School of Civil Environmental and Land Management Engineering
School of Industrial and Information Engineering

Master of Science in Geoinformatics Engineering
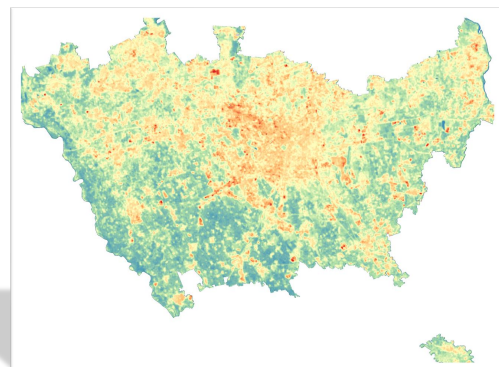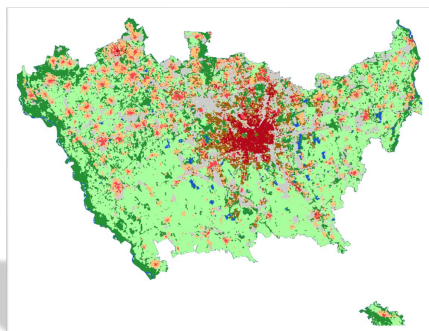
Analysis of temperature variation with respect to the LCZs using in-situ measures and satellite imagery: study case of the Metropolitan City of Milan

Author: Mathilde Puche (matr. 963894)

Supervisor: Prof. Maria Antonia Brovelli

Academic year: 2021/2022

Abstract

In this report, the Urban Heat Island (UHI) effect was studied by introducing the Local Climate Zone (LCZ) concept, which is an effective tool for the differentiation of local urban climates. With the first objective of developing a universal and accurate method for LCZ maps derivation, Random Forest supervised classification was used on remote sensing images covering the Metropolitan City of Milan. Then, satellite images of summers 2006 to 2009 and 2020 to 2022 were leveraged to study the variation of land surface temperature (LST) in the light of the urban morphology and land cover. Complementary, the air temperature (AT) variation between LCZs was investigated as well using official and crowd-sourced in-situ measures from 2018 to 2021. As a result, an overall accuracy of 94% has been achieved for the LCZ map using Landsat-8 images. However, the optimal final result was obtained by adding a building height dataset, reducing the universality of the method. With respect to temperature variations, the results confirmed that: (1) The artificial LCZs had higher LSTs than natural ones with a mean difference statistically and practically significant of $6.7\,°C$. (2) Two main factors participated in the reduction of LST: the height of building, introducing shadows on the ground, and the vegetation, increasing the portion of pervious materials. (3) In turn, vegetation was retrieved as factor of AT reduction, but the building height turned out to have the opposite effect as high buildings retain the heat and prevent air from circulating. This was illustrated by a statistically significant mean difference of $1.0\,°C$ between the Open low-rise and the Compact mid-rise classes. (4) The LCZs with higher building density were associated with slightly lower ATs, especially in winter and autumn. Finally, the analysis of records over 2018 to 2021 revealed a strong linear correlation of 0.9 between the AT and the LST. This study showed great potential of both satellite thermal data and crowdsourced information for climate research. In addition, the findings might be of great interest for urban climate specialists and urban planners in the pursuit of heat mitigation strategies.

Keywords : Local climate zone (LCZ), Land surface temperature (LST), Air temperature (AT), Urban heat island (UHI) effect, Remote sensing, Supervised classification, Crowdsourcing.

Table of Contents

List of Figures

List of Tables

1 Introduction

Nowadays, more than 55% of the world population leave in urban areas (UN [2020]) and this number reaches more than 75% for European population (EEA [2021]). As cities have expanded, urbanization has replaced green spaces with artificial structures with different thermal capacities and conductivities, affecting the land surface energy processes and regional climates (Salazar et al. [2015]). In particular, the development of urban areas led to the formation of urban heat islands (UHI), defined as the phenomenon in which temperature tends to be higher in urban regions than surrounding non-urban areas.

UHI are caused by a combination of several factors such as the ability of materials to absorb and store heat, the reduced vegetation, the anthropogenic heat and the varying urban geometry influencing wind flows (NASA [2020]). They have important implications for regional climate change, air and water quality, energy consumption, human health and urban ecosystem health (McCarthy et al. [2010], T. R. Oke [2017], Shao et al. [2006]). Therefore, the investigation of UHI is undoubtedly relevant and studies focusing on understanding the causes and effects of UHI are increasing, as shown by Almeida et al. [2021]'s review.

Two types of UHI can be differentiated: the Surface Urban Heat Islands (SUHI), representing the radiative temperature difference between impervious and natural surfaces, and the Atmospheric Urban Heat Islands (AUHI), referring to the air temperature in layers from surface to 1.5 km above it. The SUHI is mainly studied using remote sensing and thermal sensors data that enable to derive the land surface temperature (LST). On the other hand, the AUHI is traditionally measured using in situ sensors, capturing air temperature (AT) at 2m above the ground.

Due to the relevance of the issue, the UHI have been widely investigated in studies of urban climatology and observations of urban–rural air temperature differences have been reported for cities worldwide (Rizman et al. [2008]). However, traditional UHI research mostly calculates the urban-rural difference of temperature but does not consider more local scales, which over-simplifies the spatial and temporal heterogeneity of urban climates. Therefore, to address the inadequacies of urban–rural description, the "local climate zone" (LCZ) classification system has been developed (Stewart and Oke [2012]). Defined as "regions of uniform surface cover, structure, material, and human activity that span hundreds of meters to several kilometers in horizontal scale", it is a useful research tool for measuring the relation between urban configuration and the UHI phenomenon, and is growing in popularity in studies of local temperature features.

1.1 Thesis outline

In this context, the present thesis aims to study more locally the temperature variation using LCZ classification as well as satellite images and in-situ measurements.

Indeed, the first chapter is devoted to the development of an optimal method to derive LCZ maps from satellite imagery. This method intends to be universal in

order to be replicable in any city of the world.

The second chapter focuses on the land surface temperature to analyze the SUHI effect through the prism of the LCZs. After determining the LST derivation method from satellite imagery, urban morphology and land use characteristics are investigated to determine the factors that further mitigate high surface temperatures.

For its part, the third and last chapter elaborates on the AUHI effect by investigating the air temperature. The AT is firstly compared with the LST, and then studied in the light of the LCZs to identify the contributors to its decrease.

1.2   Study area and time range

The study is focusing on the Metropolitan City of Milan (MCM), located in the Lombardy region of Italy (see Fig. 2) and replacing the Province of Milan since January 2015. It covers $1575\,\mathrm{km}^2$ of geographical area, including the city of Milan and other 133 municipalities. With more than 3 million inhabitants (Eurostat [2022]), it is the second most populous Metropolitan City in the nation after the Metropolitan City of Rome.
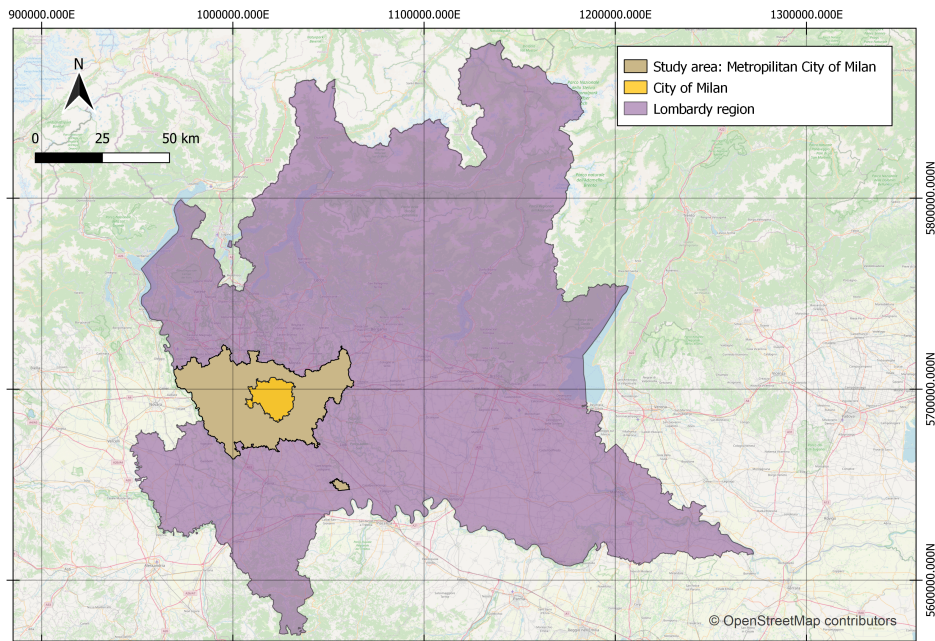
Figure 2: Location of the study area. CRS: WGS 84/Pseudo-Mercator.

The MCM is flat, with an altitude ranging from $98\,\mathrm{m}$ to $199\,\mathrm{m}$, and has a climate similar to much of Northern Italy's inland plains, where hot, humid and muggy summers and cold and wet winters prevail. In addition, Milan suffers from poor wind circulation, causing stagnation of fog and pollutants (Pichierri et al. [2012]) and intensification of the heat island effect (Bacci and Maugeri [1992]), which makes it a suitable test area for investigating local climate effects.

Furthermore, the average yearly temperature fluctuates between $2.5\,^{\circ}\mathrm{C}$ in January

and 23.6 °C in July, the hottest month in the MCM. However, maximums higher than 35 °C are becoming increasingly frequent during summer (AM [2022]), with a record temperature of 39.3 °C reached during the summer of 2003.

Also, as urbanization is mainly centered in the city of Milan with almost 40% of the total province's population, a wider range of climate zones were included by considering also the surroundings that encompass agricultural lands, forests and smaller towns.

Finally, the study focused primarily on recent years, but the period investigated depended heavily on data availability. While the LST, available on a wide range of time was exploited for both 2006-2009 and 2020-2022 periods, the AT was limited to the years from 2018 to 2021.

2   Local Climate Zones mapping

2.1   Introduction

In the future, the urban climate is likely to face increasing heat waves of higher intensity and longer duration (IPCC [2014]), giving major importance to the understanding and monitoring of urban climates. However, urban climates are usually shown through the prism of the urban-rural dichotomy, that is not adequate to relate the local heterogeneities inside the modern urban agglomerations and is not relevant from the climate perspective. Therefore, the Local Climate Zone (LCZ) classification system has been developed as a climate-centred classification system for urban temperature studies (Stewart and Oke [2012]). The LCZ concept classifies urban and natural environment into 17 classes, depending on their physical and thermal surface properties. It distinguishes between ten 'built' (1-10) and seven 'non-built' (A-G) classes (see Figure 3) and is designed to function easily in any city or region.
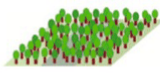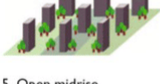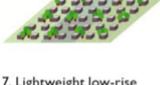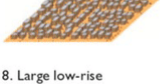
Built types	Definition	Land cover types	Definition
1. Compact high-rise	Dense mix of tall buildings to tens of stories. Few or no trees. Land cover mostly paved. Concrete, steel, stone, and glass construction materials.	A. Dense trees	Heavily wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
2. Compact midrise	Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	B. Scattered trees	Lightly wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
3. Compact low-rise	Dense mix of low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	C. Bush, scrub	Open arrangement of bushes, shrubs, and short, woody trees. Land cover mostly pervious (bare soil or sand). Zone function is natural scrubland or agriculture.
4. Open high-rise	Open arrangement of tall buildings to tens of stories. Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	D. Low plants	Featureless landscape of grass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park.
5. Open midrise	Open arrangement of midrise buildings (3–9 stories). Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	E. Bare rock or paved	Featureless landscape of rock or paved cover. Few or no trees or plants. Zone function is natural desert (rock) or urban transportation.
6. Open low-rise	Open arrangement of low-rise buildings (1–3 stories). Abundance of pervious land cover (low plants, scattered trees). Wood, brick, stone, tile, and concrete construction materials.	F. Bare soil or sand	Featureless landscape of soil or sand cover. Few or no trees or plants. Zone function is natural desert or agriculture.
7. Lightweight low-rise	Dense mix of single-story buildings. Few or no trees. Land cover mostly hard-packed. Lightweight construction materials (e.g., wood, thatch, corrugated metal).	G. Water	Large, open water bodies such as seas and lakes, or small bodies such as rivers, reservoirs, and lagoons.
8. Large low-rise	Open arrangement of large low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Steel, concrete, metal, and stone construction materials.	**VARIABLE LAND COVER PROPERTIES** Variable or ephemeral land cover properties that change significantly with synoptic weather patterns, agricultural practices, and/or seasonal cycles.	
9. Sparsely built	Sparse arrangement of small or medium-sized buildings in a natural setting. Abundance of pervious land cover (low plants, scattered trees).	b. bare trees	Leafless deciduous trees (e.g., winter). Increased sky view factor. Reduced albedo.
		s. snow cover	Snow cover >10 cm in depth. Low admittance. High albedo.
10. Heavy industry	Low-rise and midrise industrial structures (towers, tanks, stacks). Few or no trees. Land cover mostly paved or hard-packed. Metal, steel, and concrete construction materials.	d. dry ground	Parched soil. Low admittance. Large Bowen ratio. Increased albedo.
		w. wet ground	Waterlogged soil. High admittance. Small Bowen ratio. Reduced albedo.

Figure 3: Local Climate Zones classes and definitions. (Stewart and Oke [2012])

When developing the LCZ maps, numerous types of data have been used in the literature and can be categorized into remote sensing-based, GIS-based and in-situ measurement datasets. With regard to remote sensing, the most frequently used satellite imagery data are Landsat 8, 7 or 5 and Sentinel 1 and 2 (Aslam and Rana [2022]). While the first provides images with easy access for any country or region, the second is often preferred for its larger spatial resolution. Remotely sensed data is used for a supervised classification of LCZs, following the protocol established by the World Urban Database and Access Portal Tools (WUDAPT) project and formalised by Bechtel et al. [2015]. This protocol includes: selecting a spatial area that contains a city of interest, creating training areas (TAs) that illustrate each type of LCZ present, and using the available satellite imagery to classify the entire domain into LCZ types. The supervised classifier adopted for LCZs is usually the Random Forest (RF) classifier, which is an ideal compromise between the achieved accuracy and computational performance (Bechtel et al. [2015]).

In this context, this chapter aims at creating the LCZ of the Metropolitan City of Milan with the highest possible accuracy, based on multispectral images and a RF classifier. The methodology followed aims to be universal so that it can be easily replicated in other cities.

2.2 Data sources and description

2.2.1 Premilinary analysis of Land Cover

In order to determine the temporal range to consider for the LCZ mapping, a preliminary analysis of the evolution of land use was carried out. Starting from the Soil consumption data of Milan (Carta Nazionale Consumo Suolo) for the years 2006, 2012, 2015 and 2020, changes of land cover could be highligthed by computing the difference betwenn each pair of maps. It turned out that for the years 2020, 2015 and 2012, only very small changes took place, resulting in very similar soil maps. However, for the year 2006 a significant change was observed since a motorway was constructed in June 2012 and appears distinctly on the map of difference (see Fig. 4). It is the so called "Autostrada A58", also named "Tangenziale Est Esterna di Milano" (TEEM), and corresponds to the second ring road east of Milan after the A51. As this new road could modify the thermal behaviour of its local surroundings, a deeper investigation of the temperature before and after 2012 can be carried out by targeting it (see section 3.5). However, as LCZ maps delineate uniform regions that extend over hundreds of meters to kilometers (Stewart and Oke [2012]), roads are generally not represented and this change cannot be detected using these maps. Therefore, for the LCZ classification, a single year was sufficient to represent the land cover for the period 2006-2021 and in order to exploit the most recent images, the year 2021 was chosen.
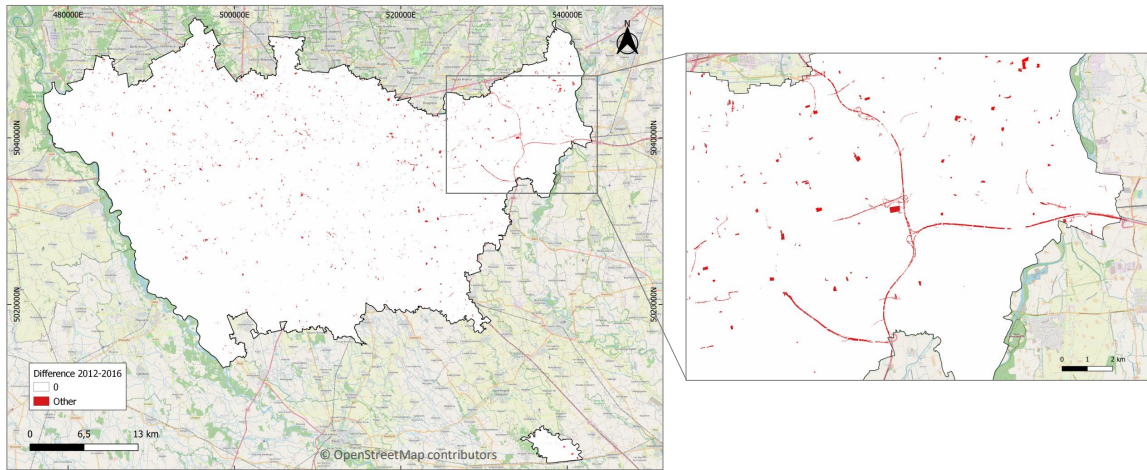
Figure 4: Differences of soil consumption between
2006 and 2012. CRS: WGS 84/UTM Zone 32N.

2.2.2   Satellite imagery

Prior to the selection, a study of the available satellite missions and their characteristics was made in order to select the most suitable for the LCZ classification. Certain criteria were imposed, including free availability and complete coverage of the area of interest, as well as high spatial resolution, necessary to detect local climates. Moreover, the LCZ classification requires multispectral images with bands in the VNIR (Visible and Near Infrared) and SWIR (Short Wave InfraRed) to exploit the widest possible spectral range. Having all these criteria set, the first state of the art on available satellite imageries resulted in the list summarized in table 1.

Several sensors were good candidates for the LCZ classification, namely Landsat 8 OLI, Sentinel-2 MSI and Terra ASTER. However, despite its high resolution appreciated for the LCZ mapping, ASTER was disqualified due to an anomaly detected on its SWIR sensor since April 2008, making six of its ten spectral bands unusable. (USGS [2008]). Thus, only Sentinel-2 and Landsat 8 missions were selected to perform the LCZ classification. A comparison of the two was made later to determine the best image for this task (see section 2.3).

Launched in February 2013, Landsat 8 is the new generation of the series of Landsat satellites. It collects images of the Earth every 16 days with two main sensors: OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor). The OLI instrument operates at nine wavelengths in the range of 0.433-2.300 µm and provides images with a spatial resolution from 15 to 30m. In this chapter, only the spectral bands 1 to 7 (30m resolution) have been exploited, as bands 8 and 9 are not relevant for classification purposes.

Available to the public since December 2020, Landsat Collection 2 (USGS [2020]) is the second major reprocessing of the Landsat archive containing both Level-1 data derived from Landsat 1-9 and science products from Landsat 4-9. They are free data open to public, downloadable via EarthExplorer and cover the entire

Mission and Sensor	Spatial resolution multispectral bands	Spatial coverage	Temporal resolution	Temporal coverage
Landsat 4, 5: TM[1]	30m	Global	16 days	7/1982 - 12/1993
Landsat 7: ETM+[2]	30m	Global	16 days	3/1984 - 01/2013
Landsat 8: OLI/TIRS[3]	30m	Global	16 days	02/2013 - Present
Landsat 1-5: MSS[4]	90m	Global	16 days	07/1972 - 01/1999
Terra: ASTER[5]	15 - 30m	Global	12 hours	03/2000 - Present
Terra & Aqua: MODIS[6]	500m	Global	16 days	04/2002 - Present
Sentinel 2A & 2B: MSI[7]	10 - 60 m	Global	5 days	07/2015 - Present and 03/2017 - Present
SNPP and NOAA 20: VIIRS[8]	375 -750m	Global	12 hours	10/2011 - Present and 11/2018 - Present

[l] TM: Thematic Mapper; [2] ETM+: Enhanced Thematic Mapper Plus; [3] OLI: Operational Land Imager, TIRS: Thermal Infrared Sensor; [4] MSS: Multispectral Scanner System; [5] ASTER: Advanced Spaceborne Thermal Emission and Reflection Radiometer; [6] MODIS: Moderate Resolution Imaging Spectroradiometer; [7] MSI: MultiSpectral Instrument; [9] VIIRS: Visible Infrared Imaging Radiometer Suite.

Table 1: Comparison of available satellite mission images.

Earth. Although Level-1 data can be rescaled easily to top of atsmosphere (TOA) radiance using radiometric rescaling coefficients, they include atmospheric effects and require additional advanced corrections to obtain TOA reflectance. In contrast, the Level-2 products are ready-to-use as they include atmospheric correction and provide Bottom Of Atmosphere (BOA) reflectance as well as science products like the surface temperature. Therefore, the Landsat 8 Collection 2 Level 2 (C2L2) images were selected for the LCZ mapping from Landsat.

In turn, Sentinel-2 consists of two identical satellites (Sentinel-2A and Sentinel-2B) operating simultaneously, achieving a revisit time of 5 days. One of the mission objectives of this satellite is to monitor changes of the Earth's surface. Its Multispectral Instrument (MSI) operates at 13 spectral bands in the range of 0.433-2.200 µm and provides images with a spatial resolution from 10 to 20 m in the VNIR and SWIR (band 2 to 8), and 60 m for its coastal aerosol and cirrus bands (bands 1 and 10, respectively). In order to take advantage of the 20 m resolution, only the bands from 2 to 8 (including 8A) have been considered in this study. Similarly to the Landsat 8 C2L2 product, the Level 2A products have been used to directly get the BOA reflectance images. They were downloaded freely from Sentinels Scientific Data Hub.

Initially, for each satellite, four dates were considered in order to take into account seasonality and increase classification accuracy, as found by Bechtel et al. [2015]. The acquisition dates of downloaded products are reported in the table 2. It can be noted that two images per date were considered for Sentinel-2 because the study area was located at the intersection of two images. Moreover, the selection of date has been strongly restricted by the availability of images with cloud coverage inferior to 5%. These scenes were used for the first LCZ classification carried out to determine the best imagery to use. Subsequently, an extra scene was added for Landsat-8 in

order to improve the accuracy of the result (see section 2.5).

	Landsat 8	**Sentinel-2**
Winter	05 December 2021 10:10AM	16 January 2021 10:23AM (2A)
Spring	19 May 2021 10:10AM	10 June 2021 10:15AM (2B)
Summer	06 July 2021 10:10AM	14 August 2021 10:20AM (2A)
Autumn	24 September 2021 10:10AM	18 October 2021 10:19AM (2A)
Additional	16 March 2021 10:10AM	-

Table 2: Selected satellite images.

2.2.3 Additional datasets

The classification performed to create the LCZ maps was a supervised classification (see section 2.3.2) which requires some training areas to learn the different classes. This is the part where urban expert knowledge plays an important role and some additional datasets were therefore needed. Starting from the work of Loftian [2016] who created a training set over the city of Milan, the vector layer has been augmented to cover the entire MCM. To that end, three types of data were used, namely the soil consumption, the building height and high resolution imageries.

The soil consumption dataset was the Soil consumption data of Milan, already presented in section 2.2.1 and is provided for free by SINAnet (Sistema Informativo Nazionale Ambientale), the national environmental information system of Italy. It divides the MCM territory into 21 land use classes, plus one class for non-built-up areas, in a 10m pixel raster. It provided a first insight of the underlying LCZ classes.

Then, the building height dataset was obtained from the Lombardy Region Topographic database, in a geopackage called *02_immobili_ed_antropizzazioni.gpkg*. The layer named *unita_volumetrica* (volumetric unit), contains the buildings of the MCM with a field portraying their height in meter. Therefore, this data was leveraged to determine the building height of the training areas.

Finally, the Google Satellite images present a very high resolution of about $0.15\,\mathrm{m}$ over Europe. Obtained through the QuickMapServices plugin of QGIS, it was exploited as a last resort to identify precisely the objects on the ground and perfect the training areas.

2.3 LCZs from Landsat-8 and Sentinel-2

Now, this section aims to perform a first classification of the LCZs using Landsat-8 and Sentinel-2 images separately and to obtain two maps for comparison. The result with the best accuracy was then used in the following chapters of this study. The different steps followed to obtain the two LCZ maps are summarised in Figure 5.
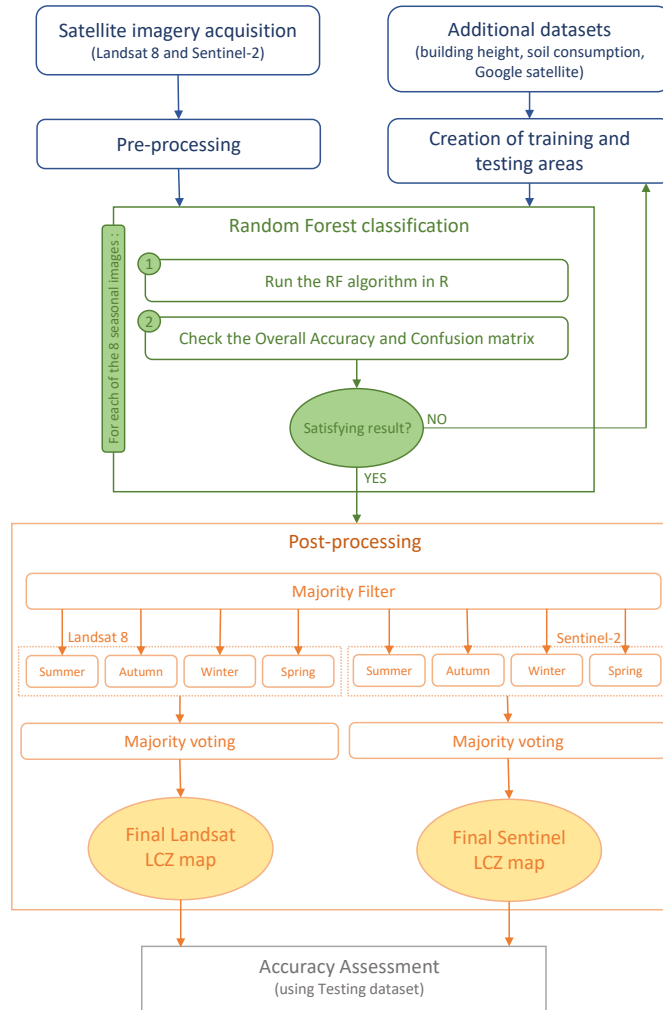
Figure 5: Flowchart for LCZ mapping from Landsat-8 and Sentinel-2

2.3.1   Data pre-processing

As C2L2 and Level 2A products of Landsat 8 and Sentinel-2, respectively, are already providing the BOA reflectance images, no additional atmospheric correction was required. Therefore, Landsat 8 images only needed the standard pre-processing such as clipping to area of interest and merging of the eight bands into one single multispectral raster, to obtain the four seasonal Landsat images.

Sentinel-2, on the other hand, required more steps, especially as the eight bands (from 2 to 8A) do not have the same resolution (10m to 20m). Arjasakusuma et al. [2022] proved in their study that for machine learning-based modeling, resampling to a coarser resolution is preferable to a higher resolution as it didn't improve the performance of the model. Therefore, SAGA GIS was used for resampling the images and obtain a homogeneous spatial resolution of 20m. Then, the same standard procedure (clip and merge) was applied to obtain the four seasonal Sentinel images.

