

서울특별시 고농도 미세먼지 자료의 다중척도 모델링

2015년 11월 4일

서울대학교 통계학과 박사과정

박선철

서울대학교 통계학과 교수

오희석

목차

1. 서론

2. 연구 자료

3. 분석 방법

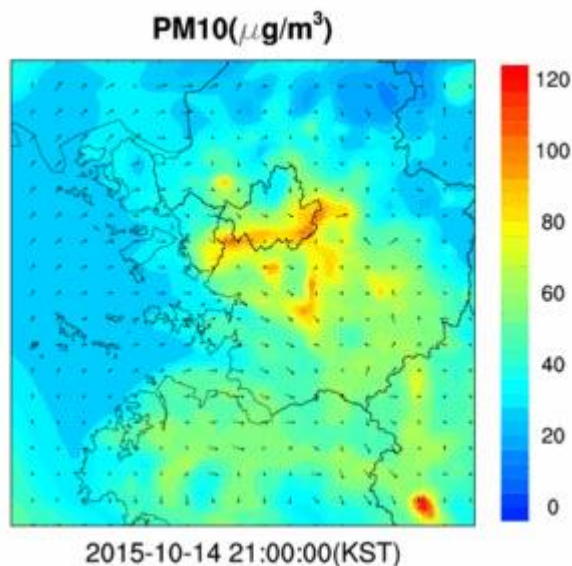
4. 분석 결과

5. 결론

왜 고농도, 다중척도인가

고농도 미세먼지에 대한 관심 증가

미세먼지 농도가 높을수록 인체에 유해



기상현상들은 다양한 지속시간을 가짐
(ex. 소나기와 장마)

고농도, 다중척도 미세먼지 모델링에 대한
선행 연구 부재

본 연구는 고농도 미세먼지가 극단 강수량처럼 장소 및 사건에 따라 각기 **다른 지속 시간 범위를** 가질 것이라는 생각에서 출발

스케일 성질과 극단값 이론을 이용해 지속 시간에 따른 고농도 미세먼지의 변화를 분석하고자 함

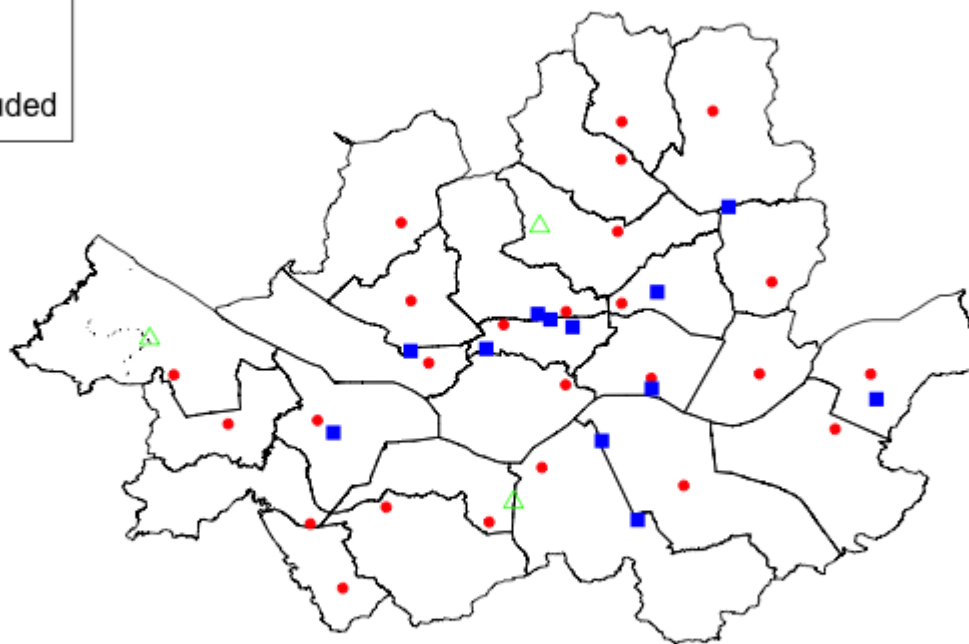
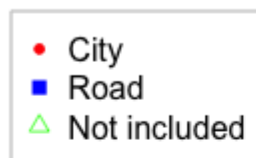
연구 자료

