수치 모델링 및 머신러닝을 이용한 대기 오염 예측

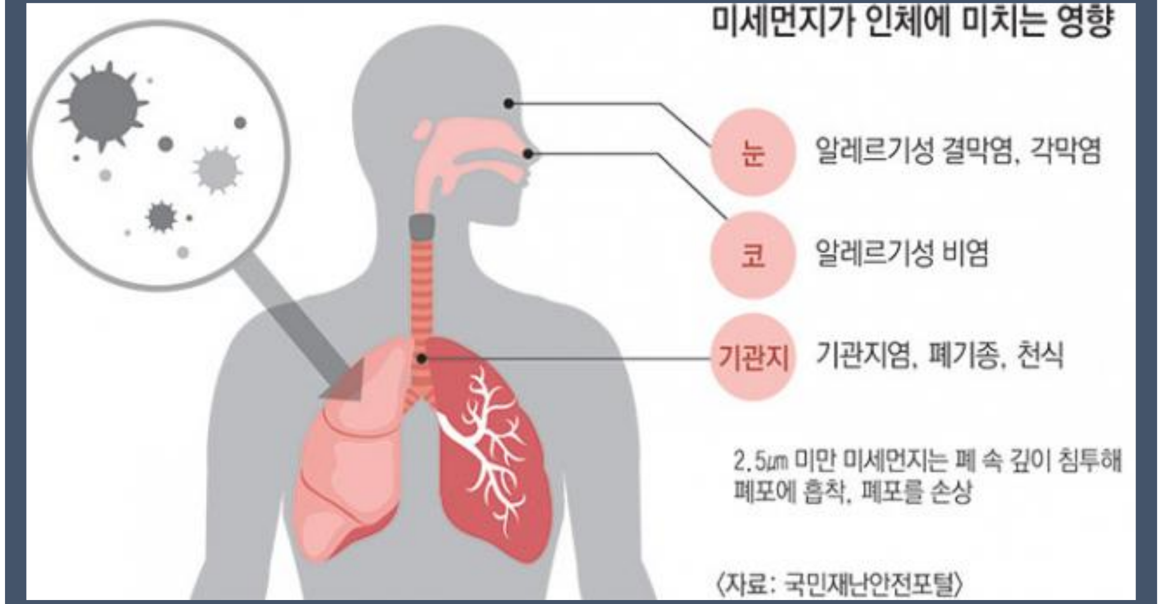
수학과 2017010698 오서영
수학과 2018010705 신영민

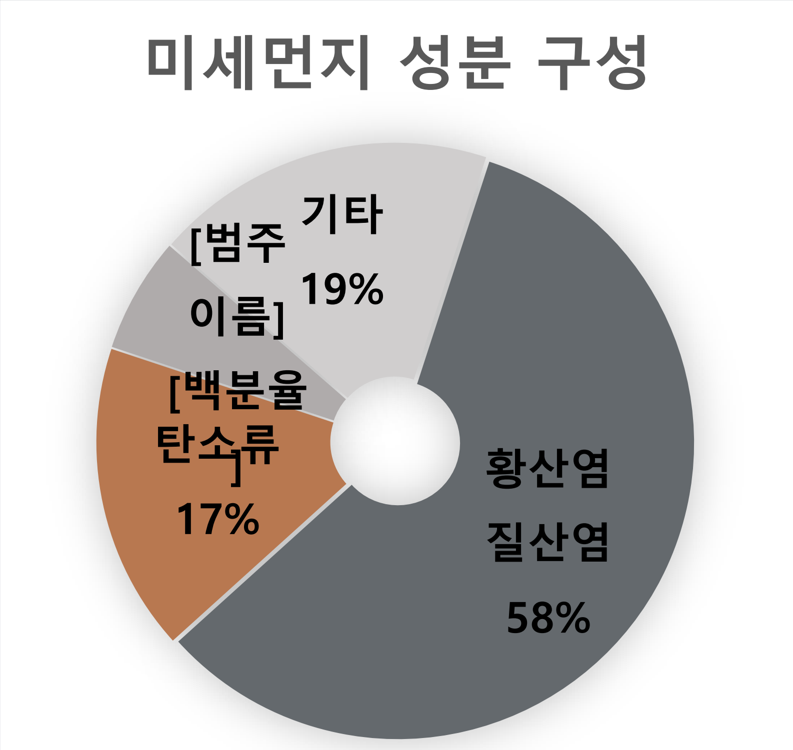
목차

1. 수행개요

목적 및 필요성

미세먼지에 의한 사망률이 증가하는 추세 → 미세먼지 농도 **예측**의 필요성





<출처 : 환경부, 국민 재난 안전 포털>

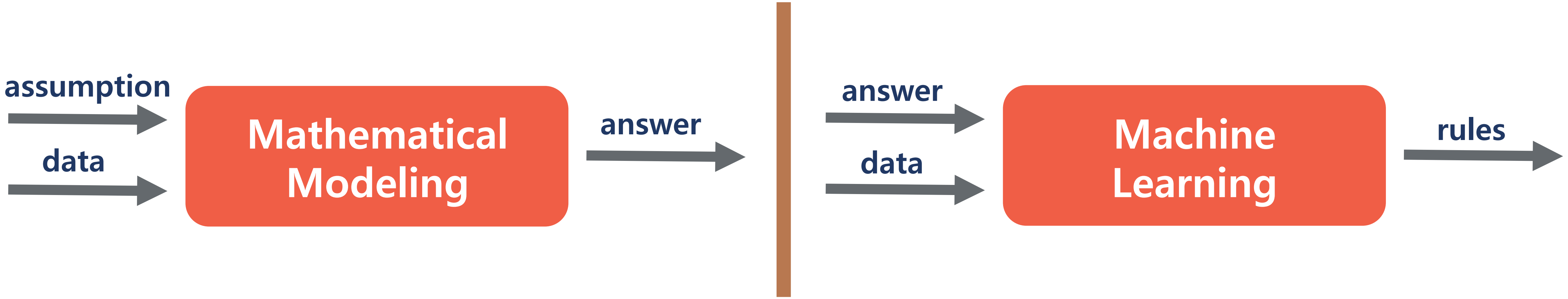
1. 수행개요

목적 및 필요성

두가지 방법론을 사용하여 미세먼지 농도 예측하기

수학적 모델링 (Mathematical modeling)

기계 학습 (Machine learning)



2. 수행과정 데이터 수집 및 정제

DATASET

지역별 시간당 **미세먼지 농도** (Pm-10)

풍속, 풍향 데이터
(2019-04-05)

일시	서울																	
	구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구	마포구	서대문구	서초구	성동구	성북구	송파구
2019-04-0501	7	27	30	27	38	27	28	28	28	26	16	27	37	28	30	33	21	25
2019-04-0502	5	25	31	28	33	32	27	27	26	26	19	30	30	29	30	27	31	23
2019-04-0503	5	24	36	30	36	29	31	29	30	30	25	30	38	35	36	32	31	24
2019-04-0504	5	26	40	29	43	34	31	28	31	33	23	32	37	37	36	34	29	25
2019-04-0505	8	31	-	33	44	37	39	33	33	36	28	35	40	41	43	38	29	32
2019-04-0506	3	35	45	36	45	46	42	34	36	44	32	38	48	46	57	48	33	38
2019-04-0507	3	50	58	46	54	54	50	42	48	57	46	51	51	58	65	73	47	50
2019-04-0508	2	65	55	62	72	76	70	55	56	59	56	55	73	67	74	80	56	60
2019-04-0509	2	73	71	85	96	84	97	76	62	84	67	67	102	96	133	110	62	74
2019-04-0510	0	78	93	92	113	112	109	77	77	111	82	78	112	113	137	124	93	96

지점	시간	풍속(m/s)	풍향(16방위)
서울(108)	2019-04-05 01:00	3.3	200
서울(108)	2019-04-05 02:00	3.1	230
서울(108)	2019-04-05 03:00	2.2	230
서울(108)	2019-04-05 04:00	1.8	270
서울(108)	2019-04-05 05:00	2.3	230
서울(108)	2019-04-05 06:00	2.7	200
서울(108)	2019-04-05 07:00	2.1	230
서울(108)	2019-04-05 08:00	1.2	230
서울(108)	2019-04-05 09:00	1.6	250
서울(108)	2019-04-05 10:00	3	250

