

풍속 자료의 공간예측

정승환¹ · 박만식² · 김기환³

¹고려대학교 경제통계학과, ²성신여자대학교 통계학과, ³고려대학교 정보통계학과

(2009년 8월 접수, 2009년 12월 채택)

요약

오래 전부터 지속적으로 제기된 기후변화에 의한 문제들은 전 세계적인 문제점으로 언급되고 있다. 이러한 환경적 위기에 처한 각 나라들의 대처방법 중에 하나는 친환경적이고 지속가능한 발전 설비를 마련하기 위한 노력과 연구를 진행하고 있다는 것이다. 그 중에서 풍력을 이용한 발전은 해외 선진국에서 오래 전부터 개발되어 발전해 오고 있고 우리나라 역시 최근에 풍력 발전에 관심을 갖고 기술개발에 노력을 기울이고 있다. 이러한 실정에서 우리나라 지역의 풍력에 대한 분석 및 예측은 천연자원의 적절한 이용이라는 관점에서 매우 중요한 연구라고 할 수 있겠다. 본 논문에서는 기상청에서 제공하는 풍속 측정 자료로 선형회귀모형에 근간을 둔 추정방법을 이용하여 주요도시지역을 제외한 남한지역의 공간적 특성을 파악할 수 있는 적절한 모형을 찾고 각 모형의 비교를 실시하였다. 이 결과를 바탕으로 남한지역의 풍속 예측지도도 구성하였다.

주요용어: 풍속, 크리깅, 공간적 연관성, 교차검증방법.

1. 서론

화석연료의 고갈로 인한 전 세계적인 에너지 대란을 미연에 방지하고 지속적으로 생산 가능한 대체 에너지를 개발하기 위하여 세계 각국은 풍력 등 친환경 재생 에너지 개발에 많은 투자와 연구를 진행하고 있다. 우리나라는 2006년 말 재생에너지가 총 1차 에너지 소비의 2.2% 수준 (에너지관리공단 신재생 에너지센터, 2007)에 불과하여, 2030년까지 신재생에너지 보급률을 11%로 올린다는 목표를 세우고 있다 (국가경영전략연구원, 2008). 현 정부에서도 목표달성을 위하여 기후변화대응 종합기본계획을 수립하고 친환경 신재생 에너지원 개발에 중점을 두고 있다 (기상청홈페이지).

풍력발전이란 바람의 운동에너지를 풍력발전 시스템에 의해 기계적 에너지로 변환하여 전기를 만들어내는 것을 말한다. 이러한 풍력발전의 특징은 공해가 없고 제한이 없는 바람의 운동에너지를 이용하기 때문에 매우 친환경적이고 국토를 효율적으로 이용할 수 있다는 것이다. 그리고 연료비가 적게 들고 대부분이 무인 원격 시스템이므로 유지비용이 적다는 점, 발전 설비 모양의 특이성으로 인해 해당 지역의 관광산업과 연결될 수 있다는 장점이 있겠다. 그러나 풍력을 이용해서 발전하려면 바람의 속도가 지정된 평균 초속 이상이어야 한다는 제한이 있기 때문에 경제성을 확보할 수 있는 입지가 제한적이라는 점 그리고 설비 이용률이 다른 발전원에 비해 낮다는 단점이 있다. 최근 풍력발전산업이 세계적으로 높은 성장률을 보이고 있지만 아직까지는 전체적인 전력생산 중에서 풍력발전이 차지하고 있는 비율은 매우 낮은 편이다. 그러나 지속적인 기술의 발달로 인해 풍력을 이용한 발전 단가가 많이 하락하여 기타 다른 발전방식과 어느 정도 효율성을 견주어 볼 수 있는 수준으로 근접해가고 있다. 이와 같이 풍력발전

²교신저자: (136-742) 서울특별시 성북구 동선동 3가 249-1, 성신여자대학교 통계학과, 조교수.

E-mail: man.sik.park@gmail.com

에 대한 기술적인 문제와 경제성에 대한 부분이 앞으로 빠르게 개선되는 추세이므로 신재생 에너지 산업 중에서 많은 장점을 갖고 있는 풍력발전에 대한 관심과 연구가 필요한 시점이라 할 수 있겠다.

이에 본 논문에서는 기상청에서 제공하고 있는 제주도 등 주요 도서지역을 제외한 460개 관측지점에서 측정한 2003년부터 2007년까지 5년 평균풍속자료를 분석에 사용하였다. 풍력발전에 활용 가능한 바람고도는 50m~100m인데 반해 기상청에서 기상예보를 위해 관측하는 바람고도는 10m로 우선 연직바람분포에 대한 관계식을 이용하여 10m의 바람을 50m와 80m의 바람으로 보정하였다 (기상청홈페이지). 본 연구에서는 보정된 80m 평균 풍속자료를 이용하여 공간적 연관구조를 가정한 선형회귀모형, 즉 크리깅(kriging)기법을 적용하였다. 크리깅기법을 적용함에 있어서 관측지점의 위치정보 뿐만 아니라 주풍향(main wind direction)을 주요 독립변수로 고려하였다. 모형의 추정 및 적합에는 잔차최대우도추정방법(restricted maximum likelihood estimation; REML)과 최대우도추정방법(maximum likelihood estimation; ML)을 사용하였으며 오차항의 공분산 함수에는 구형모형(spherical model), 가우시안모형(Gaussian model), 지수모형(exponential model) 등을 가정하였다.

최근 크리깅기법을 적용하여 실제 자료를 분석한 많은 연구들이 진행되어 왔다. 허태영 등 (2007)은 관측지점 간의 실제도로 거리를 바탕으로 일일평균 교통량을 예측하는 데 크리깅기법을 사용하였고, 정지용 등 (2008), Heo와 Park (2009)는 강수량 자료에 적용하였다. 김기환 등 (2008)은 크리깅기법으로 얻은 예측값으로 구매력 지도를 작성하였다. 구요천 등 (2005)은 크리깅기법으로 설계공간의 타당성을 높일 수 있는 방법을 제시하였고 반응면 모형(response surface model)과 정확성, 효율성, 강건성 등에 대해 비교, 평가하였다. 김성중과 박형동 (2009)은 광산 시추 조사 결과 및 지형 정보가 저장된 데이터 베이스를 이용하여 크리깅 기법 인터페이스를 제안하였다. 선창국과 정충기 (2006)은 광범위한 지역에 대한 지층 두께 분포와 지층의 동적 물성을 크리깅 방법을 이용하여 예측하였고, 안은자 (2003)는 수치고도 자료로부터 격자 형태의 DEM을 구축하기 위해 크리깅기법을 이용하였다. 이지영 (2001)은 54개 표준지점의 입지적 특성값들 간의 공간적 자기상관관계를 고려하여 조사되지 않은 지역의 특성값을 크리깅방법을 통해 분석하였다. 최광희 등 (2006)은 탄성과 탐사를 통해 확보한 습지퇴적층의 심도를 바탕으로 크리깅기법을 이용하여 화구호를 가상적으로 복원하였으며 최영석 등 (2008)은 무리말뚝으로 시공된 교량기초를 대상으로 하여 기존 예측기법에 의한 개략예측과 새로운 예측기법에 의한 정밀예측 결과를 비교하였는데 이 중 크리깅기법이 가장 우수한 것으로 판단하였다.