공간적 범위

서울특별시 대기오염 측정소 37곳 자료 활용

도시측정소 : 25곳

도로변측정소 : 12곳



시간적 범위

2010년 1월 1일 01:00 부터

2014년 12월 31일 24:00까지

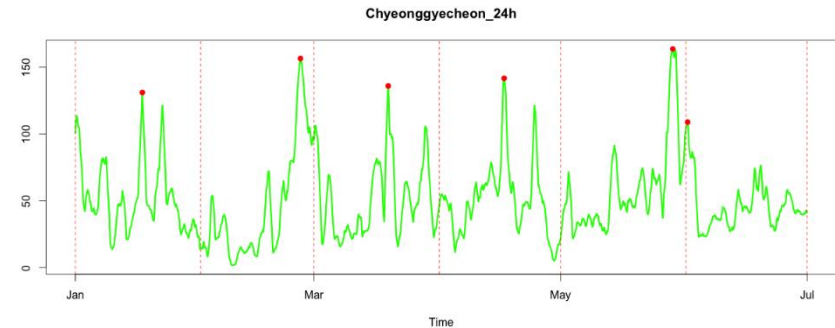
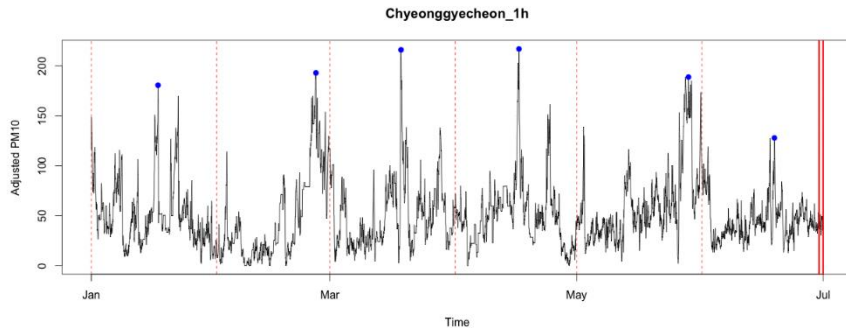
시간대별 미세먼지(PM10) 측정 자료

자료 출처

국립환경과학원

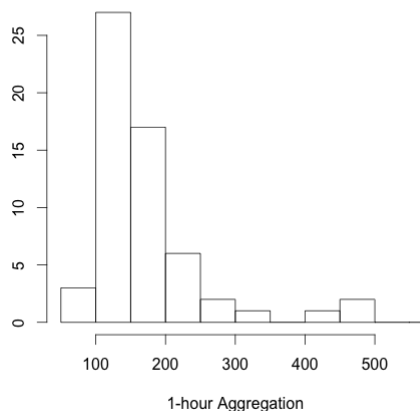
시간 집적 자료 생성

목표: 월 최대 미세먼지 농도 분포 추정과 복귀수준(분위수) 지도 작성



1시간 단위 미세먼지 자료를 이용해 임의 단위 시간 집적 미세먼지 자료 생성

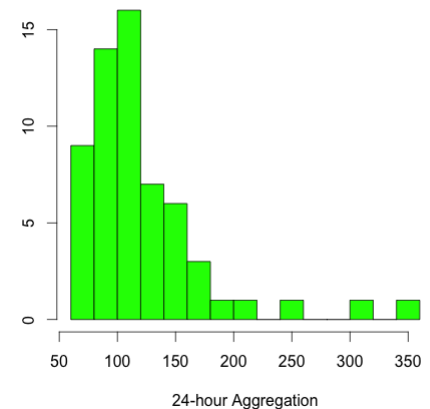
Chyeonggye_1h



Q1. 단위 시간별 월 최대 미세먼지 농도 분포를 어떻게 모델링 할 것인가?

Q2. 단위 시간의 변화와 월 최대 미세먼지 농도 분포의 모양은 어떤 관계를 갖는가?

