

# 1 지역 간 PM10 농도의 분위 상관계수와 평률 상관계수간의 비교

## 1.1 연구자료

연구자료는 기상청 기상자료개방포털(<https://data.kma.go.kr/cmmn/main.do>)에서 제공하고 있는 미세먼지 농도( $\mu\text{g}/\text{m}^3$ ) 자료를 이용하였다. 이 자료는 부유분진 측정장비(PM10)를 이용하여 미세먼지 등 대기 중에 부유하는 에어로졸 중 직경이  $10\mu\text{m}$  이하인 입자의 농도를 관측한 자료로, 황사 관측의 참고자료로 활용되고 있다. 우리의 연구에서는 기상청 기상자료 개방포털에서 제공하고 있는 PM10 농도를 2009/01/01부터 2022/12/31까지 총 15개의 지점(서울, 강화, 대관령, 수원, 천안, 추풍령, 대구, 울산, 구덕산, 진주, 안동, 진도, 광주, 전주, 고산)에 대해서 일별로 다운받았다. 대전광역시를 제외한 15개의 시도를 대표하는 지점들이며, 각 시도별 지점이 2개 이상인 강원 지역의 경우 임의로 한 지점(대관령)을 선택하였다. 결측치는 모두 제거하였다. 이후 각 관측지점에 대해서 PM10 농도의 평균값을 구하고, 주기성을 제거해 주기 위하여 PM10 값에 로그를 취해준 뒤 앞서 구한 평균값을 각각의 로그 PM10의 값에서 빼주었다. 전처리된 데이터 값을 서울을 기준으로  $\rho_{\tau}^{(q)}$  와  $\rho_{\tau}^{(e)}$  값을 각각 R의 library(quantreg)와 직접 작성된 expecreg 함수를 이용하여 계산하였다.

## 1.2 분위 상관계수의 해석

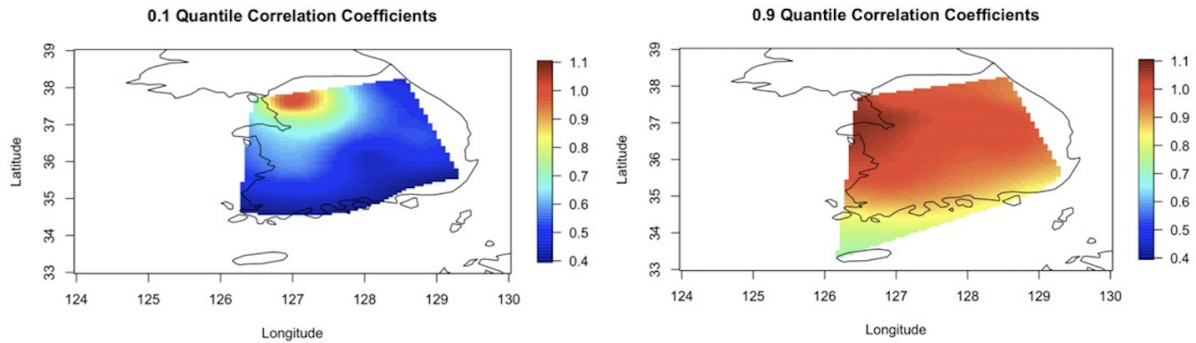


Figure 1: The sample quantile correlation coefficients  $\rho_{\tau}^{(q)}$

Figure1은  $\tau = 0.1$ ,  $\tau = 0.9$  인 상황에서 서울과 나머지 14개 지점에 대해서  $\rho_{\tau}^{(e)}$  값을 공간적으로 시각화한 그림이다. 우선 왼쪽 상황  $\tau = 0.1$ 에서는 분위수가 낮은 미세먼지의 값들, 즉 미세먼지의 농도가 낮은 경우의  $\rho_{\tau}^{(q)}$ 를 확인할 수 있다. 이는 서울과 가까운 지역들인 경기지역과 강원지역 등이 서울의 미세먼지가 “좋음”을 나타냈다면, 마찬가지로 미세먼지가 “좋음”을 나타낸 경우가 많았을 것이라는 해석을 가능하게 한다. 그러나 서울과의 거리가 먼 제주 지역, 부산 지역들의  $\rho_{\tau}^{(q)}$ 은 0.5 이하의 값을 주로 갖는다. 따라서 서울의 미세먼지의 경향성을 제주, 부산 지역들은 따라가지 않는 경우가 꽤 있었을 것이라고 해석할 수 있다. 반면 오른쪽 상황  $\tau = 0.9$ 에서는 분위수가 높은 미세먼지 값들, 즉 미세먼지의 농도가 높은 경우의  $\rho_{\tau}^{(q)}$ 를 확인할 수 있는데, 대체적으로 왼쪽 상황에 비해서 값이 (0.9 이상으로) 큰 것을 확인할 수 있다. 따라서 미세먼지가 나쁜 날에는 서울과 가까운 지역 외에도 거의 모든 지역들이 “미세먼지 나쁨” 지수를 나타냈을 것이라고 해석할 수 있다. 또한 서울과의 거리가 멀어질수록 상관계수의 값이 낮아지는 경향도 유지된다는 것이 보여진다.

