

2018년 자료를 통해 바라본 한국 미세먼지의 영향요인 분석 : 국내와 국외 요인 회귀분석을 중심으로

I. 서론

2019년 11월 21일 발표된 한중일 초미세먼지 공동연구 보고서가 발표되었다. 이 보고서에 따르면 한국 초미세먼지의 중국측 기여도는 32%, 하지만 이 연구에는 한계가 존재하는데, 중국측 연구의 투명성이 보장되지 않는다는 점이다.

공동연구라 해서 삼국의 과학자들이 한데 모여 같이 연구를 수행하는 것이 아니고, 중국은 중국, 한국은 한국, 일본은 일본 각자의 데이터를 가지고 낸 결과를 합친 것뿐이다. 여기서 중국은 한국 초미세먼지의 중국 기여도를 약 20%, 한국과 일본은 약 40%로 산정하였다. 두배에 가까운 차이에 각국 과학자들이 서로 인정하지 않아 그냥 평균값인 32%를 결론으로 냈다는 것이 이 연구의 뒷이야기이다.

이에 아이디어를 얻어 자체적으로 간단하지만 자체적으로 데이터를 수집하여 서울 미세먼지에 영향을 주는 요인들을 분석해보면 재미있겠다는 생각에 계획을 세우게 되었다.

먼저, 미세먼지는 일반적으로 우리 눈에 보이지 않을 정도로 아주 가늘고 작은 직경 $10\mu\text{m}$ 이하의 먼지 입자인 PM-10의 미세먼지, 그리고 우리가 흔히 초미세먼지라고 부르는 직경 $2.5\mu\text{m}$ 이하의 먼지 입자인 PM-2.5로 나누어진다. 우리의 데이터는 한국에서 미세먼지 수치로 서울의 하루 평균 PM-10 미세먼지 수치를 조사하여 정리하였고, 중국의 미세먼지 수치로는 한국과 가장 가까운 산둥성의 제남시의 하루 평균 PM-10, PM-2.5 데이터를 사용하였다.

한국의 미세먼지의 원인에 있어서는 기상적 요인 또한 배제할 수 없다. 왜냐하면 미세먼지의 수치는 계절에 따라 커다란 차이를 보이기 때문이다. 기상청에 따르면, 미세먼지 수치는 기압과 공기의 흐름, 강우의 유무에 영향을 주는 습도, 그리고 난방 사용에 영향을 주는 기온 등, 여러 가지 기상적 요인에 의해 커다란 영향을 받는다고 한다. 그래서 작성자는 미세먼지에 영향을 줄 수 있는 서울에서 관측되는 풍속, 기온, 그리고 습도를 변수로 사용하기로 했다.

많이 부족할 수 있겠지만, 이번 기회를 통해 나름대로 미세먼지에 대한 조금 더 객관적인 접근을 할 수 있었으면 좋겠다.

