포하고 있다. 4월과 12월은 Fig. 3a의 시간 경향 성분이 유사하여 전반적으로 매우 유사한 값을 나타내고 있다. 전반적인 경향은 달

별로 연구지역의 북서쪽, 서쪽 및 남서쪽이 높은 값을 나타내고 있으며, 동쪽으로 갈수록 상대적으로 낮은 값을 나타내고 있다. 이러한 연구지역 서쪽에서 값이 높은 경향은 편서풍 지대에 위치한 우리나라의 특성상 중국에서 발생한 오염물질의 이동과 대규모 오염원이 서쪽 연안에 위치한 지역적 특성이 반영된 것으로 판단된다. 또한 연구 지역 내의 측정소 수가 적은 관계로 내삽에 따른 평활화 양상이 나타나며, 이러한 양상은 특히 서구와 연구구에서 측정소에서 멀어질수록 두드러지고 있다. 따라서 이 연구에서는 측정소 자료만을 내삽하여 경향 성분을 추출하였지만, 광역적 범위에서 기상 자료와 토지이용 특성을

Fig. 5. (a) Experimental marginal temporal variogram with fitted model, (b) Experimental marginal spatial variogram with fitted model, and (c) Fitted product-sum space-time variogram model.

반영한 경향 성분 추출이 앞으로 필요한 것으로 판단된다.

시공간 베리오그램 모델링

측정소별로 경향 성분을 제거한 잔차 성분에 대한 실험적 시공간 베리오그램 작성에는 De Cesare et al.(2002)의 변형 GAMV 함수를 이용하였다.

이 계산된 시공간 실험적 베리오그램을 곱-합 모델로 모델링하기 위해 공간 이격거리와 시간 이격거리를 각각 0으로 놓고 계산된 시간 및 공간 베리오그램을 각각 모델링하였다. Fig. 5a의 시간 베리오그램을 살펴보면, 시간적 경향 성분이 제거된 이후 잔차 성분에서 약 10달 정도까지 시간적 자기상관성이 유지됨을 확인할 수 있으며, 이러한 경향은 지수함수로 모델링하였다. 반면, Fig. 5b의 공간 베리오그램은 시

간 베리오그램과 달리 매우 불규칙한 양상을 나타내고 있는데, 첫 번째 이격거리에서 높은 값을 나타내고 있다. 이러한 결과는 서로 인접해 있는 송림, 숲의 및 신흥(Fig. 2) 측정소 값들의 차이에 기인한 것이다. 이러한 불규칙한 공간 실험 베리오그램의 모델링에 큰 상대적 너겟효과와 구형 모델을 이용하였다. 최종적으로 곱-합 시공간 베리오그램 모델링에 필요한 시공간 문턱값은 시공간 실험적 베리오그램의 시각적 판독과 식 (10)의 조건에 따라 시간 및 공간 베리오그램의 문턱값을 고려하여 설정하였다. 논문에 제시하지는 않았지만, $t=0$ 일 때 잔차의 공간적 연속성이 낮게 나타나더라도 시간의 변화에 따른 잔차의 공간적 연속성은 상대적으로 증가하였다. 따라서 시간 및 공간 연속성의 상호 연관성을 함께 고려하기 위해 곱 모델이 아니라 곱-합 모델을 사용하였다.

Fig. 6. Scatter-plot and descriptive statistics of cross validation results, (a) spatial-only ordinary kriging and (b) space-time multi-Gaussian kriging.

Fig. 5c는 이러한 과정을 거쳐 얻어진 시공간 베리오그램 모델을 나타내고 있으며, 이 모델은 시공간 다중 가우시안 크리깅의 가중치 결정에 직접적으로 이용된다. 공간 베리오그램의 큰 상대적 너겟효과는 결과적으로 실제 크리깅 수행시 비록 전체 문턱값의 차이는 적더라도 가까운 거리에서 이웃한 측정소의 값보다 시간적 탐색창에 위치한 자료값이 가중치 결정에 크게 기여함을 의미한다.