본 논문에서는 2장에서 공간분석을 위한 공간통계학의 기본적인 개념들을 소개하였으며 공간적 연관구조를 가정한 선형회귀모형에 대해 간략히 설명하였다. 3장에는 분석에 사용한 자료의 특성들에 대해 소개하고 풍속자료를 이용한 분석결과에 대해 기술하였으며 공간적 연관구조를 가정하지 않은, 즉 전통적 선형회귀모형 하에서의 결과와 비교하였다. 4장에서는 결론 및 향후 연구 과제에 대하여 논하였다.

2. 공간적 선형회귀모형

공간통계학은 자료가 관측되는 지점의 위치정보를 자료를 분석하는 과정에서 하나의 중요한 정보로 고려하는 통계적인 모형 혹은 방법론의 집합체라고 할 수 있다. 관측지점의 위치정보는 흔히 그 지점의 위도(latitude), 경도(longitude), 또는 고도(altitude) 등이며 이 외에 그 지점에서 관측된 다른 변인들을 포함한다. 따라서, 공간통계학은 전통적인 통계적인 모형에 이러한 관측지점의 고유한 정보를 추가적으로 포함시킴으로써 지점들 간의 공간적 연관성을 모형에 반영하고자 하는 하나의 시도에서 출발하였다.

2.1. 자료 형태에 따른 구분

공간통계학의 여러 기법들을 적용하는 자료의 형태는 관측지점의 위치형태에 따라 구분되는데 크게 다

음의 세 가지로 분류된다. 첫째는 지리통계자료(geostatistical data)로서 실제적으로 가장 많이 접하게 되는 자료의 형태이다. 이 자료는 관측지점이 지도 상에 점으로 표현되며 오직 그 점에 대한 정보만을 관측하게 되며 주로 기상관측지점에서 측정하게 되는 대기오염물질의 농도나 강수량, 혹은 특정 지점에서 관측된 특정광물의 매장량 등이 이에 속한다. 이러한 자료를 분석하려는 가장 중요한 목적은 이미 측정된 지점의 정보를 이용하여 관측되지 않은, 혹은 관측하려는 임의의 지점에 대한 신뢰할 만한 정보를 예측함에 있다. 따라서, 분석결과 얻게 되는 주요 결과들 중 하나가 바로 예측지도(prediction map)이고 이를 위해 적용하는 가장 일반적인 방법이 본 연구에서 사용한 공간적 연관구조를 반영한 선형회귀모형의 적합, 즉 크리깅(kriging)기법이다.

두 번째 자료의 형태는 지역자료(areal data) 혹은 격자자료(lattice data)로서, 연구자가 정한 관심지역을 작은 소지역들로 분할하고 분할된 각 소지역 내에서 종합적으로 합산된 정보(합계, 혹은 평균)로 표현된 자료의 형태이다. 이러한 자료는 각 소지역이 지도 상에 점이 아닌 특정한 경계를 가지는 공간으로 나타나게 되고 분할된 소지역들을 합하면 애초에 연구자가 고려한 관심지역이 이루어진다. 지역자료의 예는 서울시 각 동별 특정질환으로 인한 사망자 수, 각 구별 철도이용객수 등을 들 수 있다. 이러한 자료를 분석하고자 하는 주요 목적은 인접한 소지역들 간의 연관성의 강도를 측정하고자 함이다.

마지막으로 공간점 패턴자료(spatial point-pattern data)를 들 수 있다. 앞에서 소개한 두 자료의 형태는 관측지점 혹은 관측지역이 고정인 경우이다. 이는 어떤 관측지점을 이동하거나 행정구역을 변경하지 않는 한 위치정보는 고정된 값으로 받아들여야 한다. 하지만 공간점 패턴자료는 관측지점 자체가 유동적인 경우이다. 국립공원 내에 특정 희귀 동,식물의 관측지점 등을 예로 들 수 있는데 이는 해당 동,식물의 이동행태에 따라 상당히 유동적임을 알 수 있다. 이러한 자료는 결국 관측지점의 위치가 임의(random)적인지, 군락(cluster)을 형성하는지를 밝히려는 데 그 분석목적이 있겠다.

2.2. 크리깅(Kriging)

기상관측지점과 같이 지도상에 점으로 표시되는 자료의 형태인 지리통계자료는 다음과 같이 표현된다.

$$\{Z(\mathbf{s}) : \mathbf{s} \equiv (s_1, \dots, s_d)' \in \mathbb{D} \subset \mathbb{R}^d\},$$

여기서, $Z(\mathbf{s})$ 는 연구자의 관심지역, \mathbb{D} 내의 관측지점, \mathbf{s} 에서 관측한 종속변수이고 \mathbb{R}^d 는 d 차원의 유클리드 공간(Euclidean space)이다. 일반적으로 d 는 일반적으로 1, 2, 3의 값을 가진다. 흔히 위도와 경도만을 고려하는 경우 지리통계자료는 $\{Z(\mathbf{s}_i); \mathbf{s}_i = (s_{i1}, s_{i2})', i = 1, 2, \dots, n\}$ 으로 표현되어진다. 이러한 $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$ 의 확률과정(stochastic process)은 다음과 같은 선형회귀모형으로 표현할 수 있다.