II. 우리나라의 미세먼지 발생에 영향을 주는 중국의 요인

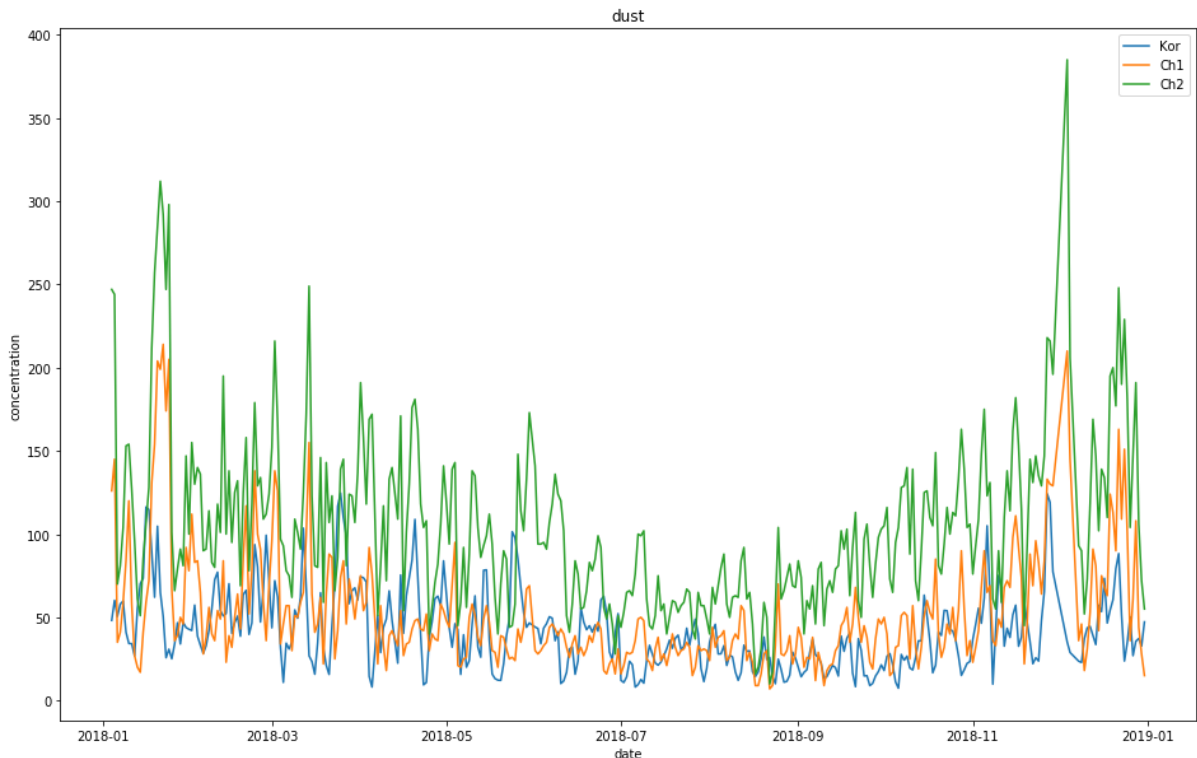
1) 국외적 요인 (중국)

세계의 공장이라는 중국은 엄청난 에너지를 소비하며, 이는 모두 미세먼지의 발생 원인이 된다. 공장의 대부분이 중국의 동쪽, 해안부근에 위치하여 한국과의 거리가 가깝다. 또한 바람의 영향으로 중국 상공의 미세먼지가 우리나라로 유입되고 있다.

- 서울의 초미세먼지 지역별 기여도: 서울 자체 기여도 21% 중국 49% 경기도, 인천 26%
- 중국 공장의 휴식기나 바람이 한국을 향하지 않을 때, 미세먼지가 없는 경우 빈번
- 중국에서 고농도 스모그 발생 2~3일 후, 한국에 미세먼지 경보 수준의 고농도 오염
- 한국 내 자체 오염원이 적은 제주도 등 도서지역에서도 미세먼지가 고농도로 관측
- 환경부에 따르면 중국 미세먼지의 영향이 연평균 30~50% 비중, 고농도 미세먼지 60~80% 비중

II-II. 우리나라의 미세먼지와 중국의 미세먼지 수치 비교

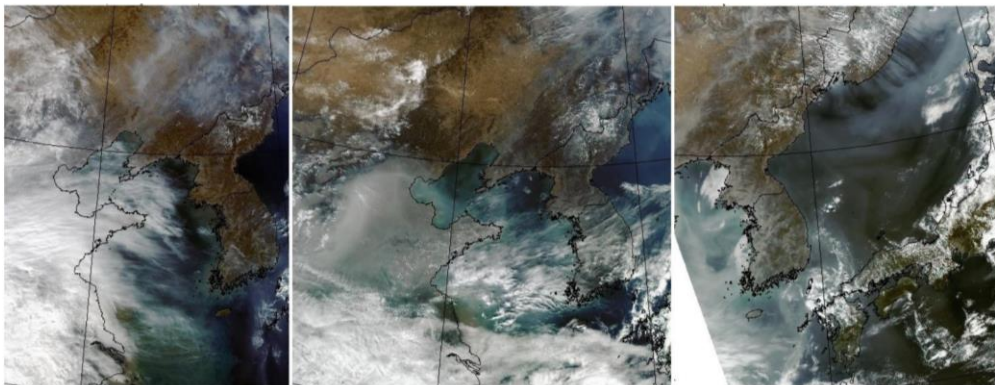
이러한 국외적 요인의 관측을 바탕으로, 우리는 한국의 미세먼지와 중국의 미세먼지와 초미세먼지의 2018년 1년 간의 시계열 그래프를 만들어 보았다.