Chyeonggye_24h

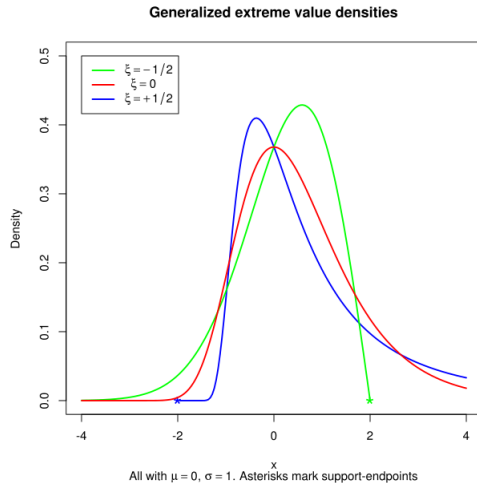


일반화 극단값 분포

일반화 극단값 분포 (generalized extreme value distribution, GEV)는 극단값 자료를 분석하는 데 사용하는 분포

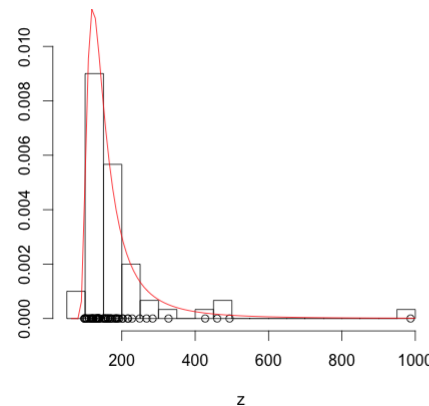
$$f(x; \mu, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} \left(1 + \left(\frac{x - \mu}{\sigma} \right) \xi \right)^{-1 - \frac{1}{\xi}} e^{-\left(1 + \left(\frac{x - \mu}{\sigma} \right) \xi \right)^{-\frac{1}{\xi}}}, & \text{if } \xi \neq 0 \\ \frac{1}{\sigma} e^{-\left[\frac{(x - \mu)}{\sigma} (\xi + 1) + e^{-\frac{(x - \mu)}{\sigma}} \right]}, & \text{if } \xi = 0 \end{cases}$$

(μ : 위치모수, σ : 척도모수, ξ : 형태모수)

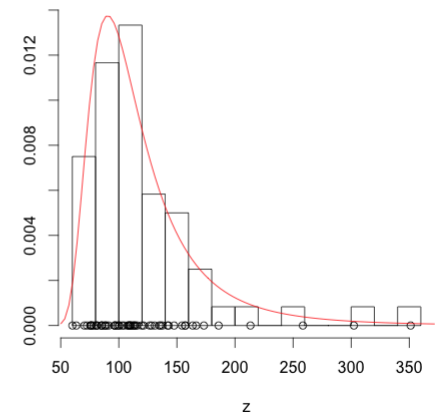


각 장소별로 단위시간을 고려한 월 최대 미세먼지 농도 분포는 GEV 분포를 이용해 모델링 가능

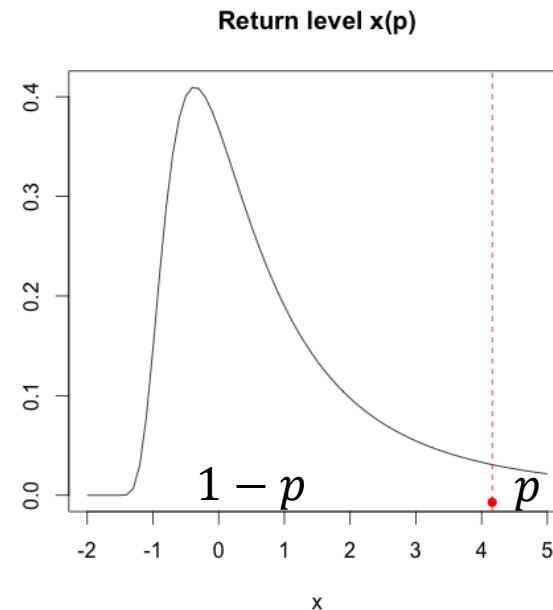
Chyeonggye_1h



Chyeonggye_24h



일반화 극단값 분포



1/p달 GEV 복귀수준(return level)은 GEV 관찰값이 평균적으로 매 1/p달마다 초과할 것으로 기대되는 값

$$x_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}], & \text{if } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1 - p)\}, & \text{if } \xi = 0 \end{cases}$$

$p = 1/12$ 로 놓음으로써 월 최대 미세먼지 농도가 일 년에 한 번 꼴로 초과할 것으로 기대되는 값을 계산할 수 있음

스케일 성질

Gupta and Waymire (1990): 극단 강수량 자료 분석에서 d_{ref} 의 복귀수준과 d 의 복귀수준 사이에 스케일 인자 $\left(\frac{d}{d_{ref}}\right)$ 와 스케일 지수(v)를 이용한 선형관계가 있음을 밝힘