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.1}$$

여기서, $\mathbf{X} = \{\mathbf{x}_1(\mathbf{s}_i), \dots, \mathbf{x}_p(\mathbf{s}_i)\}_{i=1}^n$ 는 $n \times p$ 크기인 독립변수(공변량)들로 이루어진 행렬이고 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ 는 회귀계수들로 이루어진 모수벡터이며 $\boldsymbol{\epsilon} = (\epsilon(\mathbf{s}_1), \dots, \epsilon(\mathbf{s}_n))'$ 는 오차벡터이다. 식 (2.1)에서 $\mathbf{X}\boldsymbol{\beta}$ 는 전체적인 추세나 경향을 나타내는 변동, 혹은 평균함수를 의미하고 $\boldsymbol{\epsilon}$ 는 평균함수를 제외한 비교적 작은 변동을 의미한다. 그리고 $\epsilon(\mathbf{s}_i)$ 는 평균이 0이고 2차 정상성(second-order stationarity)을 만족하는 확률변수로 가정한다. 본 논문에서 고려하는 $\boldsymbol{\epsilon}$ 는 다음과 같은 가정을 만족한다고 하자. 모든 $\mathbf{s}_i, \mathbf{s}_j \in \mathbb{D}$ 에 대하여

- (1) $E[\epsilon(\mathbf{s}_i)] = 0,$
- (2) $\text{Var}(\epsilon(\mathbf{s}_i)) = \sigma^2 < \infty,$

$$(3) \text{Cov}(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)) \equiv C(\|\mathbf{s}_i - \mathbf{s}_j\|),$$

$$(4) \mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega}).$$

여기서, $\boldsymbol{\Omega}$ 는 오차항의 분산-공분산 행렬을 의미하고 $\|\mathbf{s}_i - \mathbf{s}_j\|$ 는 두 지점 \mathbf{s}_i 와 \mathbf{s}_j 사이의 유클리드 거리를 의미한다. 따라서 공분산은 거리만의 함수인 $C(\|\mathbf{s}_i - \mathbf{s}_j\|)$ 형태로 나타내어지고 $\|\mathbf{s}_i - \mathbf{s}_j\| = 0$ 인 경우는 분산, σ^2 으로 표현할 수 있다.

식 (2.1)은 선형회귀모형의 대표적인 표현방법이며 모수적인 접근방법 하에서, 주어진 모형에 공간적인 연관성을 포함시키기 위해 오차항의 분산-공분산 행렬인 $\boldsymbol{\Omega}$ 를 다음과 같이 표현하고자 한다.

$$\boldsymbol{\Omega} = \tau^2 \mathbf{I}_n + \lambda^2 \mathbf{C}_\phi, \quad (2.2)$$

여기서, τ^2 , λ^2 , ϕ 은 공간적 연관구조를 나타내는 모수들이고 \mathbf{I}_n 는 n 차원의 단위행렬(identity matrix)이다. 이들 모수에 대해 간략히 소개하면, τ^2 는 측정오차에 의해 발생하는 분산의 일부분을 나타내는 nugget이고 λ^2 는 순수하게 공간적인 효과에 의해 발생하는 분산의 일부분이다. 그리고 ϕ 보다 큰 거리를 가지는 임의의 두 지점의 상관성이 거의 0이 된다. $\mathbf{C}_\phi = \{c_{ij}; \phi\}_{i,j=1,\dots,n}$ 는 오차항의 순수한 상관행렬로서 이를 추정하기 위해 가정하는 대표적인 모형들은 다음과 같다.

A. 구형모형(spherical model)

$$c_{ij} = \left\{ 1 - \frac{3\|\mathbf{s}_i - \mathbf{s}_j\|}{2\phi} + \frac{\|\mathbf{s}_i - \mathbf{s}_j\|^3}{2\phi^3} \right\} I(\|\mathbf{s}_i - \mathbf{s}_j\| \leq \phi),$$

B. 지수모형(exponential model)

$$c_{ij} = \exp \left[-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\phi} \right],$$

여기서, $I(\cdot)$ 는 지시함수(indicator function)이다. 이들 모형들은 구조적인 차이로 인해 거리가 길수록 줄어드는 상관계수의 경향이 다르지만 인접한 관측지점들 간에는 강한 연관성을, 멀어질수록 약한 연관성을, 혹은 무상관성을 가정한다는 공통점을 지니고 있다. 그리고 거리가 0인 경우 $\sigma^2 = \tau^2 + \lambda^2$ 으로 표현된다. 식 (2.1)로 표현되는 선형회귀모형과 공간적 연관구조를 의미하는 식 (2.2)의 분산-공분산 행렬을 기초로 우도함수(likelihood function)를 최대화하는 추정방법(일반최대우도추정법, 잔차최대우도추정법)들을 고려하고자 한다. 적합된 모형 하에서 임의의 지점에 대한 신뢰할 만한 정보를 예측하는 과정을 고려하게 되는 데 이를 크리깅이라고 한다. 이는 관측되지 않은 지점, \mathbf{s}_0 에서의 예측값은 $\{Z(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)' \boldsymbol{\beta} + \epsilon(\mathbf{s}_0)\}$ 으로 표현되고 이미 관측되어진 지점들, $\{\mathbf{s}_i\}$ 에서 얻어진 $\{Z(\mathbf{s}_i)\}$ 의 선형식으로 표현되어지며 다음과 같다.

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \gamma_i Z(\mathbf{s}_i), \quad (2.3)$$

여기서, 선형 계수 $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)'$ 는 $\boldsymbol{\gamma}' \mathbf{X} = \mathbf{x}(\mathbf{s}_0)'$ 을 만족하여야 한다. 앞에서 소개한 오차항의 가정이 충족되면 식 (2.3)은 최량선형불편성예측량(best linear unbiased predictor; BLUP)이 된다. 최대우도추정방법들을 이용하여 모수들을 추정하게 되면 이러한 선형계수들을 쉽게 얻을 수 있다. 모형의 추정과 크리깅에 대한 보다 자세한 이론은 Cressie (1993), Schabenberger와 Gotway (2005), 최승배 등 (2008) 그리고 최종근 (2008)을 참조하기 바란다.

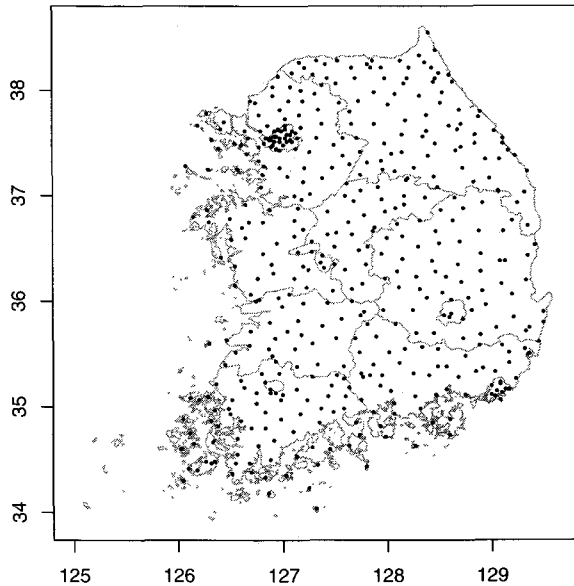


그림 3.1. 평균풍속과 주풍향을 측정하는 460개 기상관측지점

3. 사례연구

본 연구에서는 2장에서 소개한 모형을 실제 사례에 적용한 결과를 소개하고자 한다. 사용한 자료는 최근 친환경 신재생 에너지원으로 각광을 받고 있는 풍력발전에 대한 것으로 화석연료의 고갈에 따른 하나의 대안으로 급부상하고 있다. 사례연구에 사용된 프로그램으로는 ArcView GIS 3.2와 R (R Development Core Team, 2008) 등이다.