<출처 : 에어코리아, 기상자료 개방포털 >

2. 수행과정 데이터 수집 및 정제

DATASET

지역별 **위도 경도** 데이터

37.519977, 126.984509

위도 **경도**

<출처 : 구글지도 >

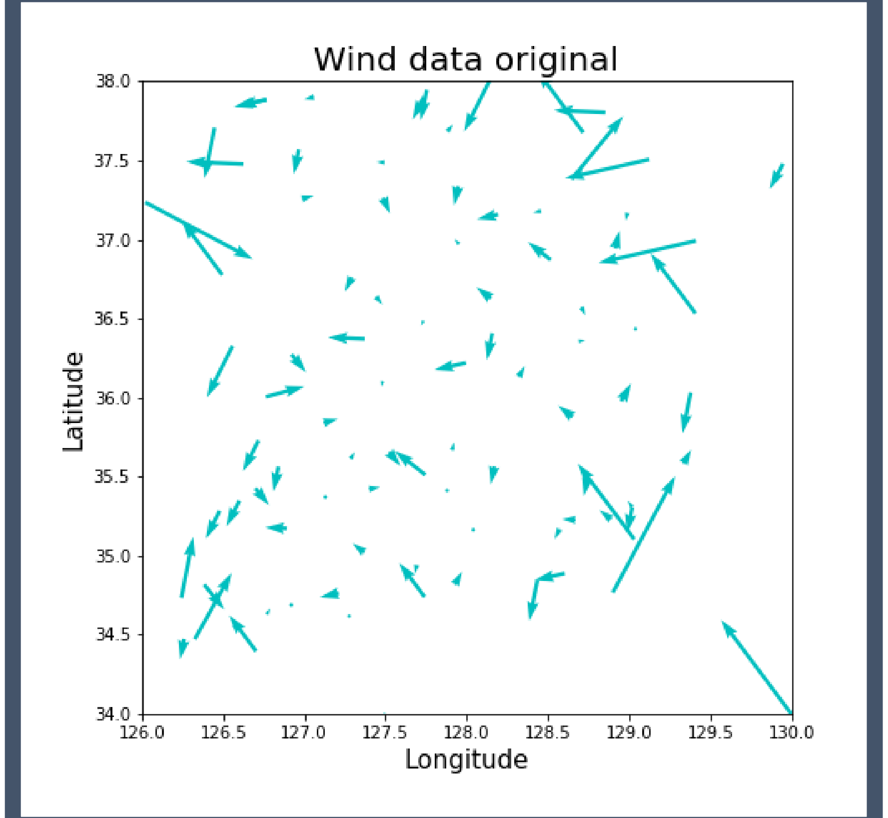
2. 수행과정 데이터 수집 및 정제

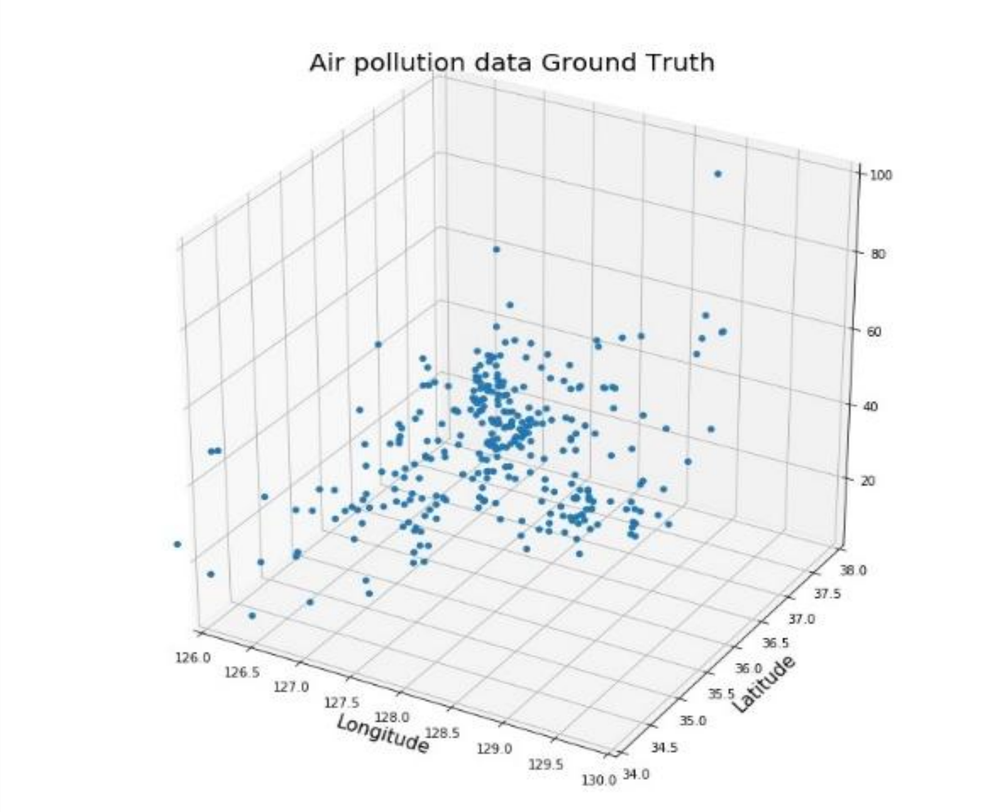
DATASET

데이터 정제 필요성

공간에 대한 통계자료를
모든 지점에서 획득하기는
현실적으로 불가능

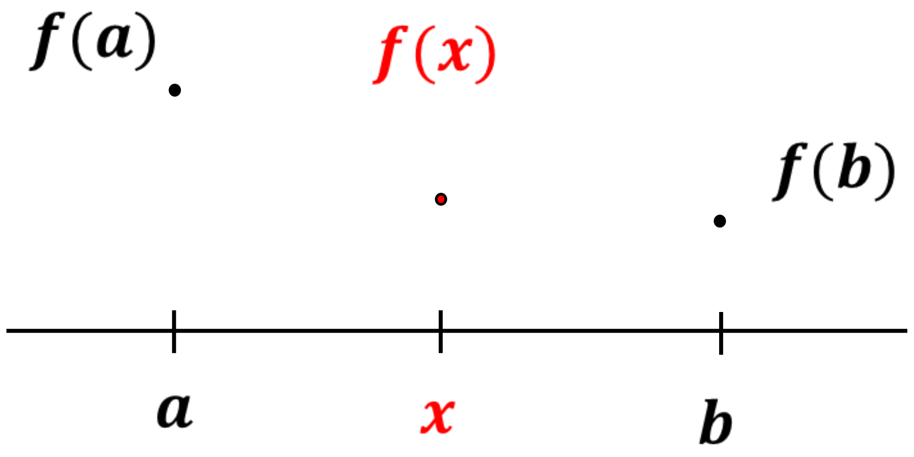
→ **보간법**





보간법 (Interpolation)

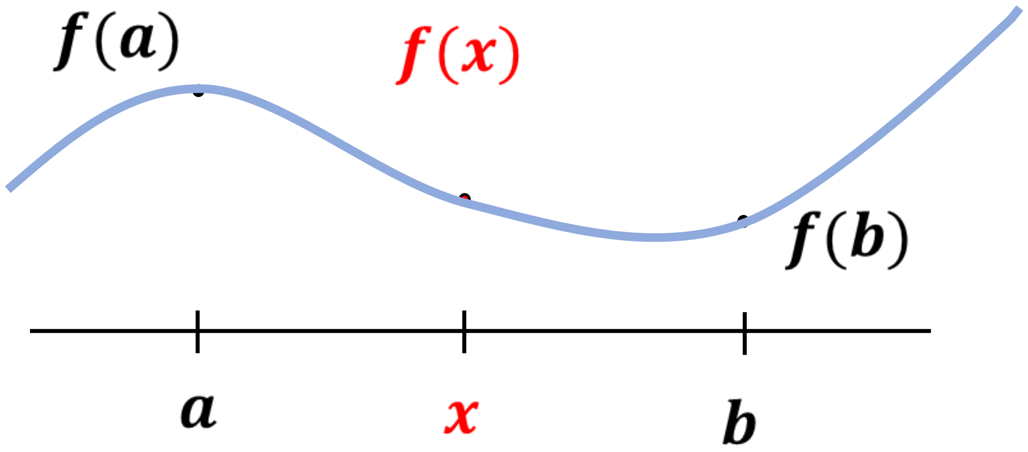
알고 있는 두 점 사이 어느 지점의 값을 추정하는 기법



Cubic Interpolation (3차 보간법)

3차 다항식을 활용하여 보간

우리가 알고있는 두 점이 3차함수 위에 있다고 가정



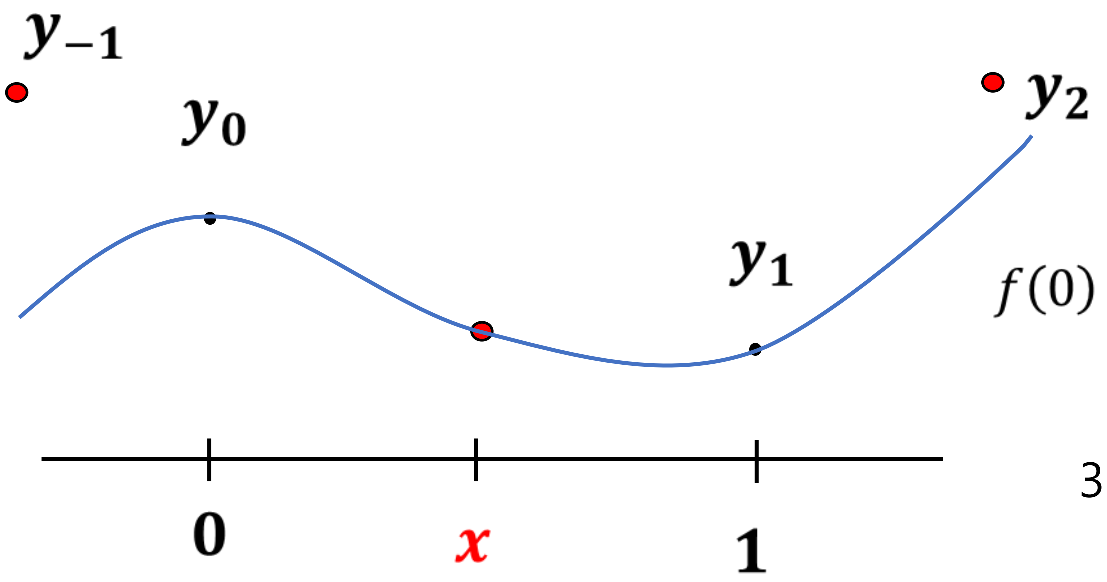
$$f(x) = mx^3 + nx^2 + jx + k$$

2. 수행과정

데이터 수집 및 정제

Cubic Interpolation (3차 보간법)

