# Utilizing urban IoT sensor data to build more advanced air temperature information using machine learning

Ayano Aida*1, Chan Park*2

*1 Dept. of Urban Planning and Design, University of Seoul *2 Dept. of Landscape Architecture, University of Seoul

## Introduction

- Extreme heat is becoming an important issue in cities around the world.
- In order for cities to adapt to extreme heat, they first need information about the hazard.
  - To obtain this information, machine learning is one of the methods used to predict and mapping air temperatures in cities
  - Urban IoT sensors are being installed around the world to obtain more information.

## Purpose

1. To understand how much Urban IoT sensors affect performance in a machine learning model for predicting urban temperatures.
2. Understand how the potential population exposed to extreme heat varies with the combination of model inputs.

## Method

### 1. Estimation of the daily maximum air temperature (Tmax) by machine learning algorithm.

**Label Data :** Two types of station networks were used: the Korea Meteorological Administration's automatic weather observation stations (AWS) and S·DoT, an Urban IoT sensors operated by Seoul Metropolitan Government.



**What is AWS**
- AWS is an automated weather station operated by the Korea Meteorological Administration (KMA).
- The AWSs are installed according to KMA's strict regulations, and the Ta sensor is placed 1.2 to 2.0 m above the installation surface.

**What is S·DoT**
- S·DoT is a ground-based observation system operated by Seoul since April 2020, and it collects ten types of data, including Ta, PM10, PM2.5, humidity, illumination, and noise.
- There are more than 855 stations in Seoul, and most of these stations are installed on CCTV poles at a height of approximately 2–3 m.

**Model Setting :** Model algorithm – Random Forest, Test size – 20%

| Label data | Common predictors | Category | Model name | Additional Predictors | Num. of Samples |
|---|---|---|---|---|---|
| AWS | X, Y : XY coordinates<br>DOY : Day of year<br>LC : Land Cover<br>Water : Distance from fresh water<br>Sea : Distance from fresh sea<br>Green : Distance from green space<br>DEM : Digital Elevation Model<br>Slope<br>Aspect | Including data from Landsat 8 | M1 | LST, NDVI | 133 |
| | | | M1_IDW | LST, NDVI, IDW_S-DoT | 133 |
| | | | M1_Near1 | LST, NDVI, One NearT_S-DoT, Dist. to One Near_S-DoT | 133 |
| | | | M1_Near3 | LST, NDVI, Three NearT_S-DoT, Dist. to Three Near_S-DoT | 133 |
| | | Including only representative NDVI values for that year | M2 | NDVIyear | 6,570 |
| | | | M2_IDW | NDVIyear, IDW_S-DoT | 6,570 |
| | | | M2_Near1 | NDVIyear, One NearT_S-DoT, Dist. to One Near_S-DoT | 6,570 |
| | | | M2_Near3 | NDVIyear, Three NearT_S-DoT, Dist. to Three Near_S-DoT | 6,570 |
| | | Without Landsat 8 data | M3 | | 7,758 |
| | | | M3_IDW | IDW_S-DoT | 7,758 |
| | | | M3_Near1 | One NearT_S-DoT, Dist. to One Near_S-DoT | 7,758 |
| | | | M3_Near3 | Three NearT_S-DoT, Dist. to Three Near_S-DoT | 7,758 |
| S·DoT | LCZ : Local Climate Zone<br>FAR : Floor area ratio<br>SR : Daily cumulative solar radiation | Including data from Landsat 8 | M4 | LST, NDVI | 6,984 |
| | | | M4_IDW | LST, NDVI, IDW_AWS | 6,984 |
| | | Including only representative NDVI values for that year | M5 | NDVIyear | 344,174 |
| | | | M5_IDW | NDVIyear, IDW_AWS | 344,174 |
| | | Without Landsat 8 data | M6 | | 455,749 |
| | | | M6_IDW | IDW_AWS | 455,749 |