3.1. 풍속자료

본 연구에서 사용한 자료는 주요 도서지역을 제외한 남한지역 460개 기상관측지점(그림 3.1)에서 측정된 2003년부터 2007년까지 5년 평균 풍속자료이다. 기상청 홈페이지 (http://www.kma.go.kr/sfc/sfc_06_02_02.jsp)에서 분석에 사용된 풍속자료를 열람할 수 있다. 풍력발전에 활용 가능한 바람고도는 50m~100m로 기상청에서 기상예보를 위해 관측하는 바람고도는 10m이다. 이를 아래의 Deacon의 공식(대기환경연구회, 2003)을 이용하여 10m에서의 풍속을 50m와 80m 높이에서의 풍속으로 변환하였고 보정된 자료 중에서 80m 하에서의 평균풍속을 본 연구에서는 사용하였다.

$$Z_s = Y_s \left(\frac{c}{10} \right)^{k_s},$$

여기서, c 는 50m(혹은 80m)이고 Y_s 는 10m 높이에서의 실제풍속이며 Z_s 는 고도 c 에서의 보정된 풍속을 의미한다. 그리고 k_s 는 다음과 같이 표현된다.

$$k_s = u + v \log(Y_s), \quad u = \frac{1}{\log(j_s/j_0)} + \frac{0.088}{1 - 0.088 \log(j_s/10)}, \quad v = -\frac{0.088}{1 - 0.088 \log(j_s/10)},$$

여기서, j_s 는 지점 s 에서의 기하학적 평균고도이고 j_0 는 주변 지형의 거칠기 고도를 의미한다(이와 관련한 보다 자세한 내용은 기상청 자료관리서비스팀(2007)을 참조하기 바란다). 이와 더불어 각 기상

관측지점의 기본적인 위치정보인 위도와 경도를 독립변수에 포함시켰으며 각 지점의 고도도 포함시키기 위해 Google Earth를 이용하였다. 위도와 경도는 지구의 곡면을 보정된 위치정보를 이용하였고 위도, 경도, 고도의 단위는 킬로미터(km)이다. 또한 분석의 최종결과인 풍속의 예측지도(prediction map)를 구성하기 위해 일정한 간격을 지닌 935개 정사각형 모양의 임의 격자지점들을 고려하였고 이들 지점들에 대한 위치정보 중 고도 역시 Google Earth를 통해 구하였다. 80m에서의 5년간 평균풍속 자료를 확인할 결과 정규성이 만족되지 않았다. 따라서 Box-Cox 변환을 통해 80m에서의 5년간 평균풍속값에 로그변환을 취하는 것이 가장 합당하다고 판단하였다.

3.2. 모형적합

공간적 연관구조를 의미하는 오차항의 분산-공분산 행렬을 앞에서 2장에서 소개한 몇 가지 공분산모형들로 추정하기 전에, 먼저 평균함수 부분을 설정하고자 한다. 먼저 로그변환된 풍속과 경도(s_{i1})의 산점도(그림 3.2(a))에서 알 수 있듯이 경도에 의한 2차형의 모형식을 고려할 수 있다. 위도(s_{i2})와의 관련성 역시 그림 3.2(b)를 통해 2차형의 모형식으로 표현되어 질 수 있음을 알 수 있다. 그리고 풍속과 고도(s_{i3})와의 관련성에 대해서는 약하긴 하지만 대략적으로 선형적인 관련성을 띤다고 판단된다(그림 3.2(c)). 이러한 위치정보들과 풍속과의 관련성을 통해 해안과 멀리 떨어져 있는 내륙지역이 평균적으로 풍속이 작고 고지대일수록 평균적으로 풍속이 크다는 것을 알 수 있다. 이를 바탕으로 다음과 같은 평균함수의 모형식을 고려하고자 한다.

$$\mathbf{x}'_i\boldsymbol{\beta} = \beta_1 + \mathbf{s}_{i1}\beta_2 + \mathbf{s}_{i2}\beta_3 + \mathbf{s}_{i1}^2\beta_4 + \mathbf{s}_{i2}^2\beta_5 + \mathbf{s}_{i1}\mathbf{s}_{i2}\beta_6 + \mathbf{s}_{i3}\beta_7. \quad (3.1)$$

그림 3.1에 나타난 460개 기상관측지점에서는 평균풍속뿐만 아니라 주풍향(main wind direction)의 정보 또한 관측하였다. 각 지점에서의 주풍향은 8개 방위(북, 북동, 동, 남동, 남, 남서, 서, 북서) 중 하나로 나타냈으며 구체적인 풍향의 각도로 측정되지는 않았다(남(S)풍이 주풍향인 지점은 없었다). 이에 주풍향에 따라 평균 풍속의 차이를 살펴본 결과 그림 3.2(d)에 나타난 바와 같이 북동(NE)풍과 북서(NW)풍인 주풍향인 관측지점에서의 평균풍속이 다른 지점들에 비해 어느 정도 크다는 것을 알 수 있다. 따라서, 주풍향에 의한 평균풍속의 차이를 평균함수 부분에 포함시키기 위해 가변수 처리하였다. 따라서 최종적인 평균함수로 식 (3.1)에 북(N)풍을 참조로 한 가변수를 추가한 형태를 고려하였고 다음과 같이 표현된다.

$$\begin{aligned} \mathbf{x}'_i\boldsymbol{\beta} = & \beta_1 + \mathbf{s}_{i1}\beta_2 + \mathbf{s}_{i2}\beta_3 + \mathbf{s}_{i1}^2\beta_4 + \mathbf{s}_{i2}^2\beta_5 + \mathbf{s}_{i1}\mathbf{s}_{i2}\beta_6 + \mathbf{s}_{i3}\beta_7 \\ & + I_i(\text{NE})\beta_8 + I_i(\text{E})\beta_9 + \cdots + I_i(\text{W})\beta_{12} + I_i(\text{NW})\beta_{13}, \end{aligned}$$

여기서, $I_i(\text{NE})$ 는 지시함수로서 s_i 의 주풍향이 북동(NE)풍이면 1, 아니면 0의 값을 갖는다.

