<div align="center">

|  |  |  |
| --- | --- | --- |
| **2019** | **Report – 2019/ 5월1주** | **Seong-gu** |

</div>

**I. Plan for this week**

***Personal Research Subject***

* 최종모델링 완성할 것

**II. Plan for next week**

### China data input 부분 모델링 개선 (**Conv2DTranspose 이용)**

* 결과정리 및 본심 논문 1차 완성
* 창의 자율 과제 계획서 제출

**III. Response on Prof. Shin's feedback last week**

* Feature를 잘 선택 (데이터 프로세싱으로 정확도 개선)
* 잔차 활용 방법 검토
* 효율성 및 정확도 면에서 트레이드 오프 분석 설계 방안을 제기

JEET에 내는 것을 목표로 함.

**IV. Milestone**

### China data input 부분 모델링 개선 (**Conv2DTranspose 이용)**

* 잔차 활용 방법 검토

**V. Reports**

* 개인연구 진행사항

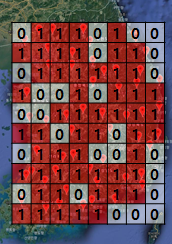
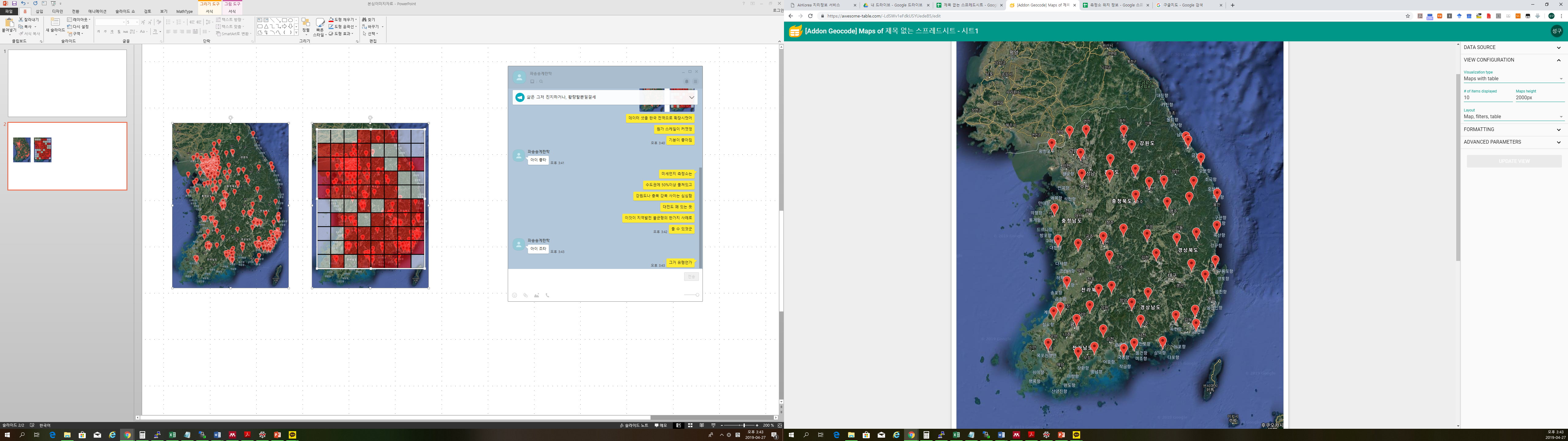
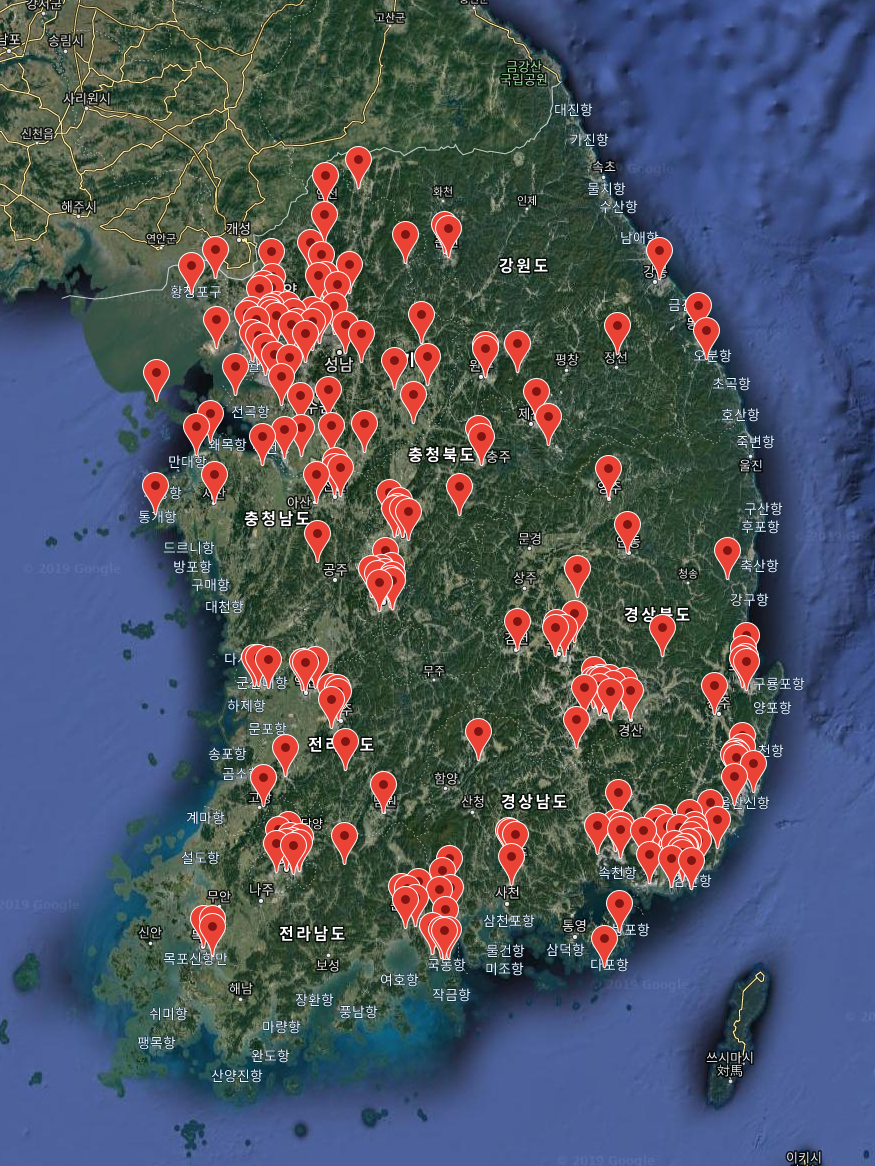
**1. 실험 설계**