미지수 4개, 식 4개 $(f(0) = y_0, f(1) = y_1)$



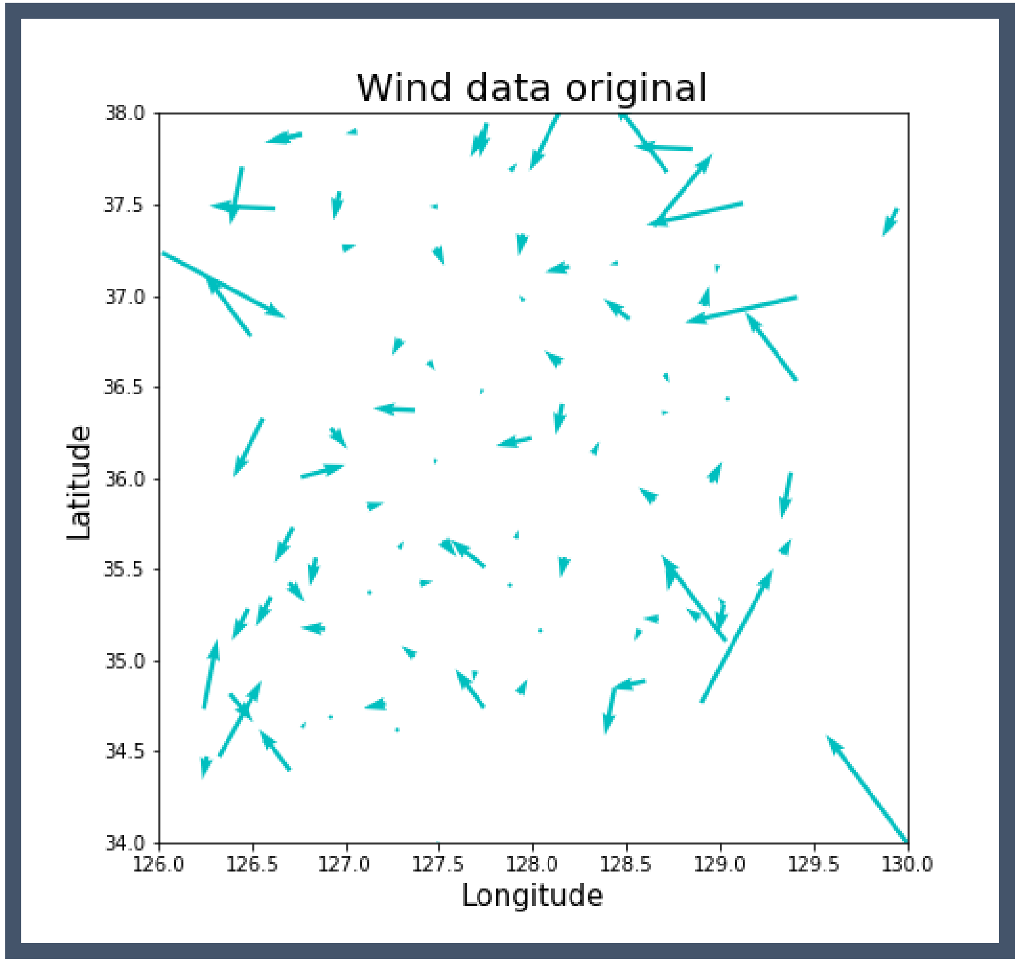
$$f(0) = y_0, f(1) = y_1, f'(0) = \frac{y_1 - y_{-1}}{2}, f'(1) = \frac{y_2 - y_0}{2}$$

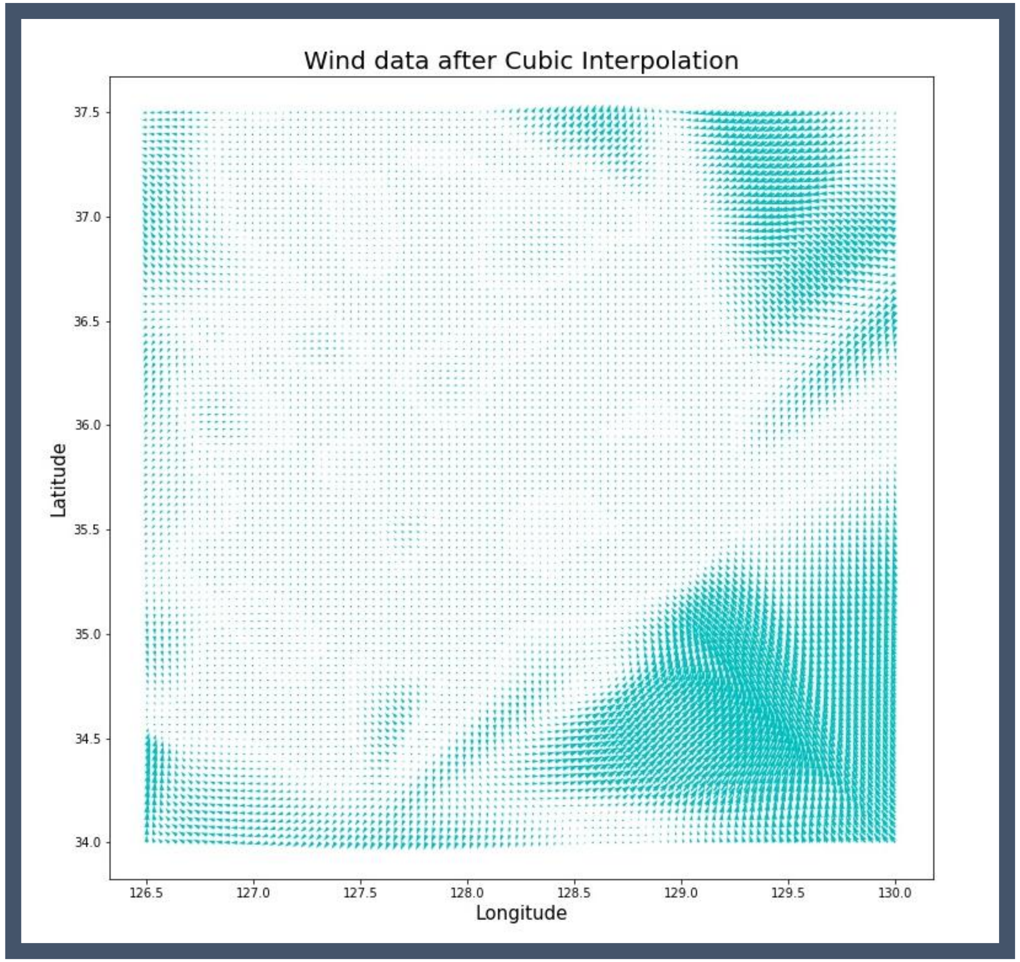
3차 스플라인 보간법은 4개의 이웃점을 필요로 함.

$f(x)$를 통해 0, 1 사이의 존재하는 어떤 점의 함수 값을 추정할 수 있다.

2. 수행과정 데이터 수집 및 정제

Cubic Interpolation (3차 보간법)





IDW(Inversed distance weighted)

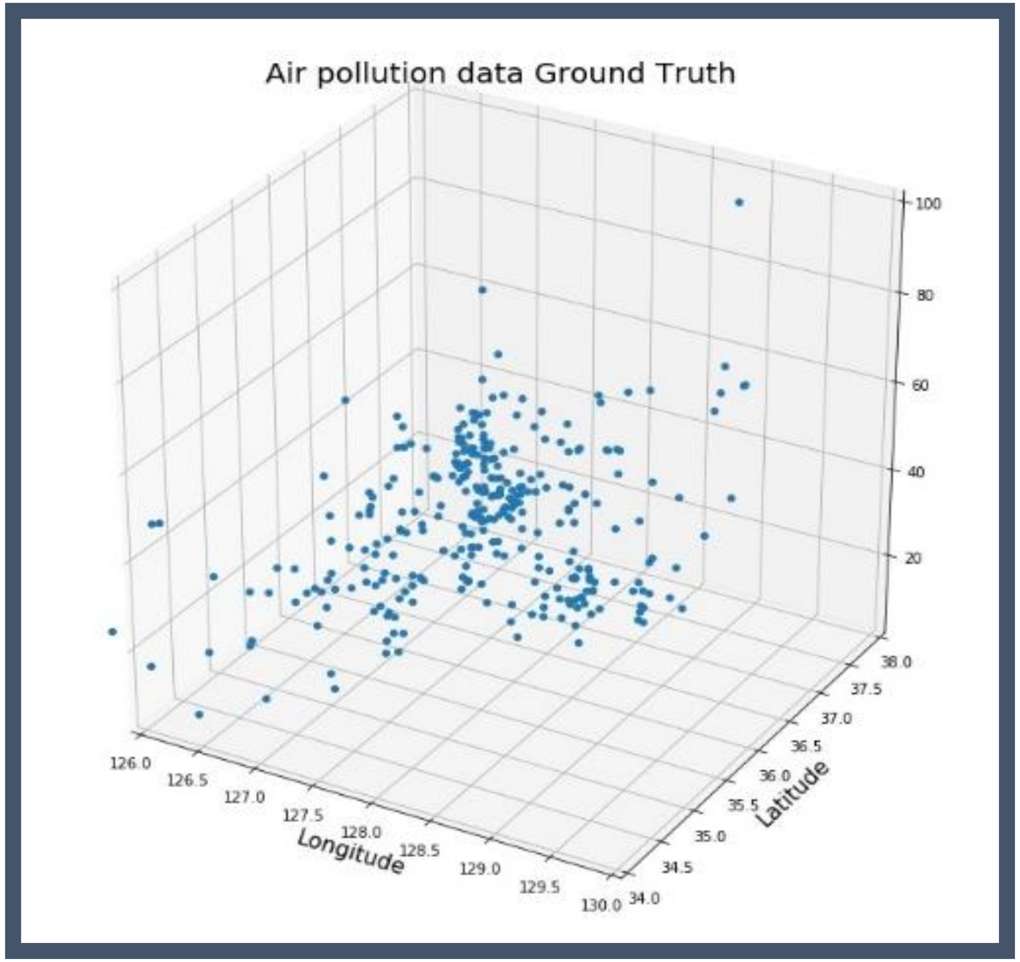
가까이 있는 실측값에 더 큰 가중 값을 주어 보간하는 방법
거리가 가까울 수록 높은 가중 값이 적용.