2.3.2 Random forest classification

In recent years, a variety of machine learning algorithm were adopted for solving problems in geosciences and remote sensing such as neural networks, support vector machines, self-organizing map, decision trees, random forests or genetic programming (Lary et al. [2016]). Their main advantage against traditional methods are that they don't require the prior knowledge of the data model and are usually faster. Indeed, they are able to learn the characteristics of target classes from training samples and to identify these learned characteristics in the unclassified data. Furthermore, supervised classifiers have been proved to be more robust than model-based approaches (Niemeyer et al. [2014]) and to provide better results for image classification.

Moreover, the use of a combination of multiple classifiers to produce a single classification is generally more accurate than any of the individual classifiers. The resulting algorithm, known as ensemble classifier, combines the predictions of its individual components to make predictions with higher precision. Over the last twenty years, the Random Forest (RF) ensemble classifier has been increasingly used due to the outstanding results obtained and its speed of processing (Belgiu and Drăguţ [2016]). It has therefore been selected as the classifier in this study and its functioning is outlined in the following.

2.3.2.1 Description of the algorithm

The Random Forest algorithm is a supervised machine learning algorithm that reduces the variance of the predictions of a single decision tree, thereby improving its performance, by combining many decision trees. Through an approach called bagging, RF performs parallel learning on multiple randomly constructed decision trees trained on different subsets of the data. The ideal number of trees, which can be several hundreds or more, is an important parameter: it is highly variable and depends on the problem. Concretely, each tree of the random forest is trained on a random subset of data according to the bagging principle, with a random subset of features. The predictions are then averaged when the data are quantitative or used for voting in the case of classification trees. This main principle is summarized in the Fig. 6.

The random forest algorithm offers several advantages over other supervised classifiers. Indeed, it is known to be one of the most efficient "out-of-the-box" classifiers (i.e. requiring little data pre-processing), it is a non-parametric method and provides fast computational performance (Oxoli et al. [2018]). Moreover, it ensures an unbiased error estimate as only 2/3 of the training dataset is randomly selected for learning and the remaining 1/3 is used to test the accuracy of the model. Thus, in principle, there is no need of additional data for the accuracy assessment. However, when dealing with spatial data, to avoid spatial correlation of testing and training sets, it is always better to use an external dataset for validation. Therefore, in this study two independent training and testing sets were created.
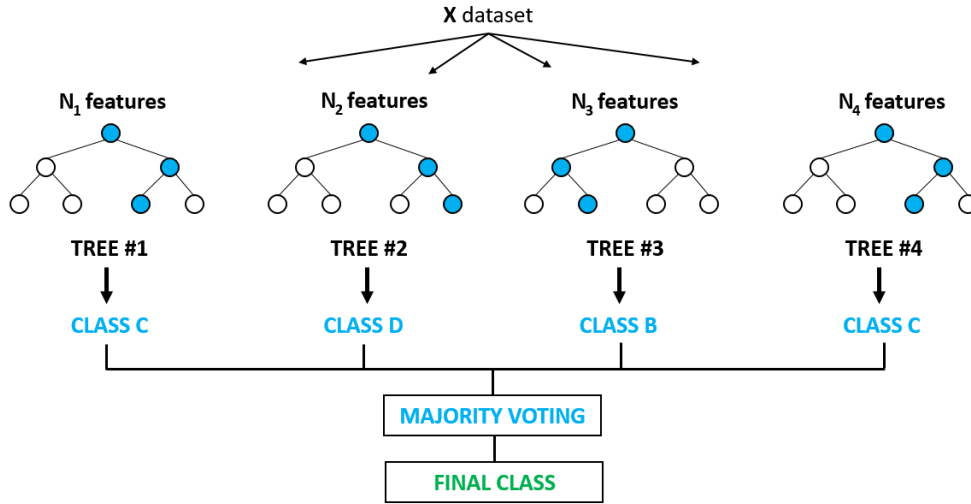
Figure 6: Main principle of the RF algorithm (from Abilash [2018])

2.3.2.2   Training and testing data

As all satellite images refer to 2021 and in general no significant change in land cover is observed over a year, one single training set has been used for all the classifications. Starting from the work of Loftian [2016], the training vector layer was augmented from 149 to 296 polygons in order to cover the entire MCM. To identify and assign the proper class to each training polygon, a combined analysis of the building height dataset and Google imagery was performed. Since the RF algorithm already estimates the classification error internally in the run, an iterative improvement of the training dataset could be carried out until a satisfying overall accuracy was achieved.

From the 17 LCZs originally defined by Stewart and Oke [2012], only 8 have been identified in the area of study, namely Compact Mid-Rise (class 2) Compact Low-Rise (class 3), Open Mid-Rise (class 5), Open Low-Rise (class 6), Large Low-Rise (class 8), Scattered Trees (class 102), Low Plants (class 104) and Water (class 107). To guarantee the best possible accuracy, a homogeneous repartition of the area per class in the training dataset has been ensured, as shown in table 3. A lower percentage can be noted for the water class (107) but as water bodies are easily well classified, this slight lack was not an issue. Similarly, an independent testing set of 129 polygons was created.

To carry out the RF classification, the algorithm implemented in R studio by Horning [2010] has been exploited. It holds the advantages of performing faster than any GIS-implemented Random Forest and enables the user to define its own parameters. In addition, it outputs the "out-of-bag" error together with the confusion matrix, useful for the iterative improvement of the training dataset. Thus, for each of the eight seasonal satellite images, the RF classification was run, resulting in eight LCZ maps (four Landsat and four Sentinel).

Class	Area (m²) in Training dataset	% of Training dataset	Area (m²) in Testing dataset	% of Testing dataset
2	5 390 302	12.9	2 316 470	14.3
3	4 242 641	10.2	1 686 754	10.4
5	4 866 078	11.7	1 714 881	10.6
6	5 234 236	12.6	2 511 317	15.5
8	5 953 483	14.3	2 307 364	14.2
102	6 512 231	15.6	1 847 137	11.4
104	6 064 292	14.6	2 039 605	12.6
107	3 378 391	8.1	1 753 408	10.8

Table 3: Repartition of classes in training and testing datasets.

2.3.3   Post-processing

To remove the so-called salt and pepper effects that are typical in the pixel-based classification, a majority filter with a 2×2 moving window was applied on each of the eight LCZ maps using SAGA GIS. Then, the four Landsat filtered maps were combined using majority voting in QGIS in order to obtain a single final Landsat LCZ map. The small percentage of pixels for which no majority was found was left with no-data value. In the same way, the final Sentinel LCZ map was obtained.

2.3.4   Accuracy assessment

The accuracy is commonly assessed using the confusion matrix which compares the pixels of a predicted class with the pixels of a ground reference class. To understand how a confusion matrix works, it is important to understand the four main terminologies: TP, TN, FP and FN. They are depicted in the Fig. 7 for the simple binary case but can be easily extended to more classes. The precise definition of each of these terms is as follows:

- TP (True Positives): cases where the prediction is positive, and the true value is actually positive.

- TN (True Negatives): cases where the prediction is negative, and the actual value is actually negative.

- FP (False Positive): cases where the prediction is positive, but the actual value is negative.

- FN (False Negative): cases where the prediction is negative, but the actual value is positive.

It enables to compute some accuracy statistics such as:

- Overall Accuracy (OA): percentage of correct classified pixels based on the total number of pixels. It is computed with Equ. 2.1.

$$OA = \frac{TP + TN}{TP + TN + FN + FP} \tag{2.1}$$

Figure 7: Confusion matrix with a binary problem.

- Producer's Accuracy (PA): percentage of classified pixels which correctly represents the true category on the ground.It is computed with Equ. 2.2 and can also be called "recall".

$$PA = \frac{TP}{TP + FN} \qquad (2.2)$$

- User's Accuracy (UA): percentage of ground truth pixels which are correctly detected in classified map. It is computed with Equ. 2.3 and can also be called "precision".

$$UA = \frac{TP}{TP + FP} \qquad (2.3)$$

Thus, the accuracy assessment was performed for the two final LCZ maps using the SCP plugin of QGIS and the independent testing set as reference.

2.3.5   Results and comparison

Figures 8 and 9 show the two final LCZ maps, derived from Landsat 8 and Sentinel-2 respectively. When looking only at these maps, very few differences could be noticed, which was a first important point as they are supposed to represent the same concept. The Sentinel LCZ map appeared however to have finer details mainly due to the higher spatial resolution (20m over 30m for Landsat 8). The confusion matrices and the derived accuracies, presented in Fig. 10 and Tab. 4 for Landsat and Fig. 11 and Tab. 5 for Sentinel, enabled to make further differentiations.

Firstly, Landsat was found to give the best output classification with 94.9% of overall accuracy against 92.7% for Sentinel-2. However, when looking at the UA and PA, except for classes 3 and 5, the two classifications revealed a similar high performance with slightly higher values for Landsat.

On the contrary, for the classes 3 and 5 Sentinel-2 gave particularly bad results with a PA of 43.6% and 30.6%, respectively. It appeared to have difficulty identifying class 3 (Compact Low-Rise) as 30% of the pixels were wrongly assigned to class 6 (Open
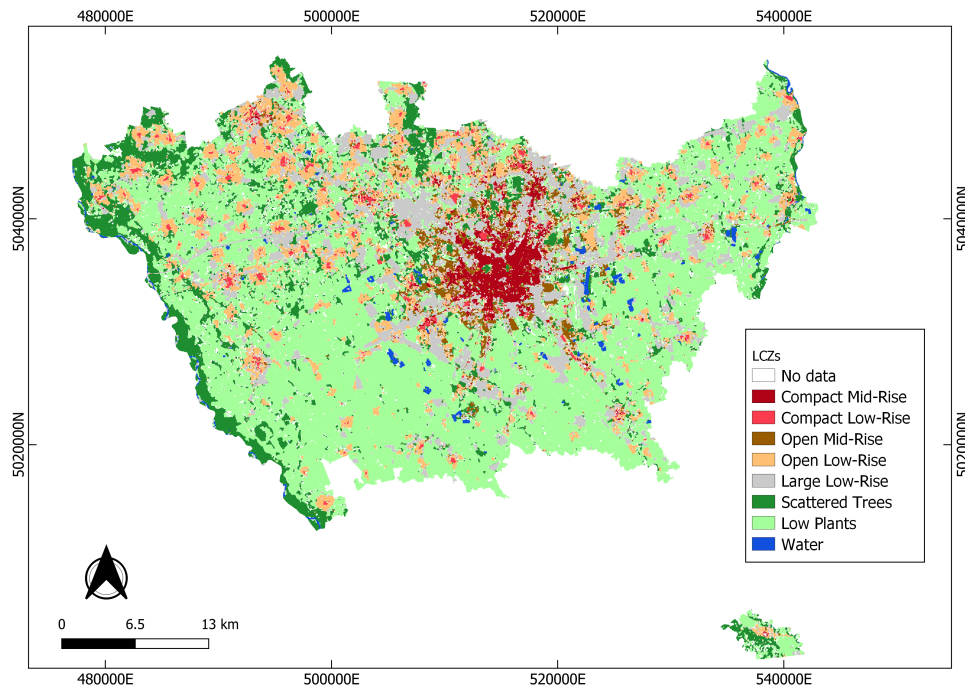
Figure 8: Final Landsat LCZ map (30m resolution).
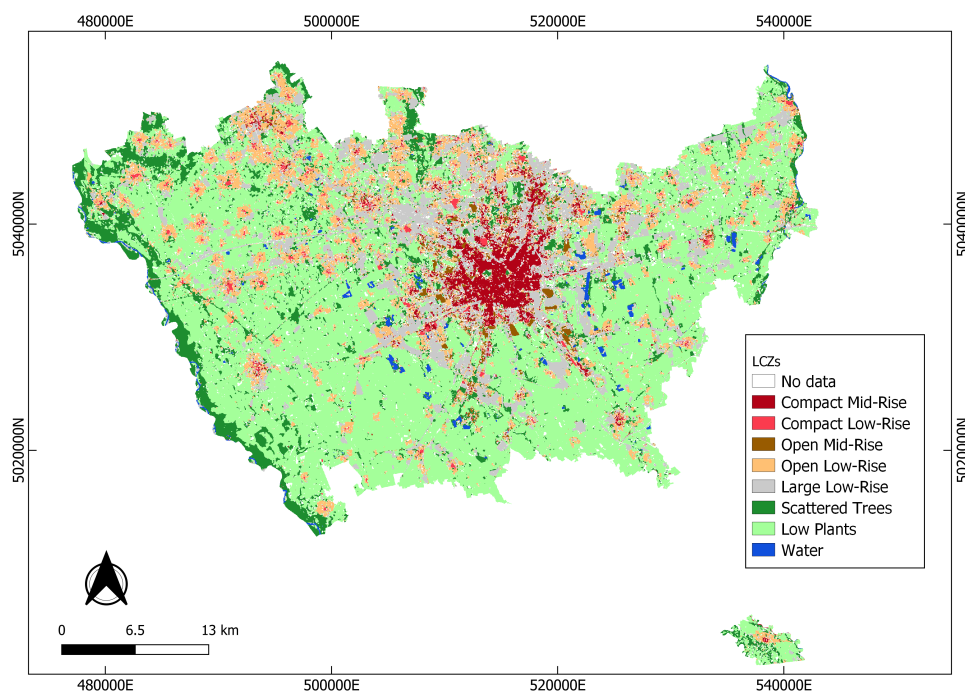
Figure 9: Final Sentinel LCZ map (20m resolution).

Low-Rise) and 16% to class 2 (Compact Mid-Rise). Also, even though Landsat was better for this class, similar confusion was observed, mainly because class 3 is exactly between classes 2 and 6. Sentinel-2 classification had even more difficulty in identifying class 5 (Open Mid-Rise) with 50% of the pixels wrongly assigned to class 6 and only 31% correctly classified. Landsat performed much better in this category with a PA of 62% but this was still very low in comparison to the other categories. This issue is mainly due to the fact that classes 3, 5 and 6 have very

similar reflectivities in all the considered bands, as shown by Oxoli et al. [2018].

Finally, independently of the comparison between the two results, both Landsat and Sentinel LCZ maps proved to hold a very good overall accuracy. As a comparison, out of the 846 available LCZ maps of the WUDAPT database, only 26 (i.e. 3%) have an accuracy higher than 90%. However, Demuzere et al. [2021] warn the users that a high OA do not automatically mean that the map is correct. It could be due to one or more of the following:

i The accuracy assessment is only performed for pixels within the training areas dataset and pixels outside these areas are not subject to quality control.

ii Too few LCZ classes in the training sample can lead to a falsely high OA.

iii Training polygons may be erroneous due to misinterpretation of the landscape by the user, such as confusion of similar classes or the use of non-persistent land covers.

The first point (i) has been avoided with the use of an external testing data for the accuracy assessment. The second (ii) has been considered in the study as Bare soil/Sand and Bare rock/Paved classes were initially added to the training categories. However, the resulting maps and accuracies were not satisfying only for these two classes, and they were not enough present on the ground to have sufficient training data. They were therefore discarded. The last point (iii) can hardly be verified or avoided as the training set is in the end always subjective and depends on the expert who created it.

In the end, the comparison of the two imageries gave similar results for both Landsat-8 and Sentinel-2, with slightly better results for the Landsat one. Therefore, in the following, the LCZ from Landsat has been used as the reference LCZ map. However, to handle the issue raised before regarding the low PA for classes 3 and 5, an attempt of improvement have been carried out and is presented in the section 2.5.

| | | Reference | | | | | | | | |
		2	3	5	6	8	102	104	107	Total
Classified	2	2486	73	50	0	21	0	0	0	2630
	3	2	689	1	22	8	3	26	0	751
	5	23	17	659	36	17	1	11	0	764
	6	3	301	328	2727	3	0	11	0	3373
	8	5	51	11	56	3711	3	0	0	3837
	102	0	0	13	7	18	3200	77	0	3315
	104	0	6	0	50	10	1	9053	0	9120
	107	0	0	0	0	0	6	0	1324	1330
	Total	2519	1137	1062	2898	3788	3214	9178	1324	25120

Figure 10: Confusion matrix for Landsat-8 LCZ classification.

2.4   Distribution of classes over the territory

Starting from the LCZ map derived from Landsat-8, an analysis of the distribution of the different classes over the MCM have been carried out. The distribution of each

Class	PA (%)	UA (%)	OA (%)
2 - Compact Mid-Rise	98.7	94.5	
3 - Compact Low-Rise	60.6	91.7	
5 - Open Mid-Rise	62.1	86.3	
6 - Open Low-Rise	94.1	80.8	94,9
8 - Large Low-Rise	98.0	96.7	
102 - Scattered Trees	99.6	96.5	
104 - Low Plants	98.6	99.3	
107 - Water	100.0	99.5	

Table 4: Accuracies for Landsat-8 LCZ classification.

		Reference							
		2	3	5	6	8	102	104	107	Total
Classified	2	5497	382	284	17	136	1	3	3	6323
	3	17	1104	8	52	13	0	0	0	1194
	5	28	33	648	27	45	6	0	0	787
	6	15	704	989	5670	22	14	29	7	7450
	8	73	122	141	194	8119	8	36	1	8694
	102	0	8	40	33	25	7037	74	0	7217
	104	0	18	11	346	35	12	20544	2	20968
	107	0	0	0	0	0	38	0	2945	2983
	Total	5630	2371	2121	6339	8395	7116	20686	2958	55616

Figure 11: Confusion matrix for Sentinel-2 LCZ classification.

Class	PA (%)	UA (%)	OA (%)
2 - Compact Mid-Rise	97.6	86.9	
3 - Compact Low-Rise	46.6	92.5	
5 - Open Mid-Rise	30.6	82.3	
6 - Open Low-Rise	89.4	76.1	92.7
8 - Large Low-Rise	96.7	93.4	
102 - Scattered Trees	98.9	97.5	
104 - Low Plants	99.3	98.0	
107 - Water	99.6	98.7	

Table 5: Accuracies for Sentinel-2 LCZ classification.

class is summarized in the Table 6. With the use of majority voting for combining the four LCZ maps, when no majority could be established the pixels were left with no-data value. As a result, the final map is composed of 7.7% of no-data pixels. To cope with this issue an improvement has been attempted in the section 2.5.

The Metropolitan City of Milan turned out to be predominantly composed of class 104 (Low Plants) with 47.6% of the territory covered by it. This reflected the extensive agricultural lands cultivated in the surroundings of Milan. With 13.8% of the territory, the second most represented class was class 8 (Large Low-Rise), accounting mainly for the industrial areas present in the MCM. Then, similar percentage was recorded for classes 6 (Open Low-Rise) and 102 (Scattered Trees) with 11.0% and 11.8% of the province, respectively. While the first disclosed the presence of

Class	Pixel count	% (with no data)	% (without no data)
2 - Compact Mid-Rise	49988	2.9	3.1
3 - Compact Low-Rise	22165	1.3	1.4
5 - Open Mid-Rise	53684	3.1	3.3
6 - Open Low-Rise	193323	11.0	12.0
8 - Large Low-Rise	240717	13.8	14.9
102 - Scattered Trees	206733	11.8	12.8
104 - Low Plants	833464	47.6	51.6
107 - Water	15756	0.9	1.0
No data	134431	7.7	-

Table 6: Distribution of classes over the MCM.

smaller cities in the surrounding of Milan, the second depicted the natural forests and woods of the MCM, especially in the west part along the Ticino (river). In turn, classes 2 (Compact Mid-Rise), 3 (Compact Low-Rise) and 5 (Open Mid-Rise) were particularly under-represented when considering the entire area of interest with less than 3.5% for all of them.

Class	Pixel count	% (with no data)	% (without no data)
2 - Compact Mid-Rise	41437	20.5	22.9
3 - Compact Low-Rise	2472	1.2	1.4
5 - Open Mid-Rise	27205	13.5	15.1
6 - Open Low-Rise	10815	5.4	6.0
8 - Large Low-Rise	40903	20.3	22.7
102 - Scattered Trees	16809	8.3	9.3
104 - Low Plants	39741	19.7	22.0
107 - Water	1187	0.6	0.7
No data	19642	9.7	-

Table 7: Distribution of classes over the city of Milan.

A second study of the classes distribution was carried out, focusing this time only on the city of Milan (see Tab. 7). The marginal class 2 into the MCM became prominent, covering 20.5% of the region's capital, as its city center is largely composed of compact mid-rise neighbourhoods. As one moves away from the center, less dense districts appear, explaining the high presence of the Open Mid-Rise class with 13.5%. In addition, the portion of class 8 (Large Low-Rise) increased, becoming the second most represented class in Milan. The numerous warehouses bordering the city may explain this phenomenon. However, the Compact Low-Rise class remained under-represented with only 1.2% of pixels over the city. Composed of a dense mix of low-rise buildings, it accounts mainly for residential areas without trees, which are quite rare in the MCM given the modern urban planning patterns. Since no large water bodies lies in the province, for both Milan and the MCM, the water class portion was found to be low with less than 1%.

2.5   Classification improvement

As highlighted in the previous sections 2.3.5 and 2.4, two main issues had to be solved regarding the LCZ classification performed on Landsat-8 images, namely the substantial percentage of no-data values in the MCM (7.7%) and the low PA obtained for classes 3 and 5.

To cope with the first point, an additional image was added to the process. Indeed, when performing the majority voting, having an even number of maps would lead more often to equality than in the odd case. Therefore, the image of 16 March 2021 was selected, pre-processed and exploited through the RF algorithm in order to produce a fifth LCZ map and carry out the majority voting on the five Landsat maps.

As the second issue was mainly due to the fact that classes 3, 5, 6 and 8 show similar reflectivities in all the considered spectral bands, an additional dataset had to be considered. Fung et al. [2022] showed in their study that adding the building height as a new band to the input raster improves significantly the accuracy of the high- and mid-rise classes in the RF classifiers. Therefore, the building height dataset has been exploited a second time and transformed into a raster of 30m resolution, by computing the average height in each cell. This raster was merged with each of the five Landsat multispectral images, giving five final rasters of eight bands (7 Landsat and 1 building height). Then, the same method as explained in section 2.3.2 was applied, with the RF classification, the majority filter and finally the majority voting.
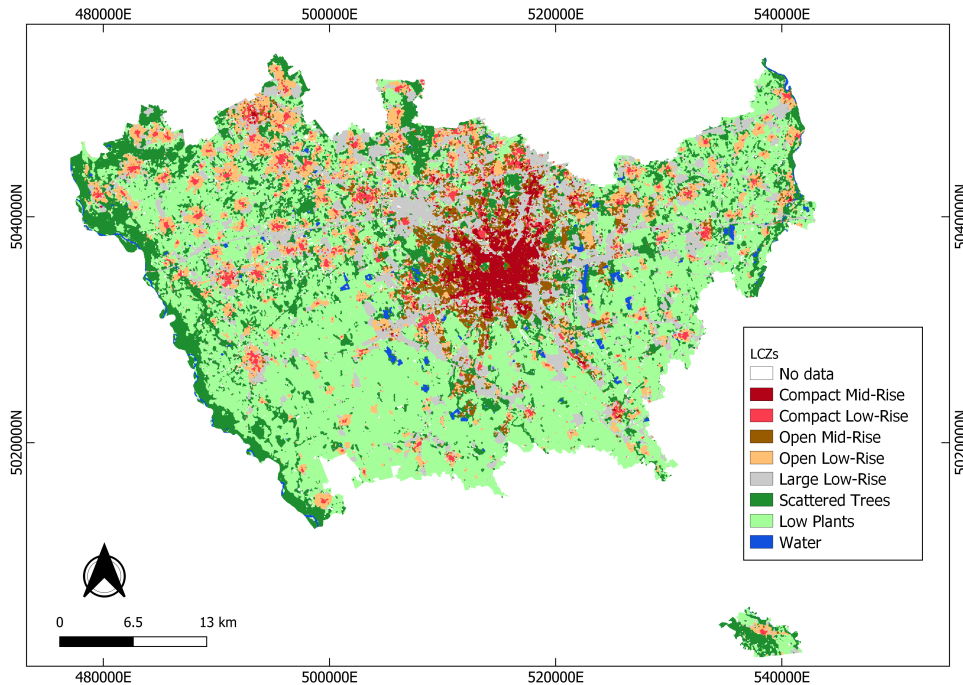
Figure 12: Final LCZ map from Landsat-8 after improvements.

The LCZ map resulting from these two improvements is presented in Fig. 12. A second analysis of the distributions over the MCM and the city of Milan was carried out using the new map, (see Tab. 8 and 9) and revealed an improvement in terms

of no-data pixels as they decreased from 7.7% to 3.7%. As regards the distribution of classes, both for the MCM and for Milan, the predominant classes remained the same as before, with a majority of Low plants, Scattered Trees, Large and Open low-rise classes in the former and a preponderance of Compact and Open mid-rise, Large low-rise and Low plants in the latter.

Then, the evaluation of accuracies was conducted. The resulting confusion matrix and accuracies can be found in Fig. 13 and Tab. 10, respectively. Firstly, the overall accuracy decreased slightly from 94.9% to 94.0% which might seems not to be an improvement. However, when looking to the PA of classes 3 and 5, a significant increase could be highlighted, passing from 60.6% to 82.8% for class 3 and from 62.1% to 82.5% for class 5. For all other classes, the PA and UA decreased slightly but remained very good with a PA and UA better than 90% and 94% respectively. Hence, containing more homogeneous and all very high accuracies (PA, UA and OA), the resulting LCZ map demonstrated a better quality and was therefore the LCZ map used later as the reference.

Class	Pixel count	% (with no data)	% (without no data)
2 - Compact Mid-Rise	50752	2.9	3.0
3 - Compact Low-Rise	45919	2.6	2.7
5 - Open Mid-Rise	75444	4.3	4.5
6 - Open Low-Rise	182048	10.4	10.8
8 - Large Low-Rise	250179	14.3	14.8
102 - Scattered Trees	314929	18.0	18.7
104 - Low Plants	749081	42.8	44.4
107 - Water	17946	1.0	1.1
No data	63963	3.7	-

Table 8: Distribution of classes over the MCM after improvements.

Class	Pixel count	% (with no data)	% (without no data)
2 - Compact Mid-Rise	44379	22.0	23.1
3 - Compact Low-Rise	4789	2.4	2.5
5 - Open Mid-Rise	36316	18.0	18.9
6 - Open Low-Rise	7961	3.9	4.1
8 - Large Low-Rise	40252	19.9	21.0
102 - Scattered Trees	24613	12.2	12.8
104 - Low Plants	32163	15.9	16.8
107 - Water	1386	0.7	0.7
No data	10075	5.0	-

Table 9: Distribution of classes over the city of Milan
after improvements.

		Reference								
		2	3	5	6	8	102	104	107	Total
Classified	2	2505	78	22	2	11	0	0	0	2618
	3	15	1536	6	137	16	5	0	0	1715
	5	29	19	1542	27	15	1	0	0	1633
	6	0	203	194	2491	5	6	0	0	2899
	8	10	19	33	54	2480	1	10	0	2607
	102	0	0	72	10	11	2036	7	1	2137
	104	0	0	1	19	1	0	2174	11	2206
	107	0	0	0	0	0	8	0	1937	1945
	Total	2559	1855	1870	2740	2539	2057	2191	1949	17760

Figure 13: Confusion matrix for the final LCZ
classification after improvements.

Class	PA (%)	UA (%)	OA (%)
2 - Compact Mid-Rise	97.9	95.7	
3 - Compact Low-Rise	82.8	89.6	
5 - Open Mid-Rise	82.5	94.4	
6 - Open Low-Rise	90.9	85.9	94.0
8 - Large Low-Rise	97.7	95.1	
102 - Scattered Trees	99.0	95.3	
104 - Low Plants	99.2	98.5	
107 - Water	99.4	99.6	

Table 10: Accuracies for the final LCZ classification
after improvements.