다중 가우시안 크리깅 적용 결과

시공간 베리오그램 모델링을 이용하여 정규점수 차원에서의 잔차에 대해 시공간 단순 크리깅을 적용하였는데, 크리깅 대상 격자의 크기는 250미터로 하였다. 크리깅 적용시 필요한 공간 탐색에는 연구지역내의 13개 모든 측정소를 사용하였으며, 시간 탐색창의 크기는 실험적으로 6달로 설정하였다. 시공간 단순 크리깅의 적용에는 Spadavecchia(2008)의 Edinburgh Space-Time Geostatistics Fortran90 프로그램을 이용하였으며, 자료 변환 및 후처리에는 Fortran 프로그램을 수행하였다.

시공간 단순 크리깅을 적용하여 연구지역의 모든 격자에서 추정값과 분산값을 얻게 되는데, 이 추정값에 경향 성분을 합한 값을 평균, 크리깅 분산값을 분산으로 하는 정규점수 차원에서의 cdf를 얻게 된다. 이 cdf로부터 식 (12)를 이용하여 추정값을 얻었다.

우선 시공간 크리깅의 예측 능력을 검증하기 위해 leave-one-out 교차 검증을 적용하였다. 또한 비교의 목적으로 공간적 자기상관성만을 고려하는 기존 공간 정규 크리깅에 대한 교차 검증을 개별 36달의 자료

에 대해 적용하였다. Fig. 6은 이 두 기법의 추정값과 참값의 산점도와 상관계수, ME, MAE 및 RMSE를 나타내고 있다. 산점도의 양상을 살펴보면, 공간 정규 크리깅에서 크리깅 예측치의 일반적인 경향인 낮은 값의 과추정과 높은 값의 저추정 양상이 상대적으로 크게 나타나는데, 특히 높은 값의 저추정 양상이 두드러진다. 원 자료와의 선형상관계수를 비교하면, 공간 정규 크리깅에 비해 시공간 크리깅의 더 큰 값을 나타내어 참값에 근사하고 있다. 교차 검증의 통계치를 살펴보면, ME는 모두 음수로 나타나서 전반적으로 저추정하고 있지만, 시공간 크리깅이 더 0에 가까운 값을 나타내어 편향 정도가 약한 것으로 나타났다. MAE와 RMSE의 경우, 시공간 크리깅이 모두 낮은 값을 나타내었는데 공간 정규 크리깅에 비해 MAE와 RMSE는 각각 39%와 35%의 향상도를 나타내었다. 이러한 결과는 공간적으로 가용할 수 있는 자료의 수가 적은 경우, 공간 자기상관성 추정에 어려움이 있을 수 있으며 또한 논문에서 제시하지는 않았지만 공간 베리오그램 모델의 상대적 너겟 효과가 대부분의 달에서 모두 크게 나타남으로써 예측 능력이 떨어진 것으로 판단된다. 따라서 시간 및 시공간 자기상관성을 크리깅 과정에 반영함으로써 보다 향상된 예측 능력을 가지는 주제도 작성이 가능함을 확인할 수 있었다.

지면상의 한계로 전술한 과정으로 제작된 2009년 4월, 8월, 10월 및 12월의 미세먼지 분포도만을 Fig. 7에 예시로 제시하였다. Fig. 4의 경향 성분과 비교하였을 때, 전반적으로는 경향 성분의 패턴이 나타나지만 확률변수로 간주한 잔차의 영향으로 국소적으로는

Fig. 7. Spatial distribution of PM₁₀ concentrations for April, August, October, and December in 2009.