$$x_{p,d} = \left(\frac{d}{d_{ref}}\right)^{-v} x_{p,d_{ref}}$$

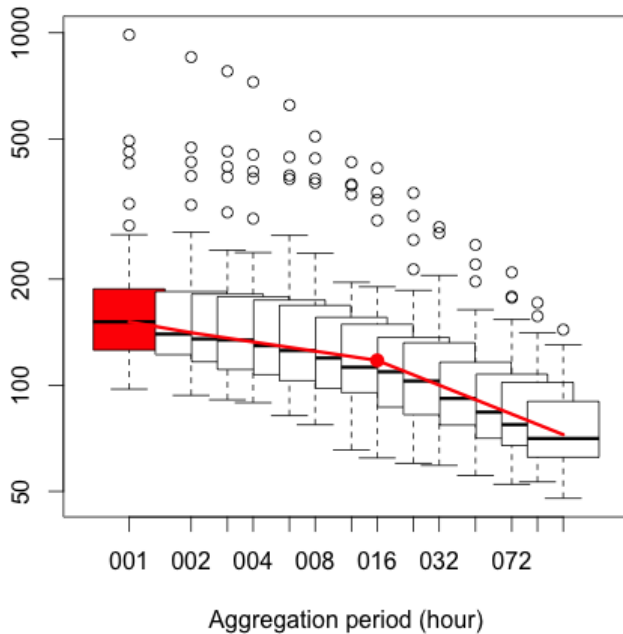
스케일 성질을 이용하면 d_{ref} 의 GEV 모수와 스케일 인자, 스케일 지수를 가지고 장소 x 에서 d 시점의 GEV 모수를 추정 가능

$$\mu_d(x) = \mu_{d_{ref}}(x) \times \left(\frac{d}{d_{ref}}\right)^{-v(x)}, \sigma_d(x) = \sigma_{d_{ref}}(x) \times \left(\frac{d}{d_{ref}}\right)^{-v(x)}, \xi(x) = \text{상수}$$

스케일 지수의 추정은 $\log(\text{지속기간 비})$ 와 $\log(\text{미세먼지 농도비})$ 의 관계를 나타낸 강도-지속시간-빈도 곡선(intensity-duration-frequency curve)의 기울기로 추정

스케일 성질

Chyeonggyecheon



문제점

그러나 서울특별시 미세먼지 농도 자료에서는

- IDF 곡선의 기울기가 선형이 아니었고
- 지속시간의 변화에 따라 GEV 분포 꼬리의 길이와 두께가 달라져

GEV 복귀 수준 모델링의 정확성이 떨어짐

해결책

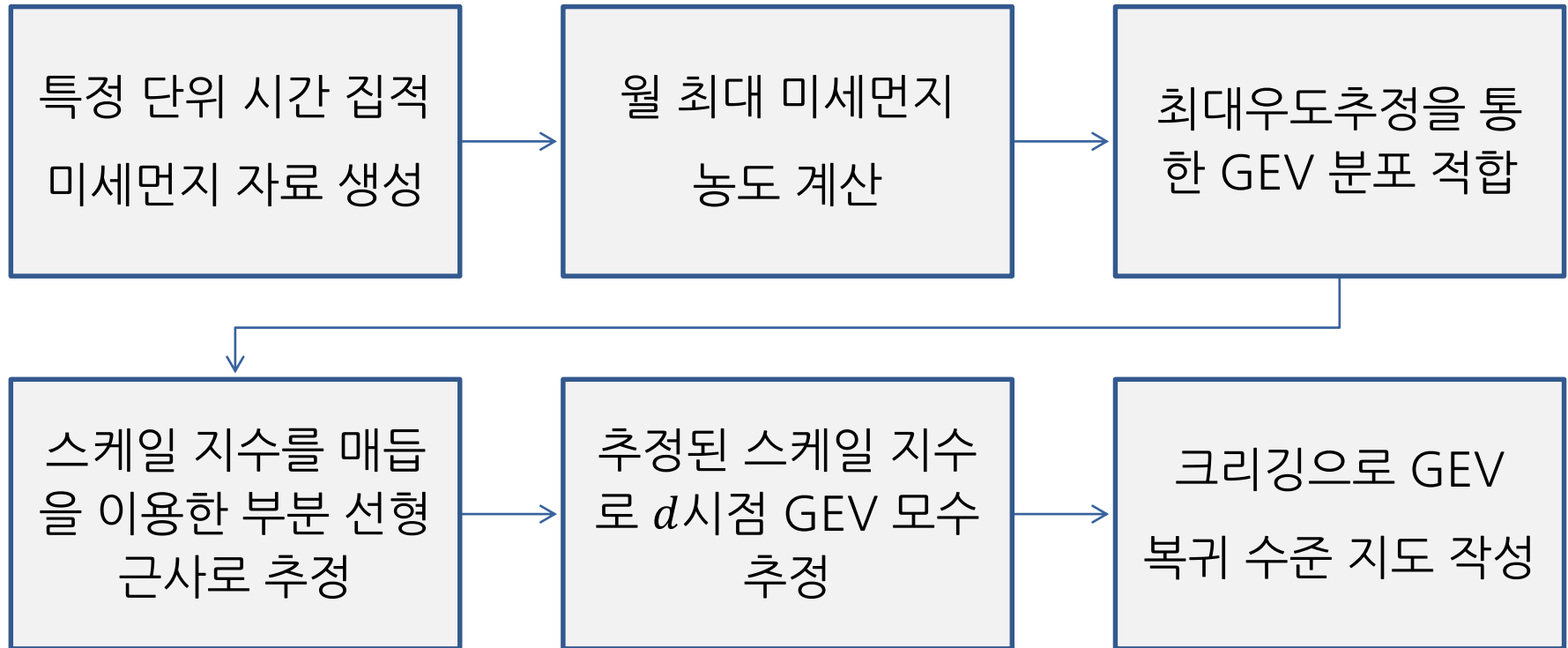
- IDF 곡선의 기울기를 매듭을 이용한 두 개의 직선으로 근사하고,
- 추가적인 스케일 지수 모수를 사용해 GEV 모수를 추정하였음

