

Local-Scale Urban Heat Island Modeling: A Machine Learning Approach Using Citizen Data

Tanguy Houget^{1,2}, Valeria Garbero³, Pietro Salizzoni^{1,2}

(1) Politecnico di Torino

(2) Ecole Centrale de Lyon

(3) Arpa Piemonte

Parallel Session WG PHY URBAN,

GM2026

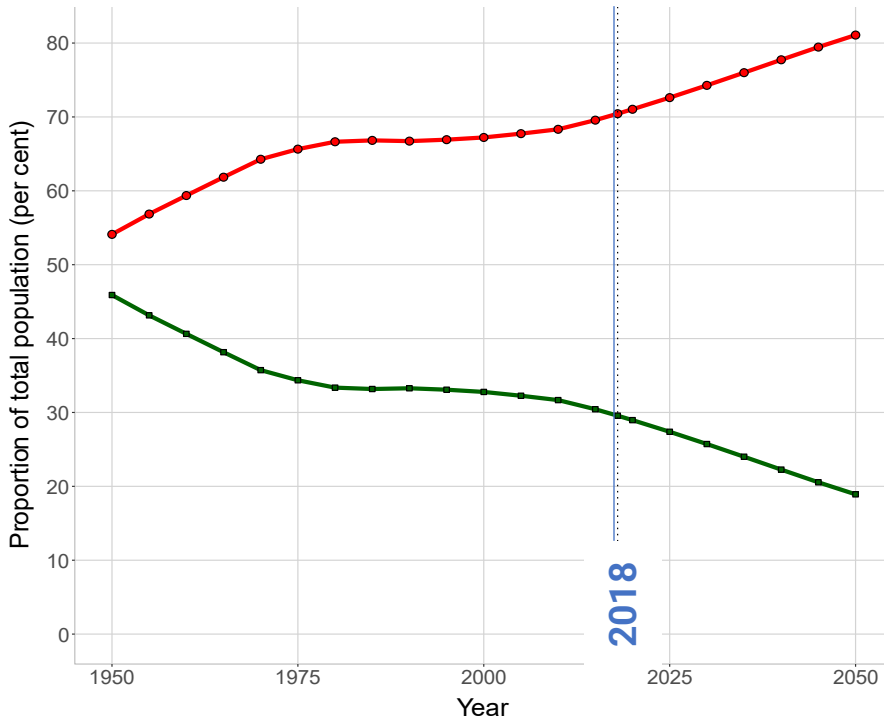
Urban Heat Waves: A Public Health Crisis

In a climate scenario characterized by increasingly intense heat waves, **cities** are the most critical environment for two reasons:

> EXPOSITION

Percentage of population in urban and rural areas Italy

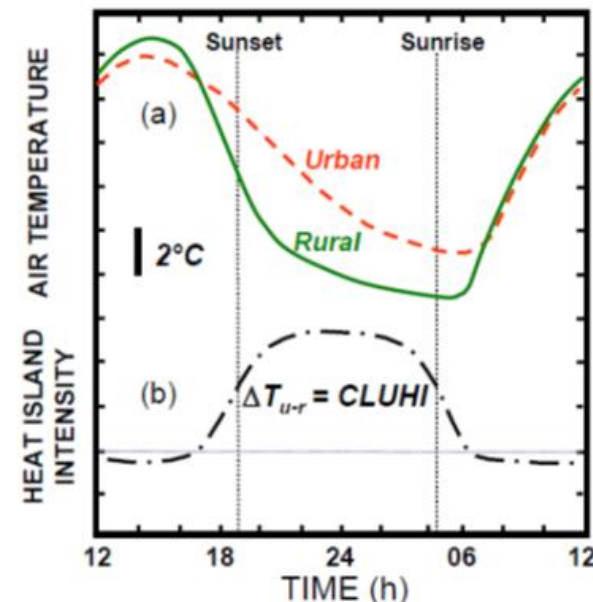
Urban Rural



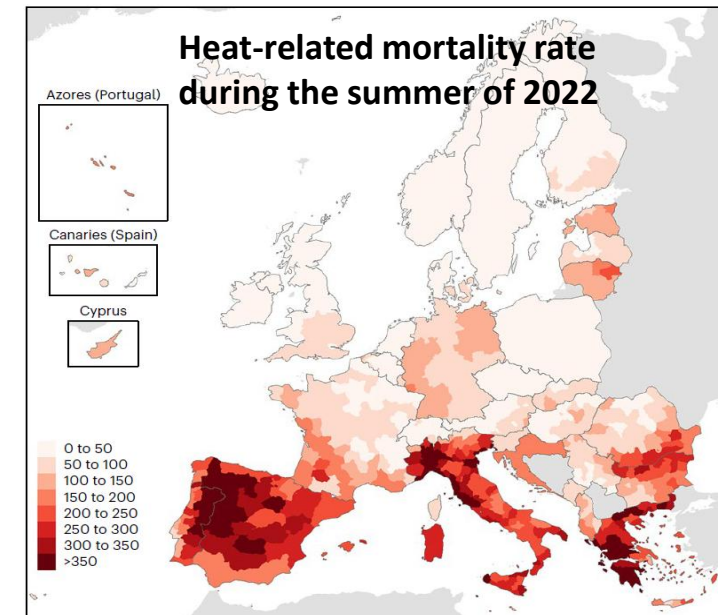
World Urbanization Prospects: The 2018 Revision
<https://population.un.org/wup/countryprofiles?country=Italy>

> HAZARD

The Urban Heat Island (UHI) amplifies heat wave impacts by reducing nighttime cooling



> RISK



The health risk associated with high temperatures for citizens is also rising!