다른 양상을 나타내고 있다. 특히 경향 성분이 거의 유사하게 나타난 4월과 12월의 경우, 12월에 연구지역의 북부 및 남쪽의 미세먼지 값이 상대적으로 높게 나타나고 있다. 이는 잔차 성분이 이 지역에서 상대적으로 크게 나타나서 시공간 크리깅 결과에 반영된 결과이다. 또한 경향 성분의 평활화 효과는 잔차 성분이 합해짐으로써 전반적으로 약해진 것을 확인할 수 있다. 4월과 12월에 상대적으로 미세먼지의 농도가 상대적으로 높게 나타났으며, 연구지역에서는 서쪽, 북서쪽 및 동춘 관측소 근방에서 높게 나타나고 있다. 이러한 양상은 8월과 10월에도 유사하게 나타나고 있다. 전반적인 값은 4월에서 8월로 갈수록 농

도가 낮아졌으며, 가을을 거쳐 겨울인 12월로 갈수록 다시 미세먼지의 농도가 증가하고 있다.

각 월별 기상 자료를 살펴보면(기상청, 2011), 최대 풍속 풍향은 4월과 12월의 경우 서북서와 북서 방향으로 나타났으며, 8월과 10월은 남남서와 북북서 방향이 우세하였다. 기존 인천 지역의 미세먼지 농도와 기상 인자의 연관성을 분석한 신문기 외(2007)에 따르면, 대규모 오염원이 서쪽 연안에 위치한 인천 지역의 특성상 서풍 계열 바람이 불 때 미세먼지 농도가 높게 나타난다는 결과와 대체로 일치한다. 또한 4월과 12월은 월 누적 강수량이 40.4 mm와 17.4 mm로 8월의 316 mm와 10월의 88.5 mm에 비해 상대적

Fig. 8. Spatial distribution of conditional variance for April, August, October, and December in 2009.

으로 작게 나타났는데, 일기 유형별로는 강우시에 보통 미세먼지 농도가 낮게 나타난다는 기존 신문기 외(2007)의 연구 보고와도 일치하고 있다. 또한 2009년에는 4월, 10월과 12월 황사 관측 일수가 각각 0.3일, 0.9일, 2.9일로 나타났다. 따라서 4월과 12월의 상대적으로 높은 미세먼지 농도값은 전술한 기상 조건과 더불어 황사에 기인한 것으로 판단된다. 반면 10월은 4월에 비해 황사 관측 일수가 많았음에도 불구하고 상대적으로 낮은 농도값을 나타내고 있는데, 이는 전술한 10월의 전반적인 기상 조건과 결합되어 즉 미세먼지가 농집할 수 있는 조건이 4월에 비해 나빠서 월평균 값 계산 과정을 통해 상대적으로 낮

게 나타난 것으로 판단된다. 이러한 기상 자료와 미세먼지 분포도와 연관성 이외에도 환경부 토지이용도와 중첩하여 전반적인 패턴을 분석하였다. 해안 도시의 특성상 연안 지역에 주로 공업단지가 위치하고 최근 택지 개발이 이루어 지역이 위치함으로써 이러한 토지이용 지역에서 대체로 높은 값을 나타내고 있다. 상대적으로는 연구지역의 가운데에 위치한 주거지역에서 상대적으로 낮게 나타나고 있지만, 연수구 지역은 택지 개발 사업 지역과 인접하여 다른 주거지역에 비해 상대적으로 높은 값을 나타내고 있다. 그러나 이러한 토지이용에 따른 패턴도 중요하지만, 지역의 특성상 편서풍 지대에 위치한다는 사실과 중

Fig. 9. (a) Spatial distribution of the probability that the PM₁₀ concentration exceeds 80 $\mu\text{g}/\text{m}^3$ for December in 2009, (b) Classification result on the basis that the e-type estimates exceed 80 $\mu\text{g}/\text{m}^3$, (c) risk α , and (d) risk β .

국으로부터의 장거리 오염원의 유입, 해염에 의한 자연 오염 효과 등도 고려해야 하기 때문에 이러한 요소들의 종합적인 해석이 추후 필요한 것으로 판단된다.