$$\mu_d(x) = \mu_{d_{ref}}(x) \times \left(\frac{d}{d_{ref}} \right)^{-(v_{\mu_1}(x)I(d \leq 16\text{hour}) + v_{\mu_2}(x)I(d > 16\text{hour}))},$$

$$\sigma_d(x) = \sigma_{d_{ref}}(x) \times \left(\frac{d}{d_{ref}} \right)^{-(v_{\sigma_1}(x)I(d \leq 16\text{hour}) + v_{\sigma_2}(x)I(d > 16\text{hour}))},$$

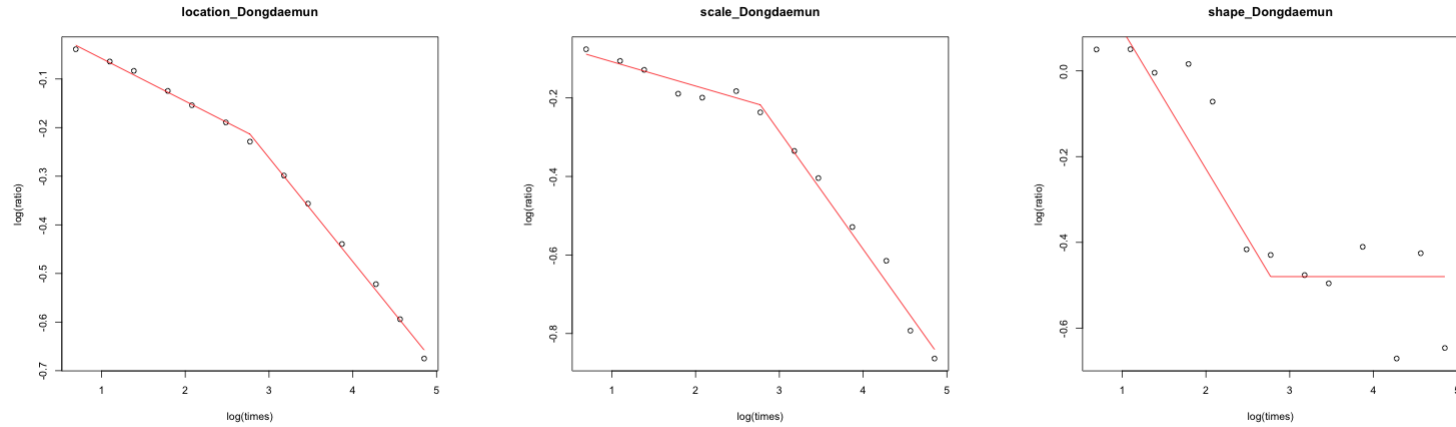
$$\xi_d(x) = \xi_{d_{ref}}(x) \times \left(\frac{d}{d_{ref}} \right)^{-(v_{\xi_1}(x)I(d \leq 16\text{hour}) + v_{\xi_2}(x)I(d > 16\text{hour}))}.$$