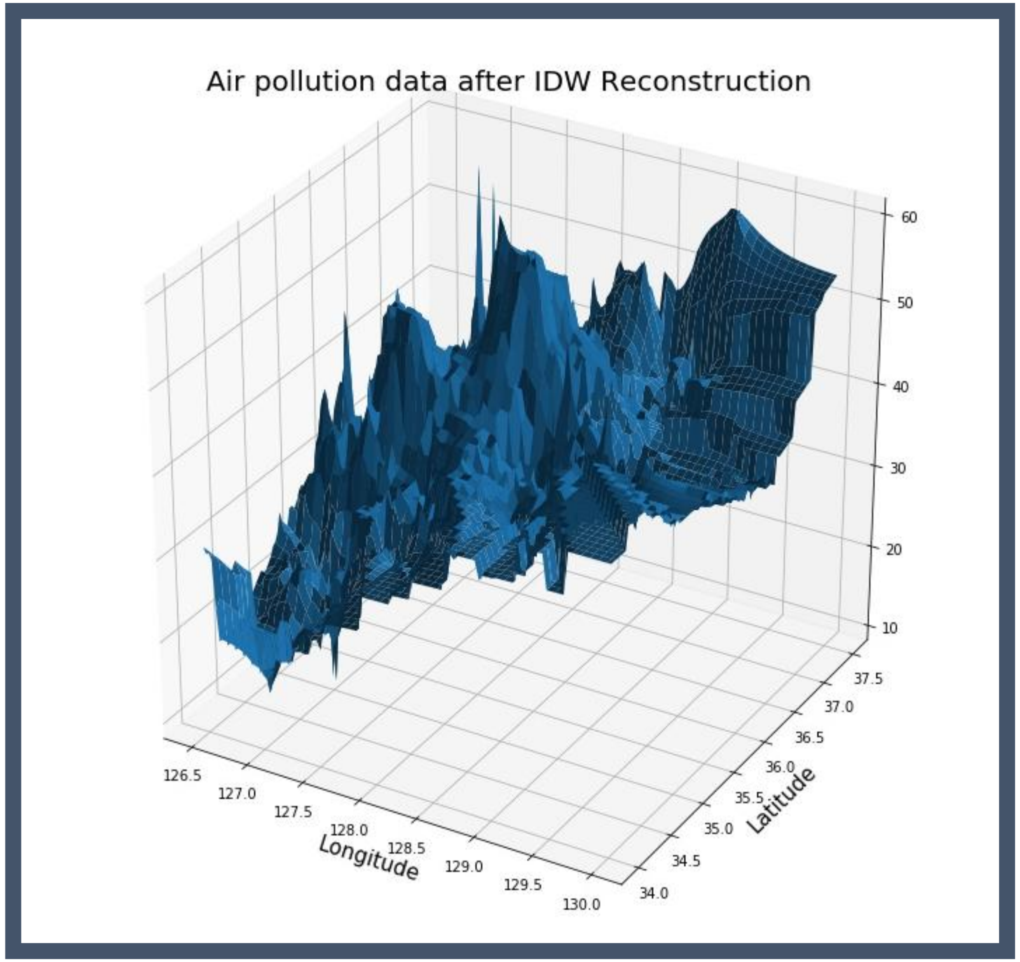
$$\bar{u}(x) = \frac{\sum_{k=0}^{N} w_k(x) u_k}{\sum_{k=0}^{N} w_k}$$

$\boldsymbol{N}$: 실측값 개수
$\boldsymbol{w_k(x)} = \frac{1}{d(x,x_k)}$: 가중치
$\bar{\boldsymbol{u}}$: 보간된 값

2. 수행과정 데이터 수집 및 정제

IDW(Inversed distance weighted)





2. 수행과정

수학적 모델링

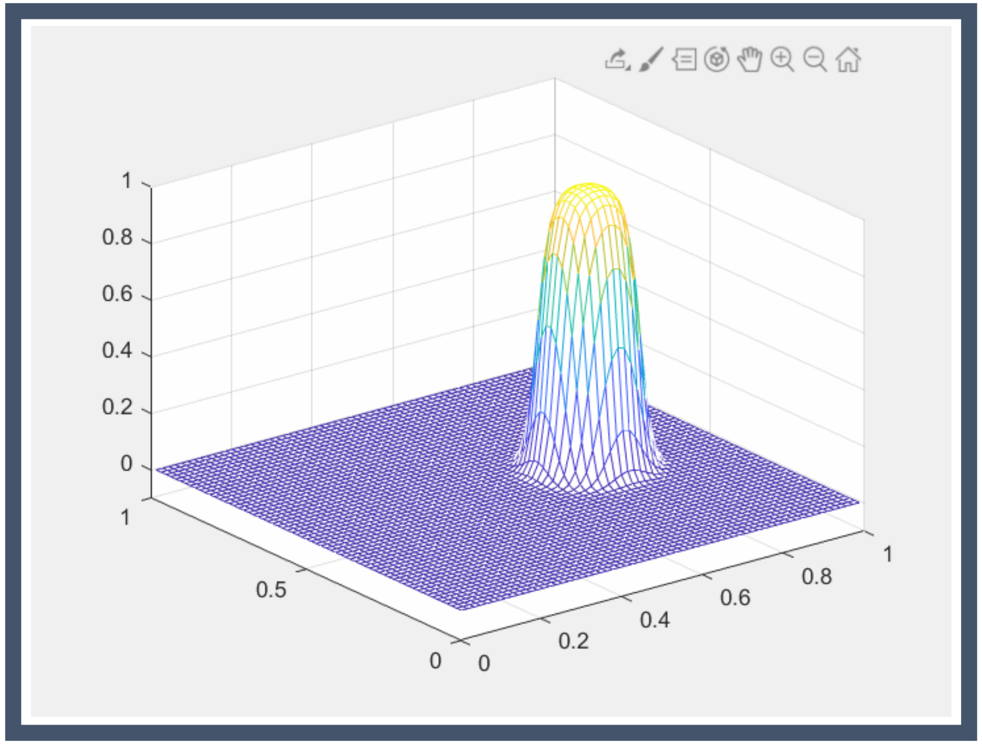
- 가정 : 물질들 사이의 화학적인 변화가 없다

공기 질 모델링 ~ 대류 + 확산

대류-확산방정식 (Convection-Diffusion equation)

대류 방정식 : $\frac{\partial c}{\partial t}(x,t) + \frac{\partial (uc)}{\partial x}(x,t) = 0$

확산 방정식 : $\frac{\partial c}{\partial t}(x,t) = D\frac{\partial^2 c}{\partial x^2}(x,t)$



2. 수행과정

수학적 모델링

대류-확산방정식 (Convection-Diffusion equation)

$c(x,y,t)$와 $(u(x,y),\ v(x,y))$를 2차원 공간 (x,y) 와 시간 t에서의
어떤 물질의 농도와 속도장이라고 하면

$$\frac{\partial c(x,y,t)}{\partial t}+\frac{\partial}{\partial x}[u(x,y)c(x,y,t)]+\frac{\partial}{\partial y}[v(x,y)c(x,y,t)]$$
$$=D[\frac{\partial^2 c(x,y,t)}{\partial x^2}+\frac{\partial^2 c(x,y,t)}{\partial y^2}]$$

2. 수행과정 수학적 모델링

대류-확산방정식 (Convection-Diffusion equation)

이산화를 위한 차분 공식

테일러 전개 : $f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(x)}{3!} + \cdots$

1계 중앙차분

$$^*f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(\xi)}{3!}$$

$$^*f(x-h) = f(x) - hf'(x) + h^2 \frac{f''(x)}{2!} - h^3 \frac{f'''(\xi)}{3!}$$

위 두식을 빼면

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2)$$

2. 수행과정 수학적 모델링

대류-확산방정식 (Convection-Diffusion equation)

이산화를 위한 차분 공식

테일러 전개 : $f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(x)}{3!} + \cdots$

2계 중앙차분

$$*f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(\xi)}{3!}$$

$$*f(x-h) = f(x) - hf'(x) + h^2 \frac{f''(x)}{2!} - h^3 \frac{f'''(\xi)}{3!}$$

위 두식을 더하면

$$f''(x) = \frac{f(x+h) - f(x) + f(x-h)}{h^2} + O(h^2)$$

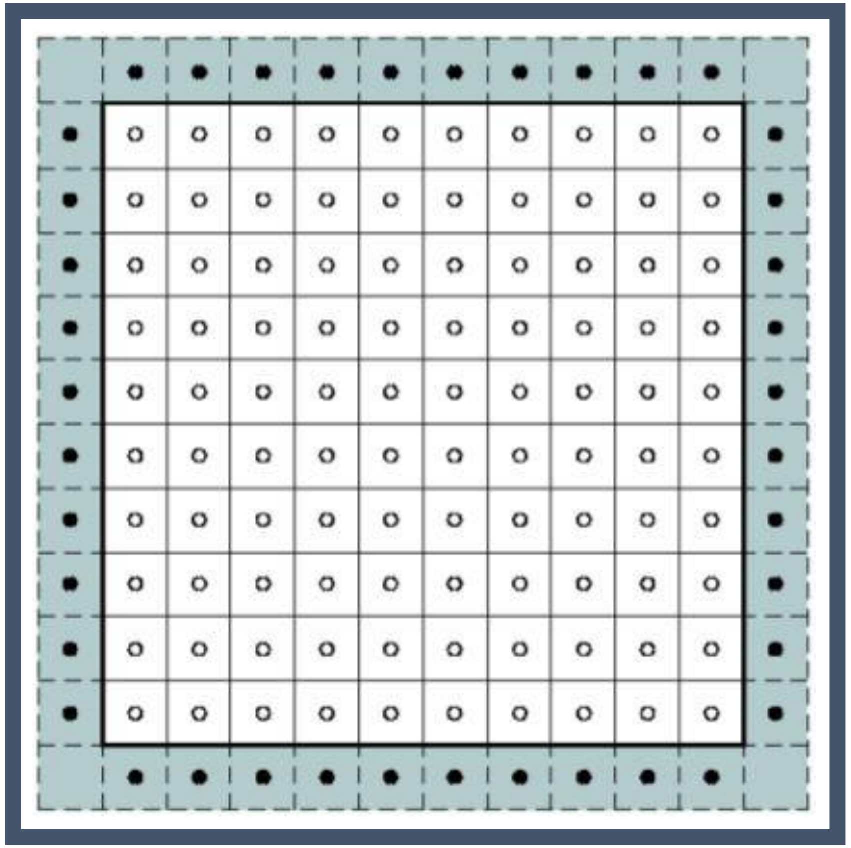
대류-확산방정식 (Convection-Diffusion equation)