불확실성 분석 결과

Fig. 8은 2009년 4월, 8월, 10월 및 12월 미세먼지의 조건부 분산값을 나타내고 있다. 전반적으로 측정소의 위치로부터 멀어지거나 측정소가 존재하지 않는 북서쪽과 남서쪽에서 조건부 분산값이 높게 나타나고

있다. 4개 달 모두에서 미세먼지 농도가 상대적으로 매우 높거나 매우 낮은 지역에서 조건부 분산값이 중간 혹은 낮은 값을 나타내고 있다. 반면, 월별로 전반적인 농도값이 중간인 지역 중에서 연구지역의 북서, 남서쪽에 위치한 지역에서 대부분 조건부 분산값이 높게 나타나고 있다. 4월과 12월에 연구 지역 북부에 위치한 검단 측정소 주변은 모두 높은 값을 나타내고 있지만, 12월에 4월에 비해 상대적으로 낮게 나타나고 있어서 이 부분의 예측 불확실성이 상대적으로 낮음을 의미한다. 즉 확률변수로 간주한 미

세먼지 농도값의 높고 낮은 편차가 상대적으로 작다는 것을 의미한다. 모든 달을 비교하였을 때에는 가장 미세먼지 농도값이 낮게 나타나는 8월에 연구지역의 동쪽에서 가장 낮은 조건부 분산값을 나타내고 있다. 이는 전반적으로 낮은 값들이 위치해서 값들의 편차가 작게 나타남을 의미한다. 이러한 조건부 분산값은 공간 정규 크리깅에서는 얻을 수 없는 값에 해당되며, 추정값의 불확실성 해석에 직접적으로 사용될 수 있는 유용한 정보에 해당된다. 예를 들어, 조건부 분산값이 높게 나온 지역은 추정값의 신뢰도가 상대적으로 떨어지는 곳을 의미하므로 추가적인 측정소 설치나 정밀 관측이 필요한 지점 파악에 이용될 수 있다. 여기서 해석상 주의할 점은 조건부 분산값으로 표현한 불확실성 척도는 확률변수로 간주한 잔차 성분 모델링 과정에서 나온 값이라는 것이다. 따라서 이 연구에서 제시한 조건부 분산값은 기상, 토지이용 등의 부가 자료 영향이 포함되지 않은 잔차 성분에 기반을 두고 있기 때문에 측정소 위치와 추정 값의 크기에 의존하게 된다. 따라서 이러한 불확실성 척도를 기상 및 토지이용 등과 같은 부가 자료와 연관지어 해석하려고 한다면, 경향 성분 모델링 과정에 부가 자료를 이용해야한다.

부가적으로 연구지역의 모든 위치에서 ccdf를 알기 때문에, 특정 미세먼지 농도 임계치를 초과할 확률 계산이 가능하다. 인천광역시 대기환경기준에 따르면 (인천광역시 보건환경연구원, 2011), 미세먼지의 기준은 연간 평균치 $50 \mu\text{g}/\text{m}^3$ 이하, 24시간 평균치 $100 \mu\text{g}/\text{m}^3$ 이하이다. 이 사례연구는 월별 미세먼지 농도를 대상으로 하므로, 직접적인 대기환경기준 적용에는 무리가 있다. 그래서 시범적으로 $80 \mu\text{g}/\text{m}^3$ 를 임계치로 설정하여 이 임계치를 초과할 확률을 계산하였다. 비록 시범 임계치를 적용하였으나, 임의의 임계치 설정이 가능하므로 일련의 임계치에 대한 초과 확률 계산이 가능하다.

Fig. 9a는 미세먼지 농도가 상대적으로 높게 나타났던 12월의 미세먼지 농도가 $80 \mu\text{g}/\text{m}^3$ 를 초과할 확률 분포도를 나타내고 있다. 12월의 경우, 북쪽의 대부분 지역에서 확률이 높게 나타나고 있으며, 남서쪽에서도 넓은 지역에 걸쳐 높은 확률값이 퍼져 있음을 확인할 수 있다. 이러한 특정 임계치를 초과할 확률은 Fig. 7의 분포도, Fig. 8의 불확실성 분포도와 함께 미세먼지 위험지역 구분에 유용하게 사용될 수 있는 정보에 해당된다. 예를 들어 Fig. 7의 12월 분