분석 프로세스

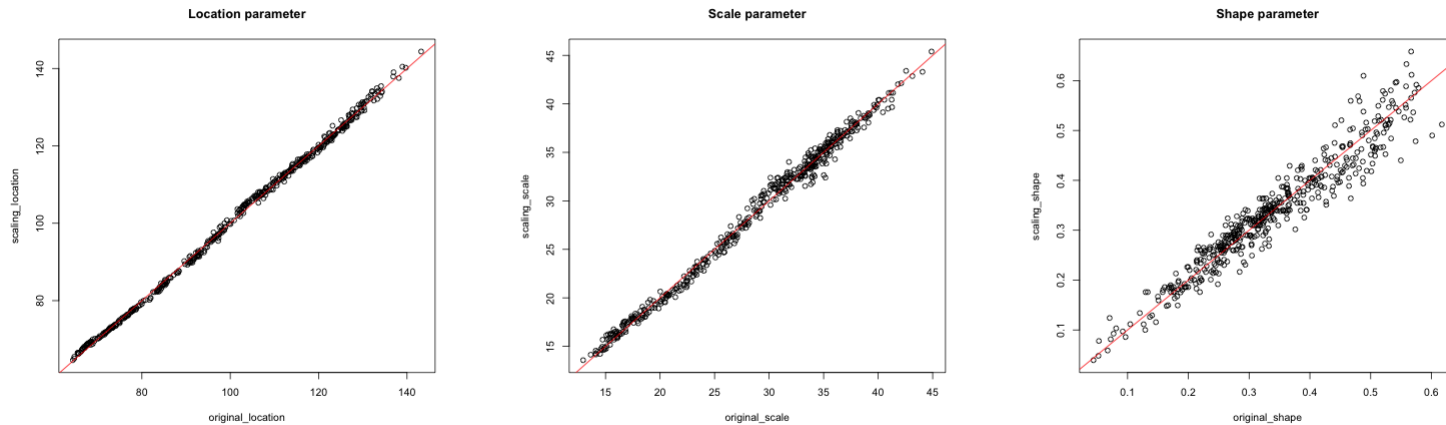


- ① 단위 시간별 GEV 모수를 직접 계산한 경우의 GEV 복귀 수준 지도와
- ② 매듭을 이용한 변형된 스케일 방법의 GEV 복귀 수준 지도를 비교

결과: 스케일 인자 추정



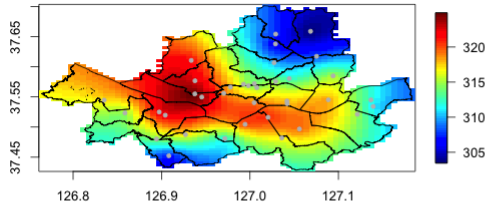
동대문 측정소의 위치모수, 척도모수, 형태모수의 스케일 인자 추정



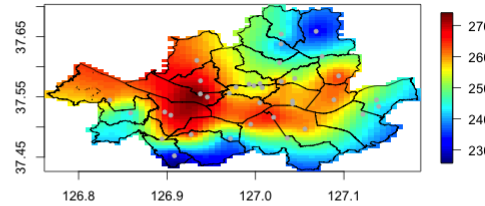
두 방법으로 구한 위치모수($\hat{\mu}_d$), 척도모수($\hat{\sigma}_d$), 형태모수($\hat{\xi}_d$)의 상관관계 그림