이산화한 결과

$$
\begin{aligned}
& c_{ij}^{n+1} \\
& = c_{ij}^{n} - \triangle t \left[\frac{cu_{i+1,j}^{n} - cu_{i-1,j}^{n}}{2h} + \frac{cv_{i,j+1}^{n} - cv_{i,j-1}^{n}}{2h} \right] \\
& + D \frac{\triangle t}{h^2} [c_{i+1,j}^{n} + c_{i,j+1}^{n} - 4c_{ij}^{n} + c_{i-1,j}^{n} + c_{i,j-1}^{n}]
\end{aligned}
$$

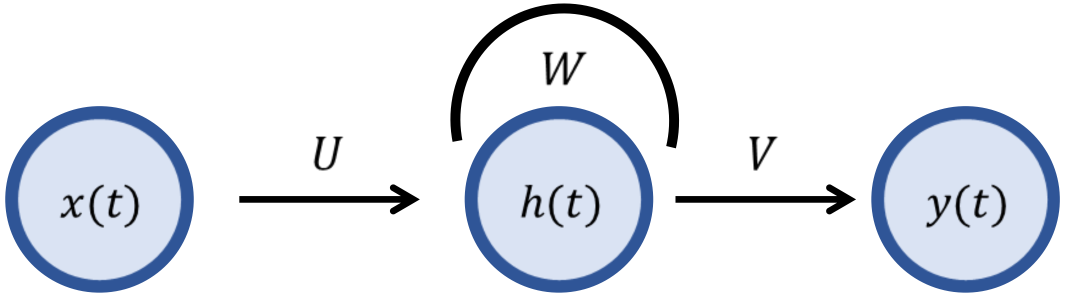
대류-확산방정식 (Convection-Diffusion equation)

경계 조건은 **노이만 경계 조건** (Neumann boundary condition) 을 사용



RNN (Recurrent Neural Network)

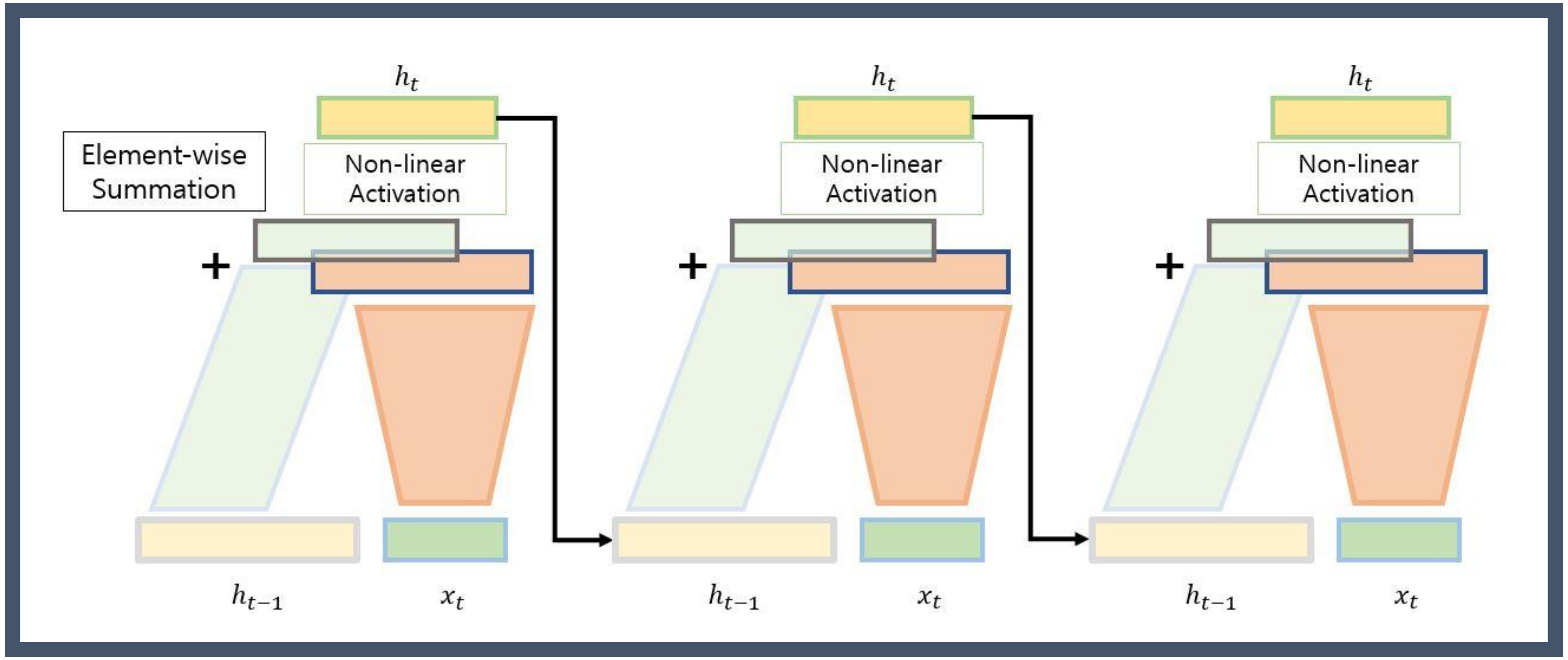
시계열 데이터 예측에 적합한 딥러닝 모델



$$h_t = f_W(h_{t-1},\ x_t)$$
$$= \tanh(w_{hh}h_{t-1} + w_{xh}x_t)$$
$$y_t = w_{hy}h_t$$

x_t : 현재 입력, h_{t-1} : 과거 기억, h_{t-1} : 현재기억

2. 수행과정 머신러닝



RNN (Recurrent Neural Network) → **LSTM (Long Short Term Memory)**

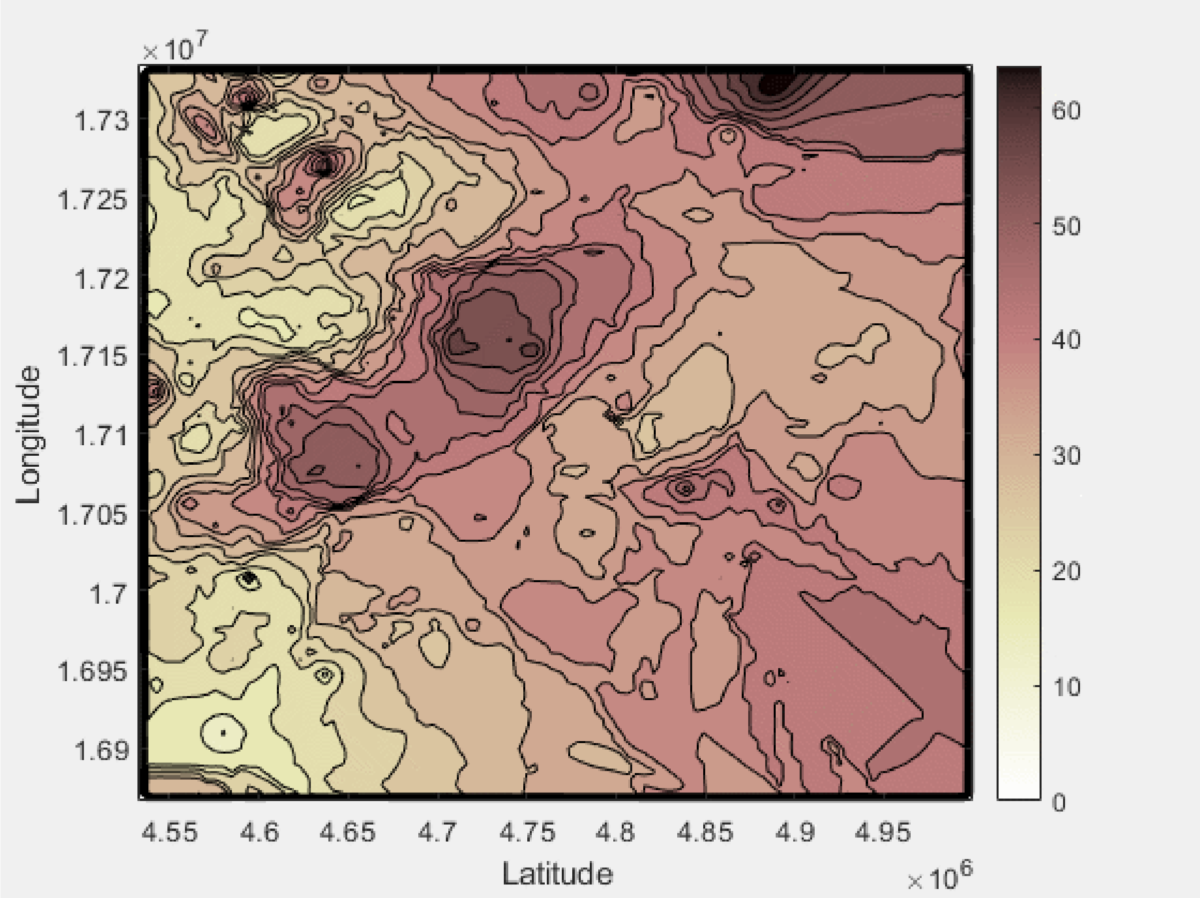
장기 의존성 문제
: 은닉층의 과거 정보가 마지막까지 전달되지 못하는 현상