포도를 $80 \mu\text{g}/\text{m}^3$ 를 기준으로 초과지역을 오염지역으로 구분할 경우(Fig. 9b), 80를 초과한 위험 지역에서 대부분 확률이 높게 나타나고 있다. 이러한 분류에 수반되는 위험성은 Fig. 9a의 확률과 결합하여 박노욱(2010b)에 제시한 방법으로 α 위험도와 β 위험도로 표현할 수 있으며, 결과는 Fig. 9c와 d에 제시하였다. α 위험도의 경우에는 확률값이 0에 근접하고 있어 안전한 지역을 위험지역으로 오분류할 확률이 매우 낮음을 의미한다(Fig. 9c). 반대로 위험지역을 안전한 지역으로 오분류하는 β 위험도 역시 경계부근을 제외하고는 매우 낮게 나타나서(Fig. 9d), 12월은 $80 \mu\text{g}/\text{m}^3$ 를 초과 여부에 따른 의사결정에 수반되는 위험도가 낮음을 확인할 수 있었다.

결론

이 연구에서는 기존 공간 영역에서 국소적 불확실성 모델링이 가능한 모수적 다중 가우시안 크리깅의 파라미터 추정시 시공간 개념을 도입한 시공간 다중 가우시안 크리깅을 제안하였으며, 인천 지역에서 3년간 13개 측정소에서 관측된 월별 미세먼지 PM_{10} 자료의 시계열 분포도 작성 사례 연구를 수행하였다.

기존 공간 크리깅과 차별화되는 과정은 결정론적 시계열 경향 성분의 모델링과 잔차의 시공간 베리오그램 모델링에 있다. 이 연구에서는 결정론적 시계열 경향 성분의 모델링에 주기 혹은 경향성과 상관없이 적용할 수 있는 일반적인 방법론으로 고려하고 있는 시간별 연구 지역 전체에서의 관측값의 평균 프로파일을 사용하였다. 잔차 성분의 내삽을 위해 시공간 자기상관성 정보를 정량화하기 위해 곱-합 모델을 적용하였다. 이 과정을 통해 각 위치에서의 ccdf 모델링이 가능하며, 이 ccdf를 이용하여 각 위치에서의 속성값과 더불어 조건부 분산에 의한 불확실성의 확률론적 모델링이 가능하다.

미세먼지 농도 시계열 분포도 작성 사례 연구 결과, 시공간 자기상관성 정보를 반영하였을 때에 기존 공간적 자기상관성 정보만을 이용한 공간 정규 크리깅에 비해 상대적으로 적은 편향과 35-39% 정도의 예측 오차의 감소 효과를 나타내어 시공간 지구통계학이 유용함을 확인하였다. 또한 조건부 분산 지도를 작성함으로써 미세먼지 농도 분포도의 해석과 관련된 부가 정보를 함께 제시하였으며, 확률론적 관점에서 특정 임계치를 초과할 확률을 계산함으로써 추후 위

협성 분석을 위한 보조 자료를 얻을 수 있었다.

서론에서 제시한 최근 지구통계학의 연구 이슈들은 시공간 지구통계학 분석은 상호 결합할 수 있다. 자료 통합 이슈와 관련하여 경향 성분 해석과 최종 분포도 해석에서 제시한 것처럼 광역 기상 인자와 토지이용 현황이 유용한 정보로 사용될 수 있다. 모델링 관점에서 이러한 부가 자료는 시공간 지구통계학적 분석에서 경향 성분 모델링 과정에서 반영되어야 하며, 이 과정에서 다변량 지구통계학 기법의 이용이 가능하다. 또한 두 번째 이슈인 다중 스케일 문제와 관련해서 상대적으로 저해상도인 기상 인자 자료 등에 대해 다중 스케일 지구통계학 기법을 적용함으로써 경향 성분 모델링 과정에 반영할 수 있다.