### 2. Estimation extreme heat exposed population by de facto population data

- The potentially exposed populations was considered as the de facto population within the grid with predicted temperatures higher than two benchmarks, 31°C and 35°C.
- The de facto population was estimated based on telecommunication company data for each output area (OA) and hour.
  - For this study, the de facto population data for July 21, 2021 at 14:00, the day with the highest temperature among the days for which all predictors could be obtained, was used.
  - The de facto population of each grid was calculated by allocating it to the developed land cover and artificial green spaces within OA.
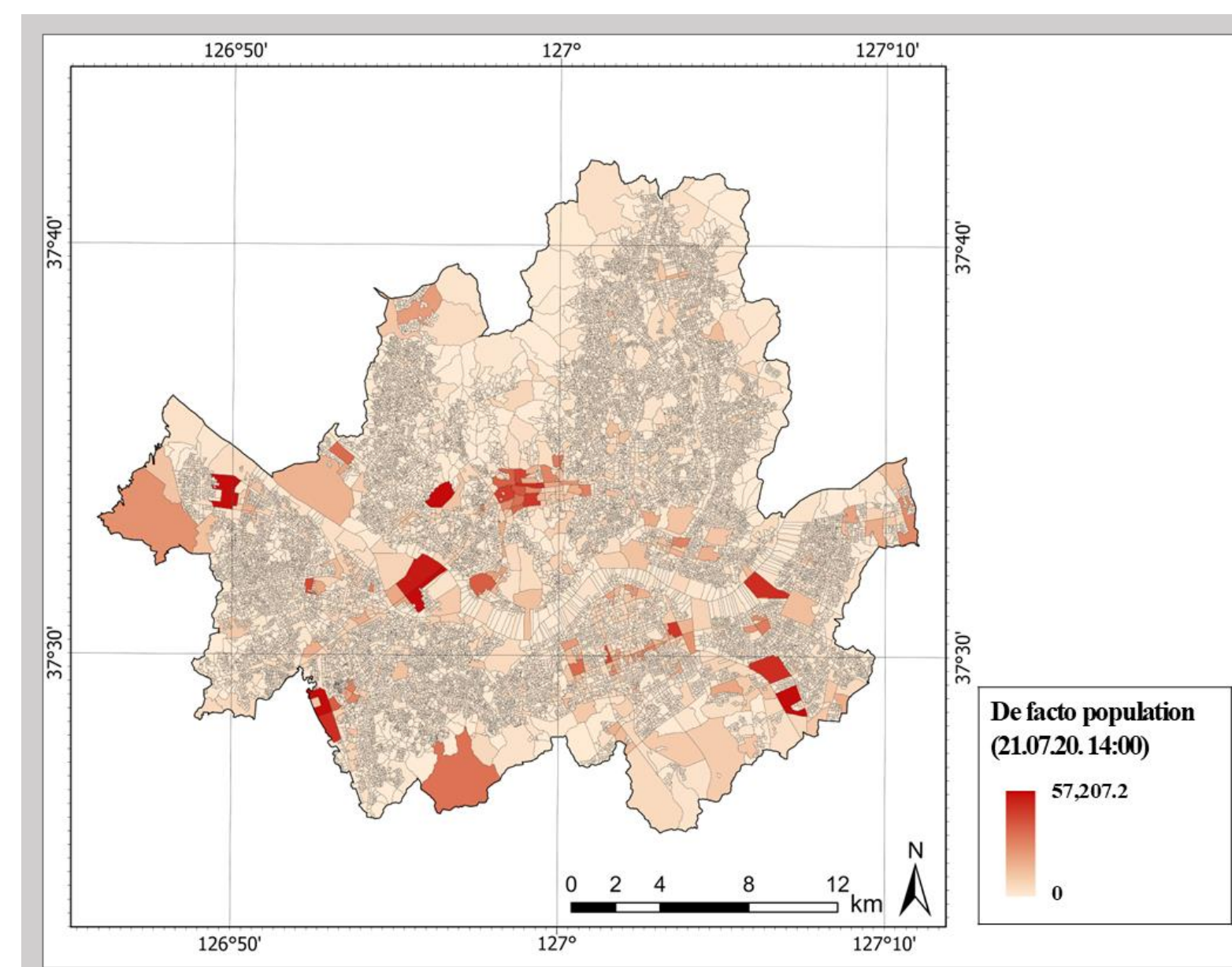


## Results

○ **Estimation of the daily maximum air temperature (Tmax) and extreme heat exposed population**