결과적으로 평균풍속이 남한 내의 소지역(예를 들면 도)들 간의 차이가 비정상성을 나타낼 가능성을 어느 정도 방지하기 위해 해발고도와 주풍향 등을 모형에 추가로 포함시켰다. 공간자료는 시계열자료와 유사한 성질을 가지고 있다. 공간자료는 시계열자료와 같이 특정 시점(지점)에서의 관측값이 인접한 시점(지점)의 관측값과 연관되어져 있고 시점(지점)들 간의 간격이 커질수록 연관성의 강도가 약화된다. 하지만 이러한 자료탐색이 가능하려면 먼저 시계열자료의 경우 적절한 차분이나 변수변환을 통해 비정상적인 자료를 정상화하는 과정이 선행되어야 한다. 이와 유사한 방법으로 공간자료의 경우 정규성을 만족하기 위해서 적절한 변환과 비정상성(혹은 국소적 정상성)을 유발하는 잠재인자인 여러 가능한 독립변수들을 모형에 포함시키게 된다. 물론 이러한 접근방법 이외에도 초기부터 국소적 정상성을 가정한 모형들을 고려할 수 있지만 이 논문에서는 잠재적인 독립변수들을 모형에 포함한 경우만을 고려하였다.

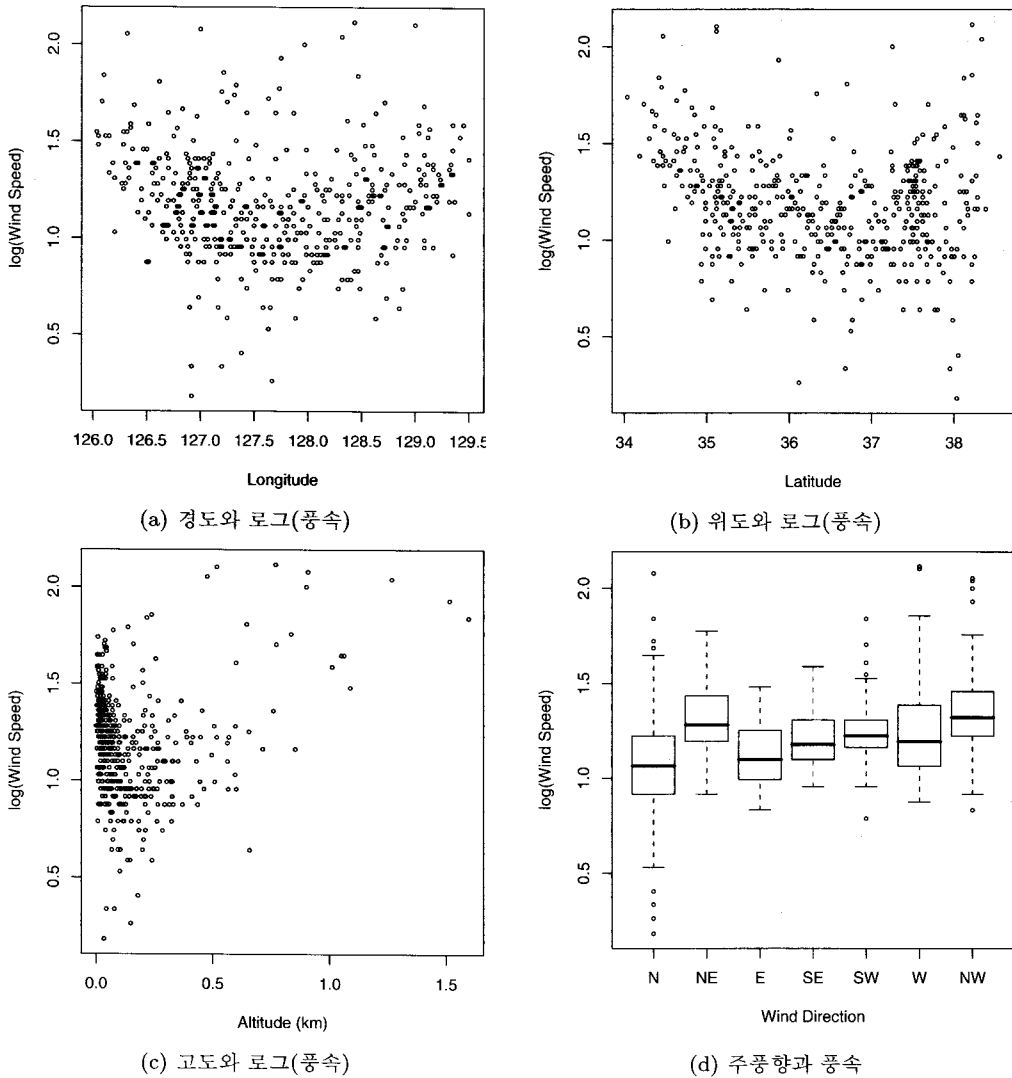


그림 3.2. 평균풍속과 위치정보(지리좌표)와 주풍향과의 산점도 및 상자그림

공분산함수를 적합하기 위해 본 연구에서는 유클리드 거리를 고려하고자 한다. 물론, 유클리드 거리 이외에 다른 종류의 거리도 적용이 가능하리라 생각된다. 우선 지점들 간의 실제거리를 고려해 볼 수 있을 것이다. 예를 들어, s_i 와 s_j 그리고 $s_{i'}$ 와 $s_{j'}$ 간의 유클리드 거리가 같다고 하더라도 s_i 와 s_j 사이에는 풍속의 변화를 줄 만한 장애지형(산맥)이 없고 $s_{i'}$ 와 $s_{j'}$ 사이에는 풍속을 감소시킬 만한 장애지형이 존재한다면 이를 고려한 거리를 생각해 볼 수 있을 것이다.

공간적 연관구조를 가진 선형회귀모형(식 (2.1)와 (2.2)를 참조)에 포함된 모수($\beta, \tau^2, \lambda^2, \phi$)들을 추정 한 결과는 표 3.1에 나타나 있다. 공간적 연관구조를 의미하는 분산-공분산 행렬을 고려한 경우 우도함수를 최대화하는 모수추정방법을 적용하였다. 다시 말해서, 잔차최대우도추정(REML)방법 하에서는 구형모형보다 지수모형을 공분산함수로 고려한 모형이 가장 큰 우도값을 보였고 일반최대우도추

표 3.1. 모수 추정값과 95% 신뢰구간 ($a^b \equiv a \times 10^{-b}$)