* 1. **실험목표**

ConvGRU기반 인공신경망 기법을 이용하여, 국내 전역의 특정 시간 뒤의 지역별 미세먼지를 예측합니다.

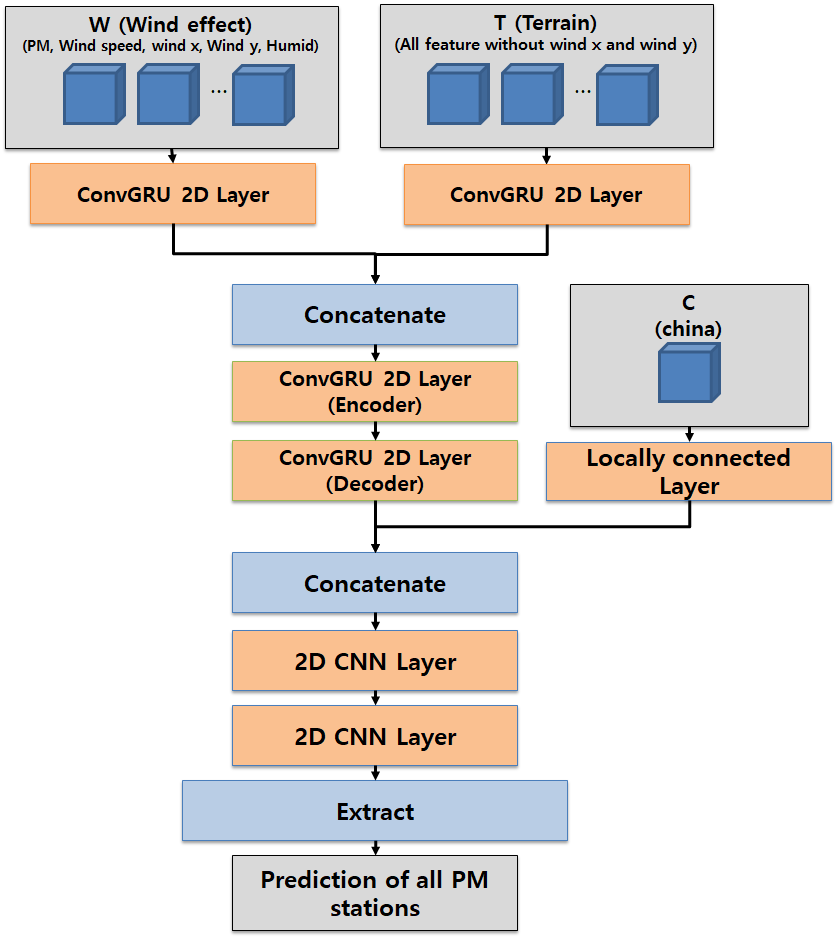
* 1. **실험데이터**

바람에 의한 유입 요인을 더 넓게 잘 반영하고 예측 문제를 쉽게 하기 위해서, 실험데이터의 공간해상도를 낮추고 범위를 국내 전역으로 넓혔으며, 8x10 grid로 구역을 나누고 각 구역에 포함된 측정소의 데이터를 평균해서 구역의 데이터를 대표하도록 하였습니다. 측정소가 하나도 포함되지 않는 구역은 주변 구역에 대해 IDW보간법을 적용하였습니다.



**왼쪽- 대기질 측정소, 오른쪽-기상측정소**

* 1. **모델링기법**



* 1. **실험가설**

(1) 지역별 미세먼지 농도를 동시에 예측하기 위해서, 기존 논문에 쓰이는 temporal한 요소만 반영한 LSTM(GRU)모델보다는 Spatial요소와 temporal 요소를 고려하는 ConvGRU나 3D CNN의 성능이 우수할 것이다.

-> 전시간대에 대해, ConvGRU 및 3D CNN과 LSTM(GRU)비교

\* 가설(1)은 현재 만족하지만 LSTM 결과가 예상보다 너무 나쁘게 나와, 모델링에 문제가 있는지 검토중입니다.

(2) 지역별 미세먼지 농도를 특정 시간대에 정확히 예측하기 위해서는, Recurrent구조를 가지면서 각 시간대별로 인접한 지역의 feature만 뽑아내면서 Spatial요소와 temporal요소를 ‘동시’에 고려하는 ConvGRU방식이 3DCNN방식보다 성능이 우수할 것이다..

-> 전시간대에 대해, ConvGRU방식을 차용한 모델(T,T+W,T+W+C)와 3D CNN 비교

(3) 풍향과 풍속 및 주변 미세먼지 농도를 적절히 feature extraction하면 국내미세먼지 유입요인을 잘 고려할 수 있기 때문에, 단기간 미세먼지 예측 성능을 높일 수 있을 것이다.

**-> 4시간~24시간 T와 T+W 비교**

(4) 중국 미세먼지 데이터와 풍향 풍속을 적절히 feature extraction하면 국외 유입요인을 고려할 수 있기 때문에, 중-장기간 미세먼지 예측 성능을 높일 수 있을 것이다.

**-> 24시간 T+W와 T+W+C 비교**

**2. 실험 결과**

**1시간 예측**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | T | T+W | T+W+C | 3D CNN | Follow |
| RMSE | 6.43 | 6.35 | 6.47 | 6.97 | 7.38 |
| MAE | 4.09 | 4.06 | 4.18 | 4.54 | 4.58 |
| IOA | 0.9854 | 0.9856 | 0.9851 | 0.9825 | 0.9812 |

가설(2)대로 ConvGRU기반 모델이 3DCNN보다 더 좋게 나왔으며, 가설(3)대로 풍향, 풍속, 주변 미세먼지를 고려한 W part를 결합한 T+W이 T보다 더 좋은 결과를 결과를 나타냈습니다.

**4시간 예측**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | T | T+W | T+W+C | 3D CNN | Follow |
| RMSE | 13.15 | 12.43 | 12.59 | 27.35 | 15.95 |
| MAE | 8.42 | 7.78 | 8.10 | 20.48 | 9.73 |
| IOA | 0.9260 | 0.9364 | 0.9389 | 0.1473 | 0.9071 |

이 결과에서는 3D CNN 결과에 문제가 있어 다시 실험할 예정입니다. T, T+W의 오차 격차가 더 벌어져 가설(3)을 만족합니다.