3   Land Surface Temperature mapping

3.1   Introduction

Land Surface Temperature (LST) is an important variable within the Earth climate system, integrating the interactions between the surface and atmosphere and all energy fluxes between the atmosphere and the land (Meng et al. [2017], Sahani [2021]). According to NASA's Goddard Institute for Space Studies (GISS) (NASA [2021]), Earth's global average surface temperature in 2020 tied with 2016 as the hottest year on record, continuing a long-term warming trend due to human activities. Consequently, an increasing number of researchers have investigated the importance and effects of LST on various topics, including urban climate (Voogt and Oke [2003]), evapotranspiration (Senay et al. [2019]), geothermal energy (Sekertekin and Arslan [2019]), forest fire monitoring (Maffei et al. [2018]) and environmental studies (Weng [2009]), among others.

LST can be estimated either from radiance measurements by in-situ meteorological stations or through remotely sensed thermal infrared (TIR) data. Although they usually offer a high resolution and wide time coverage, fixed ground stations are often few in number and therefore fail to provide large-scale thermal information, such as for an entire city or region. More and more studies are exploiting satellite remote sensing images that, on the contrary, allow temporal and spatial LST analysis on a large scale.

The radiances measured by the radiometers onboard satellites depend not only on surface parameters (temperature and emissivity) but also on atmospheric effects (Li et al. [2013]). Therefore, different methods have been used in the literature for the estimation of LST, depending on the knowledge of the Land Surface Emissivity (LSE) and the atmospheric corrections. In case LSEs are known a priori, single-channel (or mono-window) methods, multi-channel (or split-window) methods and multi-angle methods are the three most commonly used categories of algorithms. While the first uses only one TIR band, the second and third include more than one TIR band. However, in case LSEs are unknown, the stepwise recovery method, the simultaneous recovery of LSE and LST with known atmospheric information, and the simultaneous recovery method with unknown atmospheric information are applied (Sahani [2021]).

Several studies focus on the relation between LCZ and LST. Han et al. [2022] showed that the LSTs in the city of Xi'an (China) tended to decrease from urban to natural areas and highlighted a gradually decreasing LST from the city center to rural areas. In addition, Du et al. [2020] conducted their study in Nanjing (China) and found out that the warmest and coolest zones varied with seasons and that LCZs showed greater differences in LST during summer than in other seasons. They also pointed out the substantial impact of building height on LST, with low buildings giving rise to higher LSTs. Furthermore, Khoshnoodmotlagh et al. [2021] demonstrated that compact high-rise, compact mid-rise, and heavy industrial areas tended to increase the surface temperature in the city of Tehran (Iran). However, Yang et al. [2020] explored the thermal environment characteristics of several cities in the light of their

LCZs. It turned out that the proportion of LCZs are different depending on the city sizes and that higher LSTs are experienced in larger cities. In addition, even if similar patterns between cities could be outlined with highest temperatures in class "Industry" for all of them, some differences were found, highlighting that the specific thermal environment of each city as well as their size modify the distribution of LST among the LCZs.

With this background, the present chapter includes three main aims. Firstly, two methods of LST retrieval were compared to determine the most appropriate for the study of local climates in the Metropolitan City of Milan. Then, the selected method was leveraged to determine the LST variability among the LCZs of the MCM for statistical analysis and comparison. Finally, a further study was carried out on the new road highlighted in section 2.2.1, to determine whether a change in the local thermal environment was caused by the construction of such a highway.

3.2 Data sources and description

Similarly to the previous chapter, a preliminary study of the available satellite missions suitable for the LST mapping have been carried out. As this task is performed on images acquired by thermal sensors, only satellites carrying thermal sensor(s) have been considered. In addition, the free availability and the full coverage of the area of interest were imposed, together with the highest possible spatial resolution of the thermal band, necessary to detect the local temperatures. From these criteria, the table 11 was built.

Mission and Sensor	Spatial resolution thermal bands	Spatial coverage	Temporal resolution	Temporal coverage
Landsat 4, 5: TM	120m (resampled 30-60m)	Global	16 days	7/1982 - 12/1993
Landsat 7: ETM+	60m (resampled 30m)	Global	16 days	3/1984 - 01/2013
Landsat 8 and 9: OLI/TIRS	100m (resampled 30m)	Global	16 days	02/2013 - Present
Terra: ASTER	90m	Global	12 hours	03/2000 - Present
Terra & Aqua: MODIS	1km	Global	16 days	04/2002 - Present
ISS / ECOSTRESS: PHyTIR	60m	CONUS only	varies/every few days	06/2018 - Present
Sentinel 3A & 3B: SLSTR	1km	Global	1 day	O2/2016 - Present and 04/2018 - Present
SNPP and NOAA 20: VIIRS	375 -750m	Global	12 hours	10/2011 - Present and 11/2018 - Present

Table 11: Comparison of available satellite missions with thermal sensors.

Landsat-8 and Landsat-9 satellites carry specific Thermal Infrared Sensors (TIRS) measuring the land surface temperature in two thermal bands (10 and 11) with a resolution of 100m. However, due to a calibration error detected on band 11, numerous exiting studies determine the LST using only Landsat 8 band 10 (García [2021]). Their predecessor, Landsat-5, was carrying the Landsat Thematic Mapper (TM) which created images consisting of six spectral bands with a spatial resolution

of 30 m (Bands 1-5 and 7), and one thermal band (Band 6) of 120 m resolution. The Landsat 7 Enhanced Thematic Mapper (ETM+) images are usually discarded from studies because from 2003 onward the Scan Line Corrector was not working correctly, resulting in scene losses of around 22% (Hamid Reza Pourghasemi [2019]).

The analysis of available thermal images, summarized in Tab. 11, stressed that Landsat-5, 8 and 9 were the best imagery for deriving the LST map. Indeed, the Landsat Collection 2 data includes thermal images with the highest resolution (30 m) thanks to advanced data processing and resampling applied on acquired scenes (USGS [2020]). This collection contains both Level-1 data and Level-2 science products, both of which were used in this study to derive the LST with different methods (see section 3.3).

The correlation between the LCZs and the LST was chosen to be studied over the most recent period with images from 2020 to 2022. To detect the highest temperatures experienced across the study area, only images acquired during summer (from the 21st of June to the 21st of September of each year) were selected, which resulted in six different dates, reported in the table 12. The selection was restricted to dates with a maximum cloud coverage of 5% over the study area.

Level	Date	File name
L2	19 July 2020 10:10AM	LC08_L2SP_194028_20200719_20200911_02_T1
L2	05 September 2020 10:10AM	LC08_L2SP_194028_20200905_20200918_02_T1
L2	06 July 2021 10:10AM	LC08_L2SP_194028_20210706_20210713_02_T1
L2	06 July 2021 10:10AM	LC08_L2SP_194028_20210722_20210729_02_T1
L2	09 July 2022 10:10AM	LC08_L2SP_194028_20220709_20220721_02_T1
L2	17 July 2022 10:10AM	LC09_L2SP_194028_20220717_20220719_02_T1
L1	06 July 2021 10:10AM	LC08_L1TP_194028_20210706_20210713_02_T1
L1	22 July 2021 10:10AM	LC08_L1TP_194028_20210722_20210729_02_T1

Table 12: Selected Landsat 8-9 satellite images for the period 2020-2022.

Then, to evaluate the effect of the construction of the new road in 2012 in the East part of the MCM, images in the period 2006-2009 were exploited as well. The presence of clouds in most of images limited the selection to the four dates listed in table 13.

Level	Date	File name
L2	16 July 2007 10:04AM	LT05_L2SP_194028_20070716_20200830_02_T1
L2	02 September 2007 10:03AM	LT05_L2SP_194028_20070902_20211210_02_T1
L2	05 July 2009 09:59AM	LT05_L2SP_194028_20090705_20200827_02_T1
L2	06 August 2009 09:59AM	LT05_L2SP_194028_20090806_20200827_02_T1

Table 13: Selected Landsat-5 satellite images for the period 2006-2009.

Moreover, in order to validate the LST maps derived from Landsat products, a reference dataset was needed. As reported by García [2021], there exists four validated methods for establishing the reliability of a LST map. They are: in situ measurements, the radiance method (relying on atmospheric profiles), comparison with ambient temperatures near the ground and cross-reference. The latter method consists in validating a given LST by comparing it with the LST obtained by another satellite whose data are well founded. In this study, the Sentinel-3 level 2 LST

products were chosen as a reference as they directly include the LST with a spatial resolution of 1km. This validation was carried out only on images of 2021 in order to determine the most appropriate method to derive LST in this study, as explained in the following section 3.3. Therefore, only the images of the 06 July 2021 and 22 July 2022 were downloaded, both acquired at 09:46AM i.e. 24 minutes before the Landsat-8 images.

3.3   Methods for LST retrieval

As explained in the introduction to this chapter, several methods exist for deriving the LST from satellite images, two of which were exploited in this study. While the first is directly using the Collection 2 Level 2 (C2L2) products already containing ready-to-use LST maps, the second is deriving LST from Collection 2 Level 1 (C2L1) images by applying the so called single channel algorithm (SCA). In this section, the two methods are presented and compared to a reference, the Sentinel-3 level 2 LST product, in order to determine the approach to follow when computing the LST in the next sections.

3.3.1   LST from Level 2 products

The Landsat 8 Level 2 ST product is derived from the Collection 2 Level 1 TIRS band 10 and generated from the single channel algorithm (USGS [2020]). Therefore, it already contains the atmospheric corrections calculated using, among others, the ASTER Global Emissivity Database (GED) data, ASTER Normalized Difference Vegetation Index (NDVI) data, and atmospheric profiles of geo-potential height, specific humidity, and air temperature.

To retrieve the LST from the L2C2 image, a linear transformation of the digital number (DN) was performed on band 10 as shown in Eq. 3.2 and 3.2.

$$LST_{Kelvin} = 0.00341802 * DN + 149 \tag{3.1}$$

$$LST_{Celsius} = LST_{Kelvin} - 273.15 \tag{3.2}$$

As documented in Cook et al. [2014], the presence of clouds may cause large negative errors in the LST distribution. To avoid as much as possible this issue, the portions of the MCM with significant cloud coverage were removed for each of the two dates. Fig. 14 displays the resulting LST map for the 6th of July 2021, with temperatures ranging from $7.1\,°C$ to $65.9\,°C$ and an average and standard deviation of $39.8\,°C$ and $7.1\,°C$, respectively. In turn, Fig. 15 shows the result for the 22nd of July 2021, that variates from $12.0\,°C$ to $63.2\,°C$ with a mean of $40.6\,°C$ and a standard deviation of $5.6\,°C$. Thus, the latter map presents less variability and slightly higher temperature than the first. The greater range of variation for July 6 can be explained by the strong effects of clouds, which are not all detected and therefore not all eliminated, but modify the thermal response acquired by the sensor.
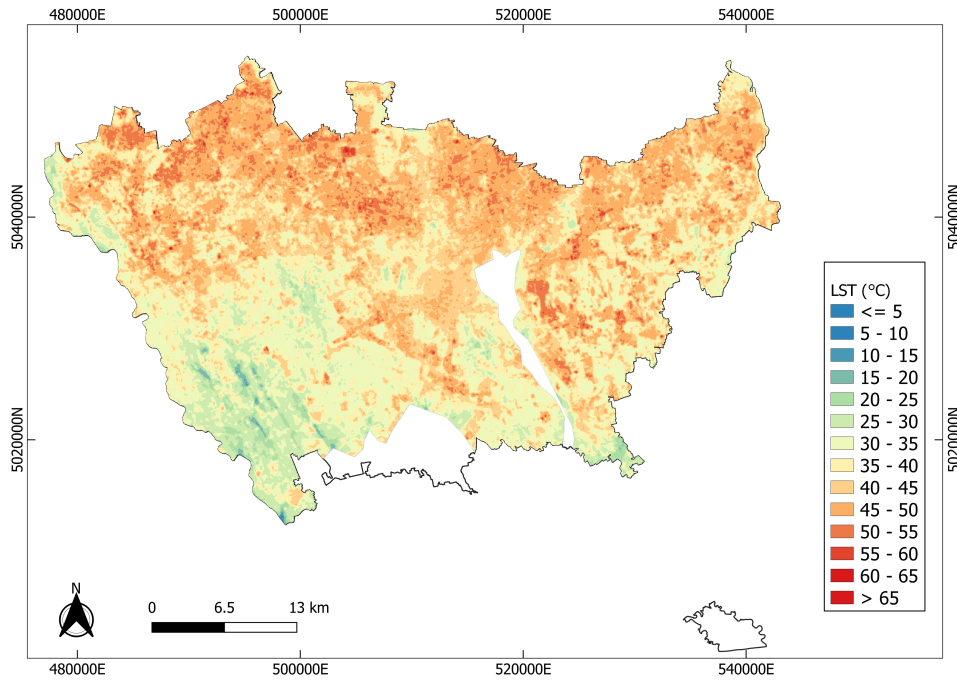
Figure 14: LST map of 06 July 2021 from C2L2 product (30m resolution).
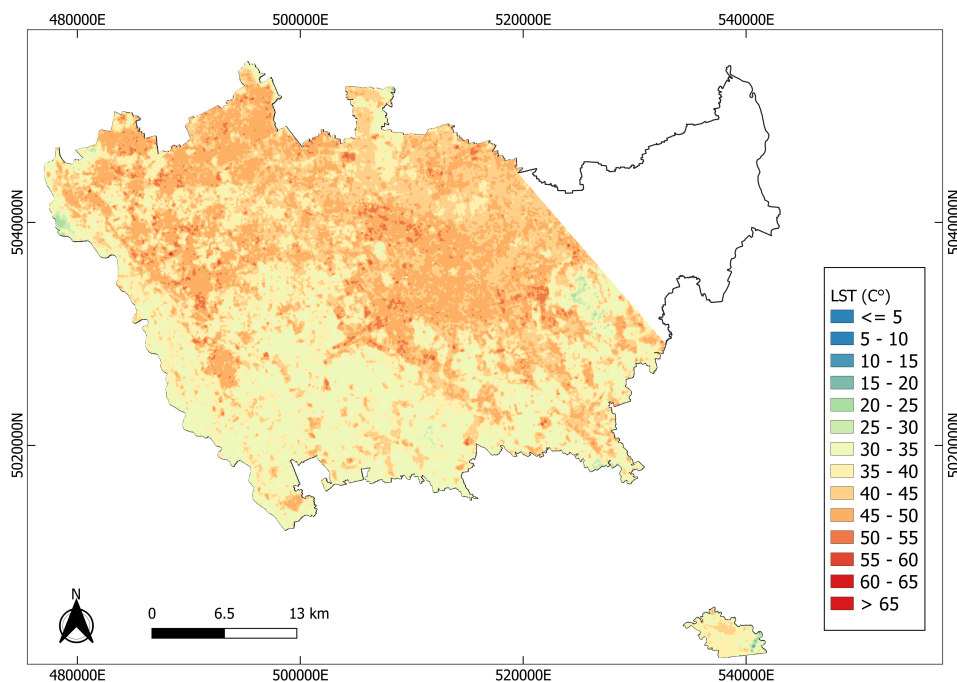
Figure 15: LST map of 22 July 2021 from C2L2 product (30m resolution).

3.3.2   LST from Level 1 data

Although the C2L2 product has the serious advantage of holding atmospheric corrections and being a ready-to-use image, most of the studies in the literature start from the C2L1 products, mainly because the C2L2 product is global and researchers prefer to apply their own and more local corrections.

Among the algorithms mentioned in the introduction of this chapter, the single

channel algorithm was used in the present study to retrieve the LST from Landsat-8 level 1 products. SCA has been widely used to extract LST using single thermal band (Duan et al. [2019], Sahani [2021], Mahato and Pal [2018]) and has been proved to be theoretically accurate for the LST retrieval (Mujabar and Rao [2018]). Also known as the model emissivity method (Hook et al. [1992]), it uses estimated atmospheric parameters to correct the atmospheric effect in a single channel to retrieve LST.

The procedure suggested by Weng et al. [2004] has been followed in this study and contains four steps that are: (1) the retrieval of TOA spectral radiance, (2) the conversion of spectral radiance to at-satellite brightness temperature, (3) the estimation of land surface emissivity and (4) the estimation of LST. These four steps were carried out using the Rasterio library in Python.

Step 1: Retrieval of TOA spectral radiance from the digital number

When the Landsat-8 TIRS sensor acquires an image, a mixture of energy reflected from the surface and the atmosphere is recorded. This is the so called Top of Atmosphere spectral reflectance (L_λ), retrieved from band 10 using the radiance rescaling factors provided in the metadata file (USGS [2016]) (see Eq. 3.3).

$$L_\lambda = M_L * DN + A_L \tag{3.3}$$

where:

- L_λ : the TOA spectral radiance for each pixel (in $\mathrm{W\,m^{-2}\,sr^{-1}\,\mu m^{-1}}$),

- M_L : the Band-specific multiplicative rescaling factor from the metadata (M_L = 0.0003342),

- A_L : the Band-specific additive rescaling factor from the metadata (A_L = 0.1000),

- DN : the digital number i.e. the pixel value.

Step 2: Conversion to At-Satellite Brightness Temperature

The Brightness temperature (T_B) is the temperature of a blackbody that would emit the same amount of radiation as the targeted body in a specified spectral band (Spampinato et al. [2011]). Thus, the spectral radiance can be modified for at-satellite brightness temperature using the thermal constants K_1 and K_2 provided in the metadata file (see Eq. 3.4).

$$T_B = \frac{K_2}{ln(\frac{K_1}{L_\lambda} + 1)} \tag{3.4}$$

where:

- T_B : At-satellite brightness temperature (in Kelvin),

- L_λ : the TOA spectral radiance for each pixel (in $\mathrm{W\,m^{-2}\,sr^{-1}\,\mu m^{-1}}$),

- K_1 and K_2 : Band-specific thermal conversion constants from the metadata (K_1 = 774.8853 and K_2 = 1321.0789).

Step 3: Estimation of land surface emissivity (ϵ)

The emissivity characterizes the ability of a body to emit radiation (Rhinane et al. [2012]) and is affected by both the surface and the wavelength. As the (T_B) obtained in step 2 refers to a black body, corrections for spectral emissivity are required. Various approaches are available for the modelling and estimation of LSE, but the NDVI-threshold method proposed by Sobrino et al. [2004] was implemented in this study, as it is a reliable, applicable and easy-to-apply approach.

The Normalized Difference Vegetation Index (NDVI) is a common and widely used remote sensing index. It ranges from -1 to +1 and assesses whether the observed target contains green healthy vegetation or not. While higher values correspond to green and dense vegetation, lower ones show sparse vegetation or even water (for -1). It is computed as the difference between near-infrared (NIR) and red (RED) reflectance divided by their sum as shown in Eq. 3.5. In this study, the Red and NIR bands of Landsat-8 OLI sensor were used, that are respectively bands 4 and 5.

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{3.5}$$

Then, the NDVI-threshold method obtains the emissivity estimation from the NDVI considering three cases:

1. **NDVI < 0.2** : In this case, the pixel is considered as bare soil and the emissivity is obtained from reflectivity values in the red region with the following equation :
$$\epsilon = 0.973 - 0.074 * RED \tag{3.6}$$

2. **NDVI > 0.5** : The pixels are considered as fully vegetated and a constant value of emissivity is fixed, typically to 0.99 :
$$\epsilon = 0.99 \tag{3.7}$$

3. **0.2 $\leq$ NDVI $\leq$ 0.5** : When the NDVI is between 0.2 and 0.5, the pixel contains a mixture of bare soil and vegetation and the emissivity is computed as defined in Eq. 3.8 :
$$\epsilon = \epsilon_v + \epsilon_s * (1 - P_v) + dE \tag{3.8}$$
   where:
   - ϵ_v : emissivity of vegetation ($\epsilon_v = 0.9863$, approximation from García [2021]),
   - ϵ_s : emissivity of soil ($\epsilon_s = 0.9668$, approximation from García [2021]),
   - P_v : proportion of vegetation or fractional vegetation cover, computed with Eq. 3.9.
   - dE : coefficient of rugosity, taken as 0 for flat surfaces.

$$P_v = (\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}})^2 \tag{3.9}$$

Step 4: estimation of land surface temperature (LST)

Finally, the LST has been calculated using Eq. 3.10, following Artis and Carnahan [1982] and Mahato and Pal [2018] later.

$$LST = \frac{T_B}{1 + (\lambda * (\frac{T_B}{\rho}) * ln(\epsilon))} \tag{3.10}$$

where:

- LST : Land Surface Temperature in Kelvin,

- T_B : brightness temperature in Kelvin,

- λ : wavelength of emitted radiance (10.895 µm for Landsat-8 band 10),

- ρ : $\rho = h * c/\sigma = 1.438 * 10^{-2}$ m K (σ = Boltzman's constant = 1.28 x 10^{-23} J K^{-1}, h = Planck's constant = 6.626 x 10^{-23} J s, c = velocity of light = 2.998 10^8 m s^{-1})

- ϵ : land surface emissivity.

The LST, obtained in Kelvin, has been converted into degree Celsius by subtracting 273.15 from each pixel value. The resulting maps for the 6th and 22nd of July 2021 are presented in Fig. 16 and 17. As a first result, it could be highlighted that the level-1 derived LST maps presented lower temperatures than the level-2 ones, with averages of 26.5 °C and 26.4 °C for 06 and 22 July respectively (against 39.8 °C and 40.6 °C from level 2 products). The variation in temperature was lower as well since the standard deviations were 3.7 °C and 2.8 °C as opposed to 7.1 °C and 5.6 °C. These numbers are summarized in the Tab. 14 and 15 of section 3.3.4, where a deeper comparison of the results was carried out.
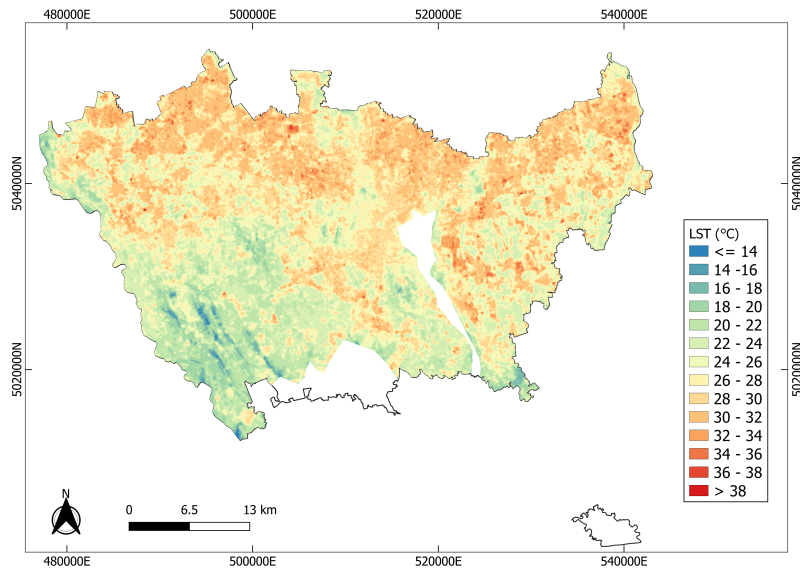
Figure 16: LST map of 06 July 2021 from C2L1 product (30m resolution).
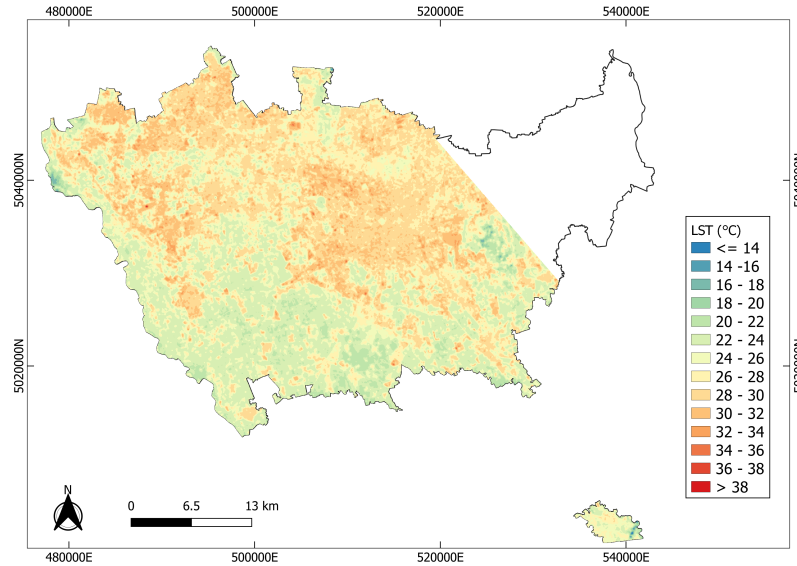
Figure 17: LST map of 22 July 2021 from
C2L1 product (30m resolution).

3.3.3 LST from Sentinel-3 products

Sentinel-3 satellites are equipped with the Sea and Land Surface Temperature Radiometer (SLSTR) whose main objective is to provide global and regional sea and land surface temperatures with a very high level of accuracy (ESA [2021]). It captures images in six spectral bands with a resolution of 500 m (bands 1 to 6) and three thermal bands that enable to determine the LST (bands 7, 8 and 9) with a spatial resolution of 1000 m.

Similarly to Landsat-8, Sentinel-3 thermal products are available for download with two levels of processing (level 1 and level 2). While level 1 data contain radiance and brightness temperatures, level-2 products directly includes the LST together with other quantities such as the Fire Radiative Power or Water Single Temperature. Therefore, the SLSTR Level-2 LST products were exploited to provide reference maps for the two previously selected dates (06 and 22 July 2021). They are shown in Fig.18 and 19 and their statistics are summarized in the Tab. 14 and 15.

3.3.4 Comparison of LST maps and methods

At this point, four Landsat-derived maps (06 July 2021 Level-1, 06 July 2021 Level-2, 22 July 2021 Level-1, 22 July 2021 Level-2) and two Sentinel-3 references (06 July 2021 and 22 July 2021) were established.

To begin with, a comparison of the box-plots for each of the two dates was made. The Fig. 20 shows the box-plots for the 6th of July 2021, displaying the distribution of data based on a five number summary (minimum, first quartile, median, third quartile, and maximum). This graph anticipated a greater proximity of Sentinel-3 to the Landsat-8 Level 2 product than to the Landsat-8 Level 1 one since its median appeared to be included in the interquartile range (IQR, 25th to the 75th
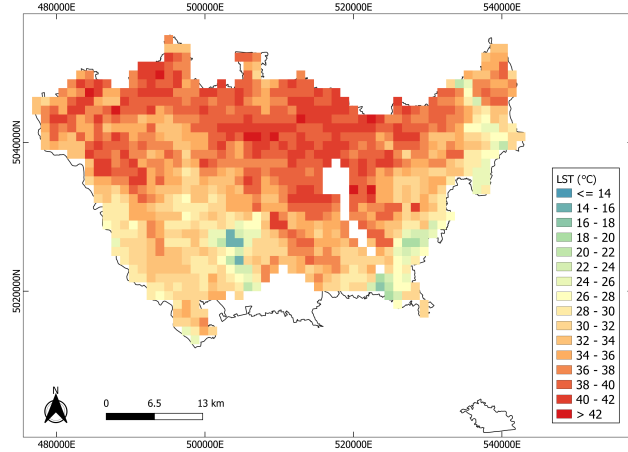
Figure 18: LST map of 06 July 2021 from
Sentinel-3 (1km resolution).
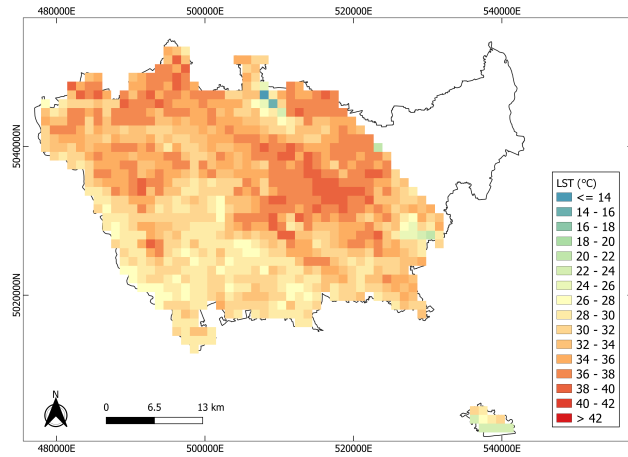
Figure 19: LST map of 22 July 2021 from
Sentinel-3 (1km resolution).