파란색 선은 한국(서울)의 미세먼지(10PM)를 의미하고, 주황색 선과 녹색 선은 각각 중국(제남)의 초미세먼지(2.5PM)와 미세먼지(10PM)를 의미한다. 우리는 이 시계열 그래프를 통해 완전히 같다고 할 수는 없지만, 한국과 중국의 미세먼지 수치 그래프가 상당히 비슷한 모양을 나타내는 것을 보아, 우리는 한국과 중국 미세먼지 수치 간에는 일정 수준의 관계가 있다는 것을 추론해 낼 수 있다. 이는 우리가 단순한 수치의 비교만으로도, 외부적 요인을 고려해 볼 가치가 있다는 것을 보여준다.

III. 연구 설계

1년 365일 하루하루를 각각 하나의 개체로 보고 분석을 하기로 하였다. 아래 사진은 중국의 미세먼지가 서풍을 타고 한반도로 유입되는 모습을 보여준다. 중국 상공의 스모그가 한국 상공으로 이동하는데 약 3일이 걸린 것을 보아, 중국 미세먼지 농도는 약 3일간의 시차를 두고 서울 미세먼지 농도에 영향을 줄 것임을 알 수 있다. 따라서 서울 미세먼지 농도 데이터와 중국 미세먼지 데이터는 3일의 시차가 있도록 조정하였다.



미 항공우주국 테라/아쿠아 위성 사진. 왼쪽부터 2월 26일, 2월 28일, 3월 1일 사진이다. 26일 한반도 상공은 중국과 달리 맑았지만, 28일부터 중국 오염물질이 들어오면서 1일에는 서해와 한반도 서쪽이 스모그로 뒤덮여 있다. [사진 기상청 홈페이지]

1. 가설

- (1) 중국 산둥성 미세먼지 농도는 3일치 시차를 두고 서울 미세먼지 농도 증가에 영향을 미칠 것이다.
- (2) 중국 산둥성 초미세먼지 농도는 3일치 시차를 두고 서울 미세먼지 농도 증가에 영향을 미칠 것이다.
- (3) 습도가 높은 날은 서울 미세먼지 농도에 (-)영향을 미칠 것이다.
- (4) 풍속이 빠른 날은 서울 미세먼지 농도에 (-)영향을 미칠 것이다.
- (5) 기온이 낮은 날은 서울 미세먼지 농도에 (+)영향을 미칠 것이다.

2. 변수의 선정

모든 관찰 값들은 정오를 기준으로 한 수치를 사용하였다.

종속변수: 서울시 미세먼지 농도 (10PM) - (에어코리아)

(데이터셋에는 월별 전국 석탄화력발전량과 서울시 경유사용량이 포함되어 있지만 분석에서는 사용하지 않았다)

기상 요인 및 control variable

기온 – 기상청

-기온은 미세먼지 수치에 커다란 영향을 미칠 수 있다. 날씨가 추워지면 난방 등 연료 사용이 급증하기 때문이다. 만약 시베리아의 찬 공기의 영향에 의해 기온이 내려갔으면, 풍향이 바뀌어 미세먼지 수치가 크게 내려가기도 한다.

풍속(최대평균풍속) – 기상청

- 풍속이 낮아지면 대기가 안정되어, 한반도 상공의 미세먼지가 날라가지 않고 한반도에 머물게 된다. 특히 최근, 기후 변화가 심해져, 한반도 겨울철의 풍속이 낮아져, 고농도 미세먼지의 영향이 받는 날이 많아졌다는 분석도 있다. 그래서 중국의 요인과는 별개로 국내의 요인으로 풍속과 우리나라 미세먼지 농도와 관련한 식을 세웠다.