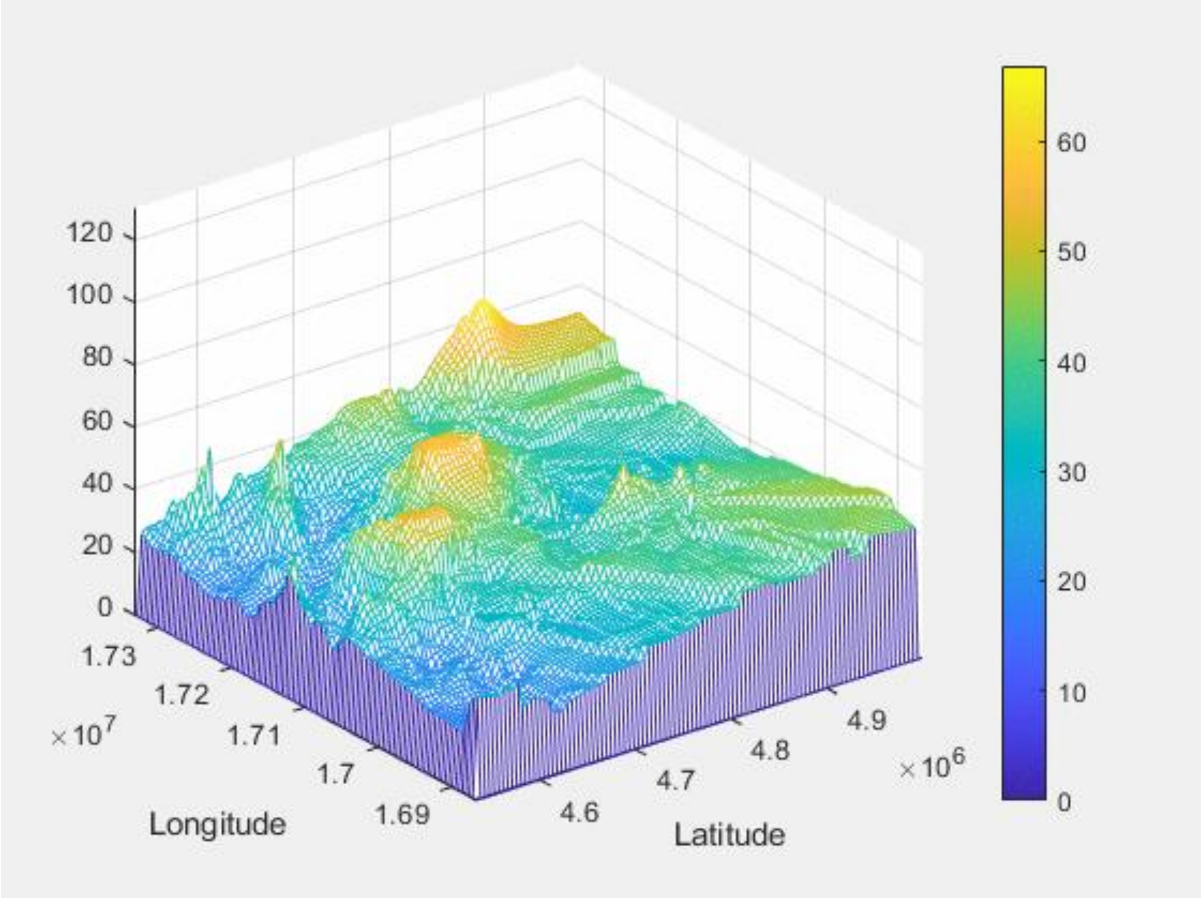
장기 의존성 학습을 할 수 있는 RNN의 한 종류

3. 결과 및 문제점

수학적 모델링 _ Convection-Diffusion model

24시간 대기오염 농도

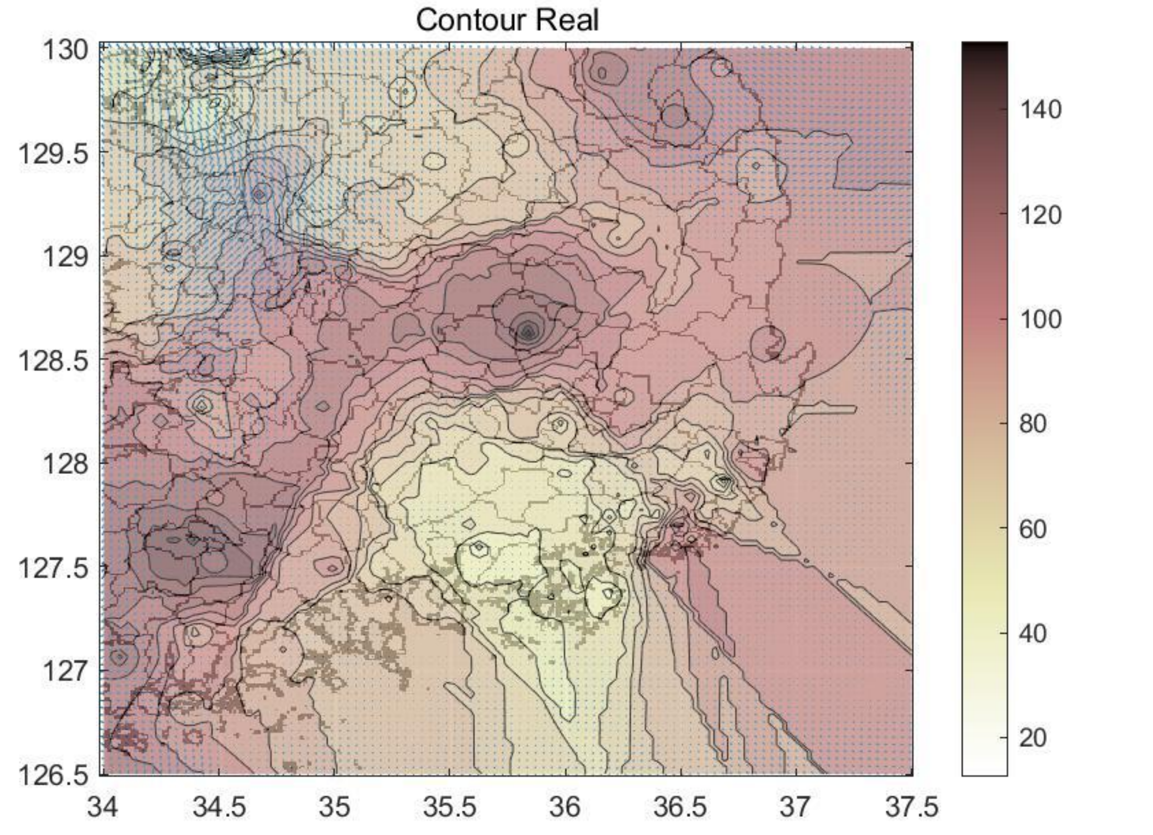


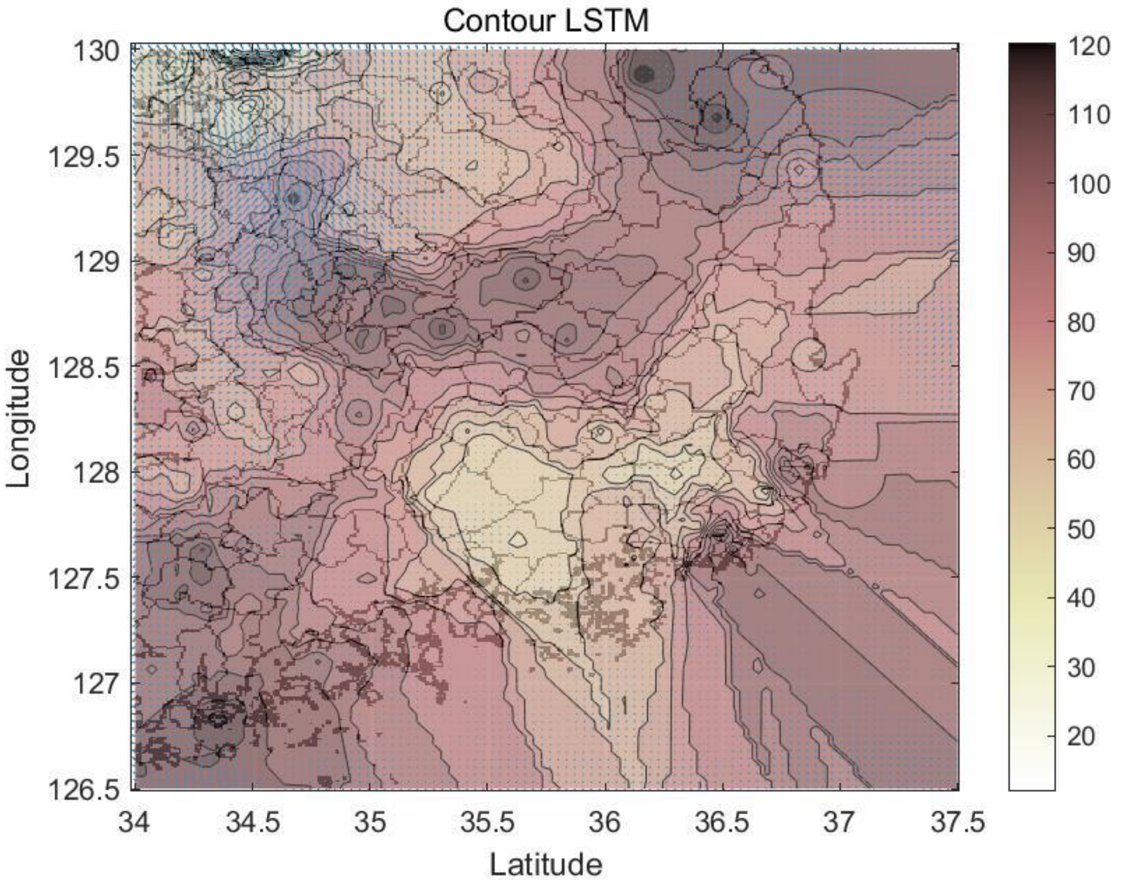


3. 결과 및 문제점

머신러닝 _ LSTM

마지막 한시간에 대한 대기오염 농도

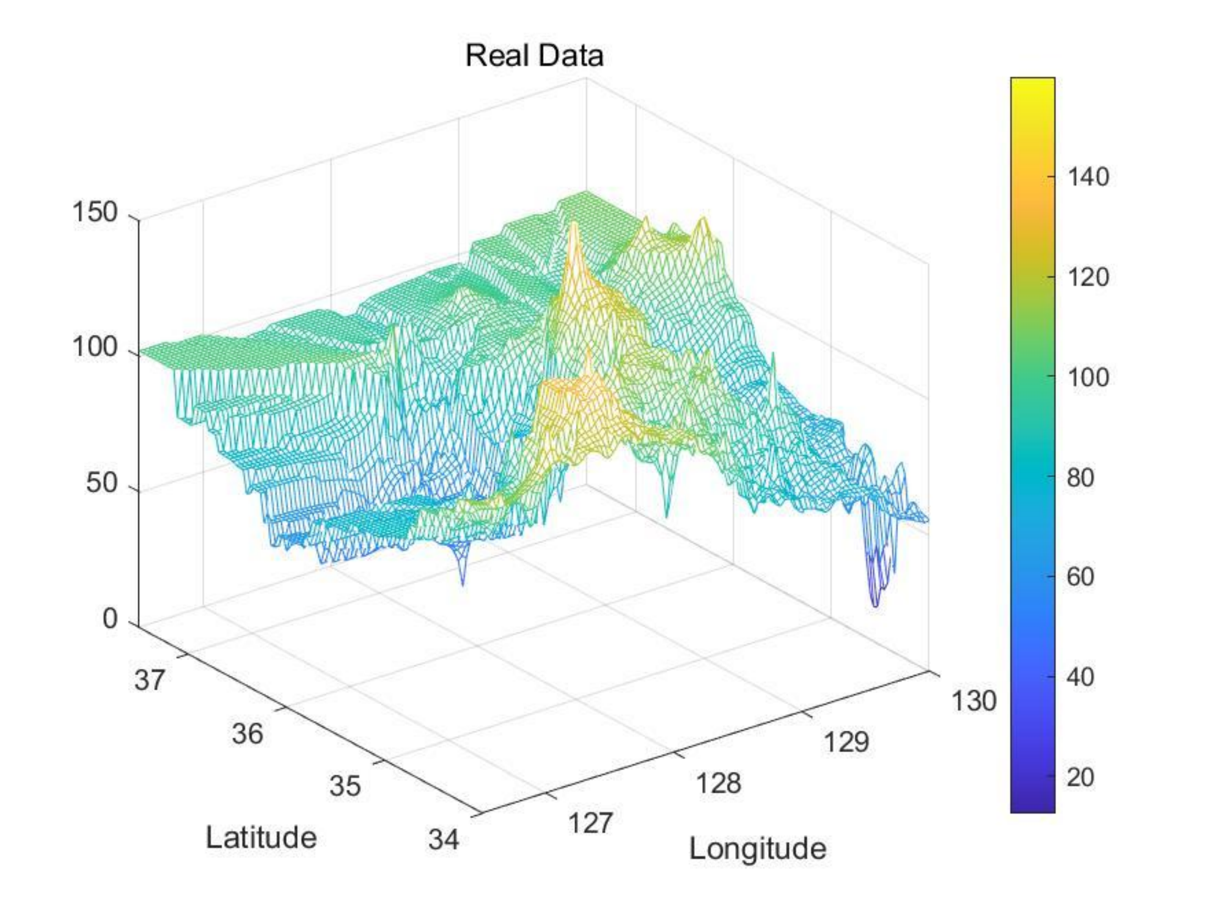