**12시간 예측**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | T | T+W | T+W+C | 3D CNN | Follow |
| RMSE | 20.36 | 19.68 | 19.80 | 20.88 | 24.04 |
| MAE | 13.37 | 12.74 | 12.69 | 13.71 | 14.79 |
| IOA | 0.7525 | 0.8013 | 0.8027 | 0.7657 | 0.7707 |

가설(2)대로 ConvGRU기반 모델이 3DCNN보다 더 좋게 나왔으며, T+W가 T보다 좋은 결과를 보이며 가설(3)을 만족합니다. 단, 12시간 예측에서도 중국데이터를 넣었을 때, 결과가 더 좋아진다는 가설(4)는 만족하지 못했습니다.

**24시간 예측**

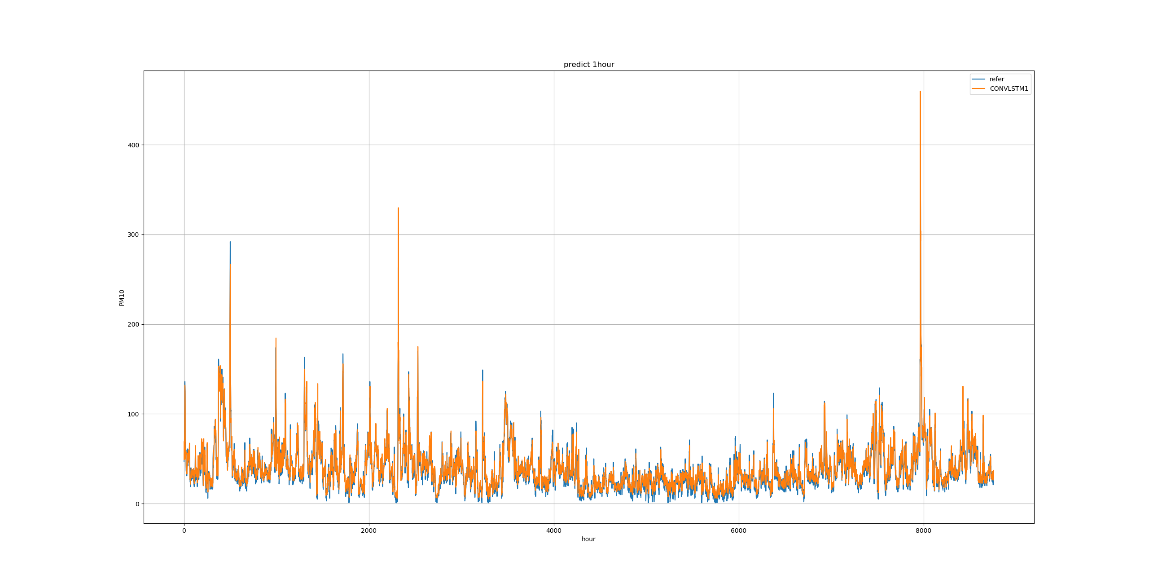
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | T | T+W | T+W+C | 3D CNN | Follow |
| RMSE | 24.73 | 23.05 | 22.70 | 23.14 | 27.58 |
| MAE | 17.20 | 16.07 | 15.84 | 15.65 | 17.47 |
| IOA | 0.6101 | 0.6482 | 0.6928 | 0.6504 | 0.6873 |

24시간 예측의 경우, 3D CNN이 T보다 좋은 결과를 나타내지만, 여전히 W+T이 더 좋은 결과를 나타내면서 가설(2)이 아직 만족한다고 볼 수 있습니다. 그러나 24시간 예측은 어려운 문제로 모델이 주어진 국내데이터로는 잘 학습하지 못해 3D CNN과 W+T의 격차가 거의 없는 것을 볼 수 있습니다. 따라서 이 경우에는 China data를 결합한 T+W+C 모델이 가장 우수한 성능을 보이면서 가설(4)를 만족합니다.