결과: 복귀 수준 지도

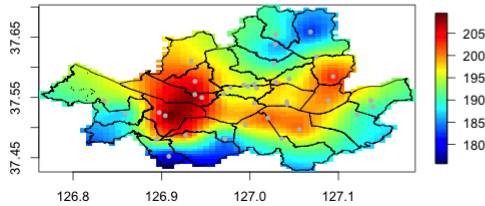
Return Level_1h



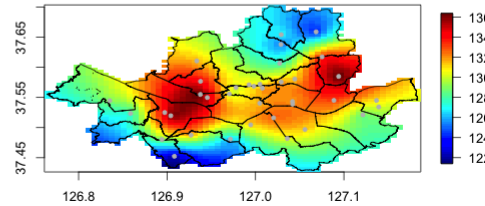
Return Level_8h



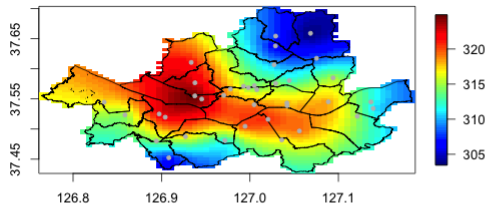
Return Level_24h



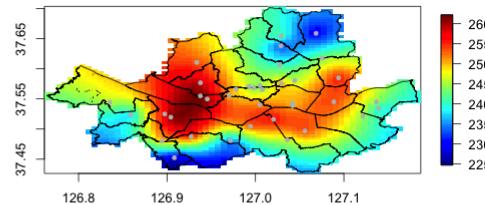
Return Level_96h



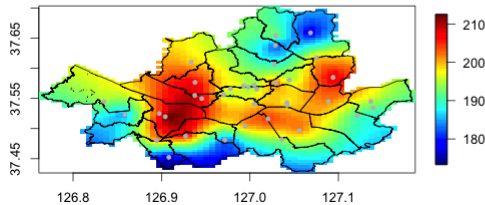
Return Level_1h



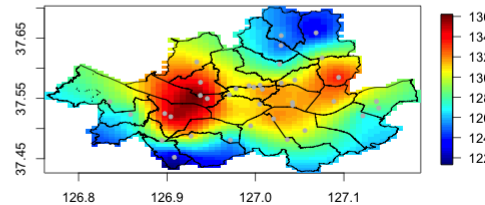
Return Level_8h_Scaling



Return Level_24h_Scaling



Return Level_96h_Scaling

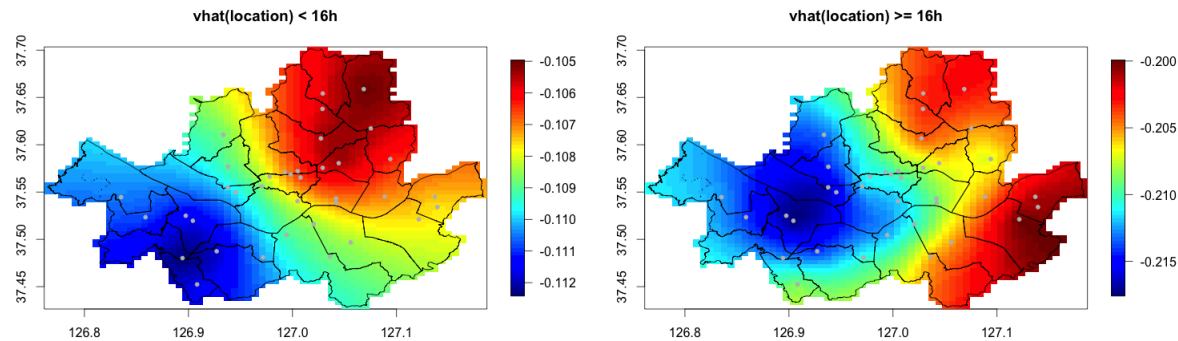


방법 ①의 GEV 복귀수준(위)과
방법 ②의 GEV 복귀수준(아래)

- 지속시간 1시간, 8시간, 24시간, 96시간의 미세먼지 12달 복귀수준 지도
- 여의도 부근의 미세먼지 복귀수준이 제일 높았으며, 중랑구 지역은 지속시간이 길어질수록 미세먼지 복귀수준이 상대적으로 높아짐
- 두 방법으로 계산한 지도가 큰 차이를 보이지 않음 → 매듭을 이용한 변형된 스케일 방법의 GEV 지도가 원 지도를 잘 묘사하고 있음

결과: 복귀 수준 지도

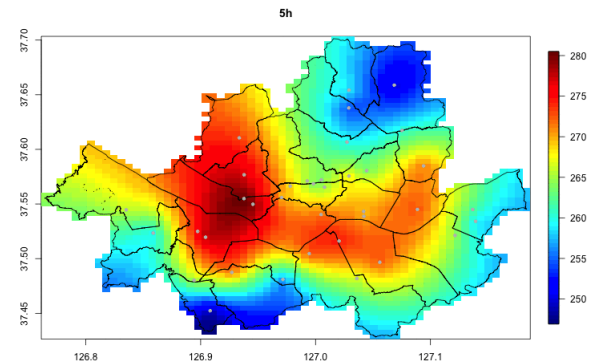
GEV 복귀 수준 변화에 가장 큰 영향을 미치는 $-\hat{v}_{\mu_1}, -\hat{v}_{\mu_2}$ 의 공간 분포 그림