LST Map	Mean (°C)	Stdev (°C)	Min (°C)	Median (°C)	Max (°C)
Landsat C2L2 (30m)	39.8	7.1	8.2	40.5	65.9
Landsat C2L1 (30m)	26.5	3.7	10.7	26.8	40.5
Sentinel-3 (1000m)	34.7	5.1	14.2	35.4	43.2

Table 14: Summarized statistics for the derived LST maps of 06 July 2021.

percentile) of the Level-2 product. However, the C2L2 map turned out to hold a large variation with its maximum and minimum values encompassing both the Level-1 values and Sentinel-3 ones, which could reveal a lower accuracy and precision in the LST estimation.

In turn, the Fig. 21 draws the box-plots for the 22nd of July 2021. This time, the median of the Sentinel-3 product ($33.2\,°C$) was not contained in any of the two other IQRs. However, its third quartile (from 50th to the 75th percentile) appeared

LST Map	Mean (°C)	Stdev (°C)	Min (°C)	Median (°C)	Max (°C)
Landsat C2L2 (30m)	40.6	5.6	12	41.2	63.2
Landsat C2L1 (30m)	26.4	2.8	12	26.7	38.4
Sentinel-3 (1000m)	33	3.1	19.1	33.2	38.7

Table 15: Summarized statistics for the derived LST maps of 22 July 2021.
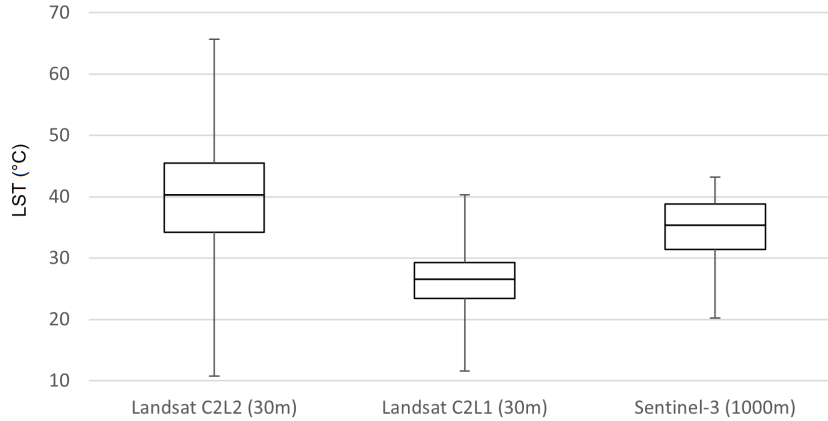
Figure 20: Box-plots of LST values per method - 06/07/2021.

to be included into the first quartile (from 25th to the 50th percentile) of the Level 2 product, which might again lead to the conclusion of a greater proximity of Sentinel-3 to the Level 2 LST map.
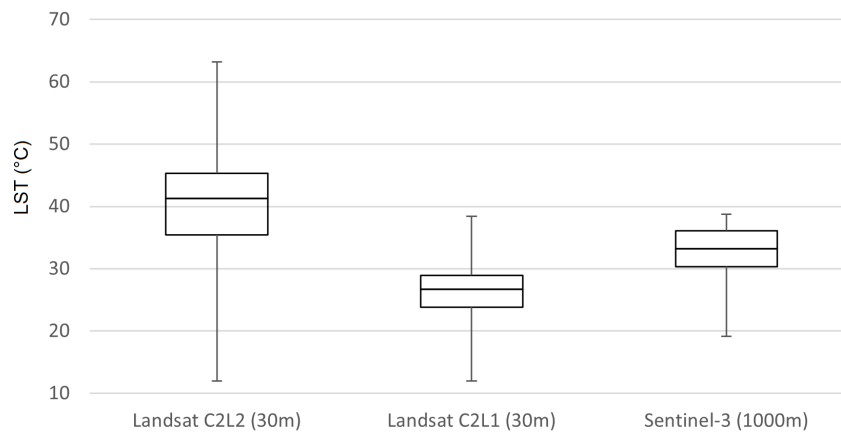
Figure 21: Box-plots of LST values per method - 22/07/2021.

These preliminaries, though, were not sufficient to draw any conclusion. To take into consideration the spatial distribution of the temperature, a further analysis of the differences per pixel was carried out. To do so, the four Landsat-derived LST maps had to be resampled to the same spatial resolution of 1000 m, which has been done using the SAGA GIS resampling tool with the *Mean value (cell area weighted)* upscaling method.

Then, the difference between the resampled Landsat maps and their corresponding Sentinel-3 reference was computed, pixel per pixel. The Tab. 16 presents the range [mean difference – standard deviation ; mean difference + standard deviation] for each date and each level versus the Sentinel-3 product. It can be noted that the computed difference was *Landsat – Sentinel*, which led to negative values in case Sentinel had higher temperatures.

	06/07/2021	22/07/2021
Landsat C2L2 range of diff. (˚C)	[0.4 ; 10.0]	[5.1 ; 10.5]
Landsat C2L2 mean diff. (˚C)	5.2	7.8
Landsat C2L1 range of diff. (˚C)	[-12.1 ; -4.1]	[-8.3 ; -4.5]
Landsat C2L1 mean diff. (˚C)	-8.1	-6.4

Table 16: Range of difference between Landsat-derived LST and Sentinel-3.

For the 6th of July 2021, the Level 2 product turned out to be closer to the reference with an absolute mean difference of 5.2 °C, against 8.1 °C for Level 1, and a range closer to 0, which would mean a zero difference. For the 22nd of July 2021, the difference between the two absolute means was smaller but the Level 1 was closer to the reference with an absolute mean of 6.4 °C, against 7.8 °C for Level 2. In addition, it is important to notice that, while Level 2 products appeared to have, in average, higher values than the Sentinel-3 ones, the Level 1 maps presented the opposite behaviour with lower temperatures.

Then, to give more meaning to these numbers, the study of García [2021] has been exploited. Indeed, he studied six different techniques to derive the LST from both Landsat-8 and Sentinel-3 images, using 10 images over the year 2017 for each mission. By averaging the derived LST maps, he could compare the results from each of the six algorithms, which resulted in the Fig. 22. While the first three box-plots correspond to algorithms using Landsat-8 products, the following three refer to algorithms using Sentinel-3 images and the last one is an in-situ reference. Moreover, the yellow box-plot was derived from the official Sentinel-3 SLSTR Level 2 product, also used in the present chapter.

These results revealed important points for the analysis in the present study. Firstly, for all algorithms the values have a large variation around the mean with more than 10 °C of difference between the 25th and the 75th percentile. More specifically, the Sentinel-3 reference product has an IQR of 14 °C. Therefore, the differences found in the present study (see Tab. 16) being smaller than 14, they are not aberrant results.

Moreover, García [2021] demonstrated that the methods based on Landsat 8 algorithms had higher mean values than Sentinel 3 with a mean difference of 3 K, this difference being even greater during dry periods (April-September). As the present study has considered the month of July 2021, that is the hottest month of the year in the MCM, it appeared coherent to find a difference between Level 2 and reference products of 5.2 °C and 7.8 °C depending on the date.

On the other hand, the Level 1 map, with values below the Sentinel-3 reference
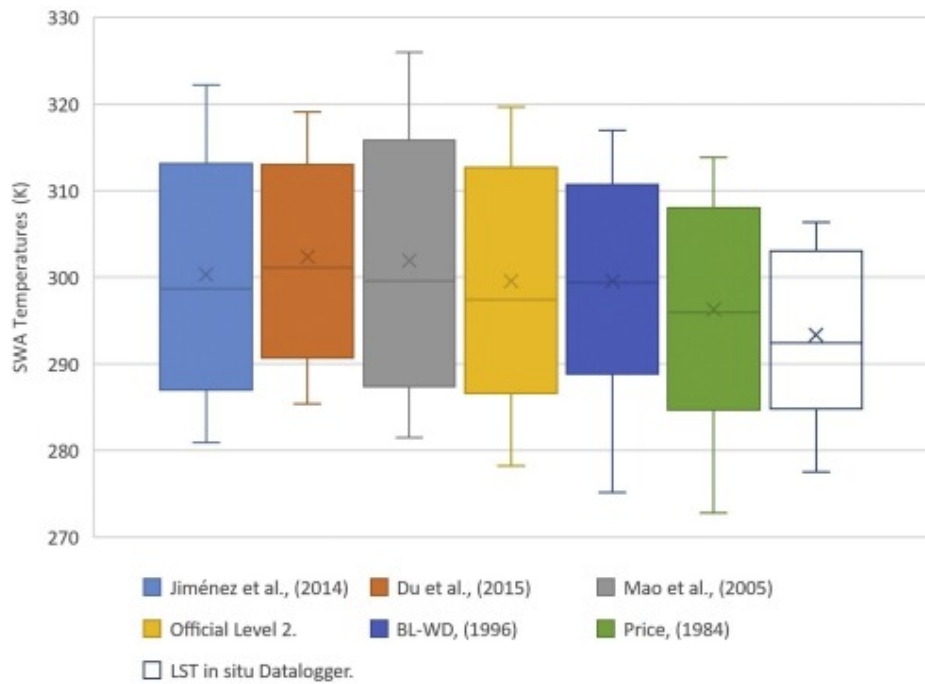
Figure 22: Statistics Landsat 8 and Sentinel 3A
LST algorithms (from García [2021])

product, deviated from the previously mentioned study. Although García [2021] also used Level 1 products, algorithms requiring more knowledge of atmospheric corrections have been applied, which could explain the significant difference in results. Thus, the Level-1 LST maps derived in the present study didn't have enough evidence of their validity to be exploited as they were.

However, when looking at the LST maps from Level 1 and Level 2 of the same day (Fig. 14 and 16 for 06 July, Fig. 15 and 17 for 22 July), many similarities could be observed. The legend scale being different for the two levels, it seemed that they both represent the same quantities, only shifted of a certain bias or linearly transformed. In order to check these hypotheses, the linear correlation between the two maps has been computed through the Pearson's coefficient (see section 4.2.3.2 for definitions), which resulted, for both dates, in a correlation coefficient of 0.99. This revealed a strong positive linear relation, consistent with the hypotheses of bias or linear transformation.

In addition, the per-pixel difference was evaluated, which led to the statistics contained in Tab. 17. With a range of difference of [9.9 °C-16.7 °C] for the 06 July 2021 and [11.3 °C-16.9 °C] for the 22 July 2021, the bias hypothesis seemed less plausible than the linear transformation to describe the relationship between Level 2 product and Level 1 derived map.

As a conclusion, this section compared the Landsat-8 Level 1 derived LST maps with the Landsat-8 Level 2 products and intended to use the Sentinel-3 official SLSTR Level 2 product to validate or reject them. However, significant differences were found between the three products for the two considered dates of 2021. In the absence of ground truth, the Sentinel-3 level 2 product cannot be considered more

	06/07/2021	22/07/2021
Range of diff. (°C)	[9.9 ; 16.7]	[11.3 ; 16.9]
Mean diff. (°C)	13.3	14.1

Table 17: Differences between Level 1 and Level 2 derived LST.

accurate than the Landsat-8 one, nor vice versa. Therefore, in the general case, both sources should be used in parallel when studying the LST without any ground validation. Nonetheless, as Landsat-8 presents a higher resolution of 30m, it should be preferred for local analyses.

Moreover, even if the Level 1 and Level 2 maps turned out to be strongly linearly correlated, the Level 2 maps have shown promise of better accuracy than Level 1. In their validation study, Laraby and Schott [2018] found that when the atmospheric transmission is at least 0.85 and clouds are more than 10 km away, the root mean squared error (RMSE) of Landsat Level 2 LST is at most 0.78 K, which is a very good accuracy of a global ST product. As a result, Landsat-8 Level 2 LST data have been used in the following to study the LST on other dates. Level 1s, on the contrary, have no longer been calculated for these additional dates. However, with an initial idea of a linear relationship between the two Landsat-8 LST levels, further analysis comparing the two LSTs could be the subject of future work.

3.4   Study of LST per LCZ

Once the method for LST retrieval was defined, the analysis of the LST variation between the LCZs could be carried out. The period of 2020-2022 was chosen to perform this task, with the six dates recalled in the Tab. 12.

3.4.1   LST maps for 2020-2022 period

Thus, four additional LST maps were computed from the Landsat-8 Level 2 LST products, using the method described in section 3.3.1. They resulted in four maps whose statistics are summarized in the Table 18.

Date	Mean °C	Stdev °C	Min °C	Max °C
2020 07 19	35.2	5.1	21.8	58.2
2020 09 15	33.8	3.3	21.2	52.2
2021 07 06	39.8	7.1	8.2	65.9
2021 07 22	40.6	5.6	12.0	63.2
2022 07 09	44.6	5.9	23.2	62.8
2022 07 17	48.0	6.2	30.0	67.9

Table 18: Statistics for derived LST maps (2020-2022 period)

However, the two maps of 2020 presented lower temperatures than the other four,

with a mean LST of $35.2\,°\text{C}$ and $33.8\,°\text{C}$. This was not initially a problem but later proved to be a hindrance as it prevented the temperature distribution within each LCZ class from being normal. Indeed, when studying the density histograms for the artificial classes, (as explained in section 3.4.3), three peaks appeared, highlighting a non-homogeneity with respect to temperature (see Fig. 23a and 23b). When the two 2020 maps were removed, these peaks disappeared, revealing that the temperature difference in that year was the cause of the non-normal distribution within the artificial classes. Consequently, the two maps of 2020 were discarded for the study of LST per LCZ.

(a) Built-up, including maps of 2020   (b) Built-up, excluding maps of 2020

(c) Natural, including maps of 2020   (d) Natural, excluding maps of 2020
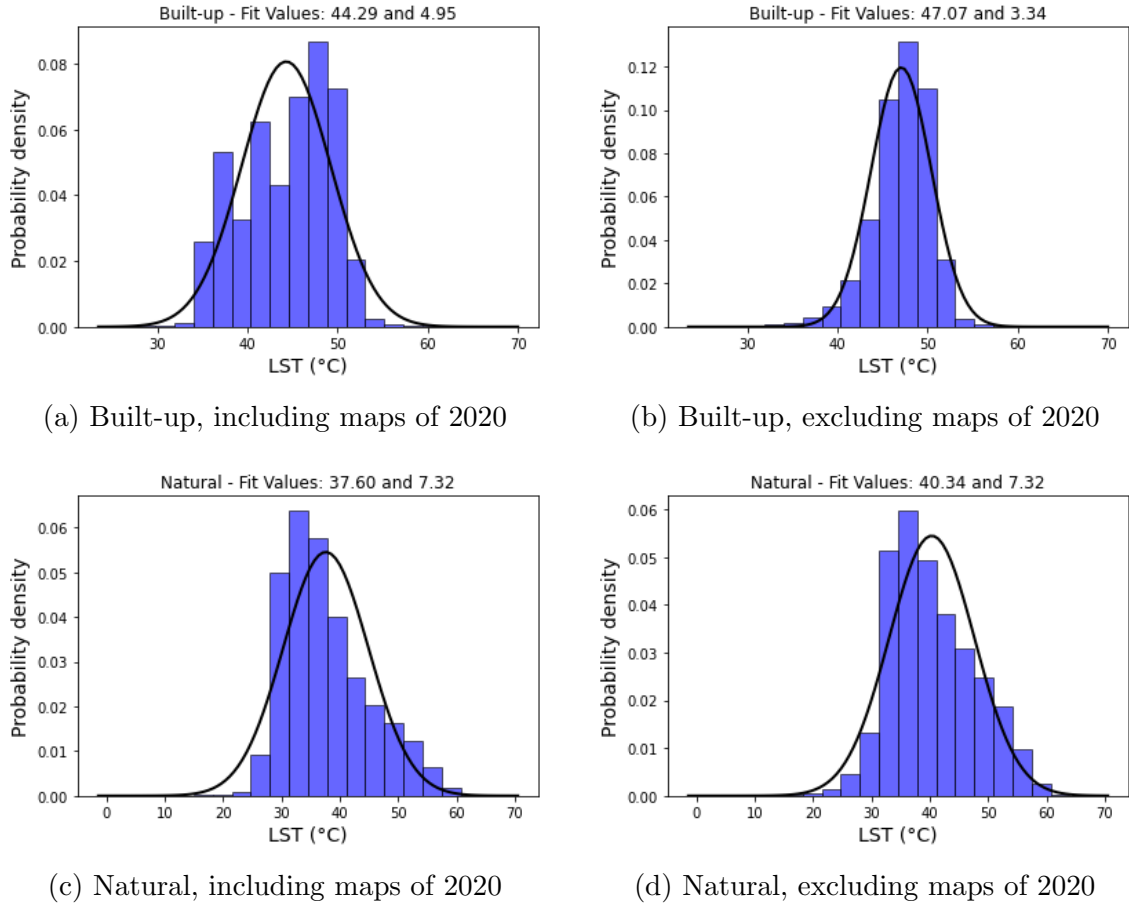
Figure 23: Density histograms of built-up and natural groups

Nevertheless, the remaining four maps, presented in Fig. 24, were used to study the temperature per class of the LCZ system.

3.4.2 LST per LCZ class

The LCZ map derived in the section 2.5 from Landsat-8 images of 2021 was exploited to compute the temperature statistics per class. Fig. 25 depicts the box-plot of each class, with the first five representing artificial zones and the last three representing natural ones.

A first observation concerning the separation of artificial classes from natural classes could be made. Indeed, the built-up climate zones turned out to experience higher

(a) 06/07/2021   (b) 22/07/2021

(c) 17/02/2022   (d) 09/07/2022
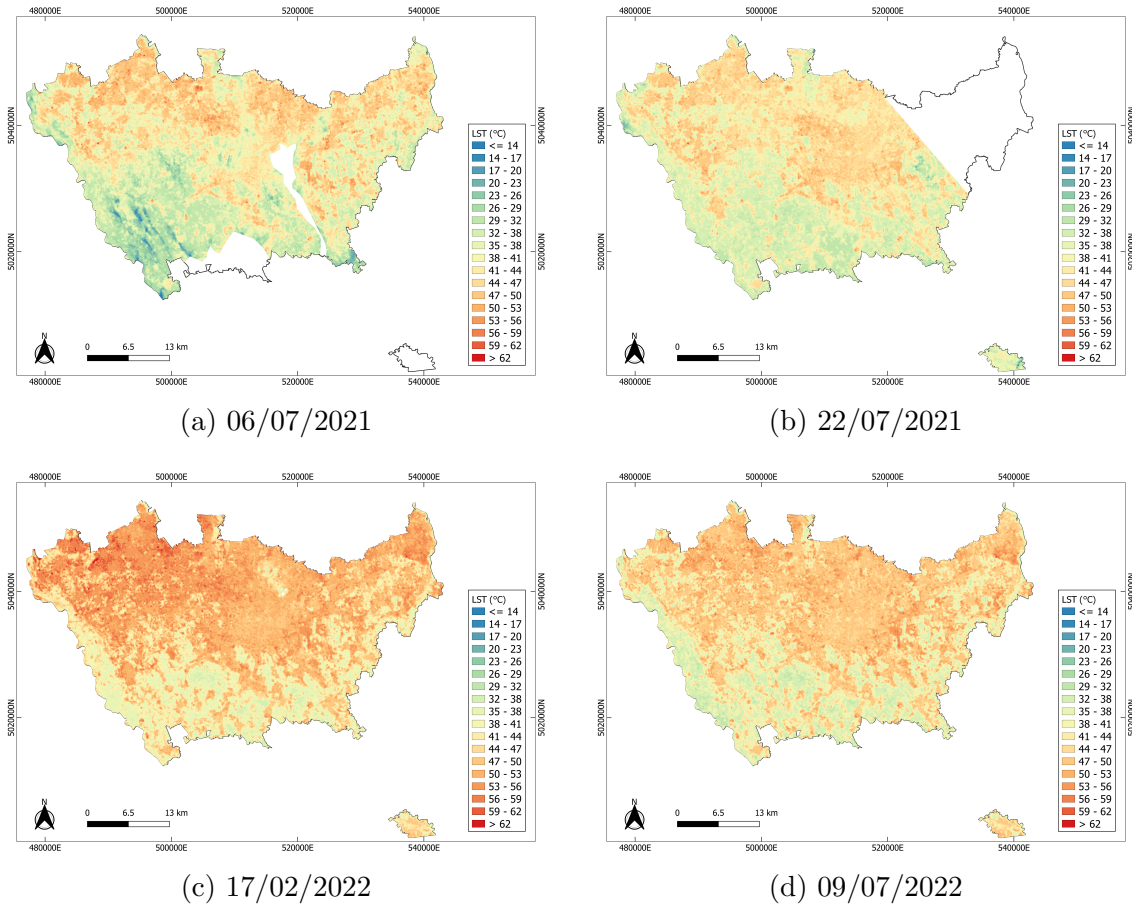
Figure 24: The four considered maps for the period 2020-2022.
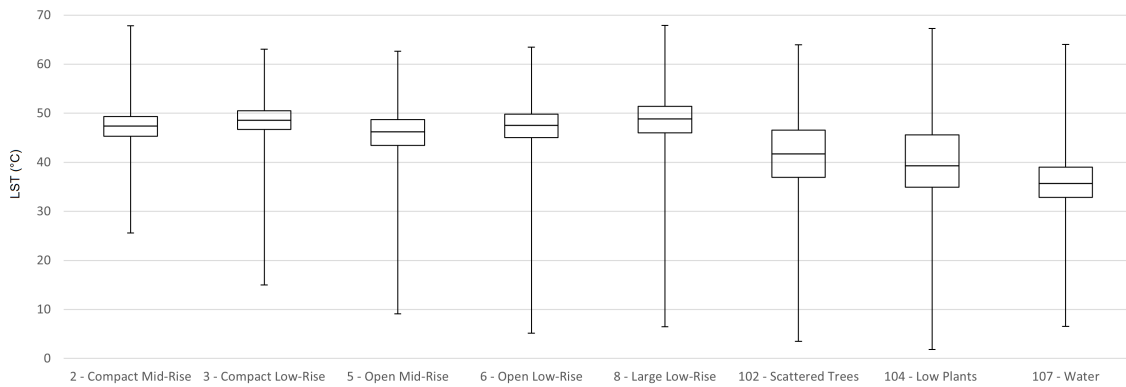
Figure 25: Boxplots of LST per LCZ class (period 2020-2022)

mean temperatures (ranging between 45.9 °C and 48.5 °C) than the non-built-up zones (where average temperatures range between 36.1 °C and 41.8 °C). Nonetheless, the variation within each class being quite large, a study of the statistical significance of the difference in means between artificial and natural classes was carried out in the section 3.4.3.

In addition, when looking only at the artificial classes (from 2 to 8), the means,

ranging between 45.9 °C and 48.5 °C, appeared to be very similar. To better understand if a difference could be detected between built-up classes, a deeper analysis was carried out in section 3.4.4.

3.4.3   Comparison of LST - Artificial VS Natural classes

In the following, the difference in means found between artificial and natural classes was studied under the prism of statistical and practical significances. The eight LCZ classes were merged and reorganized into two groups: artificial and natural areas. Thus, the mean temperature and standard deviation could be calculated for each of these two new classes and are presented in Table 19.

	Mean (°C)	Stdev (°C)
Artificial	47.1	3.3
Natural	40.3	7.3

Table 19: Statistics for Artificial and Natural groups (period 2020-2022)

The objective was then to determine the probability of the difference between the two groups. To do so, several statistical measures exist.

T-Test

The T-Test is an exact statistical test used to determine if there is a significant difference between the means of two groups. It runs on a set of three assumptions that are:

1. **Independence:**   The observations in one sample are independent of the observations in the other sample.

2. **Normality:** Both samples are approximately normally distributed.

3. **Homogeneity of Variances:** Both samples have approximately the same variance.

While the first two could be assumed as true (see the density histograms in Fig. 23b and 23d), the third one could not be validated. Indeed, the built-up group held a variance of 11.1 °C against 53.6 °C for the natural one. Therefore, the T-Test has not been applied.

Z-Test

The two-sample Z-Test is an approximative statistical test that could be used as an alternative as it only requires the first two aforementioned assumptions. It is usually applied in case of large sample size (more than 30 samples) to check if the means of two populations are different or not. This is done by establishing the null hypothesis and the alternative hypothesis and calculating the value of the z-score, as in Eq. 3.11 :

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{3.11}$$

where:

- $\bar{X}_1$ and $\bar{X}_2$: means of the two samples,

- μ_1 and μ_2 : means of the two populations,

- σ_1^2 and σ_2^2 : variances of the two populations, estimated from the data as the sample size is very large (it is the number of pixels i.e. more than 190 000 for both groups),

- n_1 and n_2 : size of the two samples.

The null hypothesis can be stated as $H_0 : \mu_1 = \mu_2$ and the alternative hypothesis as $H_A : \mu_1 \neq \mu_2$.

To check if the statement in the null hypothesis was correct, a significance level was set at 5%. Then, the p-value, or probability value, indicates the probability that the data could have occurred under the null hypothesis. While a small p-value (≤ 0.05) indicates strong evidence against the null hypothesis, a large p-value (> 0.05) indicates weak evidence against the null hypothesis i.e. an inability to reject it.

From the considered data, a very high z-score of 403.3 was obtained, corresponding to an extremely low p-value close to 0. Theoretically, this result would have given enough evidence to reject the null hypothesis and validate the alternative one, meaning that the difference of the two means is statistically significant and that natural areas mitigate high temperature more than artificial ones.

However, Lin et al. [2013] pointed out the p-value problem associated with very large samples. Indeed, with more than 10 000 samples as in the present case, p-values go quickly to zero, and solely relying on p-values can lead the researchers to claim support for results without practical significance. Far from rejecting the use of hypothesis testing, they advocate its use in combination with other considerations such as the effect size and confidence intervals, in order to give more practical meaning to the analysis.

Effect size

An effect size is a measure of the strength of an observed effect of one variable on another. Calculated from empirically observed data, it indicates the practical significance of a research outcome. In the present case, the difference of mean temperature between the natural and artificial groups is the effect size. The greater it is, the more this difference is likely to be meaningful in the real world. Contrary to statistical significance, the effect size is not influenced by the sample size, which makes it an appropriate tool to use in the present study.

One of the most common measurements of effect size is Cohen's d (Cohen [2013]), which is calculated as in Eq. 3.12 :

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}} \tag{3.12}$$

where:

- $\bar{X}_1$ and $\bar{X}_2$: means of the two samples,

- σ_1^2 and σ_2^2 : variances of the two populations, estimated again from the data.