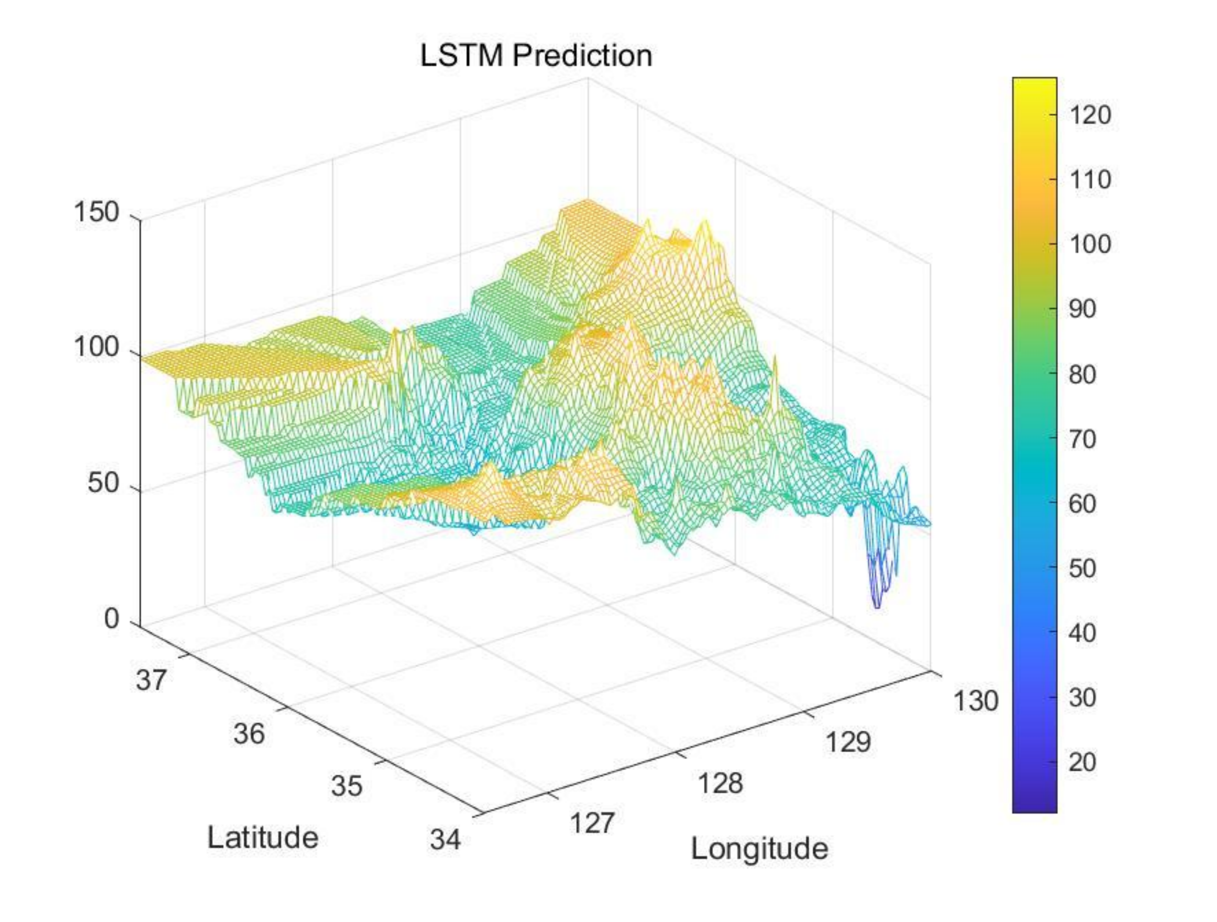


3. 결과 및 문제점

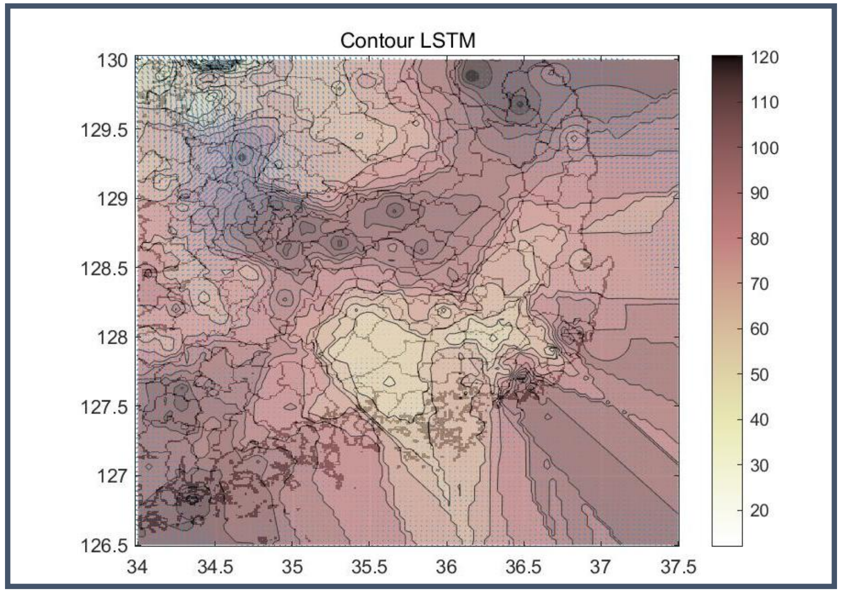
머신러닝 _ LSTM

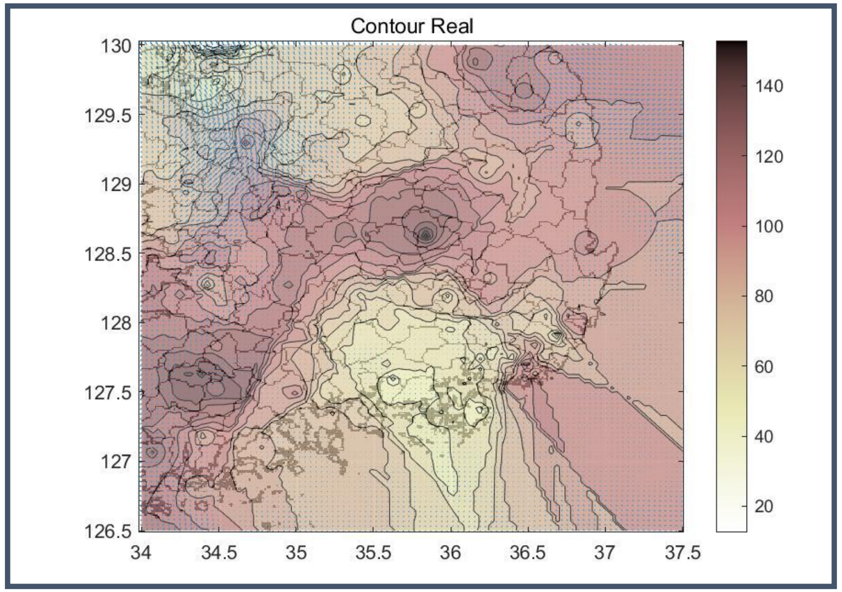
마지막 한시간에 대한 대기오염 농도

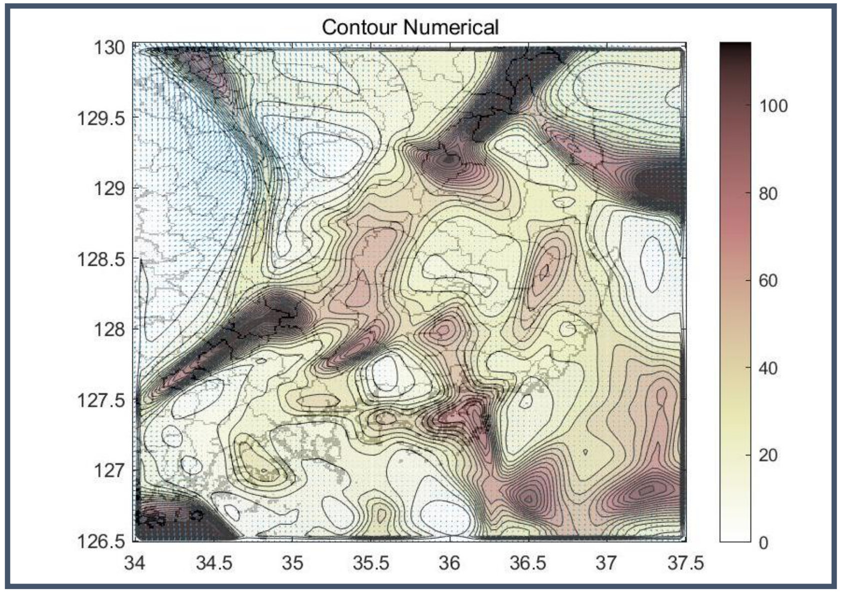


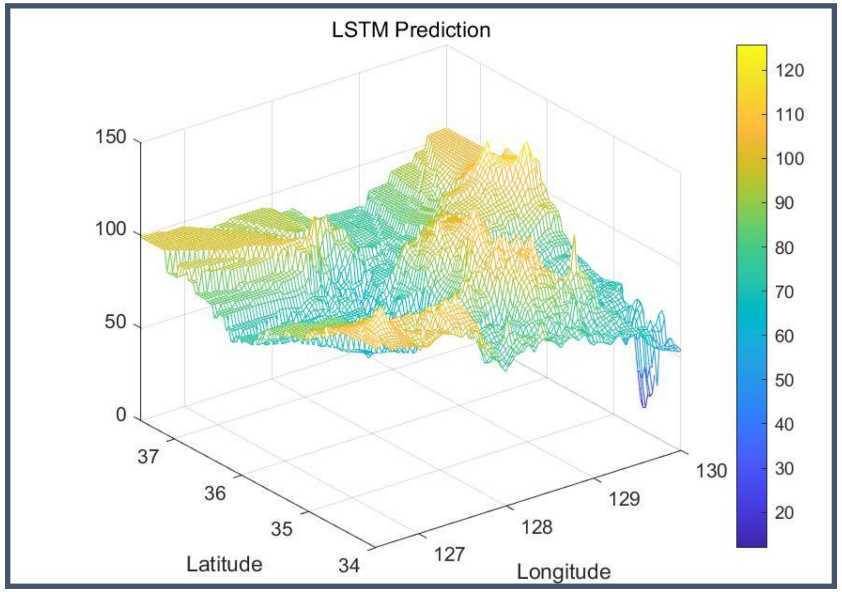


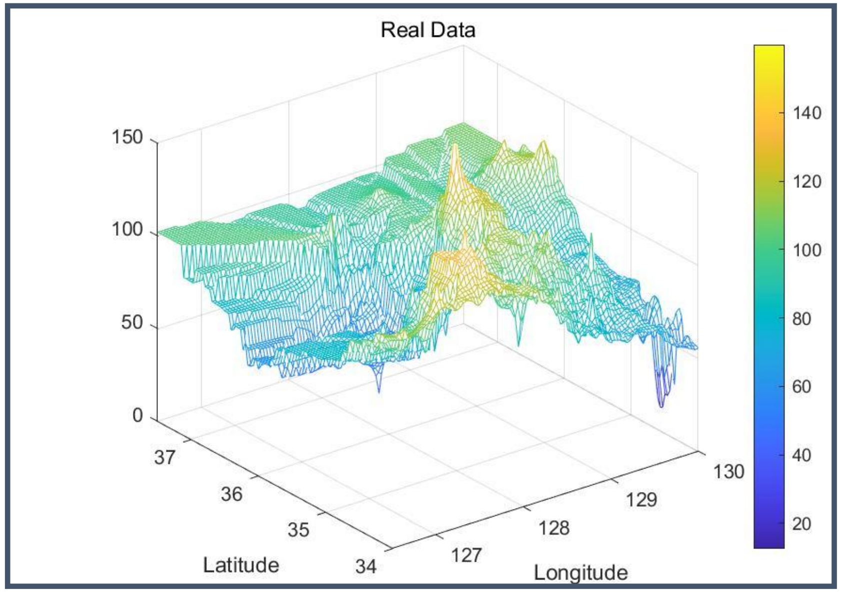
3. 결과 및 문제점