Local-scale UHI modeling is needed

Limits of Existing Models

Numerical simulations (meteorological models)

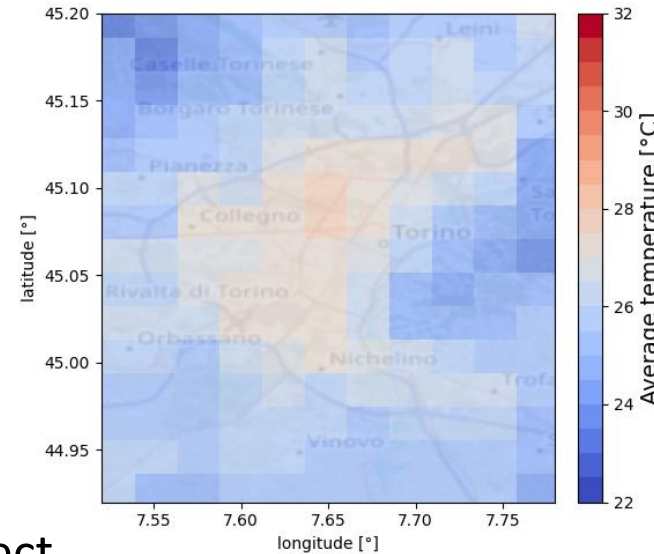
- Operational model ICON-2I @2.2 km resolution
- ICON+TERRA_URB in hindcast mode @1 km resolution

are limited in resolution :

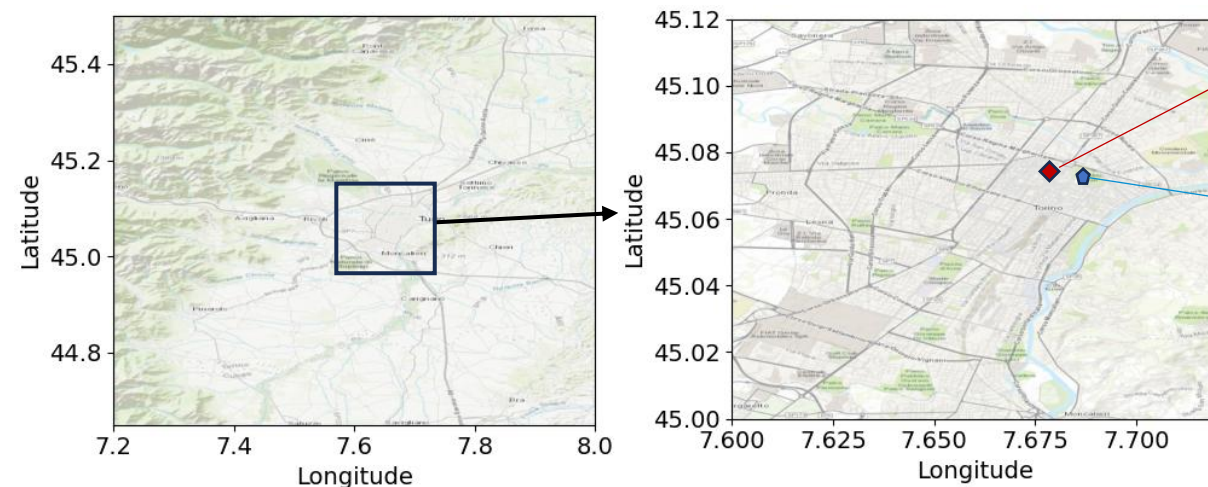
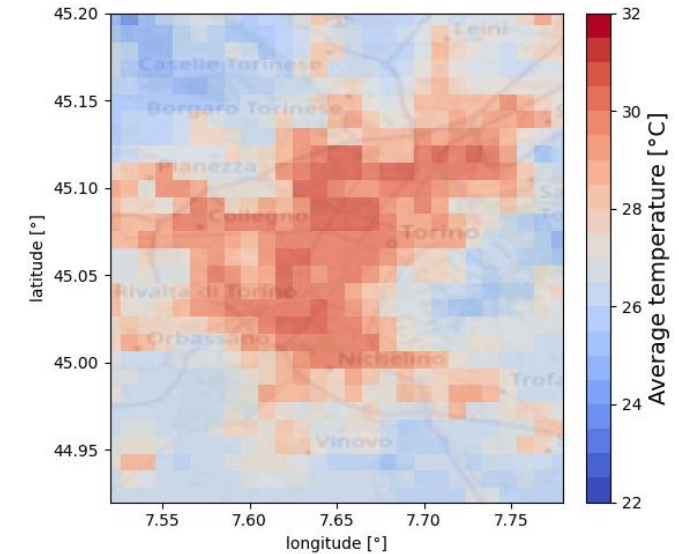
- Longer simulation times
- Need for better hardware
- Harder parameterisation

TERRA_URB improves the reproduction of UHI effect but cannot capture micro-scale spatial variability

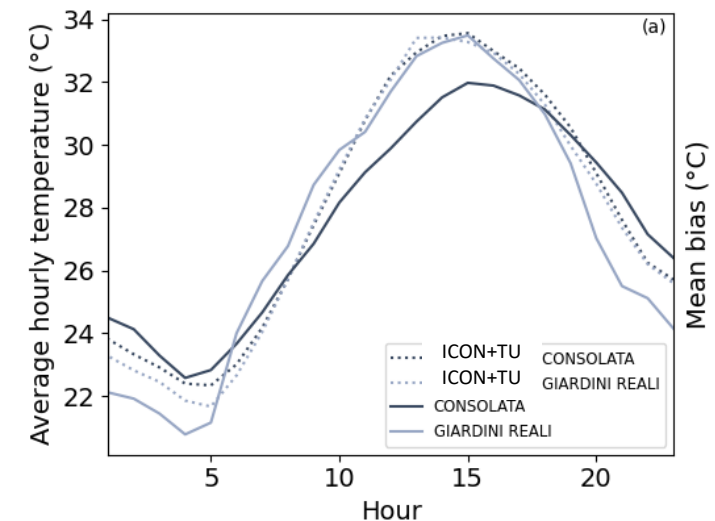
ICON-2I



ICON+TU



Consolata station, located in a densely urbanized area, and Giardini Reali station, situated within an urban park, are approximately 1 km apart