## 1.3 평률 상관계수의 해석

Figure2은 앞선 상황과 동일한 상황에서  $\rho_{\tau}^{(e)}$  값에 대해서도 공간시각화를 해본 그림이다.,  $\rho_{\tau}^{(e)}$ 의 양상은 Figure1에서의  $\rho_{\tau}^{(q)}$ 과 전체적인 양상과 매우 유사한 것을 볼 수 있다. 이는 우리의 Expectile Correlation Coefficient가 Quantile correlation coefficients의 분위수에 관한 해석을 그대로 이어받을 수 있고 그 성질 또한 유사할 것이라는 해석을 가능하게 한다.

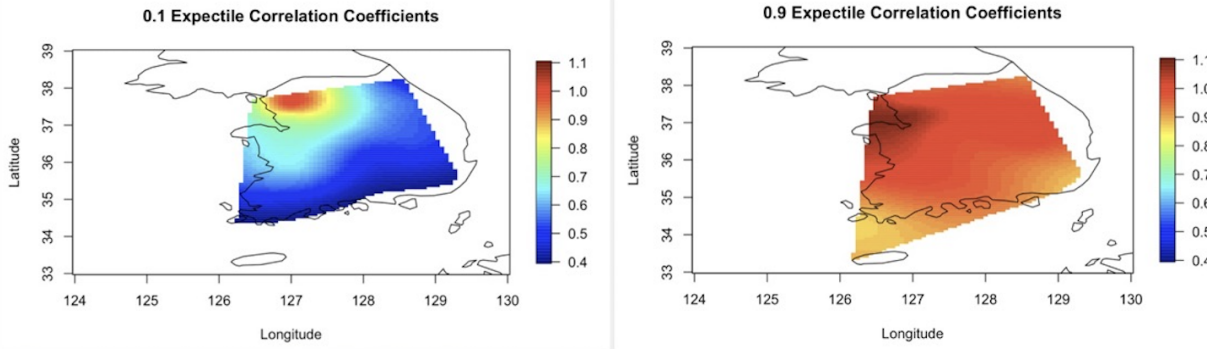


Figure 2: The sample Expectile correlation coefficients  $\rho_{0.1}^{(e)}$

#### 1.4 평률 상관계수와 분위 상관계수의 비교

Figure3은 서울과의 거리에 따라  $\rho_{\tau}^{(e)}$  값과  $\rho_{\tau}^{(q)}$  값을 비교한 그림이다. 두 경우 모두 서울과의 거리가 멀어질수록 상관계수의 값은 낮아지는 양상이다.  $\tau = 0.9$  인 상황이  $\tau = 0.1$  인 상황에 비해 상관계수가 높고, 1보다도 높은 값이 나오기도 한다. 또한  $\rho_{\tau}^{(e)}$ 의 값이  $\rho_{\tau}^{(q)}$  값보다 조금 더 큰 값이 많이 나타나고 있는 양상이다.

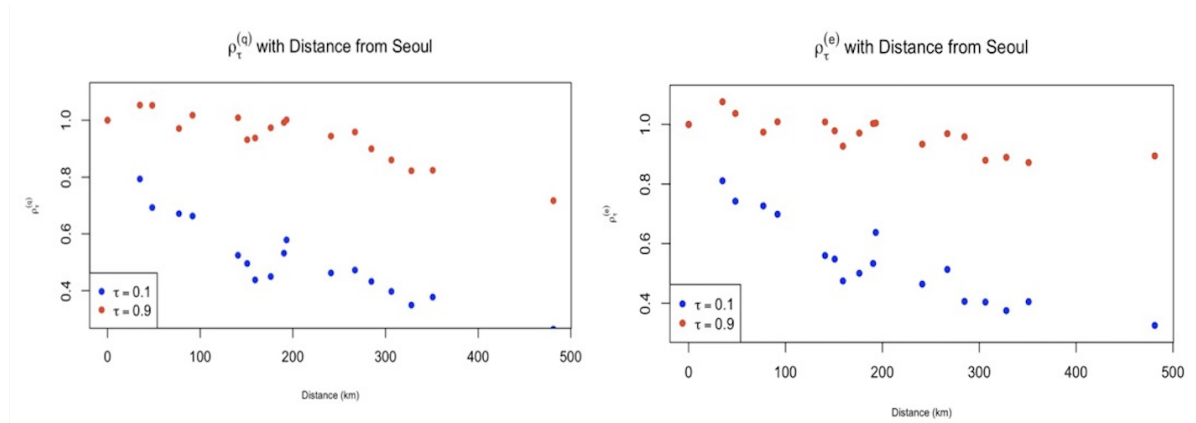
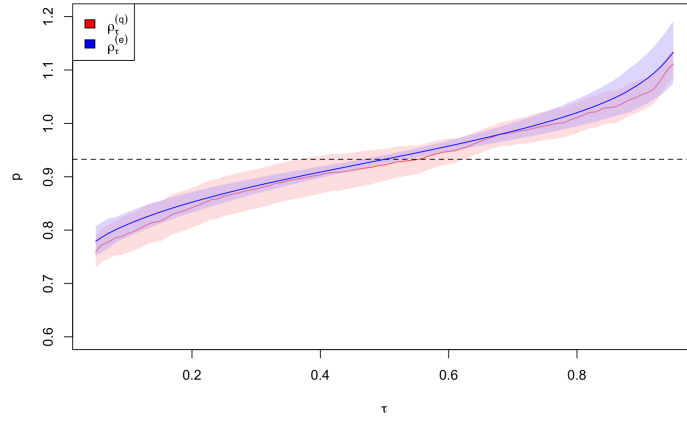