It takes the difference between two means and expresses it in standard deviation units, giving as a result how many standard deviations lie between the two means. To interpret the result, the following guidelines are set:

- $d < 0.01$ indicates a negligible effect,

- $0.01 \leq d < 0.20$ indicates a very small effect,

- $0.20 \leq d < 0.50$ indicates a small effect,

- $0.50 \leq d < 0.80$ indicates a medium effect,

- $0.80 \leq d < 1.20$ indicates a large effect,

- $1.20 \leq d < 2.00$ indicates a very large effect,

- $2.00 \leq d$ indicates a huge effect,

Thus, using the Eq. 3.12 the obtained d-value was 0.94, corresponding to a large effect. The built-up mean LST is therefore 0.94 standard deviations above the natural one and this difference was proved to have a large practical significance.

Confidence intervals

Confidence intervals (CIs) are ranges that are assumed to contain the true parameter value somewhere within them with a fixed probability (Davis [2020]). A hypothesis test can be carried out by checking if the CI includes the null hypothesis value (0 in case H_0 is looking for a difference between two means). The advantage of using CIs over the statistical significance is that CIs imply the magnitude of the effect in addition to the positive or negative response of the effect plausibility. In case of normally distributed data, the CIs for the difference of two means can be calculated using the z critical values, as shown in equ. 3.13.

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{3.13}$$

where:

- $\bar{X}_1$ and $\bar{X}_2$: means of the two samples,

- $z_{\alpha/2}$: z-score associated to a confidence level $\alpha/2$ (for a 95% CI, $\alpha = 0.05$),

- μ_1 and μ_2 : means of the two populations,

- σ_1^2 and σ_2^2 : variances of the two populations, estimated again from the data,

- n_1 and n_2 : size of the two samples.

From this equation, the obtained confidence interval was 6.71 °C-6.75 °C meaning that the difference between artificial mean LST and natural one is at 95% probability between 6.71 °C and 6.75 °C.

Consequently, the analysis carried out in this section confirmed that natural areas mitigate high temperatures more, with a difference in mean LST statistically and practically significant. This difference for the four considered dates of 2021 and 2022 lies in the range 6.71 °C and 6.75 °C with a 95% probability.

3.4.4 Comparison of LST between artificial classes

Next, this section focuses on the artificial areas to determine whether there was a significant difference in temperature between the different classes.

Class	Mean (°C)	Std. (°C)
5 - Open Mid-Rise	45.9	4.2
2 - Compact Mid-Rise	47.1	3.3
6 - Open Low-Rise	47.3	4.3
3 - Compact Low-Rise	48.5	3.7
8 - Large Low-Rise	48.5	4.4

Table 20: Ordered mean LST for artificial classes (period 2020-2022)

The Tab. 20 ranks the five classes according to their mean LST. While class 5 (open mid-rise) appeared to have the lowest mean LST with 45.9 °C, classes 3 (compact low-rise) and 8 (large low rise) exhibited the highest LSTs with a mean of 48.5 °C. The other two classes, including class 2 (compact mid-rise) and class 6 (open low-rise), resulted in an intermediary mean LST with 47.1 °C and 47.3 °C, respectively. These result were similar to other researches (Unal Cilek and Cilek [2021], Han et al. [2022]). Nonetheless, with means ranging from 45.9 °C to 48.5 °C, artificial areas have a maximum difference of 2.6 °C between pairs of built-up classes. Observing that the standard deviations are all higher than this difference (from 3.3 °C to 4.4 °C), it appeared difficult to draw straight forward conclusions on the differences in average LST. However, as in the previous section, an analysis of statistical and practical significance was carried out.

The most common statistical test used to compare the means of multiple (three or more) groups of a single independent variable is one-way ANOVA, also known as the Analysis of Variance test. Like the T-test, it helps to determine whether differences between groups of data are statistically significant. It runs on the same set of assumptions, namely independence, normality and homogeneity of variances.

The first was admitted and, as the sample sizes were large, the normality could be assumed and was confirmed by the density histograms in Fig. 26. However, here again, the variances differed too much to validate the third assumption (see table 21).

(a) Class 2   (b) Class 3   (c) Class 5

(d) Class 6   (e) Class 8
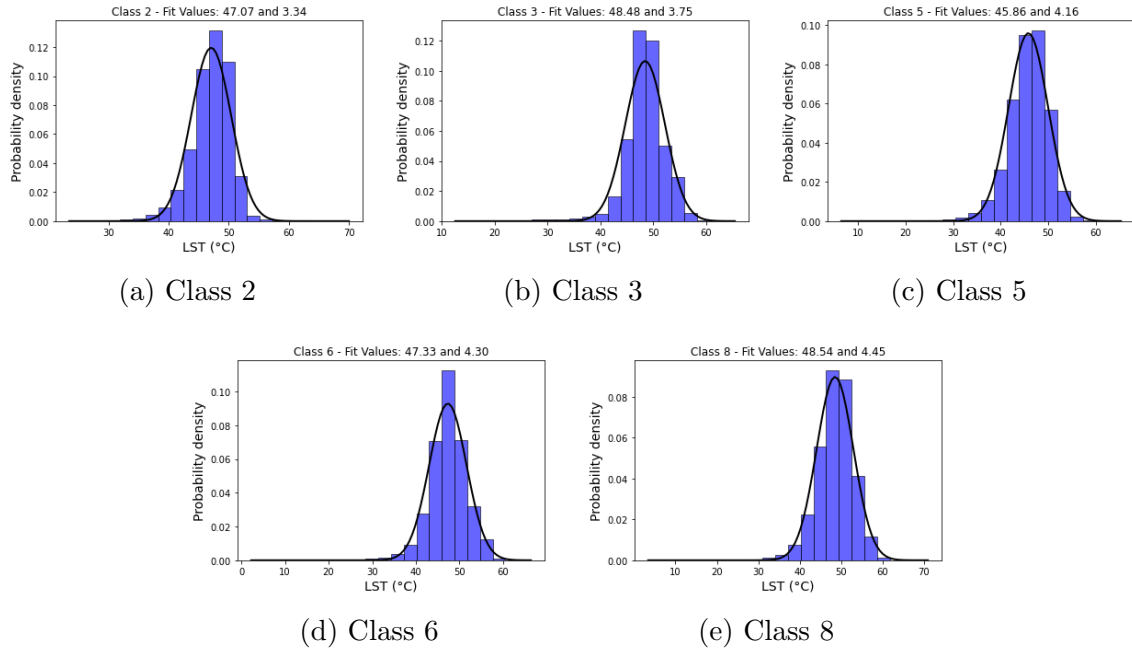
Figure 26: Density histograms for artificial classes (period 2020-2022)

Class	Variance
5 - Open Mid-Rise	11.1
2 - Compact Mid-Rise	14.0
6 - Open Low-Rise	17.3
3 - Compact Low-Rise	18.5
8 - Large Low-Rise	19.8

Table 21: Variance in LST for artificial classes (period 2020-2022)

When the data violates the assumption of homogeneity of variances, the Welch's ANOVA (W-test) can be used as an alternative. Under the only assumptions of independence and normal distribution, it enables to compare the means of three or more groups, as explained by Delacre M [2019]. However, the W-test relies on the p-value and in the same way as for the previous Z-test, the obtained p-value was very close to 0. Indeed, the sample sizes still being very large, with more than 175 000 pixels, the p-value problem persisted.

Therefore, two-by-two comparisons have been performed to detect if any pair of classes had a mean LST difference with practical significance. While Table 22 summarizes the Cohen's d of each pair and their corresponding interpretations, Tab. 23 provides the 95% CIs of their difference.

With a difference of mean LST ranging from $2.59\,°C$ to $2.65\,°C$, classes 3 (Compact low-rise) and 5 (Open mid-rise) turned out to have a d-value of 0.65, corresponding

CLASS	2	3	5	6	8
2 - Compact Mid-Rise	-	Small (0.40)	Small (0.32)	Very small (0.063)	Small (0.34)
3 - Compact Low-Rise	Small (0.40)	-	Medium (0.65)	Small (0.27)	Very small (0.015)
5 - Open Mid-Rise	Small (0.32)	Medium (0.65)	-	Small (0.35)	Medium (0.61)
6 - Open Low-Rise	Very small (0.063)	Small (0.27)	Small (0.35)	-	Small (0.28)
8 - Large Low-Rise	Small (0.34)	Very small (0.015)	Medium (0.61)	Small (0.28)	-

Table 22: Effect size for difference between artificial classes

CLASS	2	3	5	6	8
2 - Compact Mid-Rise	-	1.38 to 1.43	1.19 to 1.23	0.24 to 0.27	1.45 to 1.49
3 - Compact Low-Rise	1.38 to 1.43	-	2.59 to 2.65	1.12 to 1.16	0.04 to 0.09
5 - Open Mid-Rise	1.19 to 1.23	2.59 to 2.65	-	1.46 to 1.49	2.67 to 2.70
6 - Open Low-Rise	0.24 to 0.27	1.12 to 1.16	1.46 to 1.49	-	1.19 to 1.23
8 - Large Low-Rise	1.45 to 1.49	0.04 to 0.09	2.67 to 2.70	1.19 to 1.23	-

Table 23: CIs for difference between artificial classes (in °C)

to a medium effect. The mean LST of the compact low-rise class is therefore 0.65 standard deviations above the open mid-rise one, and this difference has a medium practical significance. This could be explained by the fact that, in contrast to class 3, class 5 contains some trees and mid-rise buildings that introduce some shadows into the area and contribute to the lowering of the LST.

Similarly, the d-value between classes 8 (Large low-rise) and 5 (Open mid-rise) revealed a medium effect size, with a difference of mean LST ranging from 2.67 °C to 2.70 °C and higher temperatures experienced in class 8. Indeed, while the Large low-rise class corresponds to a land cover mostly paved without trees, which identifies mainly industrial and storage areas in the MCM, the Open mid-rise class encompasses more permeable cover, trees and shadows.

On the contrary, all other pairs turned out to be very similar, with d-values highlighting small or very small effects.

Fortunately, this point did not prevent the conclusions of two previously argued observations from being drawn. Indeed, when looking at the order of the classes sorted by increasing average LST (Tab. 20), it appeared that two factors influenced the decrease in temperature: firstly, as expected and proven in section 3.4.3, the presence of trees and permeable cover, and secondly, the height of buildings, introducing some shadows. Indeed, on the one hand, the open mid-rise class, that

encompasses both factors, demonstrated the lowest mean LST, while on the other hand, the large low-rise and compact low-rise classes, which are deprived of both, hold the highest average LST. Moreover, class 2 (Compact Mid-Rise) that only includes mid-rise buildings but lacks of trees, ended up in the middle. The same was true for class 6 (Open Low-Rise) that, on the contrary, carries some trees but only low-rise structures. Therefore, the positive effect of one factor only compensated the lack of the other, resulting in a mean LST that could not be significantly different from classes holding both factors or none of them.

Moreover, it could be highlighted that the difference of density of buildings between classes did not demonstrate any effect on the LST. Indeed, classes 5 and 8 that are the two most different classes in terms of mean LST, both refer to areas of low density. Also, classes 3 and 6 (compact low-rise and open low-rise) that only differ by the density of buildings, were proven to have only a small effect size i.e. a difference in mean LST not practically significant.

To conclude, the analysis of LST between the five artificial classes highlighted two main factors participating in the mitigation of high temperatures. The first is the high height of buildings (over $10\,\mathrm{m}$) since LSTs tended to decrease as the building height increased. It seemed therefore that the presence of tall buildings reduces the heat captured from the ground by creating shaded areas. The second identified factor is the presence of vegetation that increases the portion of pervious materials in the area. These conclusions, that are in line with those of Du et al. [2020], must be qualified, however, as the differences observed have not been shown to be of large significance (only a medium effect at best could be observed). To further investigate the two aforementioned factors, one could take into consideration more images over a wider period of time. As the main obstacle to this achievement was the presence of clouds, a more refined way of dealing with clouds could be sought in order to increase the number of usable images and obtain a clearer result.

3.5   Effect of the new road on LST

The preliminary analysis of the evolution of land cover carried out in section 2.2.1 revealed an important change from 2006 to 2012. Indeed, the motorway called "Autostrada A58" in the east part of Milan was built in June 2012, involving land cover changes over 15 kilometres. Thus, the present section aims to analyse the effect of this new road on the LST, by comparing mean temperatures before and after its construction.

3.5.1   Area of study

To focus particularly on the road, the area of interest (AOI) has been narrowed to a rectangle of $19\,\mathrm{kilometer}$ by $15\,\mathrm{kilometer}$ as shown in Fig. 27.
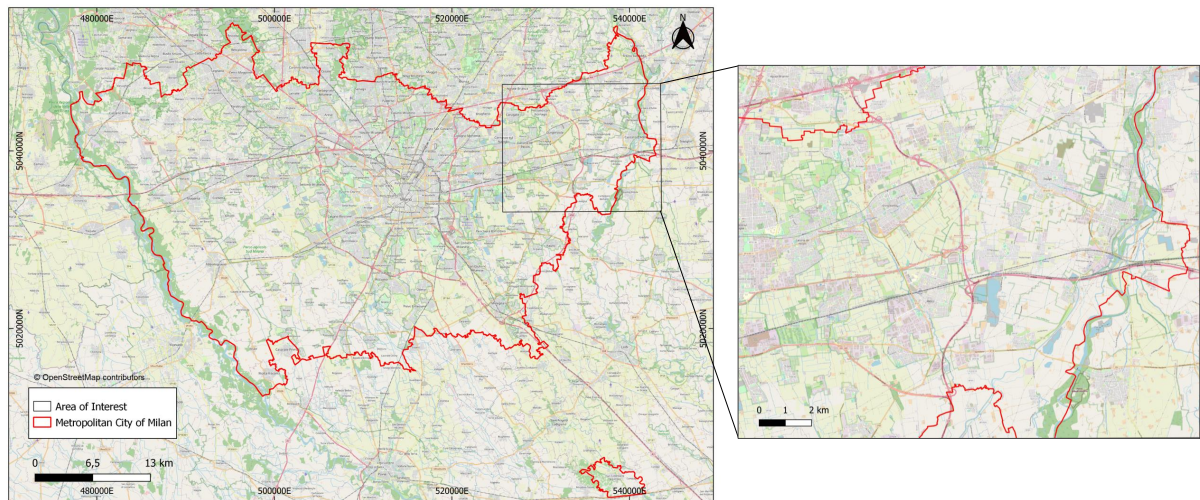
Figure 27: AOI focusing on the new road.

3.5.2   LST maps for the "before" and "after" periods

Then, to represent the period before the construction of the motorway, images over 2006-2009 have been selected. Due to the high presence of clouds during the summer of the years 2006 and 2008, no images could be exploited. However, for 2007 and 2009 four dates were presenting a cloud coverage of less 5% and have therefore been selected (see Tab. 13). The Collection 2 Level 2 LST products were exploited to derive the LST maps presented in Fig. 28, clipped to the new area of interest.

The period 2020-2022 previously analysed has been selected to represent the LST after the construction of the new road. From the six derived LST maps in section 3.4.1, only five were leveraged, namely the 19/07/2020, 05/09/2020, 06/07/2021, 09/07/2022 and 17/07/2022. The image from the 22nd of July 2021 presented clouds over the AOI and was therefore discarded.

3.5.3   Mean LST per period

From the four images of the 2007-2009 period, the mean value per pixel has been computed, giving birth to the mean LST map for 2007-2009, depicted in Fig. 29a. Likewise, the mean LST map for 2020-2022 has been derived and is presented in Fig. 29b.

3.5.4   Mean LST over the road

Then, from both resulting maps, the mean LST along the location of the new road was computed. To avoid over-interpreting the temperature difference along the motorway, which could be due to a simple overall difference over the two periods, the mean LST and its standard deviation over the entire AOI were derived as well. Fig. 30 illustrates the result from which a first observation could arise. While the

(a) 16/07/2007

(b) 02/09/2007

(c) 05/07/2009

(d) 06/08/2009
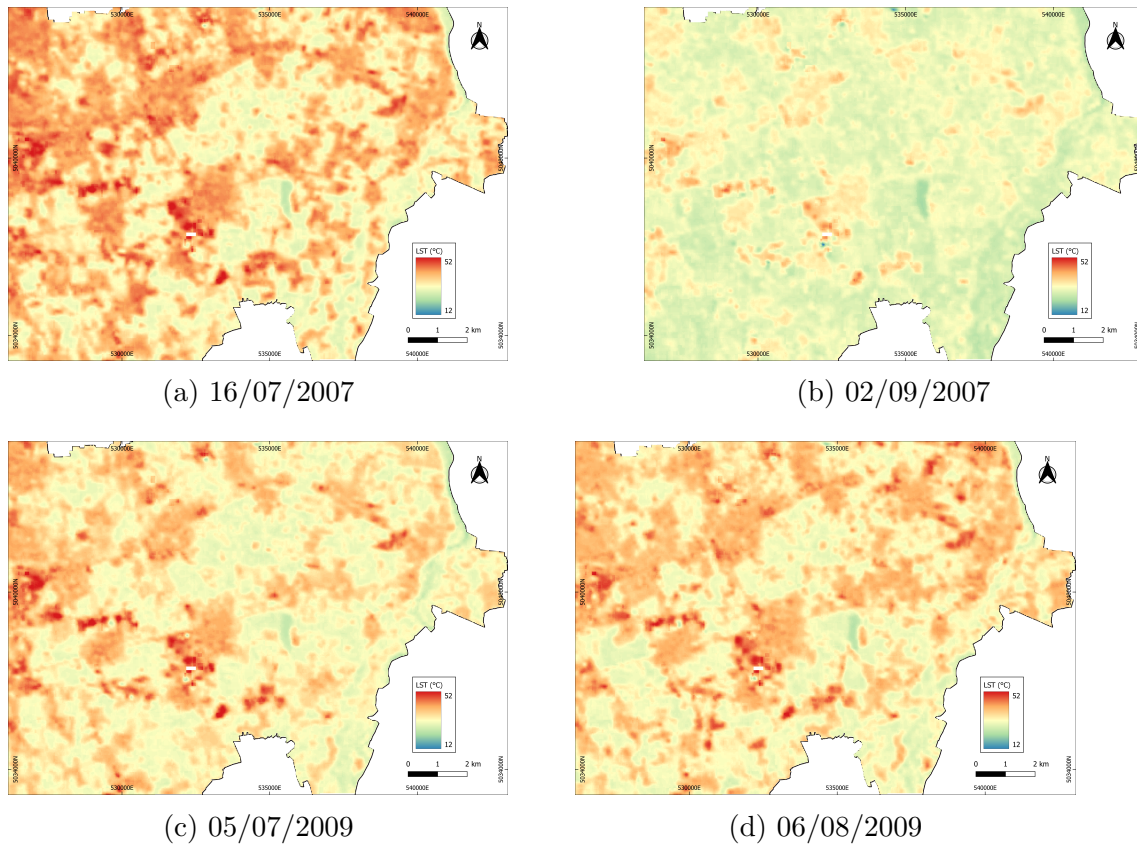
Figure 28: The four considered maps for the period 2007-2009

(a) 2007-2009 period

(b) 2020-2022 period
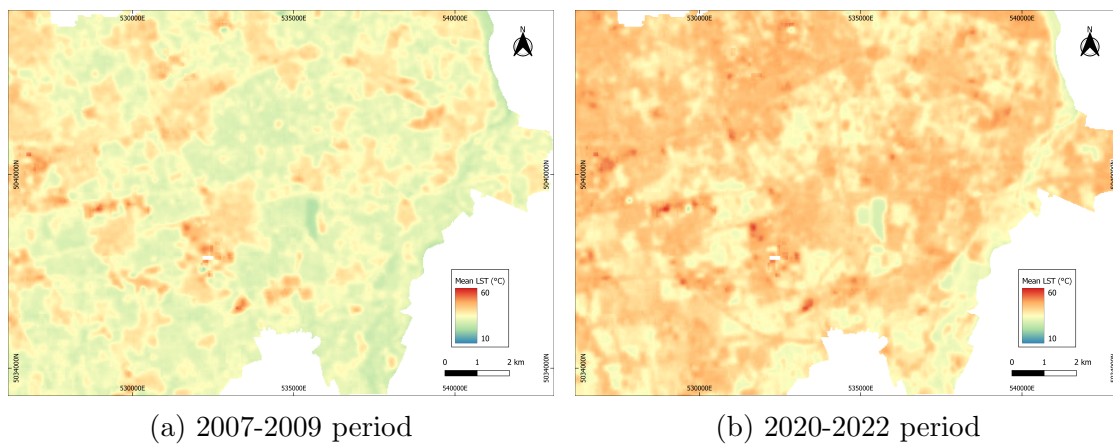
Figure 29: Mean LST maps before and after road construction

difference in mean over the whole area of study is 6.1 °C, the difference in mean over the road is 9.0 °C i.e. about 3 degrees more, thus anticipating an increase of LST due to the construction of the motorway.

3.5.5   Map of difference between the two periods

To further investigate this difference, a second analysis was carried out. The difference between the two maps presented in Fig. 29 (Mean LST over 2007-2009 and
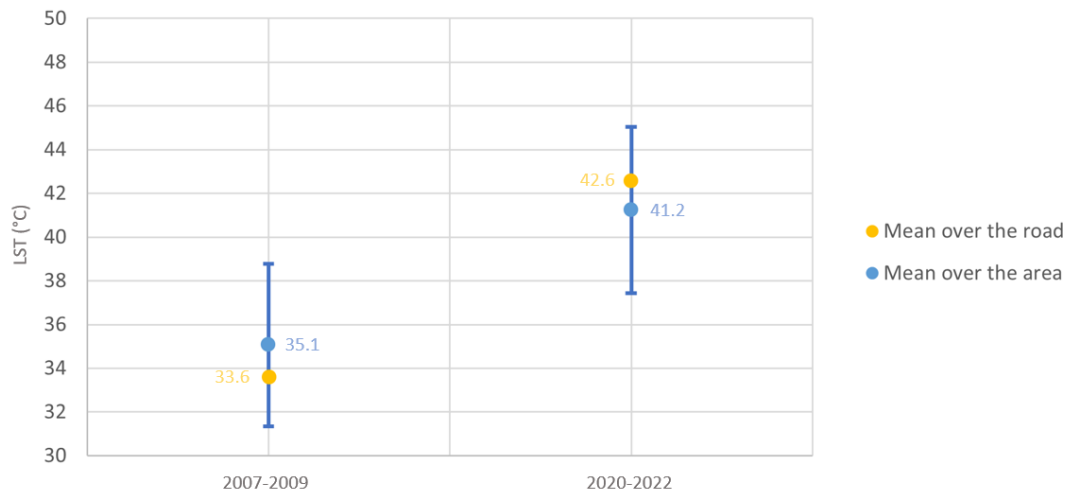
Figure 30: Mean LST per period.

Mean LST over 2020-2022) was computed, resulting in the map shown in Fig. 31. Looking at it, the new motorway could already be discerned, which demonstrates a greater temperature difference in its location.
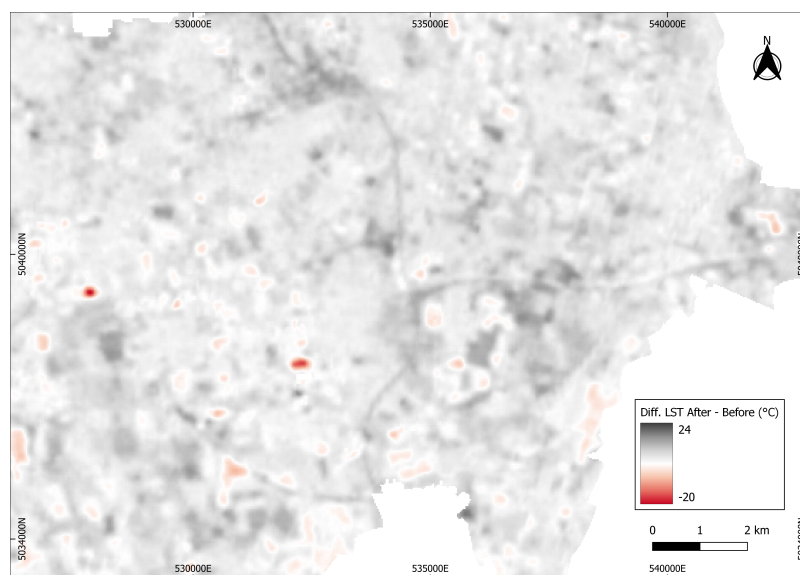
Figure 31: Map of difference in LST (After - Before)

Furthermore, from the map Fig. 31, the mean difference between the two periods over the whole area of study could be derived, as well as the mean difference over the location of the road, depicted in Fig. 32. This last investigation confirmed that soil temperatures at the site of the new motorway were higher after than before its construction. Indeed, while the LST over the AOI increased by 6.2 °C in average, the rise was even higher at the site of the new road with an average of 9.3 °C.

In conclusion, this section demonstrated the alteration of a local climate due to the construction of a motorway. Specifically, an increase of 3.1 °C on average was observed at the location of the new road. This finding could be interpreted in the lights of the conclusions drawn in section 3.4, as motorways are usually lines of
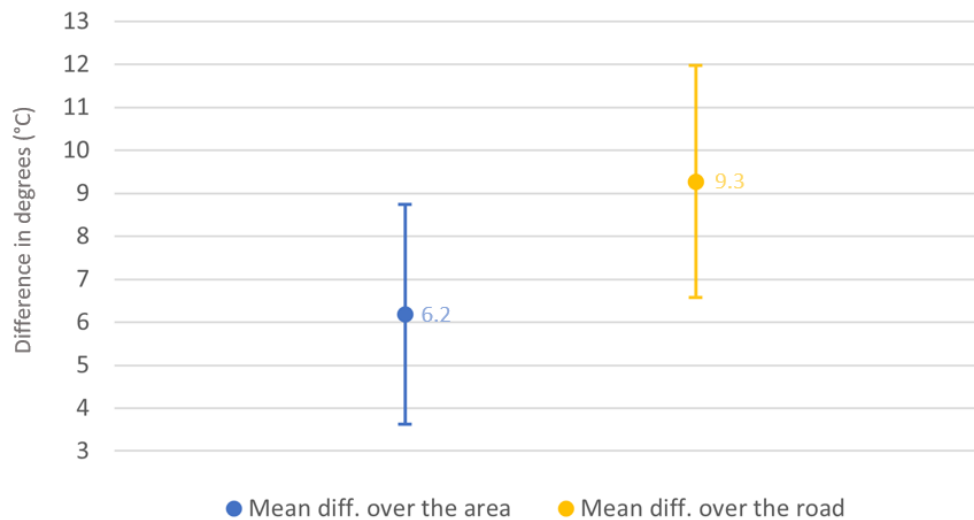
Figure 32: Difference between "mean 2020-2022" and "mean 2007-2009" maps

impermeable concrete, devoid of vegetation and shadows. Thus, they do not hold either of the two factors detected as temperature regulators, which could explain the observed increase of LST. Furthermore, the present conclusion could be used as a starting point for planners who want to study the impact of vegetation and shade surrounding a given road or area. Indeed, as this section showed that a conventional highway has a significant impact on ground temperature, one could wonder whether the presence of vegetation or mid-rise buildings on its sides could reduce it significantly.

4   Air temperature

4.1   Introduction

Air temperature (AT) is the most commonly measured meteorological parameter and is fundamental to climate and global change research. Its estimation and monitoring is useful in many fields such as meteorology, hydrology and ecology and, unlike the surface temperature, it reflects the environmental conditions directly experienced by living beings.

Most current AT measurements are collected at 2 m above ground level by standard weather stations which generally rely on regional infrastructure for weather data collection. Although they acquire data with high accuracy and temporal resolution, these stations provide only point samples whose distribution is rarely designed to effectively reflect spatial variation in air temperature (Vancutsem et al. [2010]). Moreover, official weather stations tend to be strategically located to be operational and therefore do not optimally represent all environments (Prihodko and Goward [1997]).