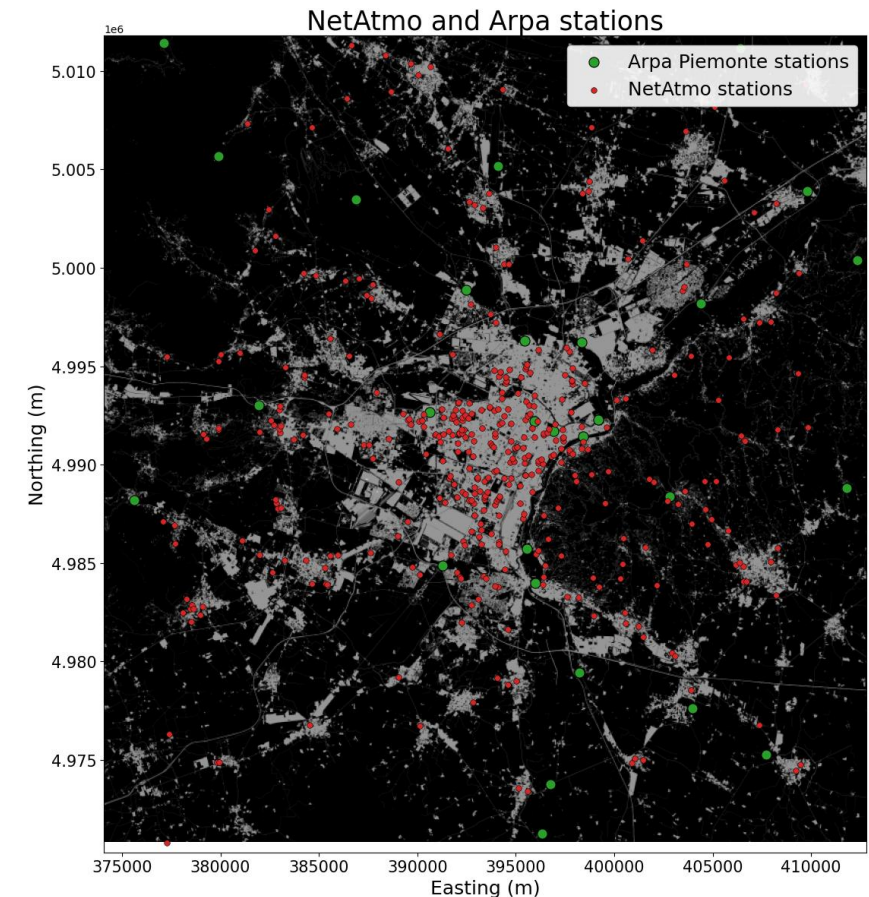
Citizen Data Meets Machine Learning

Machine Learning: trying to capture urban micro-scale spatial variability

The high density of Citizen weather stations (CWS) represents a potential to train ML models

Netatmo CWS (<https://weathermap.netatmo.com/>)

denser but.. accurate? A rigorous Quality Control (QC) procedure is needed!



- 1 A Python script to download NetAtmo data via API has been developed
- 2 **A new quality control procedure** has been developed in Python specifically for NetAtmo temperature data, based on Napoly method
- 3 It was then compared with the **CrowdQC+** QC technique developed by Fenner et al. for NetAtmo data

PP APOCS

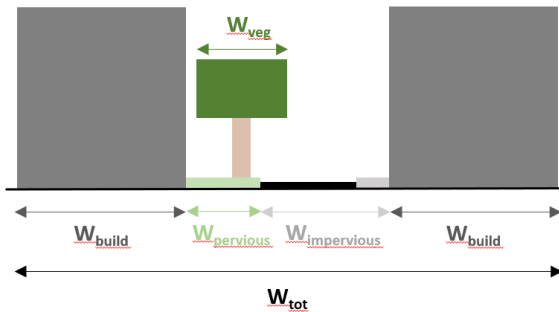
SubTask 2.3 Quality control procedure for T2m and Rh2m

2D surface parameterisation

Input features in the Machine Learning approach for local-scale UHI modeling

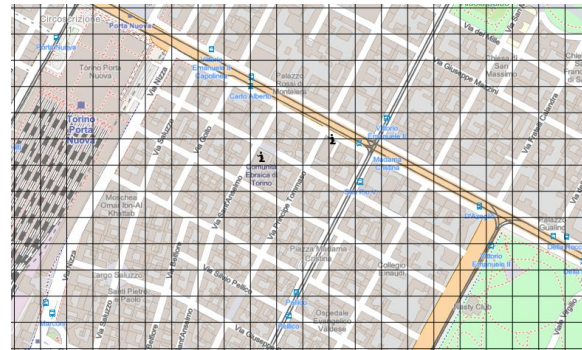
Simplification used in our work:

The urban street canyon



Parameters :

1. Geometry of built structures
2. Surface materials
3. Tree coverage



Surface parameters on a 50 m resolution grid



2D bulk parameters

- 1. Geometry of built structures**
 - ABH : Average building height
 - BUILD_SF : Building surface fraction
- 2. Surface materials**
 - IMP_SF : Impervious surface fraction
 - NDVI : Normalized Difference Vegetation Index
 - LST : Land Surface Temperature
- 3. Tree coverage**
 - VEG_SF : Vegetation surface fraction
 - SVF : Sky View Factor
(with effect of trees)

Additional parameters :

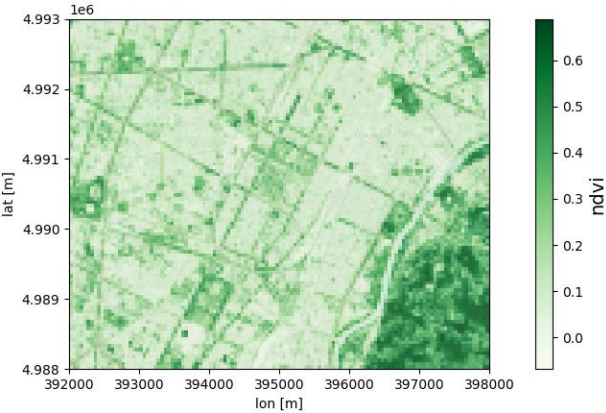
- Altitude

Total: 8 parameters

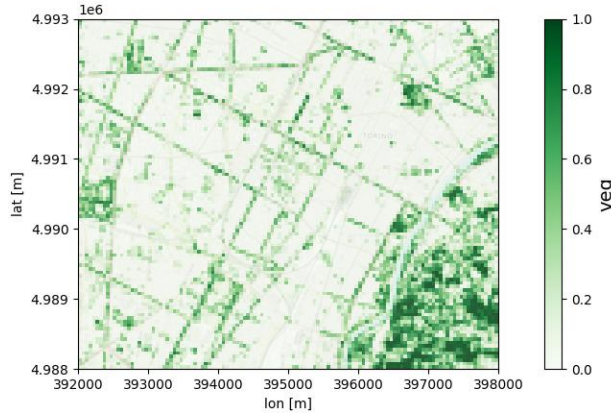
Urban parameterisation

Construction of 50m resolution maps for 8 parameters over Turin and surroundings:

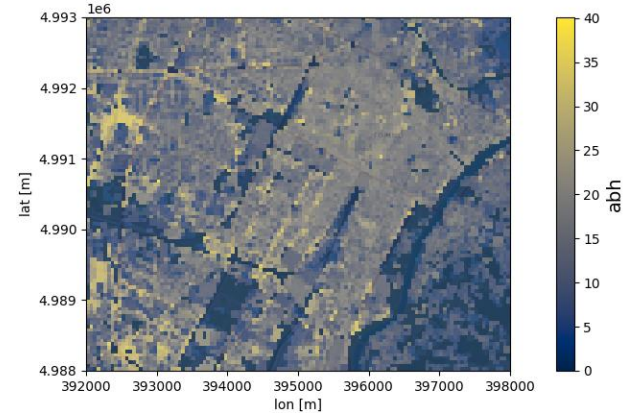
NDVI : Normalized Difference Vegetation Index



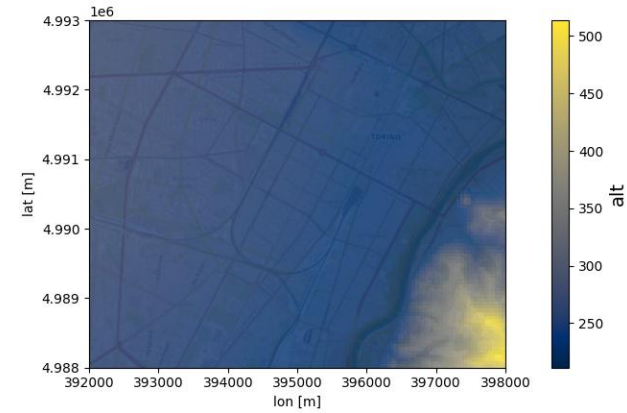
VEG_SF : Vegetation Surface Fraction



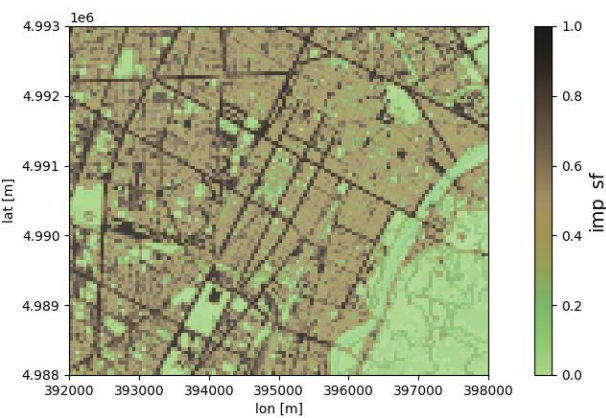
ABH : Average Building Height



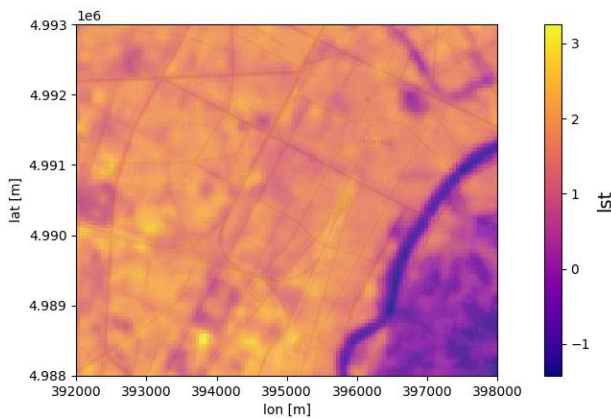
ALT : Altitude



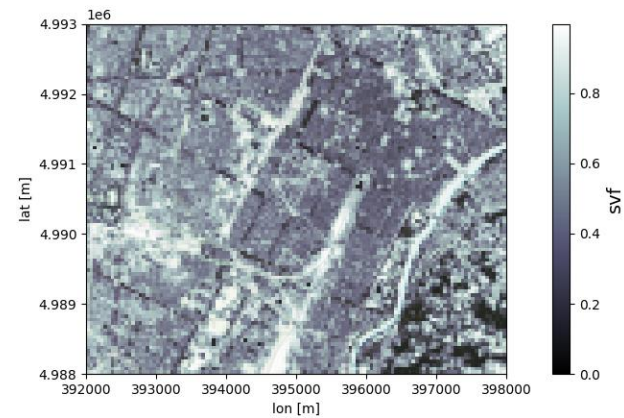
IMP_SF : Impervious Surface Fraction



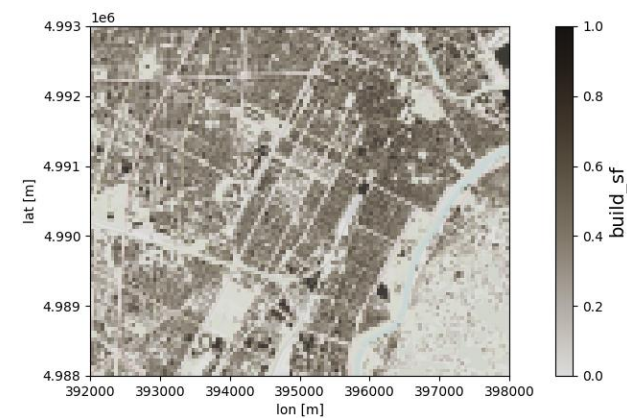
LST : Land Surface Temperature



SVF : Sky View Factor

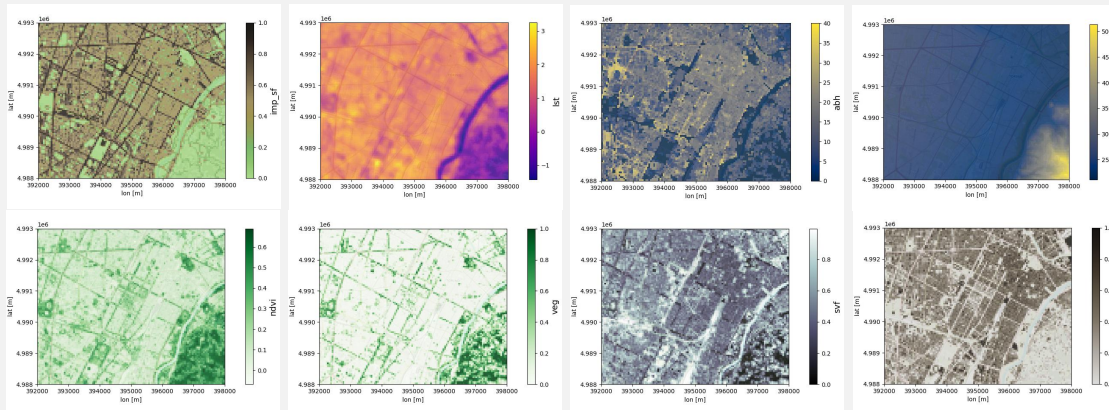


BUILD_SF : Building Surface Fraction



Simplified pipeline of the methodology

I. Urban parameterisation : 8 features



Additional training variables : 2 features

- **ICON** : Regional meteorological model
- **time**

II. Training

Model :
Convolution Neural Network

Training :
Iterative optimisation of the model based on its mistakes on training examples

but also ...

- Estimation of the error of the model on unseen data
- Estimation of the relative importance of the model inputs

III. Maps

Use of the model on unknown situations

1. **Reconstruction :**

Estimation of the temperature at locations without sensors, in the **current state** of the city