	전통적 선형회귀모형			공간적 선형회귀모형					
	Least Squares			REML(Exponential)			ML(Spherical)		
	추정값	95% 신뢰구간		추정값	95% 신뢰구간		추정값	95% 신뢰구간	
β_1	8.19 ¹	7.76 ¹	8.62 ¹	7.51 ¹	-4.74 ¹	1.98 ⁰	8.00 ¹	7.36 ¹	8.65 ¹
β_2	-2.78 ⁴	-5.40 ⁴	-1.63 ⁵	6.57 ⁴	-1.61 ³	2.92 ³	-2.10 ⁴	-6.20 ⁴	2.00 ⁴
β_3	-5.51 ⁴	-7.23 ⁴	-3.79 ⁴	-2.16 ⁴	-2.20 ³	1.77 ³	-6.09 ⁴	-8.83 ⁴	-3.34 ⁴
β_4	1.71 ⁵	1.38 ⁵	2.04 ⁵	2.12 ⁵	9.32 ⁶	3.32 ⁵	1.81 ⁵	1.33 ⁵	2.30 ⁵
β_5	5.54 ⁶	4.06 ⁶	7.02 ⁶	9.37 ⁶	1.80 ⁶	1.70 ⁵	6.67 ⁶	4.41 ⁶	8.94 ⁶
β_6	-2.90 ⁶	-5.16 ⁶	-6.34 ⁷	-3.82 ⁷	-1.09 ⁵	1.01 ⁵	-3.11 ⁶	-6.67 ⁶	4.59 ⁷
β_7	4.43 ¹	3.42 ¹	5.44 ¹	5.43 ¹	4.43 ¹	6.43 ¹	5.16 ¹	4.17 ¹	6.15 ¹
β_8	2.18 ¹	1.41 ¹	2.95 ¹	1.85 ¹	1.12 ¹	2.57 ¹	1.86 ¹	1.14 ¹	2.58 ¹
β_9	1.11 ¹	-1.27 ⁴	2.22 ¹	1.10 ¹	8.84 ³	2.12 ¹	1.07 ¹	6.25 ³	2.09 ¹
β_{10}	1.60 ¹	7.05 ²	2.50 ¹	1.53 ¹	7.02 ²	2.36 ¹	1.61 ¹	7.80 ²	2.44 ¹
β_{11}	1.58 ¹	8.55 ²	2.31 ¹	1.08 ¹	3.97 ²	1.76 ¹	1.15 ¹	4.71 ²	1.83 ¹
β_{12}	2.15 ¹	1.58 ¹	2.73 ¹	1.70 ¹	1.15 ¹	2.24 ¹	1.80 ¹	1.26 ¹	2.34 ¹
β_{13}	1.98 ¹	1.40 ¹	2.56 ¹	1.60 ¹	1.05 ¹	2.15 ¹	1.69 ¹	1.14 ¹	2.23 ¹
τ^2				2.86 ²			2.80 ²		
λ^2				4.15 ¹			1.27 ²		
ϕ				1284			57.6		

정(ML)방법 하에서는 지수모형보다 구형모형을 고려한 경우가 가장 큰 우도값을 보였다. 그리고 전통적 선형회귀모형만을 고려한 경우 최소제곱(least squares)방법을 적용하였다. 최소제곱방법과 잔차최대우도추정방법 그리고 일반최대우도추정방법들 간의 모수추정값에 대한 큰 차이는 보이지 않고 있다. 특히, 주풍향의 관점에서는 추정방법에 상관없이 북동(NE)풍과 서(W)풍 그리고 북서(NW)풍인 경우 다른 방향들에 비해 높은 평균풍속을 보이고 있는 것으로 나타났다. 오차항의 분산-공분산 행렬에 포함된 모수들을 살펴보면 τ^2 의 추정값은 지수모형을 고려한 잔차최대우도추정방법과 구형모형을 고려한 일반최대우도추정방법이 거의 유사한 값으로 추정되었으나 λ^2 과 ϕ 는 두 추정방법들 간에 다소 상이한 결과를 보이고 있다. 특히, 상관성이 거의 0이 되는 거리를 나타내는 ϕ 는 잔차최대우도추정방법 하에서는 실제 관측지점들 간의 최대거리를 훨씬 상회하는 값이 추정되었고 이는 그만큼 연관성의 강도가 꽤 먼 거리까지도 유지된다고 해석될 수 있겠다. 이와는 대조적으로 일반최대우도추정방법 하에서 ϕ 는 비교적 작은 값으로 추정되었다.

표 3.1에 나타난 우도함수를 기반으로 한 두 추정방법(REML, ML)에 의해 적합된 모형들을 근간으로 풍속에 대한 예측지도를 구성하고자 한다. 이를 위해, 먼저 이 논문에서 고려한 임의의 지점은 일정한 간격이 유지된 935개 격자지점을 고려하였다. 그리고 각 지점의 고도는 앞에서 언급한 바와 같이 Google Earth를 이용하였다. 한편, 임의의 격자지점에서의 주풍향은 그 지점에서 관측되지 않았으므로 정확히 알 수 없다. 이를 해결하기 위한 한 가지 대안으로 각 격자지점의 주풍향을 가장 가까운 거리에 위치한 실제관측지점의 주풍향으로 가정하였다. 그림 3.3(a)에서 알 수 있듯이 강원도 일대의 태백산맥과 지리산 등의 소백산맥 지역이 다른 지역에 비해 고도가 높게 나타났고 해안지역은 상당히 낮은 고도를 지닌 것으로 나타났다. 주풍향은 행정구역에 상관없이 북풍(북쪽에서 남쪽으로 부는 바람)이 가장 많이 관측되었을 것으로 가정하였다(그림 3.3(b) 참조).

임의의 격자지점의 추가적인 정보를 이용하여 각 추정방법 하에서의 적합된 모형으로 예측한 결과는 그림 3.4와 같다. 그림 3.4에서의 풍속의 예측값은 원 측정단위로 변환된 것이다. 각기 다른 공분산모형을 고려한 두 추정방법 하에서 적합된 모형들을 기반으로 구성된 80m에서의 평균풍속에 대한 예측지도