In this context, many researchers have turned toward remote-sensing techniques that solve the problem of spatial dispersion thanks to their regular and global coverage of the land surface parameters. In particular, the LST recovered from thermal sensor images has been extensively studied to derive an estimate of the near surface AT. According to Vancutsem et al. [2010], this derivation is far from straightforward and, to estimate AT from LST, two approaches may be distinguished, one is for minimum AT, another is for maximum AT.

During the night, as the thermal infrared signal is not affected by solar radiation, the LST data provides a good estimation of daily minimum temperatures. Indeed, Zhu et al. [2013] compared the the MODIS LST nocturne data with official AT records and discovered a correlation of 0.94 and a root mean squared error (RMSE) of 2.97 °C. Similarly, Shen and Leptoukh [2011] reported a correlation higher than 0.95 between minimum AT and nighttime LST.

On the other hand, for day-time temperatures, the result is significantly variable between studies. During day, the relation between LST and AT depends on many factors: solar radiation, cloud-cover, wind speed, soil moisture, surface roughness, vegetation, water vapor and season (Vancutsem et al. [2010], Shen and Leptoukh [2011]). Therefore, depending on the study, the area of interest and the period considered, the results are varying. While do Nascimento et al. [2022] and Zhu et al. [2013] found a correlation higher than 0.8 in their study focusing respectively on São Paulo (Brazil) and the Xiangride River basin (Tibet), Shen and Leptoukh [2011] found significant variation of the linear regression coefficients from one station to another in northern Eurasia regions. Nonetheless, Iqbal and Ali [2022] studied the relation between AT and LST in Pakistan through various predictive models (linear, quadratic, cubic, exponential) and reported the linear function to be the most appropriate with values of correlation of 0.78 and 0.62 for two stations during summer.

Therefore, being dependent on many local factors, the relationship between AT and LST cannot currently be globalized. Nevertheless, according to Jin and Dickinson [2010], although not similar, LST and AT are complementary in their contribution of valuable information to the study of climate change. Hence, after the study of LST conducted in Chapter 3, the following focuses on air temperature with two main aims: (i) Verify the existence and type of correlation between day-time AT and LST for the Metropolitan City of Milan, (ii) Analyse the variation of AT between the LCZs. As various in situ data were considered for the air temperature, a first description of the data sources and pre-processing is made in the following section.

4.2   In situ data sources

Ground sensors can be divided into two main categories: quality-controlled reference weather stations and crowdsourced networks. While the firsts are traditionally managed by public authorities, the seconds are hold by amateurs sharing their observations.

4.2.1   Reference and quality-checked weather stations

In each region of Italy, the official environmental and weather monitoring is carried out by the Regional Agency for Environmental Protection (ARPA, Agenzia Regionale per la Protezione dell'Ambiente) which provides free access to its data. For ARPA Lombardy, the datasets are published on the Lombardy Open Data Portal. In particular, the air temperature measures are included in the weather sensor datasets (Dati sensori meteo), downloadable yearly as CSV files.

In the Metropolitan City of Milan, 15 stations have monitored the air temperature from 1989 to present, with a frequency of 10 minutes. Their spatial distribution is reported in the Fig. 33.

4.2.2   Crowdsourcing weather stations

Despite the high quality of the reference station measurements, their limited number over the territory does not provide sufficient spatial resolution to study temperature variations at the very local scale. As a result, turning to crowdsourcing appears to be a suitable complementary solution as it offers measures of stations distributed worldwide with a high spatial density. It is even more relevant since the density is higher in urban areas, where very few reference stations are usually located to ensure precise and representative observation, not affected by urban heat noises.

However, crowdsourced atmospheric data do not benefit from the calibration, quality assurance and control of traditional data, hence data correction is a crucial step before any analysis can be carried out.

4.2.2.1   Netatmo stations

The amateur weather stations chosen to be used in this study are the Netatmo weather stations (Netatmo [2021]) as they are distributed worldwide with a high density, especially in our study area as showed by the Fig. 33, and the data is made available for free via an application programming interface (API). The stations are composed of two modules, an indoor and an outdoor module, that are easy to set up and require a minimum amount of maintenance.
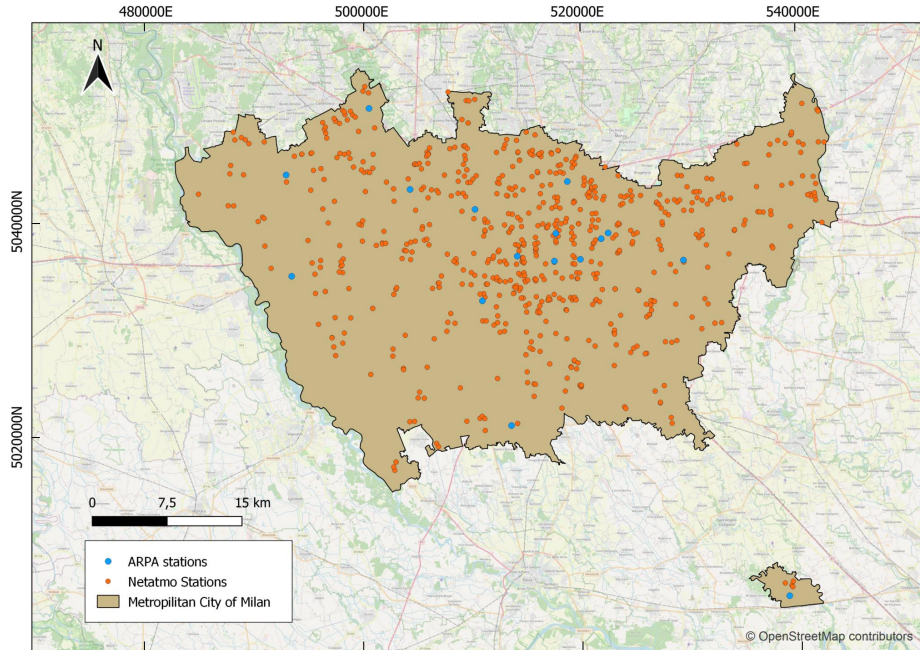
Figure 33: ARPA and Netatmo weather stations in 2021

The outdoor module measures several variables including at least the outside air temperature (specified manufacturers accuracy: $\pm 0.3\,°C$) and relative humidity ($\pm 3\%$) at 5 minutes intervals and are transferred automatically to the Netatmo server via Wi-Fi connection. Moreover, the users can choose whether their data is public or not, which in the first case makes it accessible to everyone via an API.

Using this public Netatmo API and the patatmo Python package developed by Büchau [2017], two methods have been used. First, the *Getpublicdata* method, returning instantaneous data from all stations within a specified area, along with their metadata (identifier, latitude, longitude and altitude of the station). Then, the *Getmeasure* method is returning the data for a specific station in a given time range and was thus used to extract the past records of each station. The returned data is a dictionary containing for each hour of the selected time range, the average, minimum and maximum measured temperatures, with the corresponding timestamps.

The number of available stations over the Metropolitan City of Milan depends on the year but ranges from 55 in 2014 to 688 in 2021 and increases with time. No data was recorded before 2014 in our area of study. The number of available stations before data cleaning is reported in table 24.

Year	Nb of stations
2021	688
2020	611
2019	490
2018	368
2017	287
2016	216
2015	120
2014	55

Table 24: Number of Netatmo stations in the MCM per year

4.2.3   Data cleaning

Data cleaning, also called data cleansing or data scrubbing, is an essential step before using any dataset. Essentially, garbage data in is garbage analysis out. Thus, data cleaning is the process of spotting and correcting incorrect, corrupted, duplicate, or incomplete data in a dataset and enables to improve its quality. The first step before exploiting in situ measurements has therefore been the data cleansing.

4.2.3.1   ARPA Data pre-processing

In order to clean Netatmo data, the stations of ARPA have been used as a reference. However, the downloaded ARPA data could not be used directly and needed a 3-steps pre-processing. The Fig. 34 summarizes the pre-processing flowchart of ARPA data, that have been followed for all years from 2018 to 2021.
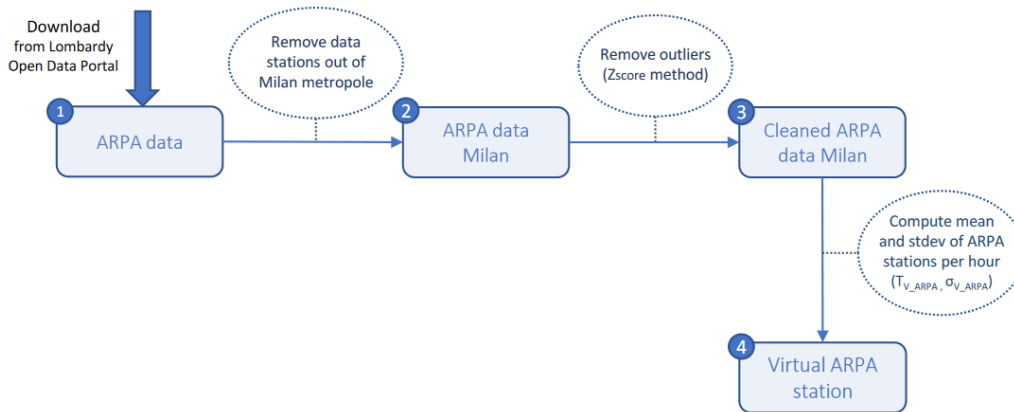
Figure 34: ARPA data pre-processing steps

Filter Milan stations

ARPA data is downloaded from Geoportale Lombardia and is provided in a CSV file containing all measures of all weather stations of the Lombardy region for the

selected year. Therefore, the first step was to filter the CSV file in order to keep only the measures of temperature and only for the 15 stations inside of the Metropolitan City of Milan. The output was saved in a file named "ARPA_data_Milan".

Remove outliers

Although their quality is checked by regional authorities and meet quality standard certification as requested by the national law, the ARPA temperature time series were found to have some outliers. For example, some stations had recorded temperatures bellow $-30\,°C$ or higher than $60\,°C$, or recorded a $30\,°C$ temperature in a winter month. Therefore, a cleaning was carried out on the ARPA_data_Milan file using the Z-score method.

The one sample Z-test is a statistical concept that helps to understand if a data value is greater or smaller than the mean and quantify how far away it is from it. More specifically, the Z-score is the number of standard deviations away from the mean that a certain data point is. It is defined as follow, for a point x, a mean μ and a standard deviation σ:

$$Z_{score} = \frac{x - \mu}{\sigma} \tag{4.1}$$

In the Z-score method for outlier detection, if the z-score of a data point is more than 3, the data point is considered to be very different from the others and therefore an outlier. In this study, the Z-score of each observation has been computed with Python using the mean and standard deviation of the corresponding month. For 2018, 2019, 2020 and 2021, the number of identified and removed outliers were 1261 (0.18%), 2999 (0.41%), 737 (0.10 %) and 1570 (0.20%), respectively. The output was saved in a file named "Cleaned_ARPA_data_Milan".

Create the ARPA virtual station

For some steps of the Netatmo data cleaning (see section 4.2.3.2), it was convenient to compare the amateur measures with the values of a single virtual ARPA station. The latter consists of a single time series whose values are the hourly averages of the measurements of the ARPA stations. The hourly minimum and maximum together with the standard deviation have also been computed and stored with the means in a file named "Virtual_ARPA_station".

4.2.3.2   Netatmo data cleaning

Being based on crowdsourced stations, the Netatmo data is de-facto provided with no quality standard certification. Therefore, several types of outliers had to be removed:

- **Outlier times series**: entire time series that is completely different from the others. They can be detected only when compared with a reference time series, in this study the measures of the ARPA virtual station. Two types of outliers times series can be distinguished:
    - **Uncorrelated time series**: Time series with low correlation with the reference station

> – **Biased time series**: Time series with high correlation with the reference station but containing a bias i.e. the mean is globally higher or lower.

- **Unrealistic values**: values out of the possible range of normal temperature for Milan (for example, 60°C or -15°C).

- **Point outliers**: value that is unusual in a specific time instant when compared either to the other values in the time series (global outlier) or to its neighboring points (local outlier).

To remove properly these outliers, two different approaches were tested. The flow-charts Fig. 35 and Fig. 36 represent their steps and the removals of measures associated to each step is summarized in the Table 26. Even if the correction was applied for years 2018 to 2021, only removals associated to 2021 were reported since similar percentages are retrieved for the four years.

Removal of uncorrelated times series

For both approach A and B, the first step has been to compute the correlation of each of the station time series with the one of the virtual ARPA station. To do so, the Pearson's coefficient (also known as Pearson's R or bivariate correlation) has been used. It is the most common way to measure a linear correlation between two sets of data. Denoted with the symbol "R", it is the ratio between the covariance of two variables and the product of their standard deviations:

$$r_{XY} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \tag{4.2}$$

where $cov(X, Y)$ is the covariance between X and Y; σ_X and σ_Y are the standard deviation of X and Y, respectively.

The Pearson's correlation coefficient returns a value between -1 and 1, where:

- -1 indicates a strong negative linear relationship
- 0 indicates no linear relationship at all
- 1 indicates a strong positive linear relationship

In this study, times series with a Pearson's coefficient lower than 0.6 were considered to have a very low correlation with the ARPA stations and have been removed.

Removal of unrealistic values

For both approach A and B, the unrealistic values were removed by using, for each month, the minimum and maximum values of the virtual ARPA station, min_{ARPA} and max_{ARPA} respectively. Every value out of the range $[min_{ARPA}-2\,°\mathrm{C}$; $max_{ARPA}+2\,°\mathrm{C}]$ were considered as unrealistic values and therefore removed.
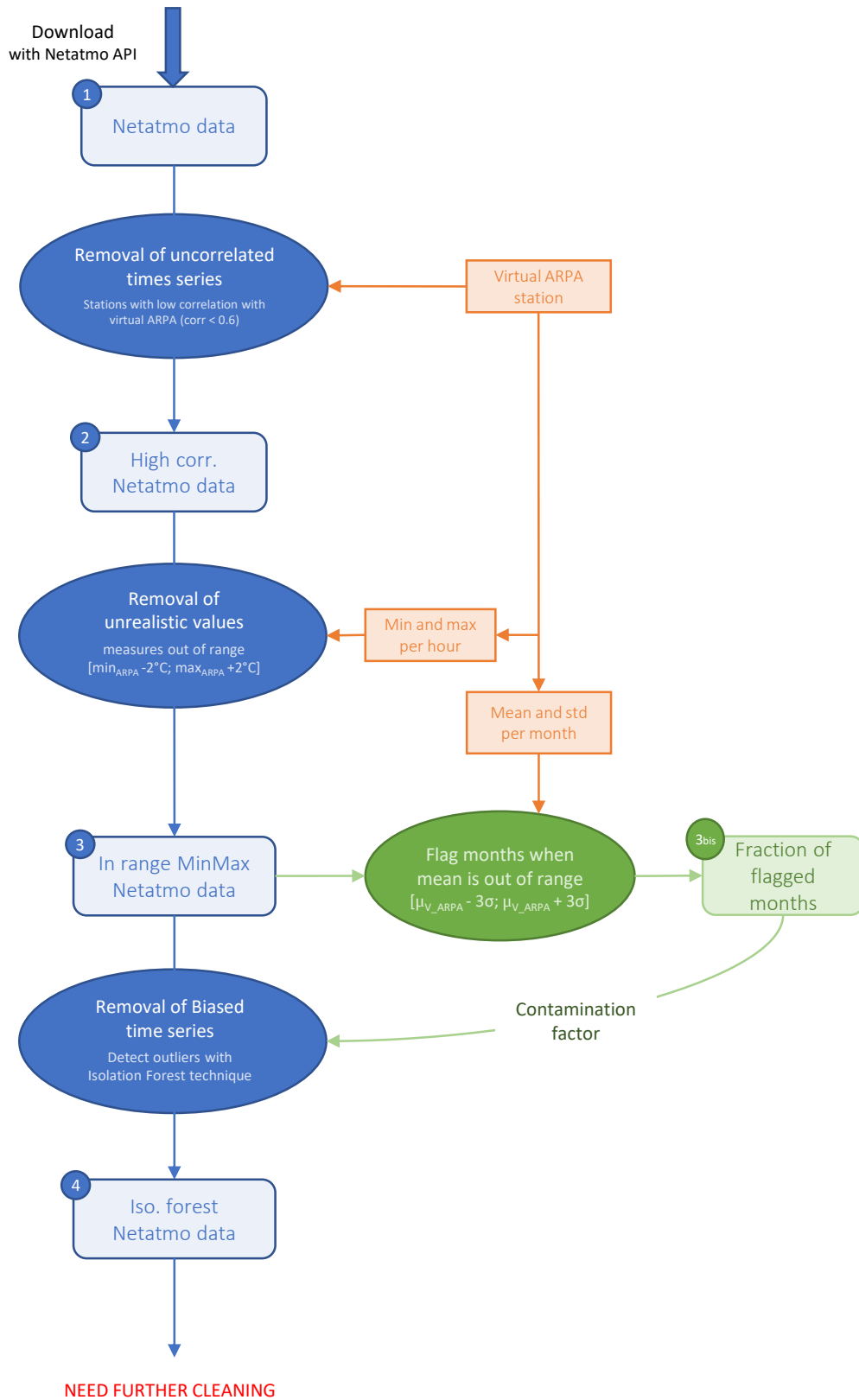
Figure 35: Netatmo cleaning steps - Approach A
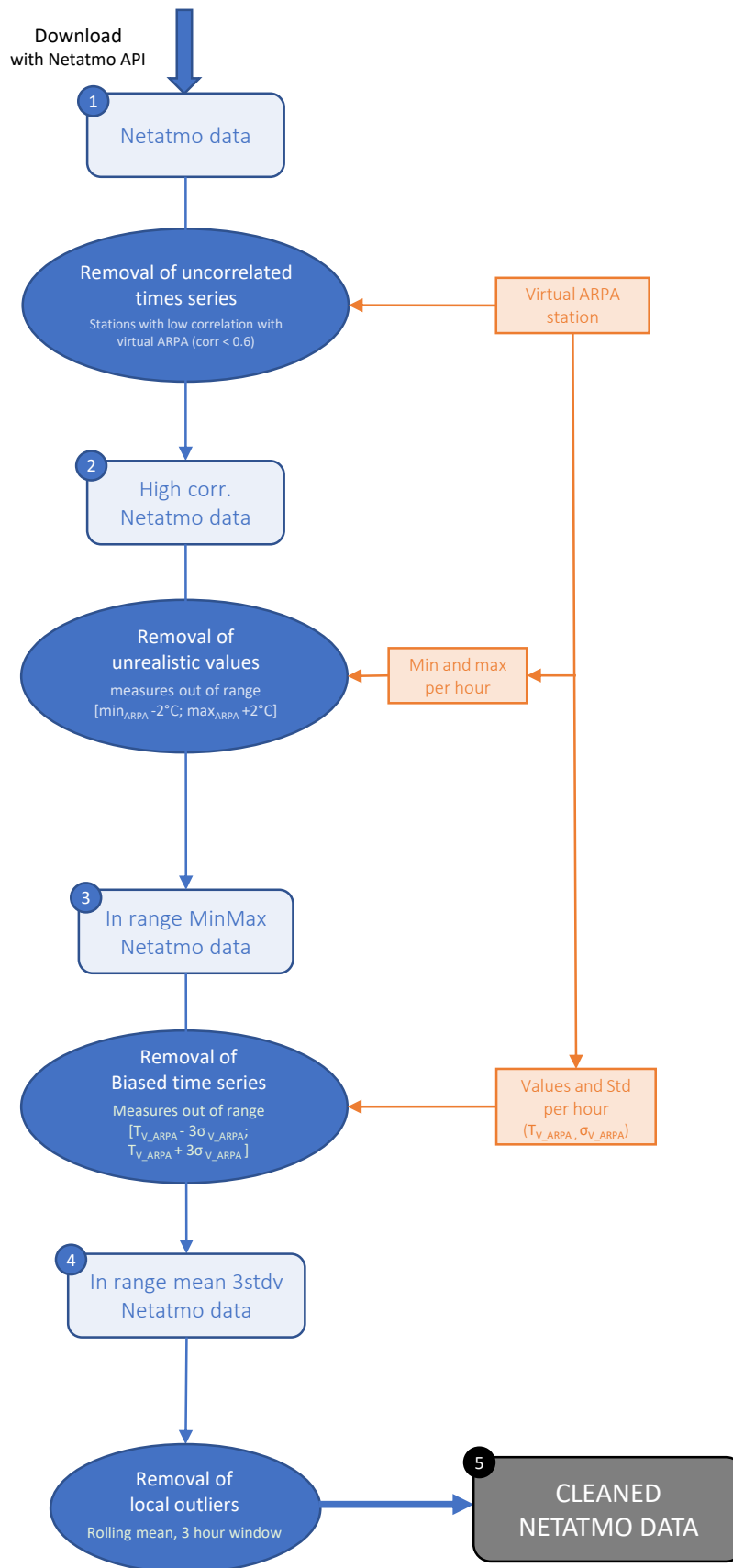
Figure 36: Netatmo cleaning steps - Approach B

Removal of Biased time series – Approach A

In the first approach, the idea was to apply an automated anomaly detection algorithm in order to remove, for each month, the time series with strong bias with respect to the others. One of the most widely used algorithms for anomaly detection is Isolation Forest, which uses a tree-based unsupervised learning approach to detect unusual points. It has the advantage of being fast with very big datasets like in this study, and efficient to detect values that are few and different.

Isolation Forest can directly detect anomalies using isolation, i.e. estimating how far a data point is from the rest of the data. However, Isolation Forest models require hyper-parameter tuning to generate their best results, particularly the "contamination" value. This parameter sets the fraction of points in the data to be anomalous. By default, this parameter is set as "auto", meaning that a generic threshold is chosen, but this may affect a lot the model results.

Therefore, an intermediate step was added to estimate the contamination factor for each month. The virtual ARPA station measures were aggregated month by month into a single mean (μ_{ARPA}) together with the corresponding standard deviation (σ) in order to be compared with the mean of each Netatmo station. Each crowdsourced mean value outside the interval [μ_{ARPA}-3σ ; μ_{ARPA}+3σ] was flagged as an outlier. Thus, the fraction of flagged months could be used as an estimation for the contamination factor. The results for 2021 is shown in Table 25. It can be noted that for the months where no station was flagged, the contamination parameter was arbitrarily set to 0.001.

Month	Nb of flagged stations	Total nb of stations	% of flagged stations	Contamination factor
January	5	426	1.12	0.012
February	1	437	0.23	0.002
March	0	449	0.00	0.001
April	0	442	0.00	0.001
May	0	433	0.00	0.001
June	0	439	0.00	0.001
July	0	446	0.00	0.001
August	0	375	0.00	0.001
September	0	370	0.00	0.001
October	0	447	0.00	0.001
November	3	456	0.66	0.007
December	2	486	0.41	0.004
All year	11	5206	0.21	-

Table 25: Removals and contamination factors used for 2021.

Once the contamination factor estimated, the Isolation Forest algorithm could be run for each month. The result for 2021 is depicted on the Fig. 37. By looking in details at it, it turned out that all previously flagged stations were removed, plus some more months (28 months, i.e 0.35% of the data for 2021). Moreover, a closer examination of the months removed by Isolation Forest but not previously flagged

was carried out and revealed that they were effectively strange and probably outliers. The algorithm was therefore satisfying at this point.
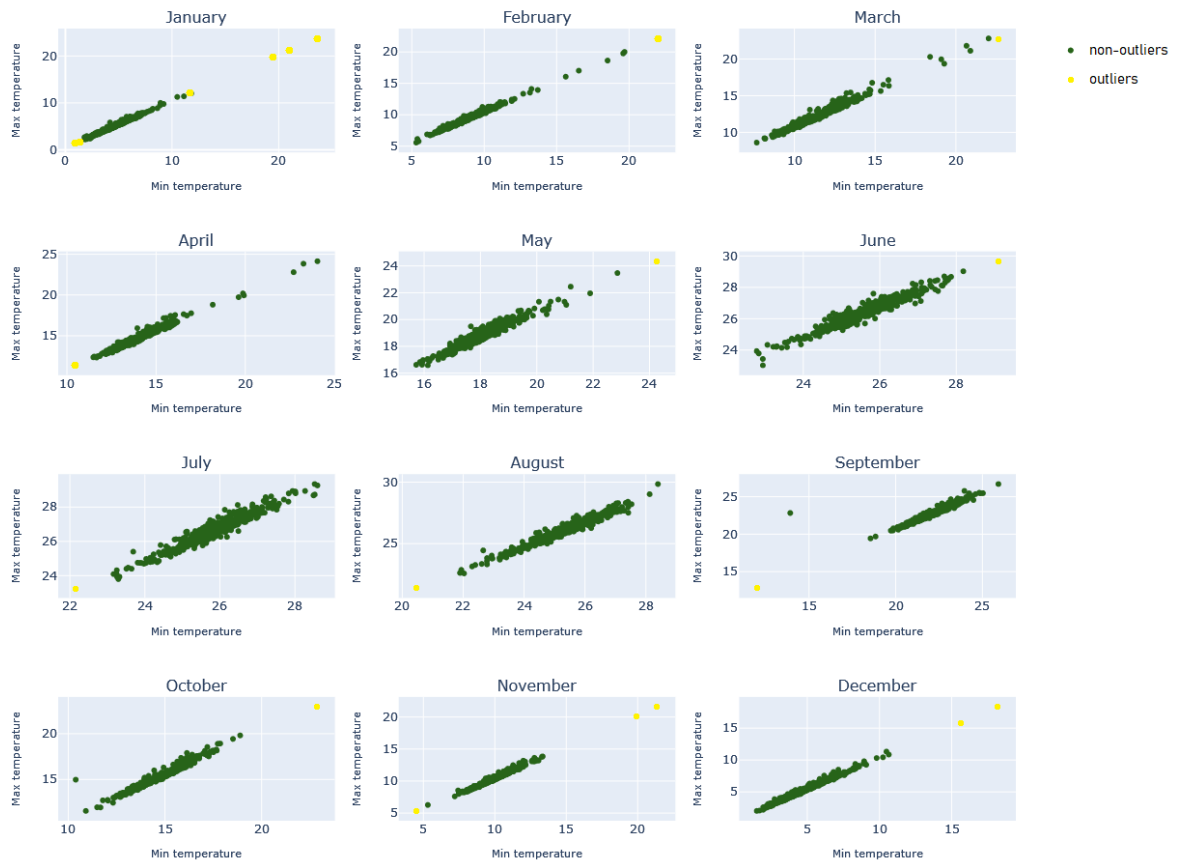
Figure 37: Result of Isolation forest - Flagged outliers per month (2021)

Removal of point outliers – Approach A

After the Isolation Forest cleaning, point outliers were still to be removed. Several anomaly detection and prediction approaches were tested, such as Local Outlier Factor (LOF), Isolation Forest, Density-Based Spatial clustering of applications with noise (DBScan) and the Mean Absolute Deviation (MAD) but none of them were satisfactory. The main issue was that the periodicity of the temperature was not well taken into account and therefore only high peaks were removed, even for heat peaks that were probably accurate when compared to the ARPA station values. Another approach, more classical, was therefore needed.

Removal of Biased time series – Approach B

In the second approach, the biased time series were removed by using the hourly values of the virtual ARPA station. For each station and each hour h, all measures outside the interval $[T_{ARPA}(h)\text{-}3\sigma \; ; \; T_{ARPA}(h)\text{+}3\sigma]$ were flagged as outliers and therefore removed. This step caused the removal of a significant amount of data for each considered year (35.6%, 19.5%, 29.8% and 18.17% for 2018, 2019, 2020 and 2021 respectively) but allowed for the removal of biased time series with high certainty.