**Model Performance with different predictor combinations**

| Model name | Label data | Additional Predictors | Num. of Samples | MSE | RMSE | R² | Rank | At predicted Tmax of 31°C Pct. of EP (%) | At predicted Tmax of 31°C Pct. of EP aged 70 and over (%) | At predicted Tmax of 35°C Pct. of EP (%) | At predicted Tmax of 35°C Pct. of EP aged 70 and over (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| m1_base | AWS | LST, NDVI | 133 | 0.85 | 0.92 | 0.83 | | 0.67 | 0.87 | 0 | 0 |
| m1_IDW | AWS | LST, NDVI, IDW_S-DoT | 133 | 0.97 | 0.98 | 0.80 | | 0.72 | 0.98 | 0 | 0 |
| m1_Near1 | AWS | LST, NDVI, One NearT_S-DoT, Dist. to One Near_S-DoT | 133 | 0.79 | 0.89 | 0.84 | | 0.7 | 0.95 | 0 | 0 |
| m1_Near3 | AWS | LST, NDVI, Three NearT_S-DoT, Dist. to Three Near_S-DoT | 133 | 1.00 | 1.00 | 0.80 | | 0.71 | 0.96 | 0 | 0 |
| m2_base | AWS | NDVIyear | 6,570 | 5.06 | 2.25 | 0.56 | | 0.48 | 0.61 | 0 | 0 |
| m2_IDW | AWS | NDVIyear, IDW_S-DoT | 6,570 | 0.79 | 0.89 | 0.93 | 3rd– AWS | 0.72 | 0.97 | 0.01 | 0.01 |
| m2_Near1 | AWS | NDVIyear, One NearT_S-DoT, Dist. to One Near_S-DoT | 6,570 | 1.06 | 1.03 | 0.91 | | 0.69 | 0.93 | 0.11 | 0.16 |
| m2_Near3 | AWS | NDVIyear, Three NearT_S-DoT, Dist. to Three Near_S-DoT | 6,570 | 0.78 | 0.88 | 0.93 | 2nd– AWS | 0.71 | 0.96 | 0.01 | 0.02 |
| m3_base | AWS | | 7,758 | 12.12 | 3.48 | −0.08 | | 0.51 | 0.68 | 0 | 0 |
| m3_IDW | AWS | IDW_S-DoT | 7,758 | 0.83 | 0.91 | 0.93 | | 0.73 | 0.98 | 0.01 | 0.01 |
| m3_Near1 | AWS | One NearT_S-DoT, Dist. to One Near_S-DoT | 7,758 | 0.88 | 0.94 | 0.92 | | 0.7 | 0.94 | 0.06 | 0.08 |
| m3_Near3 | AWS | Three NearT_S-DoT, Dist. to Three Near_S-DoT | 7,758 | 0.56 | 0.75 | 0.95 | 1st– AWS | 0.72 | 0.96 | 0.01 | 0.02 |
| m4_base | S·DoT | LST, NDVI | 6,984 | 1.29 | 1.14 | 0.69 | | 0.69 | 0.93 | 0.01 | 0.02 |
| m4_IDW | S·DoT | LST, NDVI, IDW_AWS | 6,984 | 1.16 | 1.08 | 0.72 | 3rd–S·DoT | 0.72 | 0.98 | 0.01 | 0.01 |
| m5_base | S·DoT | NDVIyear | 344,174 | 7.50 | 2.74 | 0.39 | | 0.58 | 0.75 | 0.01 | 0.01 |
| m5_IDW | S·DoT | NDVIyear, IDW_AWS | 344,174 | 1.02 | 1.01 | 0.92 | 1st–S·DoT | 0.71 | 0.96 | 0.01 | 0.02 |
| m6_base | S·DoT | | 455,749 | 12.93 | 3.60 | −0.01 | | 0.66 | 0.87 | 0 | 0 |
| m6_IDW | S·DoT | IDW_AWS | 455,749 | 1.11 | 1.05 | 0.91 | 2nd–S·DoT | 0.72 | 0.97 | 0.02 | 0.03 |