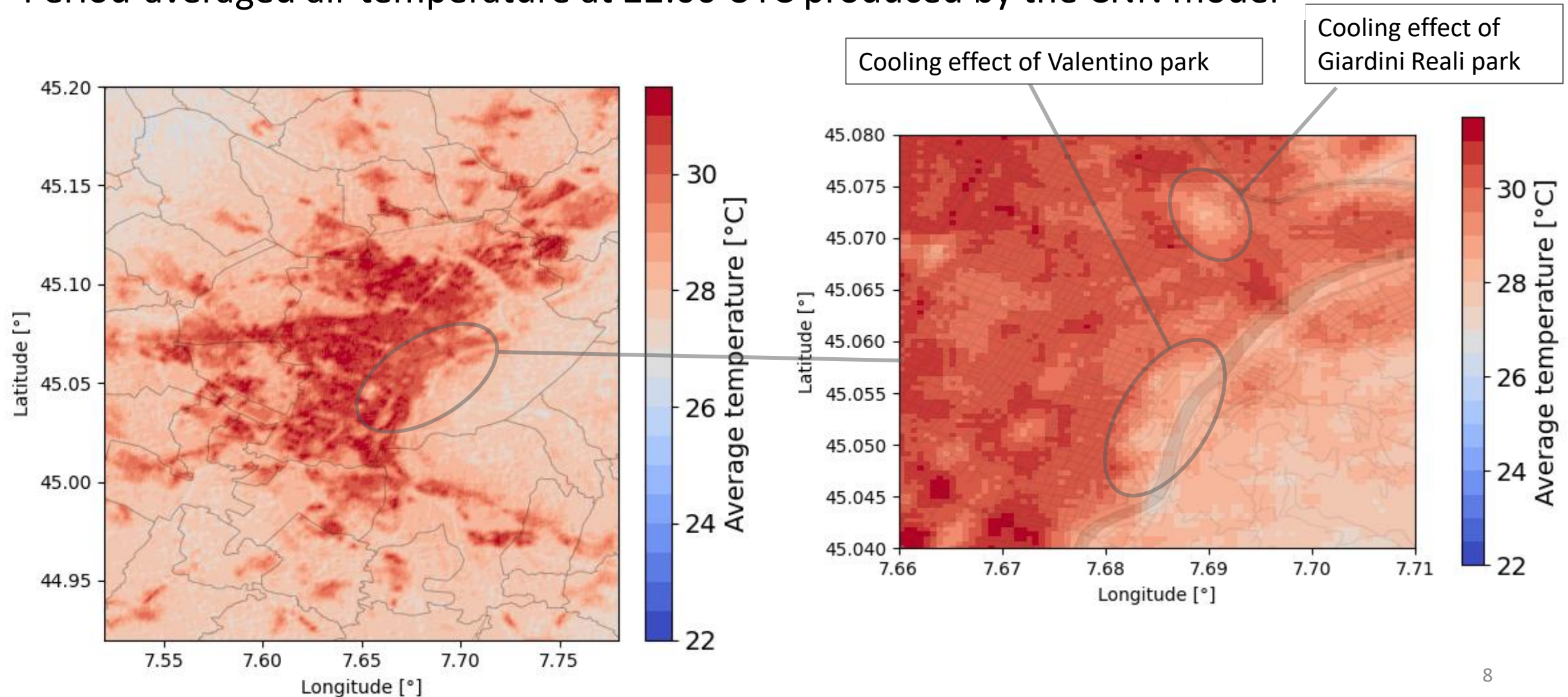
2. **Impact of scenarios :**

Estimation of the temperature everywhere, in a **modified state** of the city

Reconstruction of air temperature maps

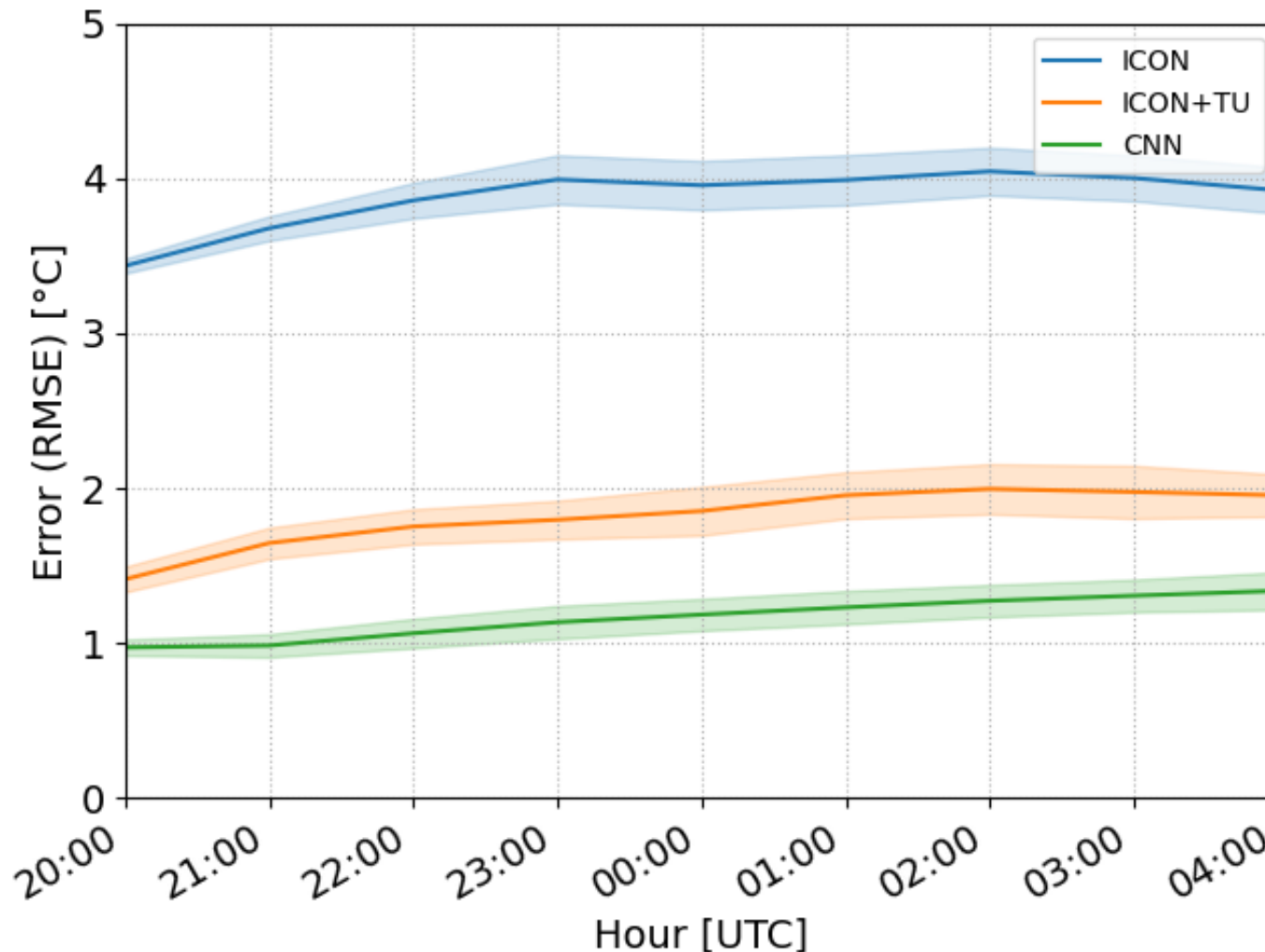
..during a past heat wave 22 to 29 June 2025

Period-averaged air temperature at 22:00 UTC produced by the CNN model



Error of the models