빨간색 지역: 지속시간이 길어질수록 GEV 복귀 수준이 상대적으로 **적게** 감소

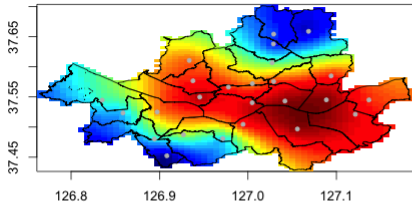
파란색 지역: 지속시간이 길어질수록 GEV 복귀 수준이 상대적으로 **많이** 감소

변형된 스케일 성질을 이용해 GEV 모델링을 할 경우 임의의 지속시간의 GEV 복귀수준 지도 또한 작성 가능

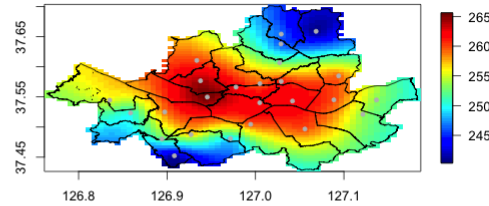


결과: 도시대기측정소만 활용시

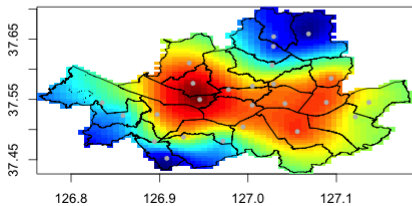
Return Level_1h_City



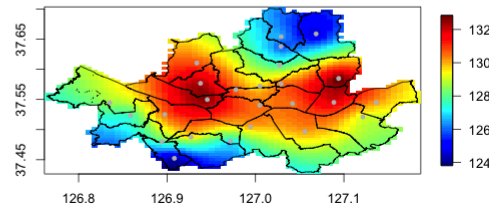
Return Level_8h_City



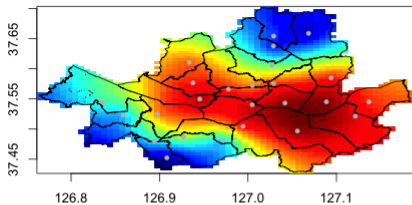
Return Level_24h_City



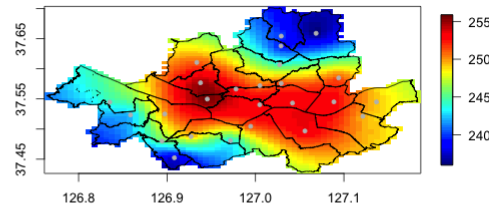
Return Level_96h_City



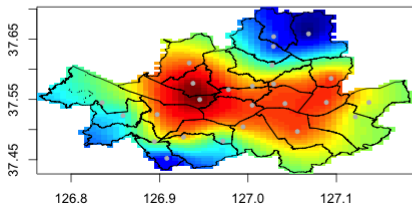
Return Level_1h_City



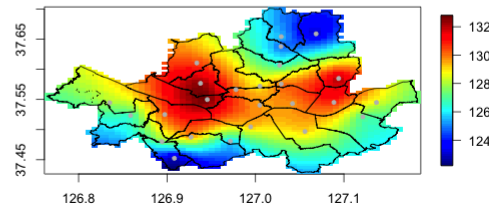
Return Level_8h_City_Scaling



Return Level_24h_City_Scaling



Return Level_96h_City_Scaling



전체적인 트렌드는 비슷하나

- 지속시간이 짧은 경우와
 - 서울 남동부 지역
- 에서 차이를 보였음

측정소	1시간 복귀수준	24시간 복귀수준
강변북로	298.4077	195.3969
성동구	331.3642	194.3296
양재로	307.1686	192.6147
강남구	335.1758	204.9765

모든 측정소의 정보를 활용하면서
도시/도로변 대기측정소의 특성을
반영한 다중척도 모델링 방법 필요

방법 ①의 GEV 복귀수준(위)과
방법 ②의 GEV 복귀수준(아래)

결론 및 시사점

결론

서울특별시 월 최대 미세먼지 분포는 GEV로 모델링 가능

각기 다른 지속시간을 갖는 GEV 모수 사이에는 부분 선형 관계 성립

변형된 스케일 성질을 이용한 모델링으로 임의의 지속시간의 GEV 복귀 수준 지도 작성 가능

시사점

지역 맞춤형 미세먼지 농도 관리 필요

고농도 미세먼지 예보시 지속시간에 대한 예보 필요

측정소 특징을 고려한 다중 척도 모델링 필요

감사합니다