습도 – 기상청

- 습도가 높으면 미세먼지 농도가 낮다는 분석이 있기에 습도도 하나의 변수로 포함시켰다.

중국측 요인

산동성 제남시의 초미세먼지 농도(PM2.5)

산동성 제남시의 미세먼지 농도(PM10) 중국의 미세먼지 수치는 한국에서 제일 가까운 성인 산동성의 대표적인 도시 제남시의 미세먼지와 초미세먼지 수치를 이용하기로 했다.

3. 데이터 분석

수집한 데이터를 바탕으로 회귀분석 모형을 만들어 보았는데, 본 조사의 회귀분석 모형

은 다음 식과 같다.

$$\text{서울시 미세먼지 농도} = \alpha + \beta_1(\text{중국요인}) + \beta_2(\text{control variables}) + \mu$$

작성자는 본 조사의 회귀분석 모형에서 앞(변수의 선정)에서 언급한 한국의 미세먼지 농도를 종속변수, 그리고 중국의 미세먼지 농도(PM10, PM2.5), 풍속, 습도, 기온을 독립변수로 사용하기로 했다.

```
In [20]: model_sm = sm.OLS.from_formula("K1 ~ tem + vel + hum + Ch1 + Ch2", data = b)
result_sm = model_sm.fit()
print(result_sm.summary())
```

OLS Regression Results						
=====						
Dep. Variable:	K1	R-squared:	0.160			
Model:	OLS	Adj. R-squared:	0.148			
Method:	Least Squares	F-statistic:	13.21			
Date:	Wed, 22 Jan 2020	Prob (F-statistic):	8.69e-12			
Time:	10:40:43	Log-Likelihood:	-1585.7			
No. Observations:	353	AIC:	3183.			
Df Residuals:	347	BIC:	3207.			
Df Model:	5					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	47.5165	7.213	6.588	0.000	33.330	61.703
tem	-0.2510	0.131	-1.915	0.056	-0.509	0.007
vel	-1.3624	0.561	-2.428	0.016	-2.466	-0.259
hum	-0.1528	0.088	-1.732	0.084	-0.326	0.021
Ch1	-0.0567	0.080	-0.705	0.481	-0.215	0.102
Ch2	0.1686	0.054	3.093	0.002	0.061	0.276
=====						
Omnibus:	54.189	Durbin-Watson:	0.708			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	80.646			
Skew:	0.970	Prob(JB):	3.08e-18			
Kurtosis:	4.313	Cond. No.	883.			
=====						

1. 기온(tem)의 경우 서울 미세먼지 농도에 (-)영향이 있으며, 이는 주로 미세먼지가 심한 시기가 겨울, 봄 등 기온이 낮은 계절이라고 해석할 수 있다(p-value 0.056으로 null hypothesis가 참이고 다른 모든 가정이 유효하다면 적어도 현재 얻은 결과와 같은 크기의 결과를 얻을 확률이 5.6%라는 것을 의미).

2. 풍속(vel)의 경우 서울 미세먼지 농도에 (-)영향이 있으며, 이는 풍속이 낮을 때 대기가 정체되어 미세먼지 농도가 높아진다고 해석할 수 있다.

3. 습도(hum)의 경우 서울 미세먼지 농도에 (-)영향이 있으며, 이는 비가 내려 습도가 높을 때 대

기중의 미세먼지가 씻겨 나갔다고 해석할 수 있다.

4. 중국 미세먼지(Ch1)의 경우 서울 미세먼지 농도에 (-)영향이 있다고 나타났지만, p-value가 0.481로, 통계적으로 유의미하지 않다.

5. 중국 초미세먼지(Ch2)의 경우 서울 미세먼지 농도에 (+)영향이 있다고 나타났다.