Comparison of errors between the operational ICON-2I model, the ICON+TERRA_URB run at 1 km resolution, and the CNN model



- CNN model improves the UHI mapping by reducing the RMSE error ✓
- The CNN model includes ICON-2I t2m as an additional feature ✓
- Future work will integrate ICON+TU t2m

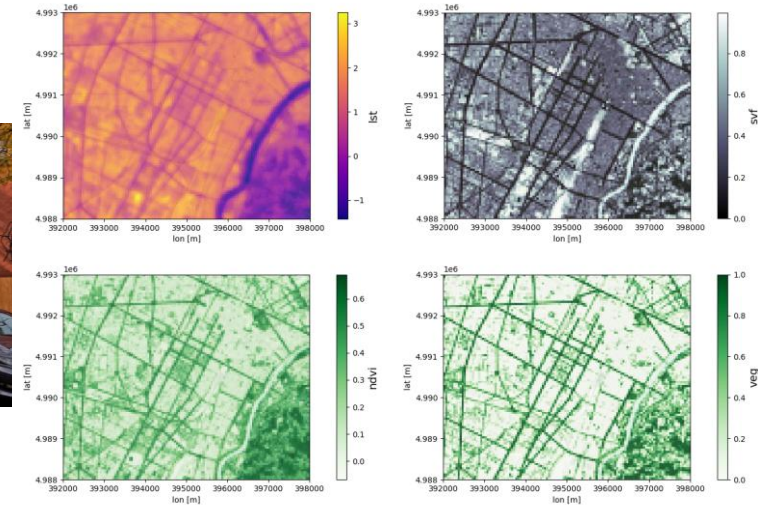
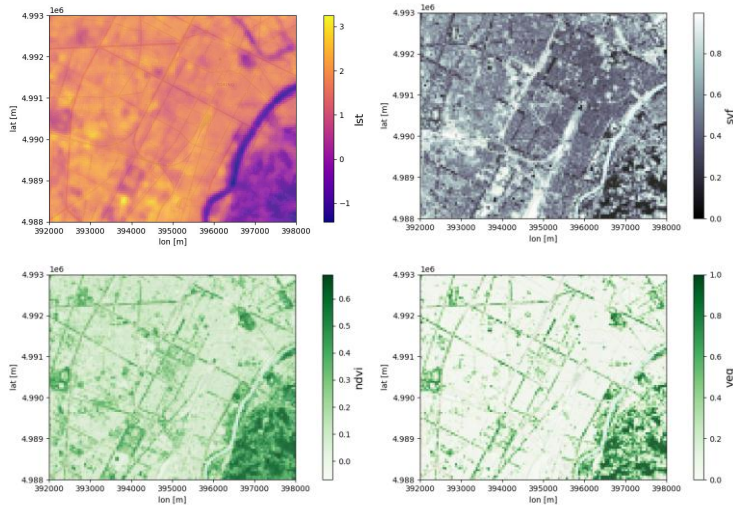
Urban parameterisation : Tree planting scenario