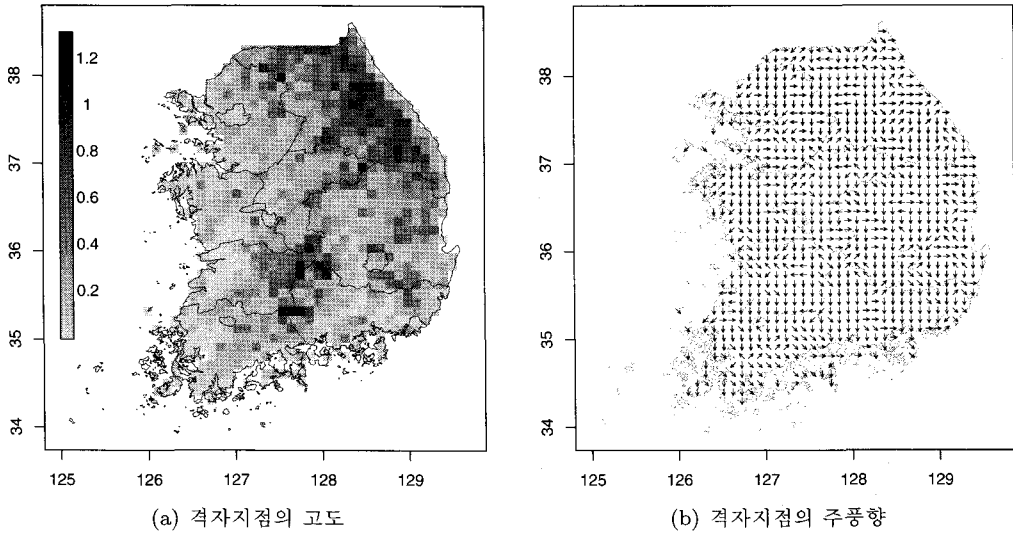


그림 3.3. 예측지도를 구성하기 위한 추가적인 정보

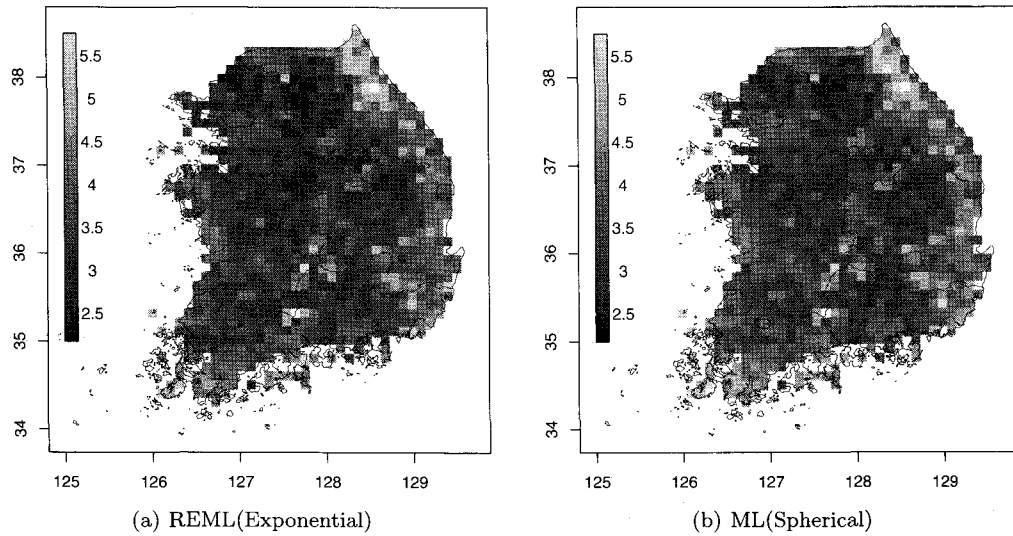


그림 3.4. 적합된 모형들에 기반한 풍속의 예측지도

는 거의 유사하지만 구형모형을 고려한 일반최대우도추정방법 하에서의 적합모형이 다소 높은 예측값의 경향을 보이고 있다. 예측지도를 좀더 자세히 살펴보면, 경기도 동부지역에서부터 태백산맥 서쪽 그리고 충청북도, 경상북도 북서지역에 이르는 내륙지역의 평균풍속이 비교적 낮게 예측되었고 설악산과 오대산이 자리잡고 있는 태백산맥 동쪽지역과 동해안, 경상남북도 동쪽지역, 남서해안지역 그리고 지리산 일대가 비교적 큰 풍속을 보이고 있는 것으로 예측되었다. 한 가지 특이할 만한 점은 전라북도, 경상남도, 경상북도, 충청북도의 접경지대에 위치한 덕유산과 가야산일대는 해안지역과 떨어져 있는, 다소 내륙지역에 위치해 있으나 고도의 영향으로 인해 풍속이 높게 예측되었다는 것이다.

마지막으로 주어진 자료를 분석함에 있어서 공간적 연관구조를 모형에 포함하는 것이 얼마나 신뢰성 있

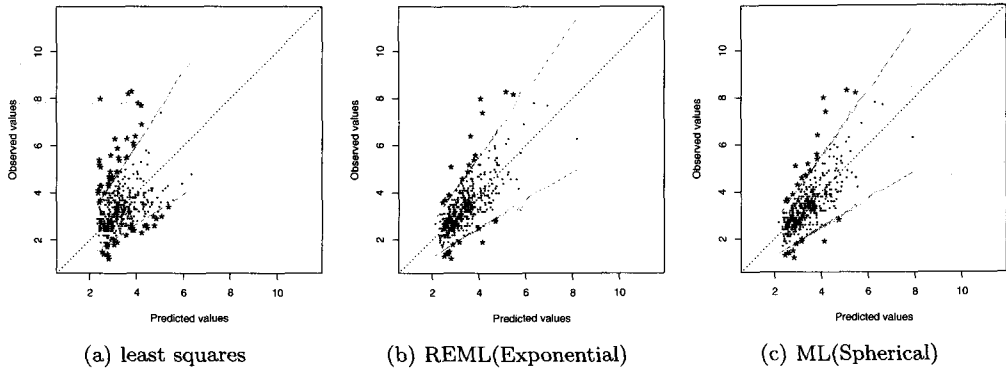


그림 3.5. 교차검증방법(CV(1))에 의한 예측값과 실제값의 산점도 및 95% 예측구간

는 결과를 도출하는지를 알아보았다. 이를 위해 공간적 연관구조를 모형에 포함한 경우와 그렇지 않은 경우를 각각 고려하여 교차검증방법(one-leave-out cross-validation; CV(1))을 적용하여 각 지점에 대한 예측구간을 구하였으며 실제 풍속값이 해당 예측구간에 얼마나 많이 포함되어지는지를 비교 및 평가하였다. 그림 3.5를 통해 알 수 있듯이 오차항을 서로 독립이며 동일한 분산을 가지는, 즉 공간적 연관성을 배제한 전통적 선형회귀모형 하에서는 실제값과 예측값의 일치성이 전혀 유지되지 못할 뿐만 아니라 어떠한 관련성도 보이지 않을 것이다. 또한 교차검증방법에 의해 계산된 95% 예측구간 내에 포함된 지점이 전체 자료의 약 87%인 것으로 나타났다. 따라서 모형의 타당성이 심각한 수준임을 알 수 있다. 이와는 대조적으로 공간적 연관구조를 모형에 포함한 경우, 즉 공간적 선형회귀모형 하에서는 추정방법에 상관없이 실제값과 예측값의 일치성이 어느 정도 유지되고 있으며 전체 지점의 약 5%만이 교차검증방법에 의한 95% 예측구간 밖에 놓이는 것으로 나타났다. 물론, 예측값이 실제값에 비해 다소 작게 산출된 경향을 보이지만 미미한 정도인 것으로 판단된다. 결과적으로 풍속자료를 분석함에 있어서 공간적 연관구조를 포함하는 선형회귀모형이 그렇지 않은 경우에 비해 훨씬 타당하고 월등히 우수한 예측결과를 보이고 있음을 알 수 있다.