경향 성분 모델링과 관련하여 이미 전술한 것처럼 기본 성분 함수의 정의에 따라 다양한 알고리즘의 적용이 가능하다. 이 연구에서는 주어진 자료에 기반을 둔 일종의 hindcast 모드의 경향 성분을 이용하였는데, 예측(forecast) 모드로는 활용할 수 없다. 만약 주기성과 경향성을 포함하는 기본 성분 함수를 이용할 경우, 미래로의 투영이 가능하여 예측 모드로 사용이 가능할 것으로 판단된다.

환경 변수는 보통 자동 관측에 의해 얻게 되는데, 측정소별로 자료 획득율이 상이하게 나타나 결측 자료의 수도 달라지게 된다. 이 경우 연구지역의 공간 평균 시간 프로파일을 성분 함수로 사용할 수 없으므로, 주기성과 경향성을 포함하는 기본 성분 함수를 이용함으로써 측정소에서의 결측 자료 예측이 가능할 것으로 보인다. 비록 장기간 관측 자료가 필요하다는 전제조건이 있지만 현재 우리나라에서는 다양한 환경 변수에 대한 장기간 관측 자료의 이용이 가능하므로 보다 광역 지역에서 이러한 주기성과 경향성으로 경향 성분을 모델링하는 사례 연구가 필요할 것으로 보인다. 다만 이 과정에서 절편, 기울기, 진폭 및 위상의 4개 계수들의 상호 상관성을 유지하면서 내삽을 수행해야 하기 때문에 모델링 과정이 다소 복잡해진다.

기존 지구통계학적 불확실성 모델링은 이 연구에서 적용한 모수적 방법론 이외에 지시자 크리깅을 이용한 비모수적 방법론의 적용도 가능하다. 그러나 공간 영역을 시공간 영역으로 확장할 경우 시공간 지시자 크리깅은 각 임계치별로 시공간 베리오그램 모델링을 수행해야 하는 번거로움이 있는 반면 이 논문에서 제안한 시공간 다중 가우시안 크리깅은 계산상 간편

한 장점이 있다. 그러나 정규점수변환이 다중점 확률 분포 함수의 정규 분포를 완전히 보장하지는 않으므로, 앞으로 비모수적 시공간 불확실성 모델링 방법론과의 비교 분석이 필요하다. 또한 국소적 불확실성 모델링과 더불어 공간 불확실성 모델링과 관련해서 확률론적 시뮬레이션 적용 연구도 수행할 예정이다.

감사의 글

이 연구는 국토해양부 첨단도시기술개발사업-지능형국토정보기술혁신사업과제의 연구비지원 (06국토정보B01)에 의해 수행되었으며, 일부는 인하대학교의 지원을 받았습니다. 사례연구 결과 해석을 보완하는데 도움을 주신 익명의 심사위원께 감사드립니다.

참고문헌

- 기상청, 2011, 기상월보 2009년 4월, 8월, 10월, 12월. <http://www.kma.go.kr> (검색일: 2011. 4. 20.)
- 박노옥, 2010a, 다중 스케일 지구통계학을 이용한 원격탐사 자료 기반 주제도의 다운스케일링. 대한원격탐사학회지, 26, 29-38.
- 박노옥, 2010b, 지화학 자료의 확률론적 불확실성 및 위험성 분석을 위한 지시자 지구통계학의 응용. 한국지구과학회지, 31, 301-312.
- 박노옥, 장동호, 2011, 시계열 기온 분포도 작성에서 시공간 자기상관성 정보의 결합. 심사중.
- 신문기, 이충대, 하현섭, 최준석, 김용희, 2007, 기상인자가 미세먼지 농도에 미치는 영향. 한국대기환경학회지, 23, 322-331.
- 오석훈, 2005, 암반등급 해석을 위한 비선형 지시자 변환과 3차원 크리깅 기술의 물리탐사 및 시추자료에 대한 적용. 한국지구과학회지, 26, 429-435.
- 오석훈, 한성미, 2010, 물리탐사 자료의 해상도 향상을 위한 지구통계학적 다운스케일링. 한국지구과학회지, 31, 681-690.
- 인천광역시 보건환경연구원, 2011, 환경정보공개시스템. <http://air.incheon.go.kr> (검색일: 2011. 4. 15.)
- Chilés, J.-P. and Delfiner, P., 1999, Geostatistics: Modeling Spatial Uncertainty. Wiley and Sons, NY, USA, 720 p.
- De Cesare, L., Myers, D.E., and Posa, D., 2001, Estimating and modeling space-time correlation structures. Statistics and Probability Letters, 51, 9-14.
- De Cesare, L., Myers, D.E., and Posa, D., 2002, FORTRAN programs for space-time modeling. Computers and Geosciences, 28, 205-212.
- De Iaco, S., Myers, D.E., and Posa, D., 2002, Space-time variograms and a functional form for total air pollution measurements. Computational Statistics and Data Analy-