Figure 3: Correlation Coefficient with Distance from Seoul

#### 1.5 평률 상관계수와 분위 상관계수의 신뢰구간 길이 비교

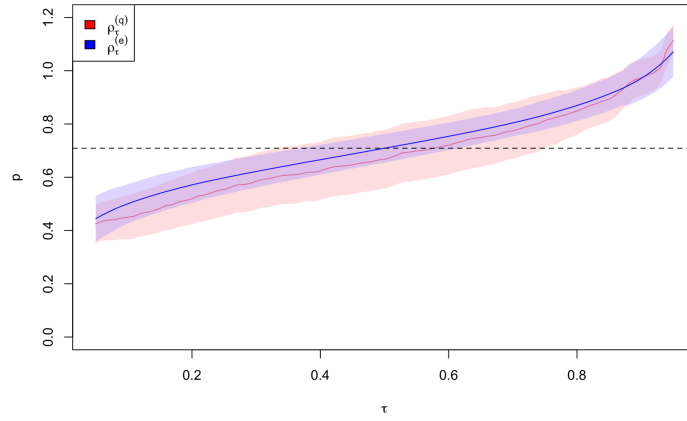
아래 그림은  $\rho_{\tau}^{(q)}$  과  $\rho_{\tau}^{(e)}$  값을  $\tau = 0.05, 0.06, \dots, 0.95$ 에 따라 추정한 값과 95% 신뢰구간을 시각화 한 그림이다.  $\rho_{\tau}^{(q)}$ 의 표준편차의 추정과정에서 조건부밀도함수의 추정방법은  $f_{BH}$ 를 사용하였다.  $\rho_{\tau}^{(q)}$ 의 표준편차를 구하는 과정에는 조건부 밀도함수 추정 과정이 필요하기 때문에 신뢰구간을 구하는 과정이  $\rho_{\tau}^{(e)}$ 에 비해 조금 더 까다롭고, 시간이 오래 걸린다. 반면  $\rho_{\tau}^{(e)}$ 은 계산이 간편하고 신뢰구간의 길이가 짧은 것을 확인 할 수 있다.

$\rho_{\tau}^{(a)}$  and  $\rho_{\tau}^{(e)}$  95% pointwise confidence interval (CI) between Seoul and Gyeonggido



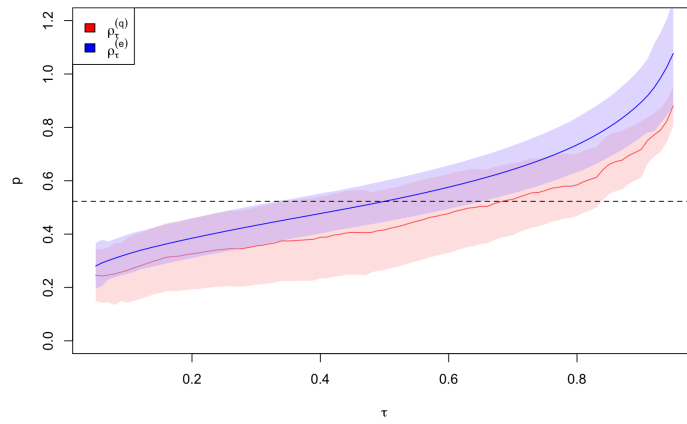
(a) Confidence intervals for Gyeonggi Province

$\rho_{\tau}^{(a)}$  and  $\rho_{\tau}^{(e)}$  95% pointwise confidence interval (CI) between Seoul and Chungcheongbukdo



(b) Confidence intervals for Chungcheongbukdo

$\rho_{\tau}^{(a)}$  and  $\rho_{\tau}^{(e)}$  95% pointwise confidence interval (CI) between Seoul and Jeju



(c) Confidence intervals for Jeju