Vegetalisation scenario :

- Trees planted in Turin municipality only
- Trees planted in streets of 15m wide or more
- Effect of vegetalisation on satellite imagery calibrated on streets with trees

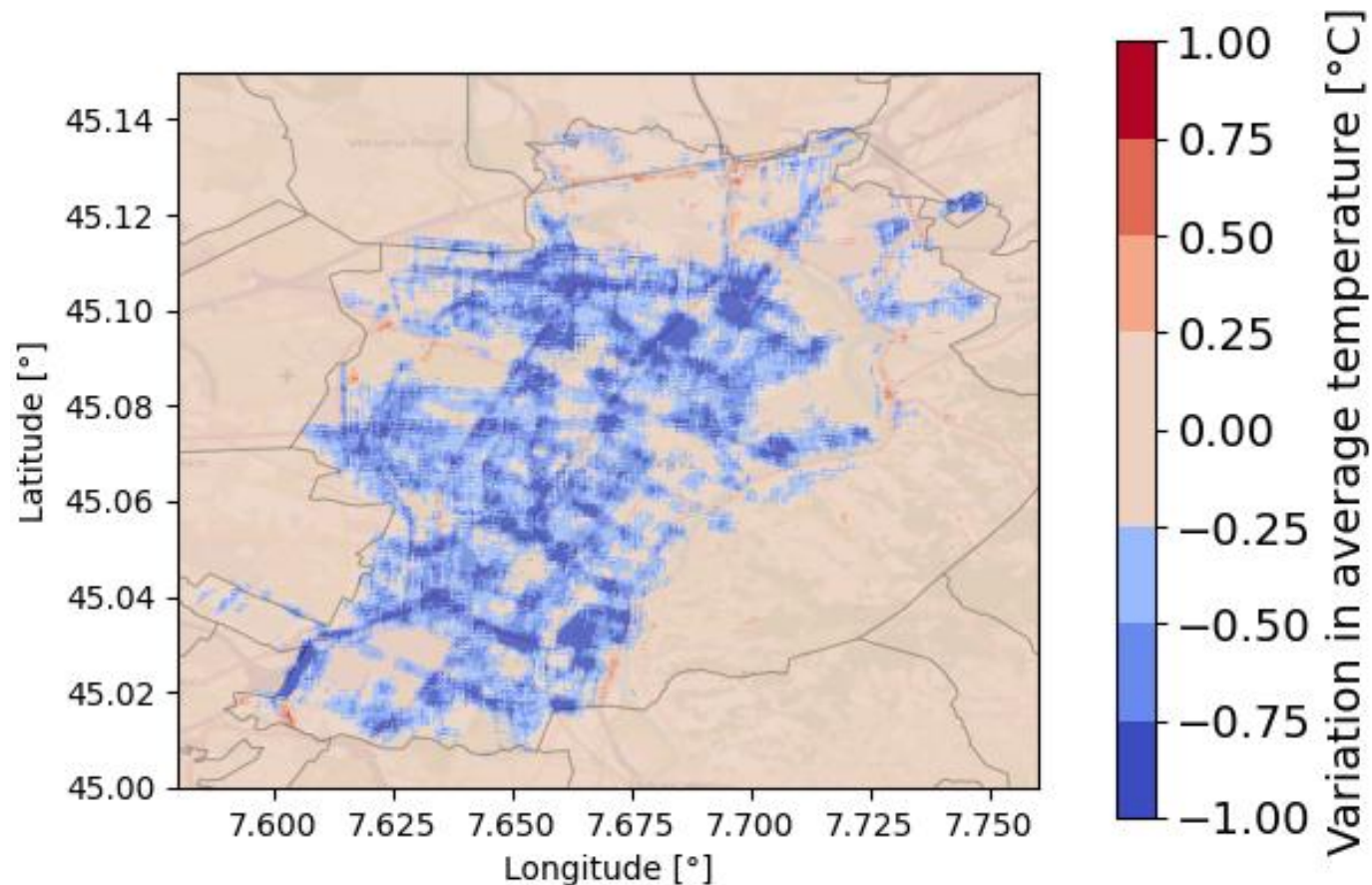
With vegetalisation, 4 parameters change:

- ABH : Average Building Height
- ALT : Altitude
- BUILD_SF : Building Surface Fraction
- IMP_SF : Impervious Surface Fraction
- LST : Land Surface Temperature
- NDVI : Normalized Difference Vegetation Index
- SVF : Sky View Factor
- VEG_SF : Vegetation Surface Fraction