Removal of point outliers – Approach B

Finally, to eliminate point outliers, a moving average was applied to the time series of each station. Also called rolling mean, the moving average is a type of arithmetic average calculated by taking the mean of a given set of values over a specified period. For this study, a 3 hour window was considered. This last step resulted in the final cleaned Netatmo dataset including measurements from the years 2018 to 2021.

Step	Nb measures	% removed by the step	% removed from inital
Initial data	3881171	-	-
High correlation	3875048	0.16	0.16
In range Min-Max	3871169	0.10	0.26
A. Isolation forest	3857792	0.35	0.60
B. In range mean 3std	3167683	18.17	18.38
B. Rolling mean	3167683	0	18.38

Table 26: Cleaning removals per step, for Netatmo data (2021)

4.3   Correlation between LST and Air temperature

Once the data was cleaned, the first objective of this chapter was to study the correlation between the land surface temperature (LST) and the air temperature (AT). As the number of measurements and available Netatmo stations decreases when going back in time, only the period 2018-2021 was considered to ensure a maximum coverage of the MCM.

4.3.1   LST retrieval for 2018-2021

The LST maps derived in the chapter 3 only covered four dates of the summers 2020 and 2021. Therefore, in order to build a more complete LST time series over the period 2018-2021, additional dates were considered. This time the selection was not restricted to the summer season and all images with less than 5% cloud cover were chosen. It resulted in 36 different dates, reported in the table 27. All images were acquired at approximatively 10:10AM and correspond to the Collection 2 Level 2 Landsat-8 products. Then, they were converted into 36 LST maps using the method explained in section 3.3.1.

4.3.2   Correlation with ARPA stations

The correlation between LST and AT was first studied using the official data of the fifteen ARPA stations. To do so, the measures to consider had to coincide with the acquisition date and time of the LST maps. Therefore, for each of the 36 selected dates and each ARPA station, the average temperature between 9:00 AM

2018	2019	2020	2021
19/01/2018	06/01/2019	14/04/2020	11/01/2021
08/03/2018	22/01/2019	01/06/2020	28/02/2021
25/04/2018	27/03/2019	19/07/2020	16/03/2021
27/05/2018	28/04/2019	20/08/2020	19/05/2021
28/06/2018	30/05/2019	05/09/2020	06/07/2021
30/07/2018	17/07/2019	24/11/2020	22/07/2021
15/08/2018	18/08/2019	26/12/2020	24/09/2021
02/10/2018	03/09/2019		10/10/2021
18/10/2018	05/10/2019		26/10/2021
			05/12/2021
			13/12/2021

Table 27: Selected Landsat-8 images for the period 2018-2021.

and 11:00 AM has been calculated, which resulted in AT time series of 36 dates for each station. Then, the LST at each station location was extracted from the 36 maps, which generated a LST time series of 36 dates for each station.

To compute the linear correlation between AT and LST, the Pearson's coefficient was used and revealed a strong correlation ranging from 0.91 to 0.99 for the fifteen ARPA monitoring points. This result agrees with the linear correlation between LST and AT found by Iqbal and Ali [2022] and the linear coefficients of Colombi et al. [2007], higher than 0.75 and often higher than 0.90.

4.3.3 Correlation with Netatmo stations

In a second step, this correlation was also evaluated using Netatmo data. In the same way as for ARPA, the average of the measurements from 9:00 AM to 11:00 AM of each date was calculated for each station and the corresponding LST extracted from the 36 maps. In order to avoid inconsistent results, stations with less than 10 measurements out of the 36 dates were excluded from this analysis. Thus, the data from 351 stations could be exploited and were also found to have a strong linear correlation with the LST with a Pearson's coefficient ranging from 0.85 to 0.99. In fact, only 3 stations were below 0.9, which showed a consistency with the results found for the ARPA stations.

However, this strong correlation does not imply that the LST and the AT are identical. Indeed, with a Mean Error (ME, equ. 4.3) of 8.98 °C and the Root Mean Square Error (RMSE, equ. 4.4) of 10.54 °C, the LST appeared to be higher than AT and a transformation should be considered in order to use it as an estimate of AT. As reported in the introduction of this chapter, this transformation is not straight forward and is not analyzed further in this study.

$$ME = \frac{1}{n} \sum (LST - AT) \tag{4.3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum (LST - AT)^2} \qquad (4.4)$$

4.4   Air temperature per LCZ

The second objective of this chapter was to analyze the air temperature variation between the LCZs. To do so, the ARPA and Netatmo in situ measures from January 2018 to December 2021 were exploited together with the LCZ map built in the section 2.5. Firstly, the LCZ class corresponding to each in situ station was extracted. The number of station per class over the period 2018-2021 is reported in the table 28. It can be noted that no station is present in the water areas and very few are placed in natural zones, mainly because Netatmo users tend to place their device at their house location.

Class	2	3	5	6	8	102	104	107	
Name	Compact Mid-Rise	Compact Low-Rise	Open Mid-Rise	Open Low-Rise	Large Low-Rise	Scattered Trees	Low Plants	Water	Total
Nb. stations	49	57	95	189	44	5	8	0	447

Table 28: Number of stations per class (2018-2021)

Several periods of time and seasons were considered and enabled to draw different conclusions, that are presented in the following. Nonetheless, for each evaluated period, only stations containing more than 70% of data were kept. This avoided biased results from stations measuring for only few dates over of the period. In the remainder of this chapter, Netatmo and ARPA stations are combined and conflated as "stations".

4.4.1   Period 2018-2021, all seasons

First of all, all four years and all seasons were considered together. Since the number of Netatmo stations has increased from year to year, considering the average temperature over the four years for all stations would have led to non-comparable values. Indeed, the recent stations would have included only a few months while the older stations would have a value covering 4 full years. Therefore, the yearly average temperature recorded at each station from 2018 to 2021 has been computed, enabling to get for each station at maximum four AT values and one LCZ class. The number of values per class is reported in Tab. 29. Then, the statistics per class could be computed and are summarized by the box-plots in Fig. 38.

The natural classes turned out to have the lowest AT means with $14.9\,°C$ and $15.0\,°C$ for Scattered Trees and Low plants, respectively. However, very few stations are located in these areas in comparison to the artificial ones, preventing any conclusion from being drawn on the difference between natural and built-up classes.

Class	2	3	5	6	8	102	104	
Name	Compact Mid-Rise	Compact Low-Rise	Open Mid-Rise	Open Low-Rise	Large Low-Rise	Scattered Trees	Low Plants	Total
Nb. values	42	80	101	290	56	24	16	609

Table 29: Number of values per class (2018-2021).
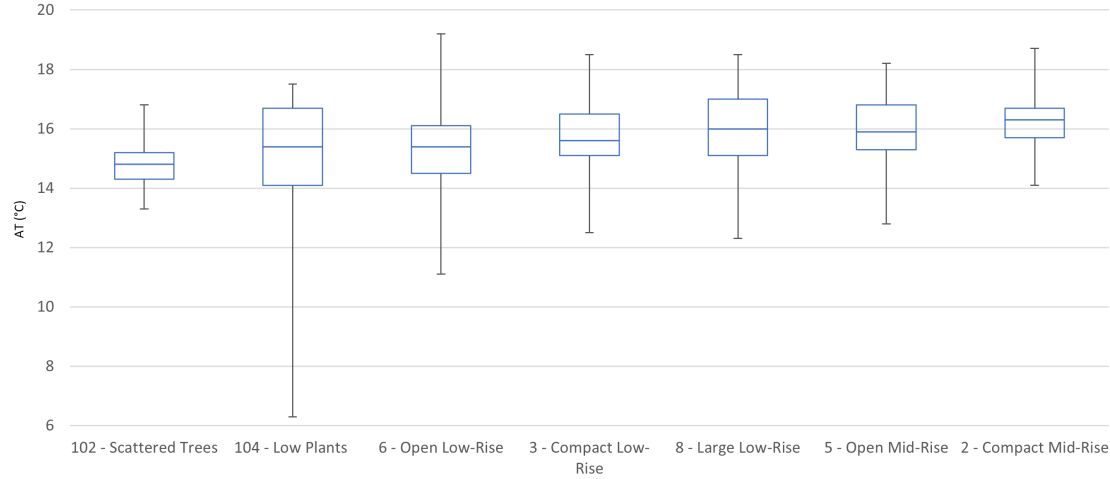
Figure 38: Box-plots of AT values per class (2018-2021)

Secondly, the Open low-rise class was found very close to the natural classes with a mean of 15.3 °C and proved to be the coolest artificial area. This zone is characterized by the presence of vegetation, a low density and small buildings.

Then, the Compact low-rise, Large low-rise and Open mid-rise classes revealed a very similar average temperature with 15.7 °C, 15.9 °C and 16.0 °C, respectively. While the Open mid-rise zones hold some vegetation and buildings of medium height, the Compact low-rise areas encompass the opposite lands with no vegetation and small buildings. Large low-rise class is also composed of soils without vegetation and rather low buildings, but corresponds more to industrial and storage areas.

Finally, the Compact mid-rise zones experienced the highest air temperatures with an average of 16.3 °C, i.e. 1.0 °C more than the lowest mean in artificial areas (Open low-rise). It is characterized by the absence of vegetation, high density and buildings of medium to high height.

As a result, it appeared that, while high (or medium) buildings are contributing to increase the air temperature, vegetation is participating to its decreasing. However, as vegetation is subject to seasonality, a high variability is observed in classes containing green grounds like class 104 or class 6. Therefore, a further analysis season per season was carried out.

4.4.2   Study of seasonality, period 2018-2021

For each year from 2018 to 2021, the average temperature recorded by each station in each season was computed. The number of values considered per class and per season is reported in the Table 30 and considers, here again, a maximum of four values per station.

Class	2	3	5	6	8	102	104	
Name	Compact Mid-Rise	Compact Low-Rise	Open Mid-Rise	Open Low-Rise	Large Low-Rise	Scattered Trees	Low Plants	Total
Summer	68	89	136	305	61	23	17	699
Autumn	58	96	123	326	62	24	19	708
Spring	47	77	124	300	60	24	16	648
Winter	55	80	119	285	71	24	16	650

Table 30: Number of values per class and per season (2018-2021).

4.4.2.1   Autumn

The Fig. 39 summarizes the air temperature statistics by class for the autumn season. Similar observations as for the hole years can be done, with specifically a difference between Open low-rise (the coldest built-up) and Compact mid-rise (the hottest built-up) of 1.3 °C.
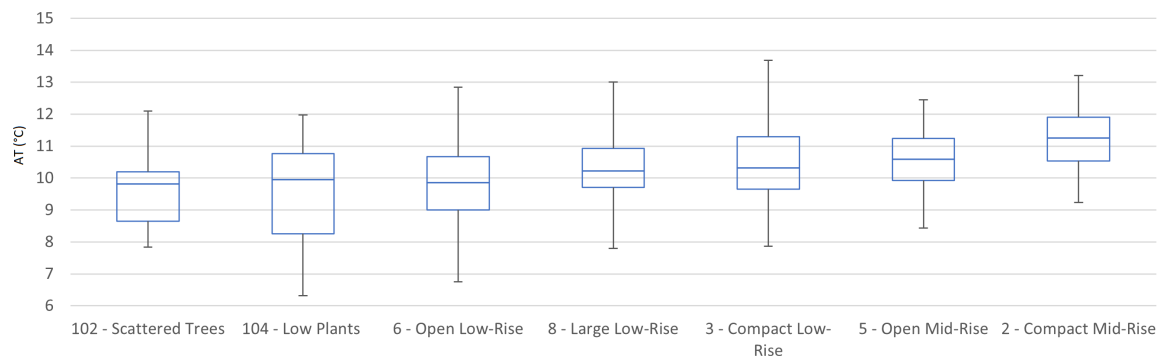
Figure 39: Box-plots of AT values per class in autumn (2018-2021)

4.4.2.2   Winter

Then, the Fig. 40 depicts the box-plots of AT per class for the winter season. While the Scattered trees class still holds the lowest mean, the Low plants areas revealed higher AT than Open low-rise ones. This might be due to the fact that in winter Low plants has almost no vegetation. Moreover, the difference in average between Open low-rise (the coldest) and Compact mid-rise (the hottest) is of 1.6 °C.

4.4.2.3   Spring

The analysis of AT for spring season led to similar observations as for winter and is illustrated in Fig. 41. Indeed, the Low plants zones also experienced higher temperatures than Open low-rise ones, maybe because the vegetation was growing.
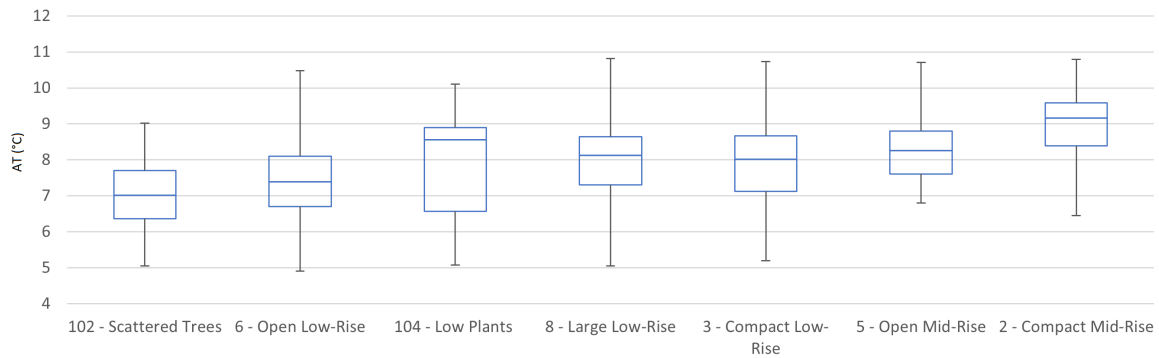
Figure 40: Box-plots of AT values per class in winter (2018-2021)

Moreover, the Low Plants class holds a large variability with a standard deviation of $1.6\,°C$, that might be due to the changes in vegetation occurring during spring season. The difference between Open low-rise (the coldest) and Compact mid-rise (the hottest) is of $0.8\,°C$ which is a bit lower than in other seasons.
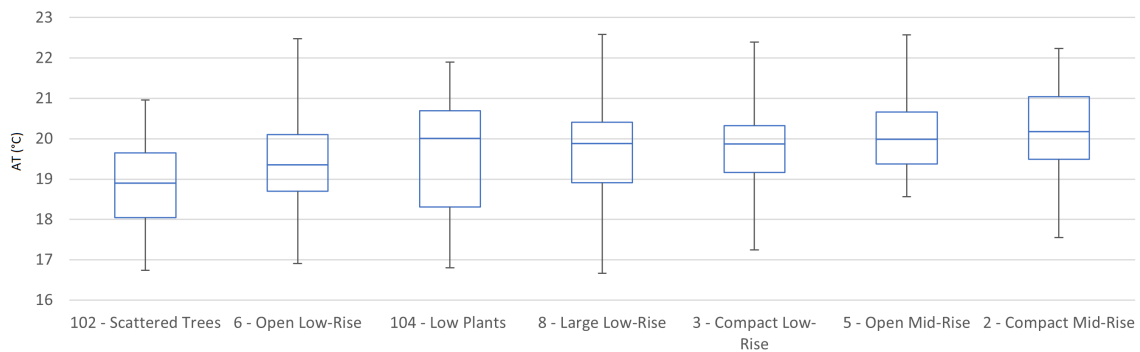
Figure 41: Box-plots of AT values per class in spring (2018-2021)

4.4.2.4   Summer

Finally, the Fig. 42 reports the box-plots for the summer season. The Scattered trees class is still the lowest with a mean temperature of $24.2\,°C$. However, unlike what would be expected for the summer season that has a flurishing vegetation, Low plants is not the second coldest class. Indeed, Open low-rise is the second with $24.5\,°C$, preceding Compact low-rise, Low plants, Large low-rise and Open mid-rise that all have almost the same mean ($25.0\,°C$, $25.0\,°C$, $25.1\,°C$ and $25.2\,°C$). The highest mean AT is still for Compact mid-mise areas with $25.6\,°C$, that is $1.1\,°C$ warmer than Open low-rise zones.

Hence, whatever the season considered, the lowest temperatures was experienced in the Scattered trees areas and the highest ones in the Compact mid-rise zones, which is consistent with the results previously obtained over the period 2018-2021. In addition, the difference between the coolest artificial class (Open low-rise) and the warmest one (Compact low-rise) was found to vary among the seasons but appeared higher than $1.1\,°C$ during winter, autumn and summer.

Furthermore, seasonality seems to have an impact on the air temperature of the
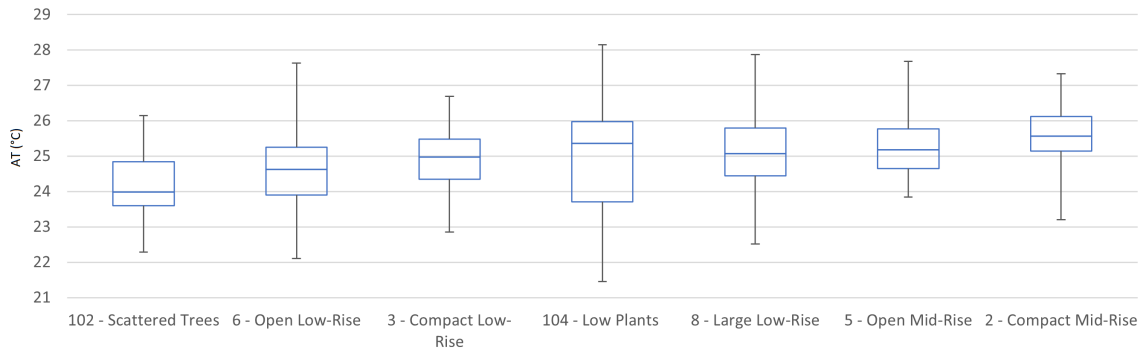
Figure 42: Box-plots of AT values per class in summer (2018-2021)

Low plant class since during winter and spring Low plants is among the classes with middle temperatures while in autumn and summer it is among the coolest. However, this result should be qualified by the fact that very few stations are present in the natural classes.

Nonetheless, as the behaviour of the Low plants class appeared to be surprising during summer, an additional study was carried out, by separating the summer values year by year. But, since the number of station is decreasing when going back in time, only the year 2021 was studied, enabling at the same time to focus on the same period as the one considered for LST.

4.4.3 Summer 2021

For each in situ station, the mean temperature over the summer 2021 was computed. The number of values considered per LCZ class is reported in the Table 31. It can be noted that less values were available as only one mean per station was calculated (only one year).

Class	2	3	5	6	8	102	104	
Name	Compact Mid-Rise	Compact Low-Rise	Open Mid-Rise	Open Low-Rise	Large Low-Rise	Scattered Trees	Low Plants	Total
Nb values	24	32	40	97	21	5	6	225

Table 31: Number of values per class (summer 2021).

As shown by the Fig. 43, the analysis of AT per LCZ highlighted again Scattered trees and Compact mid-rise as the classes experiencing the lowest and highest mean temperatures, respectively. Then, Low plants and Open low-rise have the same mean and median (24.2 °C) and are far bellow the Large low-rise and Compact low-rise zones (24.7 °C).

Therefore, for the summer 2021, the Low plants zones didn't demonstrate the surprising behaviour notified when analysing the four summers all together. Indeed, it stands among the classes with the lowest mean temperatures, together with the Open low-rise areas that also contain vegetal cover. This might be due to the fact that very few data are available for the Low plants class, making every measure hav-

ing a strong impact on the resulting mean. If at the location of one of the stations in a low plant zone, during one of the summer a change happened (no culture in the agricultural land, change of vegetation, de-calibrated sensor...), the result would be strongly impacted. This is why, in general, the conclusions drawn with regard to the AT in natural areas must be qualified and should be the subject of a more in-depth study containing more data.
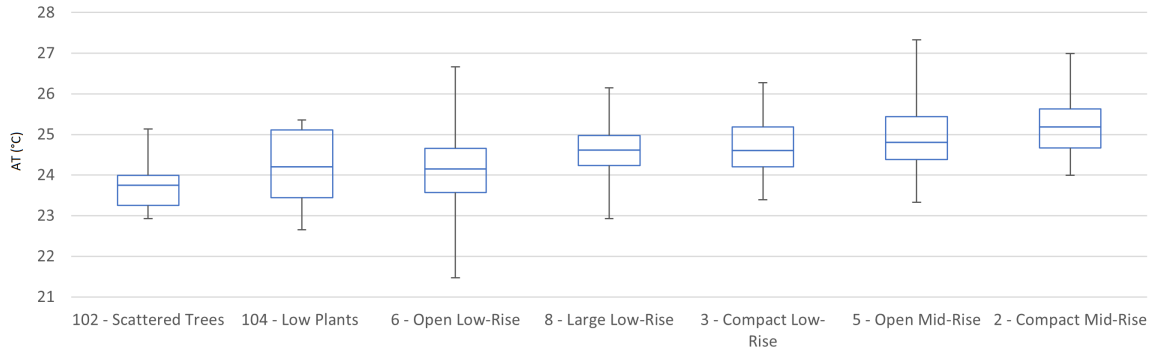
Figure 43: Box-plots of AT values per class in summer 2021

Nevertheless, the previous observations over different periods and seasons led to the following inferences:

1. While the lowest mean temperatures were recorded in the Scattered trees zones, the highest ones were in the Compact mid-rise areas, whatever the period or season considered.

2. When looking only at artificial classes, the coolest and warmest classes were, in all cases, the Open low-rise and Compact mid-rise, respectively. In average, a difference of $1.0\,°\mathrm{C}$ is observed between them and this difference is even higher than $1.3\,°\mathrm{C}$ when separating seasons (for winter and autumn).

3. When investigating the characteristics of classes with higher/lower temperatures, it appeared that high buildings are contributing to increase the AT, which is the opposite of LST, and that vegetation tends to decrease it.

4. Seasonality may have an impact on the AT of the Low plants class that shows higher temperatures during winter and spring than in autumn and summer with respect to the other classes.

In the following, these findings have been examined in the light of statistical and practical significance. In particular, the analysis focused on the difference observed between artificial classes.

4.4.4   Statistical and practical significance

In order to statistically evaluate the observations previously made, the first two cases were considered: (1) all years and seasons combined and (2) season by season.

4.4.4.1   Period 2018-2021, all seasons

The first objective was to determine whether there was a significant difference of AT between the artificial classes, considering the full period 2018-2021. Natural classes were discarded due to the lack of data.

ANOVA test

As explained in the section 3.4.4, the most commonly used statistical test to compare means of three or more groups is the one-way ANOVA, operating under three assumptions: independence, normality and homogeneity of variances. While the first was admitted, the second was checked with the help of density histograms. Depicted in Fig. 44, these histograms revealed a distribution that can be approximated as normal for classes 3, 5 and 6. However, for classes 2 and 8, it was necessary to carry out an additional test, the D'Agostino's K^2 test. Named from Ralph D'Agostino, it calculates summary statistics from the data, namely kurtosis and skewness, to determine if the data distribution departs from the normal one. For all the artificial classes, the resulting p-value was higher than 0.05, meaning that their distribution could be assumed to be normal.

(a) Class 2

(b) Class 3

(c) Class 5

(d) Class 6

(e) Class 8
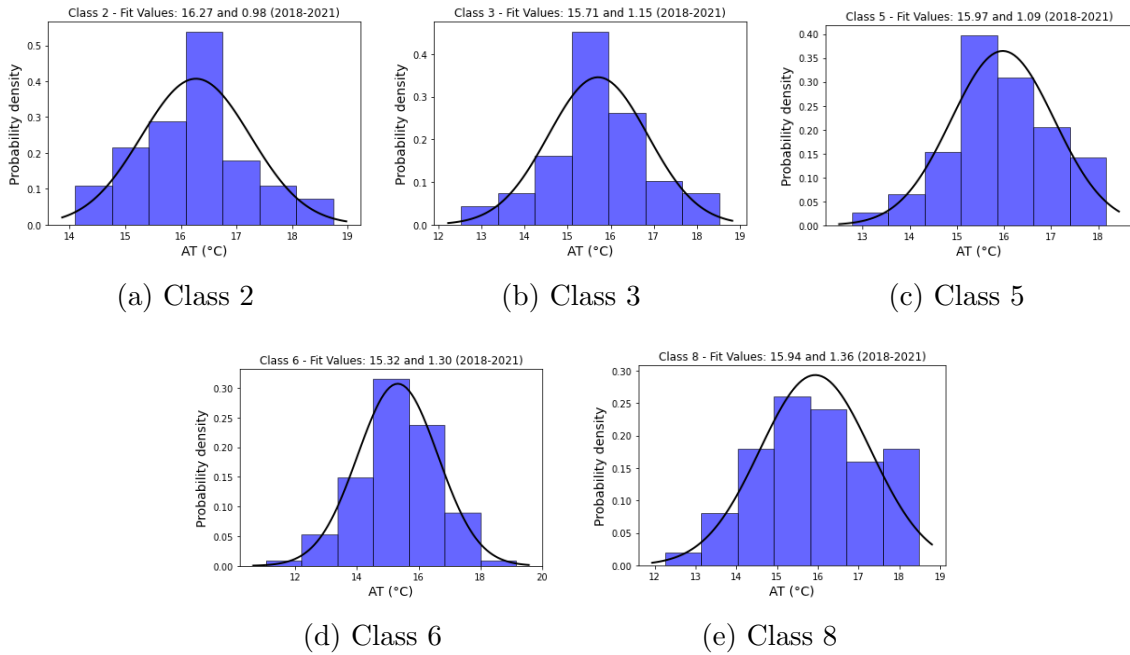
Figure 44: Density histograms of AT for artificial classes (2018-2021)

Then, since the variance of AT in artificial classes ranges from 0.8 to 1.9, the homogeneity of variance could be assumed, and was confirmed through a Levene's test. Hence, the ANOVA test could be performed on classes 2, 3, 5, 6 and 8.

The one-way ANOVA has one categorical independent variable (here, the LCZ classes) and a normally distributed continuous dependent variable (here, the AT). The goal of the ANOVA test is to check for variability within the groups as well as the variability among the groups and its test statistic is given by the F-test. The null hypothesis tested is $H_0 : \mu_1 = \mu_2 = \ldots = \mu$, i.e. all AT means are equal among

the classes. As there are several components to the ANOVA formula, they have been organized into a table (see Table 32).

Source Of Variation	Sum Of Squares	Degrees Of Freedom	Mean Squares	F Value
Between Groups	$SSB = \sum n_j(\bar{X}_j - \bar{X})^2$	$df_1 = k-1$	$MSB = \frac{SSB}{k-1}$	$f = \frac{MSB}{MSE}$
Error	$SSE = \sum\sum(X - \bar{X}_j)^2$	$df_2 = N-k$	$MSE = \frac{SSE}{N-k}$	
Total	$SST = SSB + SSE$	$df_3 = N-1$		

SSB = sum of squares between groups ; SSE = sum of squares of errors; $\bar{X}_j - \bar{X}$ = mean of the jth group; $X - \bar{X}_j$ = overall mean; n_j = sample size of the jth group; X = each data point in the jth group (individual observation); N = total number of observations; k = number of groups; SST = Total sum of squares.

Table 32: ANOVA table

When applying the ANOVA formula on the built-up classes, the resulting F-value is 10.17 which corresponds to a p-value of $5.83 * 10^{-8}$. With a significance level of 5%, the null hypothesis was therefore rejected, which confirmed that not all group means are equal. However, the one-way ANOVA cannot identify the specific groups that are statistically different from one another. Thus, an additional test had to be conducted.