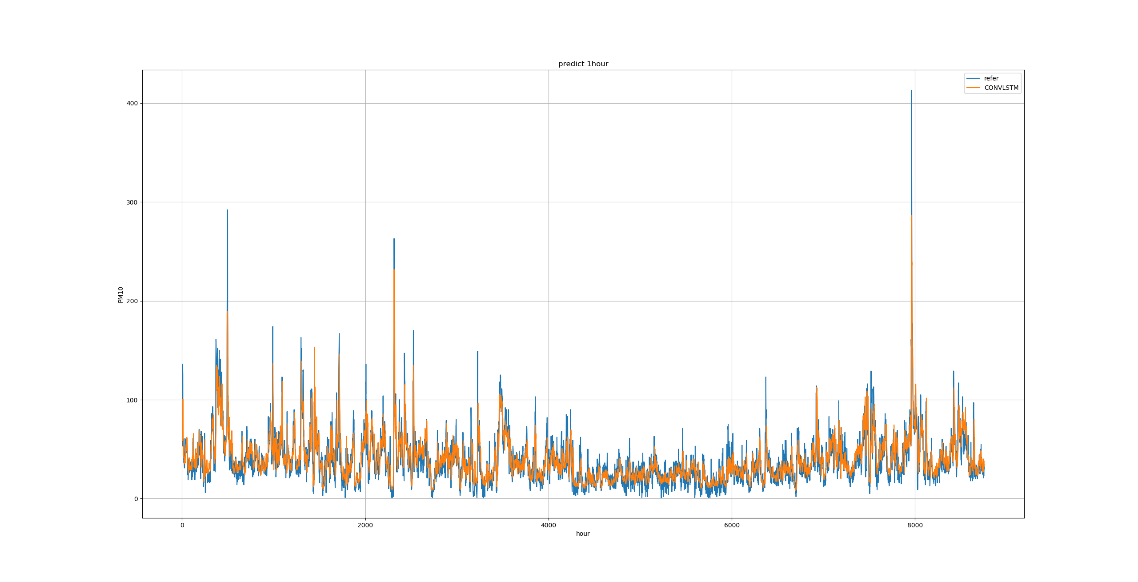
아래는 비교 그래프입니다.

2018.1.1-2018.12.31 예측 그래프 (파란색-reference, 주황색-가장 예측 잘된 모델)