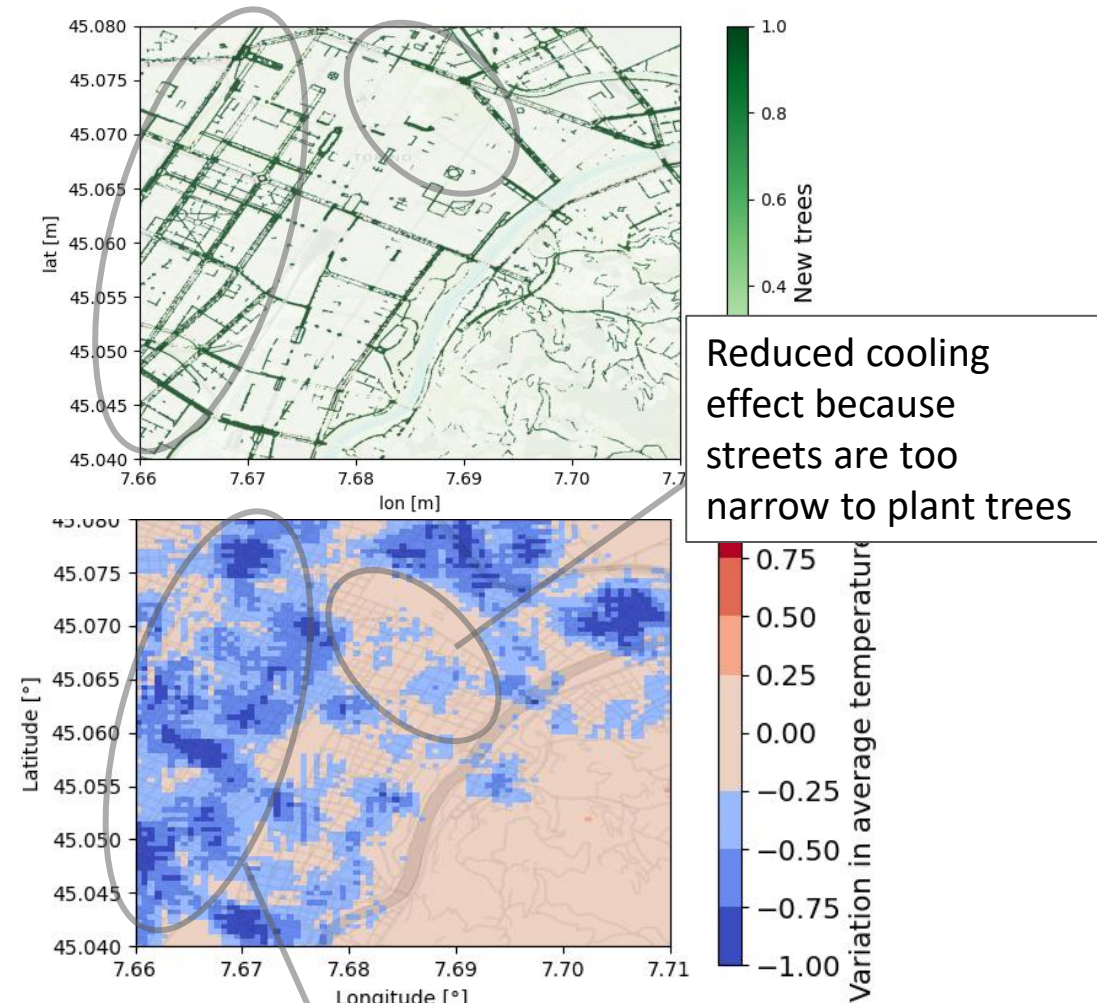


Impact of vegetalisation on air temperature

Cooling effect of the scenario on air temperature at 22:00 local time produced by the CNN model



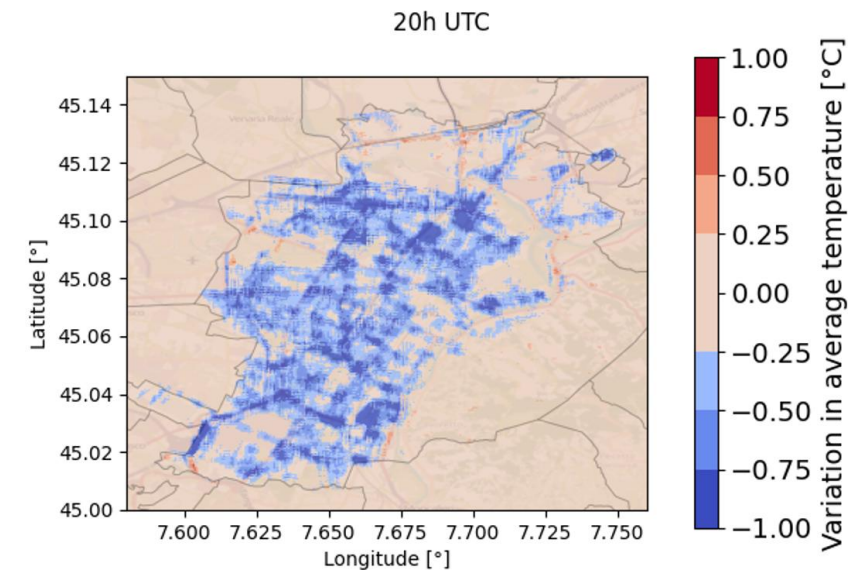
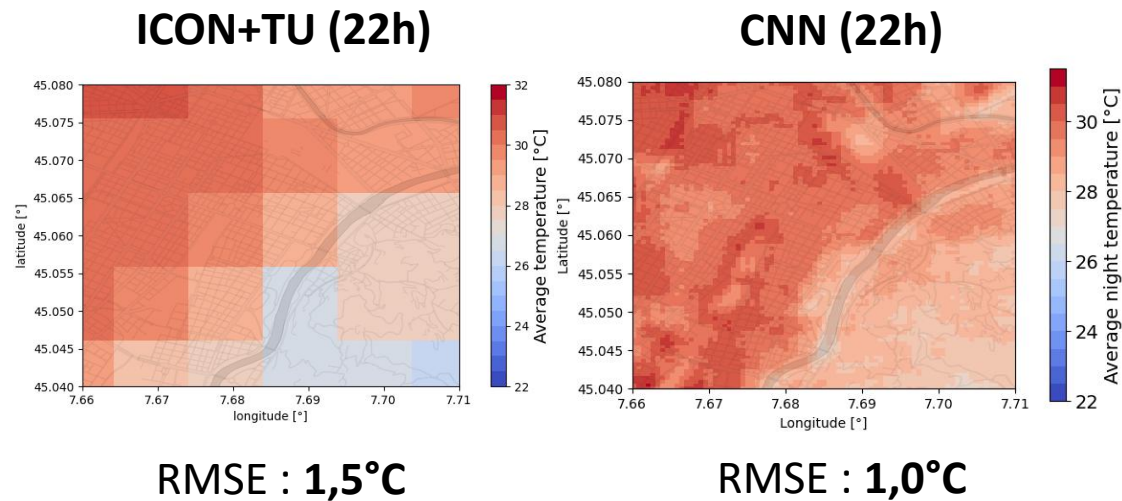
New trees density in the scenario



Conclusions

Machine Learning measure-based models reduce errors in mapping the urban air temperatures

These models make it possible to assess the impact of heat mitigation strategies with low computational cost



For questions regarding the ML approach, please contact tanguy.houget@polito.it