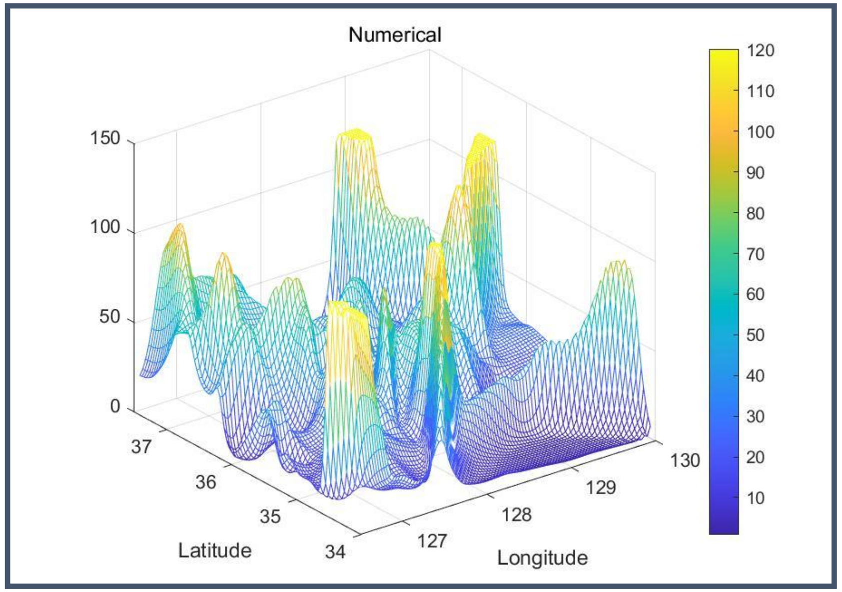

Contour LSTM


Contour Real


Contour Numerical


LSTM Prediction
Latitude
Longitude


Real Data
Latitude
Longitude


Numerical
Latitude
Longitude

4. 결론

두 방법론을
함께 사용하여
모델구상

데이터 측정 비용을 줄일 수 있고,
실시간 미세먼지 변화 양상을
누구나 쉽게 확인 가능

많은 사람들이 예측을
활용할 수 있도록
모델을 어플과 웹에
연동시키는 방안 고려

질의응답