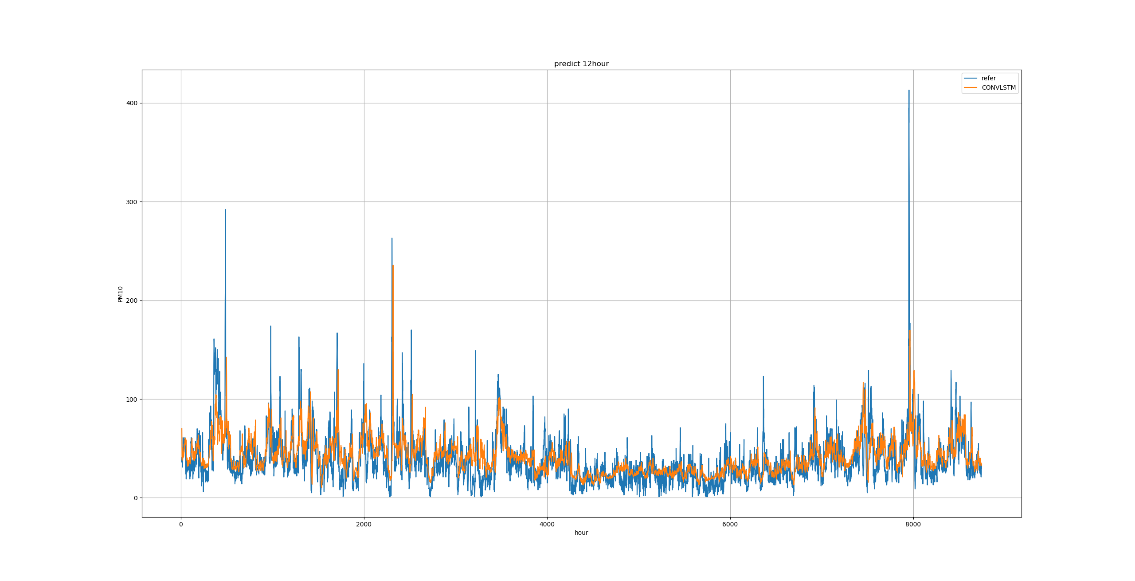
**1시간**

****

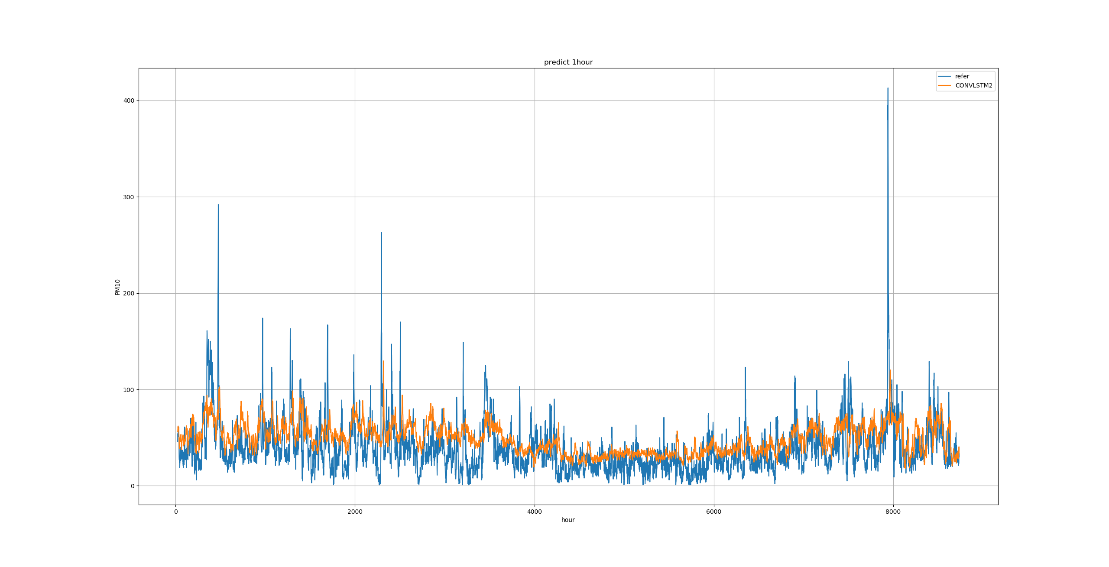
**4시간**

****

**12시간**

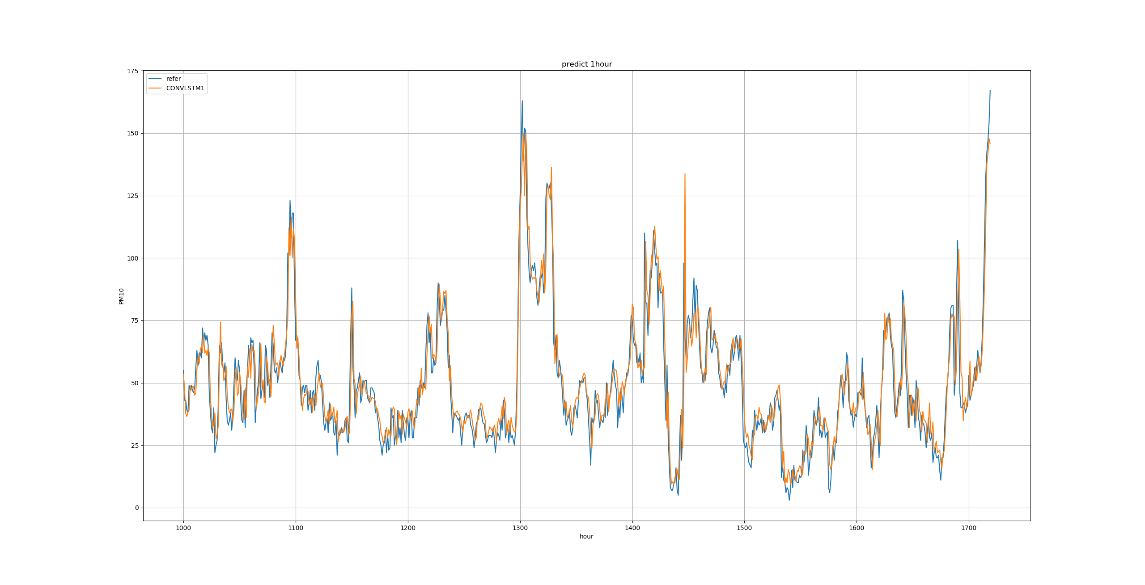
****

**24시간**

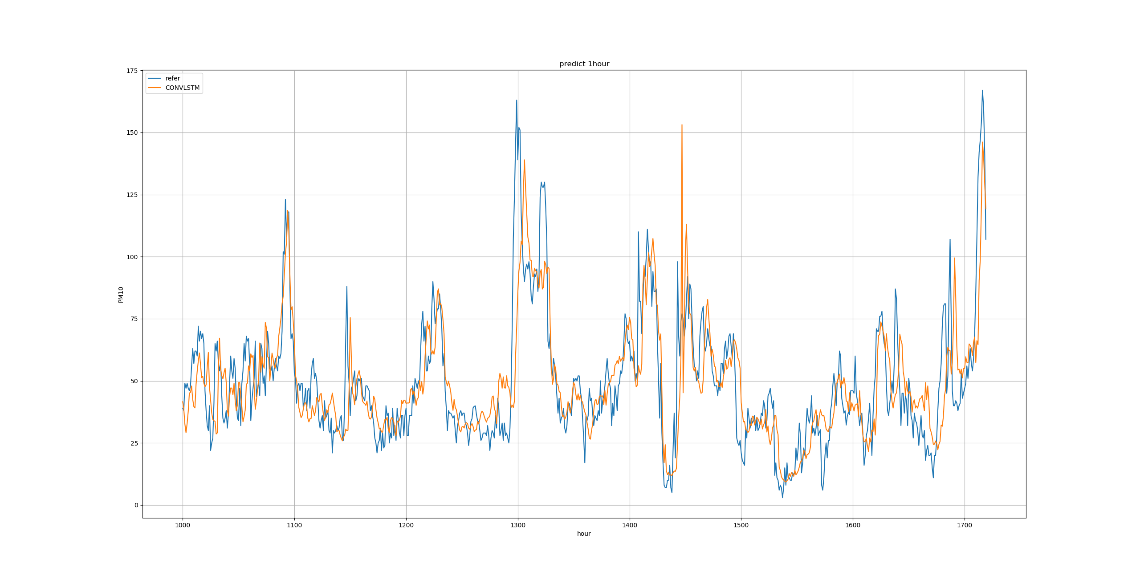