4. 결론

본 연구에서는 풍속에 영향을 미치는 요인들로는 관측지점의 위치정보(위도, 경도, 고도)와 주풍향을 고려하였다. 이를 바탕으로 최대우도추정방법을 이용하여 모형을 적합하였으며 적합된 모형을 토대로 풍속의 예측지도도를 구성하였다. 공간적 연관구조의 중요성을 알아보기 위해 교차검증방법을 사용하였으며 이를 통한 예측값과 실제값 사이의 일치성 정도 및 예측구간 내의 포함 여부 등에 주안점을 두었다. 공간적 연관구조를 포함하지 않은 경우와 비교한 결과 공간적 선형회귀모형이 훨씬 신뢰성있는 예측성능을 보였다. 물론 실제값에 비해 평균 풍속이 다소 낮게 예측되어진 경향이 있다. 이는 고도 정보가 보다 정확하게 측정되고 주풍향이 정확한 각도로 산출되어 진다면 과소추정의 경향은 해소되리라 판단된다.

본 연구는 기상청 산하 전국 460여 지점에서 관측한 5년 평균 80m 고도의 평균풍속을 이용하여 모수적인 접근방법인, 공간적 공분산함수를 고려한 선형회귀모형 하에서 공간 예측한 내용을 다루었다. 공간 예측기법인 크리깅은 이미 널리 알려져 있고 R과 SAS 프로그램에서 잘 구현되어져 있다. 하지만, 다른 논문들과 차별화될 수 있는 점을 든다면 공간적 예측력을 향상시키기 위해 해발고도와 주풍향이라는 독립인자들을 고려하였다는 것이다. 물론, 해발고도의 경우 국립지리정보원에서 유료로 제공하는 실측

지도에 의한 것이 아니라 Google Earth라는 프로그램으로 일일이 구한 값이므로 정확한 고도와는 다소 차이가 있을 수 있다. 또한 예측지점에서의 주 풍향 역시 실제로 관측되어지지 않아서 가장 가까운 실측 지점의 정보를 이용할 수 밖에 없었다. 그럼에도 불구하고 정확한 이 두 정보가 실제로 주어졌을 때를 가정한 상태에서 적용할 수 있는 모형을 이 논문에서는 고려하였다.

참고문헌

- 구요천, 전용희, 김유신, 이동호 (2005). 반응면 기법과 크리깅 기법을 이용한 설계공간의 타당성 향상, <한국항공우주학회지>, **33**, 32-38.
- 국가경영전략연구원 (2008). <제1차 국가에너지 기본계획>, 국가경영전략연구원.
- 기상청 자료관리서비스팀 (2007). <풍력자원지도 개발 연구보고서>, 기상청.
- 기상청 홈페이지. <http://www.kma.go.kr>.
- 김기환, 윤연상, 정승환 (2008). 대형 활인점 매출 데이터를 이용한 구매력 지도 작성에 관한 연구, <한국자료분석학회지>, **10**, 223-234.
- 김성중, 박형동 (2009). 베리오그램을 이용한 광채 평가 및 3차원 모델링 소프트웨어 개발, <한국지구시스템공학회지>, **46**, 151-159.
- 대기환경연구회 (2003). <대기환경개론>, 신광문화사, 179-181.
- 선창국, 정충기 (2006). GIS를 이용한 지반-지식 기반 지반 정보화 시스템 구축 기법의 개발 및 적용, <한국지반공학회논문집>, **22**, 55-68.
- 안은자 (2003). <공간통계 분석을 이용한 DEM 오차 패턴 연구>, 경희대학교, 석사학위논문.
- 에너지관리공단 신재생에너지센터 (2007). <2006년 신재생에너지 통계>, 에너지관리공단.
- 이지영 (2001). <GIS의 공간통계분석을 활용한 지가분석>, 경희대학교, 석사학위논문.
- 정지용, 진서훈, 박만식 (2008). 공간적 상관구조를 포함하는 선형회귀모형을 이용한 강수량 자료 분석, <응용통계연구>, **21**, 1093-1107.
- 최광희, 윤광성, 김종욱 (2006). 공간통계기법을 이용한 하논화산의 화구호 복원, <대한지리학회지>, **41**, 391-403.
- 최승배, 문승호, 강찬원, 조장식, 이정형 (2008). <SAS/STAT를 이용한 공간예측>, 자유아카데미.
- 최영석, 임형준, 장학성 (2008). 다양한 예측기법을 이용한 현장타설말뚝의 최적길이 산정, <유신기술회보>, **15**, 22-31.
- 최종근 (2008). <지구통계학>, 시그마프레스.
- 허태영, 박만식, 엄진기, 오주삼 (2007). 최단경로 기반 교통량 공간 예측에 관한 연구, <응용통계연구>, **20**, 459-473.
- Cressie, N. A. C. (1993). *Statistics For Spatial Data*, John Wiley & Sons, New York.
- Heo, T. Y. and Park, M. S. (2009). Bayesian Spatial Modeling of Precipitation Data, *The Korean Journal of Applied Statistics*, **22**, 425-433.
- R Development Core Team (2008). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, <http://www.R-project.org>.
- Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods For Spatial Data Analysis*, Chapman & Hall/CRC, Boca Raton.

Spatial Prediction of Wind Speed Data

Seung Hwan Jeong¹ · Man Sik Park² · Kee Whan Kim³

¹Department of Economics and Statistics, Korea University

²Department of Statistics, Sungshin Women's University

³Department of Information and Statistics, Korea University

(Received August 2009; accepted December 2009)

Abstract

In this paper, we introduce the linear regression model taking the parametric spatial association structure into account and employ it to five-year averaged wind speed data measured at 460 meteorological monitoring stations in South Korea. From the prediction map obtained by the model with spatial association parameters, we can see that inland area has smaller wind speed than coastal regions. When comparing the spatial linear regression model with classical one by using one-leave-out cross-validation, the former outperforms the latter in terms of similarity between the observations and the corresponding predictions and coverage rate of 95% prediction intervals.

Keywords: Wind speed, kriging, spatial association, one-leave-out cross-validation.

²Corresponding author: Assistant professor, Department of Statistics, Sungshin Women's University 249-1 Dongseon-Dong 3-Ga, Seongbuk-Gu, Seoul 136-742, Korea. E-mail: mansikpark@sungshin.ac.kr