**M1_Baseline model**



**M3_Near3 model**
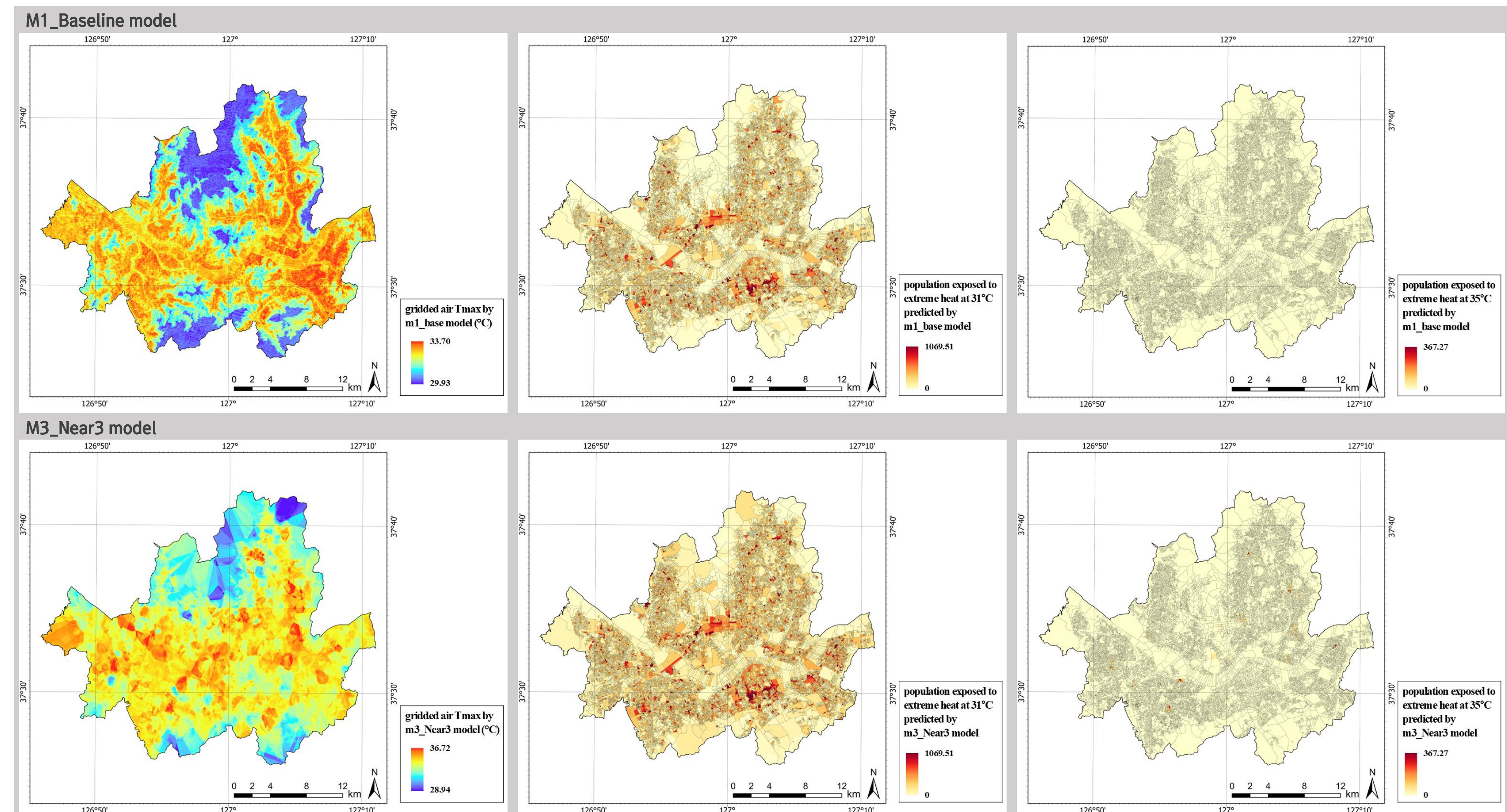


## Discussion & Conclusion

- The model performance improved when data from Urban IoT sensors was used as a predictor rather than when LST derived from satellite images was used.
- The population potentially exposed to heat was derived more when IoT sensors were used than when LST was used.
- These results show the usefulness of installing Urban IoT sensors.
- We plan to confirm the performance of each station and the importance of each variable of the label data.