위 데이터는 가설(1)을 reject하며, 중국 미세먼지는 서울 미세먼지 농도에 통계적으로 유의미한 영향을 주지 않는다고 말하고 있다. 이는 일반적인 우리들의 생각과는 다른 모습인데, 실제로 미세먼지는 입자의 크기가 크고, 국지적인 영향을 많이 받는다고 한다. 우리나라도 88올림픽 이전에는 미세먼지가 매우 심각했지만 저감조치 시행 이후 나아진 것이다.

유의할 점은, 데이터의 수와 사용한 변수의 수가 적기 때문에 분석 결과로 나온 수치 그 자체가 어떤 특별한 의미가 있다고 해석하기는 어렵다는 점이다. 단지 그 양상만 참고하는 수준이다.

IV. 한계

1. 사실, 한국의 경유사용량과 석탄의 화력발전 사용량을 국내적 요인으로 사용해, 한국의 미세먼지 농도에 대한 조금 더 심도 있는 회귀분석을 진행하고 싶었지만, 한 번에 많이 수입해 일일 사용량을 추적하기 힘든 경유와 석탄의 사용량의 특성 때문에, 데이터가 월별로 밖에 존재하지 않아서, 월평균 데이터 값을 그 달의 일수로 나누어 사용하면 데이터 본래의 성질을 잃어버릴 것이라 생각하여 제외하였다. 따라서 이번 회귀분석을 국외적 요인을 대표하는 중국의 미세먼지 수치와 국내 기상 요인만을 변수로 사용하고, 이번 회귀분석을 마무리 할 수 밖에 없었다.

2. 또한 국내 초미세먼지 데이터가 누락되어, 요인들과 국내 초미세먼지와의 관계를 분석하지 못하였다. 사실 미세먼지보다 초미세먼지가 더 건강에 부정적인 영향을 주기 때문에 설게 초기에 잘못이 있다고 생각한다.

3. 수집한 데이터가 1년 분량밖에 되지 않아 정확한 분석을 하기 어렵다는 태생적 문제가 있다. 신뢰성을 높이려면 수십년간의 데이터가 필요하다.

완전한 데이터, 연구 주제에 대한 전문적인 지식(expert knowledge), 모든 변수에 대한 완벽한 이해, 연구 환경의 완벽한 통제가 없으면 우리가 원하는 완벽한 결과는 얻을 수 없다. 현실적으로 불가능하기 때문에 우리는 오류를 줄여 나가서 연구의 신뢰성을 높이는 방향으로 계획을 세워야 한다.

대부분의 사회과학 연구는 실험연구(experimental study)가 불가능하다. 게다가 여러 국가, 기상, 산업 등이 연관된 미세먼지는 실험연구가 더욱 불가능하다. 불완전 하겠지만

위의 데이터 분석을 좀 더 신뢰할 만한 연구로 수정하고자 한다면 다음과 같은 해결방법들을 생각해 볼 수 있다.

1. 석탄화력발전, 경유사용량 등의 데이터가 월별로 있으니, 수집한 데이터 또한 월평균 데이터로 가공하여 월별 data set으로 만든다. 다만 이 때는 데이터 수집 기간을 매우 많이 늘려야 한다.
2. 국내요인에 도로 교통량을 추가한다. 미세먼지의 주요 원인 중 하나는 자동차의 매연이므로 교통량 또한 주요한 요인이 될 수 있다.

이번 보고서의 분석 방향을 정할 때, 작성자 나름대로 다양한 변수를 정해서 의미 있는 결과물을 만들어 보려고 했으나, 일부 데이터의 수집에 있어서, 데이터를 찾을 수 없는 현실적인 문제에 부딪혀, 국내적 요인은 배제한 채 이번 보고서를 마무리할 수밖에 없어 매우 아쉽다고 생각한다. 전체 과정 중에서 회귀분석을 수행하는 것 자체는 시간적으로도 수 초면 되고, 그리 중요한 과정도 아니다. 정말로 중요한 것은 연구를 설계할 때 사용할 변수를 선정하는 것, 데이터를 수집하는 것, 데이터 분석 툴 사용법, 데이터를 분석에 사용할 수 있도록 가공하는 전처리(preprocessing)과정임을 강조하면서 글을 마친다.