T-test

Running on the same assumptions as ANOVA, the T-test could be performed on pairs of classes in order to identify the groups significantly different from others. This test estimates the true difference between two group means using the ratio of the difference in means over the pooled standard error of both groups. The formula is given by equation 4.5 and the tested null hypothesis is $H_0 : \mu_1 = \mu_2 = \mu$, i.e. the means of the two populations are equal.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(s^2(\frac{1}{n_1} + \frac{1}{n_2})\right)}} \tag{4.5}$$

where:

- t is the t-value,

- $\bar{X}_1$ and $\bar{X}_2$ are the means of the two groups being compared,

- s^2 is the pooled standard error of the two groups, computed with equ. 4.6,

- n_1 and n_2 are the number of observations in each of the groups.

$$s^2 = \frac{\sum(X - \bar{X}_1)^2 + \sum(X - \bar{X}_2)^2}{n_1 + n_2 - 2} \tag{4.6}$$

A larger t-value indicates a more significant difference between the groups and is associated to a smaller p-value. Thus, the t-value was computed for all pairs of classes among the artificial classes and the corresponding p-value was reported in the table 33. In addition, in order to check also the practical significance of the

Classes	2 and 3	2 and 5	2 and 6	2 and 8	3 and 5	3 and 6	3 and 8	5 and 6	5 and 8	6 and 8
Mean diff.	0.56	0.30	0.95	0.33	0.27	0.38	0.23	0.65	0.03	0.62
p-value	0.01	0.13	0.00	0.18	0.12	0.02	0.29	0.00	0.17	0.00
H_0	NO	OK	NO	OK	OK	NO	OK	NO	OK	NO
Cohen's d	0.51	0.28	0.75	0.27	0.24	0.30	0.18	0.52	0.03	0.47
Size effect	Medium	Small	Medium	Small	Small	Small	Very small / Small	Medium	Very small	Small / Medium

Table 33: T-test and Cohen's d for pairs of classes (2018-2021)

differences, the Cohen's d value was calculated, enabling to estimate the magnitude of the effect.

As a result, the hypothesis of equal mean was rejected in five cases (see the 'NO' in table 33). The previously observed difference between classes 2 (Compact mid-rise) and 6 (Open low-rise) was found to be statistically and practically significant with a medium effect size and a mean difference of 0.96 °C. In the same way, Open mid-rise (class 5) and Large low-rise (class 8) zones have both also proved to be statistically and practically different from class 6, with a mean difference of 0.65 °C and 0.61 °C, respectively. The effect size for class 8 and 6 is only small but is very close to the limit of the medium effect ($d \geq 0.50$). These three points validated the previously stated deductions, i.e. tall buildings increase the AT and vegetation tends to decrease it. In addition, the hypothesis of equal mean was rejected as well for classes 3 (Compact low-rise) and 2 (Compact mid-rise), which differ only in building height, confirming once again the previous conclusions.

Furthermore, classes 3 and 6 also turned out to be statistically different, which could introduce a new consideration: the density of buildings. Indeed, even if Open low-rise zones introduce some vegetation, it differs from Compact low-rise areas mainly by the density of their buildings. The rejection of the null hypothesis suggests therefore that the mean difference 0.38 °C is statistically significant and that air temperature increases with building density. The effect size, however, is only small, showing that density has less impact than the previous two highlighted factors.

On the other hand, no difference could be clearly stated from the statistical point of view between classes 2 and 5, and 3 and 5. Indeed, Open mid-rise zones (class 5) contain both factors of warmth (medium size buildings) and coolness (vegetation and low density) that compensate each other and give to class 5 an intermediate mean temperature. Similarly, class 8 (Large low-rise) encompasses areas with low buildings but no vegetation and is also in an intermediate position with no significant difference from classes 2, 3 and 5.

4.4.4.2 Period 2018-2021, with seasonality

The same statistical and practical tests were performed on the AT of artificial classes considering this time seasonality. The D'Agostino's K^2 and Levene's tests validated the assumptions of normality and equal variance for each season over 2018-2021.

Thus, the t-tests on pairs of classes together with the Cohen's d coefficients were computed and are summarized in the tables 34, 35, 36 and 37.

Classes	2 and 3	2 and 5	2 and 6	2 and 8	3 and 5	3 and 6	3 and 8	5 and 6	5 and 8	6 and 8
Mean diff.	1.01	0.80	1.62	1.02	0.21	0.61	0.01	0.83	0.22	0.60
p-value	0.00	0.00	0.00	0.00	0.15	0.00	0.97	0.00	0.14	0.00
H_0	NO	NO	NO	NO	OK	NO	OK	NO	OK	NO
Cohen's d	0.96	0.85	1.55	1.02	0.21	0.56	0.01	0.80	0.23	0.57
Size effect	Large	Large	Very large	Large	Small	Medium	Negligible	Large	Small	Medium

Table 34: T-test and Cohen's d for pairs of classes - Winter (2018-2021)

Classes	2 and 3	2 and 5	2 and 6	2 and 8	3 and 5	3 and 6	3 and 8	5 and 6	5 and 8	6 and 8
Mean diff.	0.48	0.19	0.83	0.55	0.29	0.35	0.07	0.63	0.36	0.28
p-value	0.01	0.23	0.00	0.01	0.04	0.01	0.69	0.00	0.05	0.06
H_0	NO	OK	NO	NO	NO	NO	OK	NO	OK	OK
Cohen's d	0.47	0.21	0.78	0.54	0.30	0.33	0.07	0.62	0.37	0.26
Size effect	Small / Medium	Small	Medium	Medium	Small	Small	Very small	Medium	Small	Small

Table 35: T-test and Cohen's d for pairs of classes - Spring (2018-2021)

Classes	2 and 3	2 and 5	2 and 6	2 and 8	3 and 5	3 and 6	3 and 8	5 and 6	5 and 8	6 and 8
Mean diff.	0.68	0.41	1.03	0.59	0.28	0.35	0.09	0.63	0.18	0.44
p-value	0.00	0.00	0.00	0.00	0.02	0.00	0.53	0.00	0.18	0.00
H_0	NO	NO	NO	NO	NO	NO	OK	NO	OK	NO
Cohen's d	0.79	0.47	1.09	0.66	0.32	0.37	0.10	0.67	0.21	0.46
Size effect	Medium / Large	Small	Large	Medium	Small	Small	Very small	Medium	Small	Small

Table 36: T-test and Cohen's d for pairs of classes - Summer (2018-2021)

Classes	2 and 3	2 and 5	2 and 6	2 and 8	3 and 5	3 and 6	3 and 8	5 and 6	5 and 8	6 and 8
Mean diff.	0.80	0.61	1.29	0.91	0.19	0.48	0.11	0.68	0.30	0.38
p-value	0.00	0.00	0.00	0.00	0.18	0.00	0.56	0.00	0.05	0.02
H_0	NO	NO	NO	NO	OK	NO	OK	NO	OK	NO
Cohen's d	0.74	0.65	1.17	0.93	0.18	0.42	0.09	0.62	0.30	0.34
Size effect	Medium	Medium	Large	Large	Very small	Small	Very small	Medium	Small	Small

Table 37: T-test and Cohen's d for pairs of classes - Autumn (2018-2021)

When considering seasonality, more differences were reported as statistically significant, as shown by the red cells highlighting the pairs for which a different conclusion from the previous analysis was found.

For all seasons, the AT difference between classes 2 (Compact mid-rise) and 8 (Large low-rise) became statistically significant, with a size effect from Medium to Large. This result once again reinforced the idea of a warming effect of building height, not only yearly but also during each season.

Similarly, for all seasons except the spring, the difference between classes 2 (compact mid-rise) and 5 (open mid-rise) became statistically significant. However, the size effect turned out to strongly vary between the seasons with a small effect in summer, a medium one in autumn and a large effect in winter. Therefore, density, which was previously considered as a weak factor of coolness, turned out to have a greater impact during cooler periods.

In addition, H_0 was rejected for the pair of classes 3 (compact low-rise) and 5 (open mid-rise) in spring and summer, emphasising the significance of the higher AT recorded in the compact low-rise areas. This result appeared to be surprising since more vegetation is present in class 5 than class 3. But, it revealed that the building height factor is more impacting than density or vegetation during summer and spring (i.e. when temperatures are higher) inside artificial areas. The same conclusion can be drawn from the fact that during spring, difference between 6 (open low-rise) and 8 (large low rise) is no longer statistically significant.

Finally, in section 4.4.2, an accentuation of the temperature difference between zones with the lowest and highest AT (Open low-rise and Compact mid-rise) was observed when considering seasonality. It has been confirmed there since a larger practical significance during summer, autumn and winter was reported.

4.4.5 Conclusions

In summary, the analysis of AT variation between the artificial LCZs highlighted two main factors of freshness: low-rise buildings and the presence of vegetation. Indeed, in all considered cases, the coolest and warmest classes were respectively the Open low-rise and Compact mid-rise. Their mean difference of $1.0\,°C$ was proven to be statistically and practically significant and to increase when separating records season per season, with more than $1.3\,°C$ in winter and autumn.

In addition, the statistical analysis brought to light density as complementary parameter impacting the AT. In facts, for all pairs of classes holding the same characteristics but the density of building, the higher mean AT was found in the denser class. This factor seemed to have a smaller impact than the two others since only small size effects were assigned to it when considering all seasons together, but turned out to have greater impact during cooler periods (autumn and winter).

In comparison to the conclusions made for LST, the vegetation as mitigating factor was retrieved. However, while the height of buildings is reducing the surface temperature thanks to shadows, it appeared to increase the mean air temperature, probably because buildings retain the heat and prevent air from circulating. Also, the building density that did not prove to have any significant impact on the LST, was added as a small air heat factor.

Due to the lack of data in the natural areas, even if a lower average AT was recorded by the few stations inside them, no conclusion could be clearly stated. A further work containing a larger number of in situ measures inside natural zones could be carried out to establish and quantify the air temperature decrease induced by the natural areas.

5 Conclusion and recommendations for further work

This thesis aimed to study the impact of urban morphology and land covers on the temperatures experienced in the Metropolitan City of Milan. Thus, the spatial variations of both land surface temperature and air temperature in the LCZs were investigated through a study divided into three chapters.

Firstly, the LCZ map of the MCM was derived from Landsat-8 and Sentinel-2 satellite imagery using the Random Forest classification. Offering a spatial resolution of 30m, the Landsat-8 LCZ map was found to be better than the Sentinel-2 one for all considered accuracies (OA, UA and PA). However, as the classification tended to confuse the compact low-rise, open mid-rise and open low-rise classes with a PA of less than 62% for the first two, an additional dataset containing the building height was added and proved to be a good solution to the problem. The final LCZ map holds a very high OA of 94%.

In the second phase, the LST was investigated. From the Landsat-8 Collection 2 dataset, both the Level 1 and Level 2 images were exploited to retrieve the LST and were compared with a reference, the Sentinel-3 SLSTR product. As a result, significant differences were found between the three LST maps of each considered days. Indeed, while Level 2 maps showed higher temperatures than the reference with $+5.2\,°C$ and $+7.8\,°C$ degrees on average, Level 1 maps, on the contrary, illustrated lower temperatures with $-8.1\,°C$ and $-6.4\,°C$ on average. However, in the continuity of the work of García [2021], who also reported higher mean values for Landsat Level 2 than Sentinel 3 SLSTR, the Level 2 results were stated to be more accuracte than Level 1's. Nevertheless, the two Landsat levels were found to have a strong linear correlation with a Pearson coefficient of 0.99. This relationship has not been further investigated and may be the subject of future research to understand the origin and nature of the observed difference.

Then, the Level 2 product of four dates in the summers 2021 and 2022 were leveraged to study the variation of LST between the LCZ classes. First, the natural areas were proved to mitigate high temperature more than artificial zones, with a difference in mean LST statistically and practically significant of $6.7\,°C$. Secondly, smaller differences were found among the built-up classes with a maximum difference of $2.6\,°C$ between Open mid-rise and Large low-rise zones. However, two main factors participating in the mitigation of high temperatures emerged: the height of buildings and the presence of vegetation. Indeed, the higher the buildings and the greener the zones, the lower the reported LST was. In addition, the investigation of the LST on the location of a new motorway built in 2012 highlighted an average increase of $3.1\,°C$ between the 2007-2009 and 2020-2022 periods. This finding came to confirm the previous conclusions as motorways are generally lines of impermeable concrete, devoid of vegetation and shade.

The third part moved to a more perceptible quantity: the air temperature. Since official weather monitoring stations are few and far between, the dense network of amateur Netatmo stations was exploited. Being crowd-sourced data, the Netatmo

measures required a cleaning that was carried out using the regional ARPA stations as a reference. This pre-processing removed around 20% of the initial dataset.

Then, considering 36 dates from 2018 to 2021, a correlation ranging from 0.85 to 0.99 between day-time AT and LST was reported for both Netatmo and ARPA stations. This stands among the highest correlations found in the literature, together with the results of Colombi et al. [2007] who reported coefficients higher than 0.9. However, since the relation between day-time AT and LST is strongly impacted by several factors such as surface cover, wind speed, water vapour or seasonality, the transformation from one to the other is not straightforward and should be analysed in more detail in further studies, as initiated by Iqbal and Ali [2022] who suggest that research be directed towards a linear relationship whose coefficients depend on the season.

Finally, in-situ measures from 2018 to 2021 were exploited to study the variation of AT between the LCZ classes. When focusing on artificial zones, the maximum difference in mean AT was found between the Open low-rise and Compact mid-rise classes, with the latter $1.0\,°C$ warmer than the first. This difference appeared even larger when considering seasonality, with more than $1.3\,°C$ in autumn and winter. Moreover, by analysing the characteristics of warmer/cooler zones, two main factors of air freshness were identified, namely the presence of vegetation (as for LST) and low-rise buildings (as opposed to LST). The low density of buildings has also been noticed as a third element of less significance and was proved to have greater impact during cooler periods (autumn and winter). For its part, the study of AT variations in natural areas has been limited by the insufficient number of stations located there and would merit further research with more data.

As future work, several points could be considered:

- A major issue associated with the usage of LST products was the cloud coverage that reduced significantly the number of exploitable images. One could therefore try to solve this issue by considering a more refined way of dealing with clouds.

- The lack of ground truth for the LST has prevented the validation of any LST map derived in this study. A further analysis considering on-site measures of soil temperature would enable to understand the differences observed, but as such in-situ data are quite rare, specific measurements would have to be conducted by researchers and make this validation more complicated and costly.

- Taking into account the strong linear correlation established in this study, the relationship between LST and air temperature could be further investigated by considering other factors such as wind speed, water vapor or soil moisture.

- The crowd-sourced data used in this study revealed a great potential for the use of citizen data. In particular, only temperature data has been exploited, but Netatmo stations measure more parameters such as humidity or air quality that could also be explored. However, two main drawbacks of Netatmo stations have been identified. Firstly, they are not (or very rarely) located in natural areas, and are therefore more suitable for urban data analysis. One

should consider additional datasets provided for example by associations managing weather stations in extra-urban areas. Secondly, the Netatmo network is quite recent and enabled at maximum to carry out analyses from 2018. Older networks could be considered, but crowd-sourcing being a recent phenomenon, it is not easy to find many old data. One could exploit the urban-scale temperature data provided by the Copernicus Climate Change Service (Hooyberghs et al. [2019]) at a spatial resolution of 100 meters, that cover the years 2008 to 2017. This dataset has great potential but only encompasses European cities and no extra-urban areas.

Bibliography

Abilash, R. (2018). Applying random forest (classification) - machine learning algorithm from scratch with real datasets.

Almeida, C. R. d., Teodoro, A. C., and Gonçalves, A. (2021). Study of the urban heat island (uhi) using remote sensing data/techniques: A systematic review. *Environments*, 8(10).

AM, M. (2022). Tabelle climatiche 1971-2000 della stazione meteorologica di milano linate dall'atlante climatico 1971-2000 del servizio meteorologico dell'aeronautica militare. `http://www.meteoam.it/`.

Arjasakusuma, S., Kusuma, S. S., Vetrita, Y., Prasasti, I., and Arief, R. (2022). Monthly burned-area mapping using multi-sensor integration of sentinel-1 and sentinel-2 and machine learning: Case study of 2019's fire events in south sumatra province, indonesia. *Remote Sensing Applications: Society and Environment*, 27:100790.

Artis, D. A. and Carnahan, W. H. (1982). Survey of emissivity variability in thermography of urban areas. *Remote Sensing of Environment*, 12(4):313–329.

Aslam, A. and Rana, I. A. (2022). The use of local climate zones in the urban environment: A systematic review of data sources, methods, and themes. *Urban Climate*, 42:101120.

Bacci, P. and Maugeri, M. (1992). The urban heat island of milan. *Il Nuovo Cimento C*, 15:417–424.

Bechtel, B., Alexander, P. J., Böhner, J., Ching, J., Conrad, O., Feddema, J., Mills, G., See, L., and Stewart, I. (2015). Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS International Journal of Geo-Information*, 4(1):199–219.

Belgiu, M. and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31.

Büchau, Y. (2017). Patatmo, easy access to the netatmo weather api. `https://nobodyinperson.gitlab.io/python3-patatmo/`.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Colombi, A., Michele, C. D., Pepe, M., and Rampini, A. (2007). Estimation of daily mean air temperature from modis lst in alpine areas. *New Developments and Challenges in Remote Sensing*, 6.

Cook, M., Schott, J. R., Mandel, J., and Raqueno, N. (2014). Development of an operational calibration methodology for the landsat thermal data archive and initial testing of the atmospheric compensation component of a land surface temperature (lst) product from the archive. *Remote Sensing*, 6(11):11244–11266.

Davis, J. (2020). *Writing for Statistics and Data Science.* Distance Learning Edition.

Delacre M, Leys C, M. Y. L. D. (2019). Taking parametric assumptions seriously: Arguments for the use of welch's f-test instead of the classical f-test in one-way anova. *International Review of Social Psychology.*

Demuzere, M., Kittner, J., and Bechtel, B. (2021). Lcz generator: A web application to create local climate zone maps. *Frontiers in Environmental Science*, 9.

do Nascimento, A. C. L., Galvani, E., Gobo, J. P. A., and Wollmann, C. A. (2022). Comparison between air temperature and land surface temperature for the city of são paulo, brazil. *Atmosphere*, 13(3).

Du, P., Chen, J., Bai, X., and Han, W. (2020). Understanding the seasonal variations of land surface temperature in nanjing urban area based on local climate zone. *Urban Climate*, 33:100657.

Duan, S.-B., Li, Z.-L., Wang, C., Zhang, S., Tang, B.-H., Leng, P., and Gao, M.-F. (2019). Land-surface temperature retrieval from landsat 8 single-channel thermal infrared data in combination with ncep reanalysis data and aster ged product. *International Journal of Remote Sensing*, 40(5-6):1763–1778.

EEA (2021). Urban sustainability. `https://www.eea.europa.eu/themes/sustainability-transitions/urban-environment`.

ESA (2021). Sentinel-3 slstr user guide. `https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr`.

Eurostat (2022). Population on 1 january by age groups and sex - cities and greater cities.

Fung, K. Y., Yang, Z.-L., and Niyogi, D. (2022). Improving the local climate zone classification with building height, imperviousness, and machine learning for urban models.

García, D. H. (2021). Analysis and precision of the terrestrial surface temperature using landsat 8 and sentinel 3 images: Study applied to the city of granada (spain). *Sustainable Cities and Society*, 71:102980.

Hamid Reza Pourghasemi, C. G. (2019). *Spatial Modeling in GIS and R for Earth and Environmental Sciences.*

Han, B., Luo, Z., Liu, Y., Zhang, T., and Yang, L. (2022). Using local climate zones to investigate spatio-temporal evolution of thermal environment at the urban regional level: A case study in xi'an, china. *Sustainable Cities and Society*, 76:103495.

Hook, S. J., Gabell, A., Green, A., and Kealy, P. (1992). A comparison of techniques for extracting emissivity information from thermal infrared data for geologic studies. *Remote Sensing of Environment*, 42(2):123–135.

Hooyberghs, H., Berckmans, J., Lefebre, F., and Ridder, K. D. (2019). Climate variables for cities from 2008 to 2017.

Horning, N. (2010). Random forests: An algorithm for image classification and generation of continuous fields data sets.

IPCC (2014). *Summary for Policymakers*, page 1–30. Cambridge University Press.

Iqbal, B. and Ali, M. (2022). Estimation of spatio-temporal air temperature from satellite based lst under semi-arid to arid environment in peshawar basin, northwest pakistan. *Advances in Space Research*, 70(4):961–975.

Jin, M. and Dickinson, R. E. (2010). Land surface skin temperature climatology: benefitting from the strengths of satellite observations. *Environmental Research Letters*, 5(4):044004.

Khoshnoodmotlagh, S., Daneshi, A., Gharari, S., Verrelst, J., Mirzaei, M., and Omrani, H. (2021). Urban morphology detection and it's linking with land surface temperature: A case study for tehran metropolis, iran. *Sustainable Cities and Society*, 74:103228.

Laraby, K. G. and Schott, J. R. (2018). Uncertainty estimation method and landsat 7 global validation for the landsat surface temperature product. *Remote Sensing of Environment*.

Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10. Special Issue: Progress of Machine Learning in Geosciences.

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I. F., and Sobrino, J. A. (2013). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, 131:14–37.

Lin, M., Lucas, H., and Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research*, 24:906–917.

Loftian, M. (2016). Urban climate modeling, case study of milan city.

Maffei, C., Alfieri, S. M., and Menenti, M. (2018). Relating spatiotemporal patterns of forest fires burned area and duration to diurnal land surface temperature anomalies. *Remote Sensing*, 10(11).

Mahato, S. and Pal, S. (2018). Changing land surface temperature of a rural rarh tract river basin of india. *Remote Sensing Applications: Society and Environment*, 10:209–223.

McCarthy, M., Best, M., and Betts, R. (2010). Climate change in cities due to global warming and urban effects. *Geophys. Res. Lett*, 37.

Meng, X., Cheng, J., and Liang, S. (2017). Estimating land surface temperature from feng yun-3c/mersi data using a new land surface emissivity scheme. *Remote Sensing*, 9(12).

Mujabar, S. and Rao, V. (2018). Estimation and analysis of land surface temperature of jubail industrial city, saudi arabia, by using remote sensing and gis technologies. *Arabian Journal of Geosciences*, 11:1–13.

NASA (2020).   Training arset - satellite remote sensing for urban
    heat islands. `https://appliedsciences.nasa.gov/join-mission/training/`
    `english/arset-satellite-remote-sensing-urban-heat-islands`.

NASA (2021). Giss key indicators - global temperature. `https://climate.nasa.`
    `gov/vital-signs/global-temperature/`.

Netatmo (2021). Same home, just smarter. `https://www.netatmo.com/en-us`.

Niemeyer, J., Rottensteiner, F., and Soergel, U. (2014). Contextual classification
    of lidar data and building object detection in urban areas. *ISPRS Journal of
    Photogrammetry and Remote Sensing*, 87:152–165.

Oxoli, D., Ronchetti, G., Minghini, M., Molinari, M. E., Lotfian, M., Sona, G., and
    Brovelli, M. A. (2018). Measuring urban land cover influence on air temperature
    through multiple geo-data—the case of milan, italy. *ISPRS International Journal
    of Geo-Information*, 7(11).

Pichierri, M., Bonafoni, S., and Biondi, R. (2012). Satellite air temperature esti-
    mation for monitoring the canopy layer heat island of milan. *Remote Sensing of
    Environment*, 127:130–138.

Prihodko, L. and Goward, S. N. (1997). Estimation of air temperature from remotely
    sensed surface observations. *Remote Sensing of Environment*, 60(3):335–346.

Rhinane, H., Hilali, A., Bahi, H., and Berrada, A. (2012). Contribution of landsat
    tm data for the detection ofurban heat islands areas case of casablanca. *Journal
    of Geographic Information System*, 2012:20–26.

Rizman, A. M., Dennis, L. Y., and Liu, C. (2008). A review on the generation,
    determination and mitigation of urban heat island. *Journal of Environmental
    Sciences*, 20(1):120–128.

Sahani, N. (2021). Assessment of spatio-temporal changes of land surface temper-
    ature (lst) in kanchenjunga biosphere reserve (kbr), india using landsat satellite
    image and single channel algorithm. *Remote Sensing Applications: Society and
    Environment*, 24:100659.

Salazar, A., Baldi, G., Hirota, M., Syktus, J., and McAlpine, C. (2015). Land use
    and land cover change impacts on the regional climate of non-amazonian south
    america: A review. *Global and Planetary Change*, 128:103–119.

Sekertekin, A. and Arslan, N. (2019). Monitoring thermal anomaly and radiative
    heat flux using thermal infrared satellite imagery – a case study at tuzla geother-
    mal region. *Geothermics*, 78:243–254.

Senay, G. B., Schauer, M., Velpuri, N. M., Singh, R. K., Kagone, S., Friedrichs, M.,
    Litvak, M. E., and Douglas-Mankin, K. R. (2019). Long-term (1986–2015) crop
    water use characterization over the upper rio grande basin of united states and
    mexico using landsat-based evapotranspiration. *Remote Sensing*, 11(13).

Shao, M., Tang, X., Zhang, Y., and Li, W. (2006). City clusters in china: air and
    surface water pollution. *Frontiers in Ecology and the Environment*, 4(7):353–361.

Shen, S. and Leptoukh, G. G. (2011). Estimation of surface air temperature over central and eastern eurasia from modis land surface temperature. *Environmental Research Letters*, 6(4):045206.

Sobrino, J. A., Jiménez-Muñoz, J. C., and Paolini, L. (2004). Land surface temperature retrieval from landsat tm 5. *Remote Sensing of Environment*, 90(4):434–440.

Spampinato, L., Calvari, S., Oppenheimer, C., and Boschi, E. (2011). Volcano surveillance using infrared cameras. *Earth-Science Reviews*, 106(1):63–91.

Stewart, I. and Oke, T. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93:1879–1900.

T. R. Oke, G. Mills, A. C. A. V. (2017). Urban climates. Cambridge University Press.

UN (2020). World urbanization prospects: The 2020 revision. `https://www.worldbank.org/en/topic/urbandevelopment/overview`.

Unal Cilek, M. and Cilek, A. (2021). Analyses of land surface temperature (lst) variability among local climate zones (lczs) comparing landsat-8 and envi-met model data. *Sustainable Cities and Society*, 69:102877.

USGS (2008). Aster swir user advisory. `https://lpdaac.usgs.gov/documents/217/ASTER_SWIR_User_Advisory_July_18_08.pdf`.

USGS (2016). Using the usgs landsat8 product. *US Department of the Interior-US Geological Survey–NASA*.

USGS (2020). Landsat collection 2. `https://www.usgs.gov/landsat-missions/landsat-collection-2`.

Vancutsem, C., Ceccato, P., Dinku, T., and Connor, S. J. (2010). Evaluation of modis land surface temperature data to estimate air temperature in different ecosystems over africa. *Remote Sensing of Environment*, 114(2):449–465.

Voogt, J. and Oke, T. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment*, 86(3):370–384. Urban Remote Sensing.

Weng, Q. (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4):335–344.

Weng, Q., Lu, D., and Schubring, J. (2004). Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment*, 89(4):467–483.

Yang, J., Zhan, Y., Xiao, X., Xia, J. C., Sun, W., and Li, X. (2020). Investigating the diversity of land surface temperature characteristics in different scale cities based on local climate zones. *Urban Climate*, 34:100700.

Zhu, W., Lű, A., and Jia, S. (2013). Estimation of daily maximum and minimum air temperature using modis land surface temperature products. *Remote Sensing of Environment*, 130:62–73.