- sis, 41, 311-328.
- Deutsch, C.V. and Journel, A.G., 1998, GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, NY, USA, 369 p.
- Diggle, P.J. and Ribeiro Jr, P.J., 2007, Model-Based Geostatistics. Springer, NY, USA, 228 p.
- Dimitrakopoulos, R. and Luo, X., 1994, Spatiotemporal modeling: Covariances and ordinary kriging systems. In Dimitrakopoulos, R. (ed.), Geostatistics for the Next Century. Springer, NY, USA, 88-93.
- Fernandez-Cortes, A., Calaforra, J.M., Jimenez-Espinosa, R., and Sanchez-Martos, F., 2006, Geostatistical spatiotemporal analysis of air temperature as an aid to delineating thermal stability zones in a potential show cave: Implications for environment management. Journal of Environmental Management, 81, 371-383.
- Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation. Oxford University Press, NY, USA, 483 p.
- Goovaerts, P., 2010, Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. Mathematical Geosciences, 42, 535-554.
- Kyriakidis, P.C., 2004, A geostatistical framework for area-to-point spatial interpolation. Geographical Analysis, 36, 259-289.
- Kyriakidis, P.C. and Journel, A.G., 1999, Geostatistical space-time models. Mathematical Geology, 31, 651-684.
- Kyriakidis, P.C. and Journel, A.G., 2001, Stochastic modeling of atmospheric pollution: A spatial time-series framework. Part II: Application to monitoring monthly sulfate deposition over Europe. Atmospheric Environment, 35, 2339-2348.
- Kyriakidis, P.C., Miller, N.L., and Kim, J., 2004, A spatial time series framework for simulating daily precipitation at regional scales. Journal of Hydrology, 297, 236-255.
- Oh, S., 2000, Geostatistical approach to Bayesian inversion of geophysical data. Ph.D. Thesis, Seoul National University, 99 p.
- Park, G., 2008, Geostatistical integration of multi-parametric geophysical data to enhance spatial resolution. Ph.D. Thesis, Seoul National University, 120 p.
- Park, N.-W., 2004, Multi-source spatial data fusion with geostatistical uncertainty assessment: Applications to landslide susceptibility analysis and land-cover classification. Ph.D. Thesis, Seoul National University, 173 p.
- Park, N.-W., Jang, D.-H., and Chi, K.-H., 2009, Integration of IKONOS imagery for geostatistical mapping of sediment grain size at Baramarae beach, Korea. International Journal of Remote Sensing, 30, 5703-5724.
- Spadavecchia, L., 2008, Estimation of landscape carbon budgets: Combining geostatistical and data assimilation approaches. Ph.D. Thesis, University of Edinburgh, 172-227.
- Spadavecchia, L. and Williams, M., 2009, Can spatio-temporal geostatistical models improve high resolution regionalisation of meteorological variables? Agricultural and Forest Meteorology, 149, 1105-1117.

2011년 5월 3일 접수
2011년 5월 30일 수정원고 접수
2011년 5월 31일 채택