****

위 예측 그래프에서 일부 구간(한달)의 예측 그래프 (파란색-reference, 주황색-가장 예측 잘된 모델)

**1시간**

****

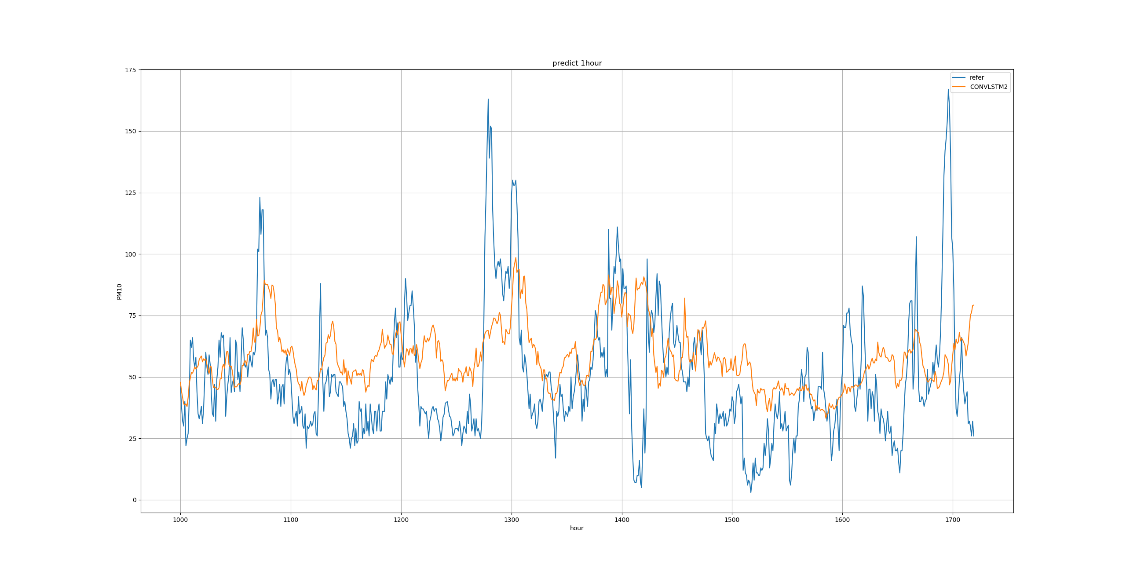
**4시간**

****

**12시간**

****

**24시간**

****