Article

Measuring Urban Land Cover Influence on Air Temperature through Multiple Geo-Data—The Case of Milan, Italy

Daniele Oxoli [†] , **Giulia Ronchetti** *,[†] , **Marco Minghini** , **Monia Elisa Molinari,**
Maryam Lotfian, Giovanna Sona and Maria Antonia Brovelli

Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32,
20133 Milano, Italy; daniele.oxoli@polimi.it (D.O.); marco.minghini@polimi.it (M.M.);
moniaelisa.molinari@polimi.it (M.E.M.); maryam.lotfian@polimi.it (M.L.);
giovanna.sona@polimi.it (G.S.); maria.brovelli@polimi.it (M.A.B.)
* Correspondence: giulia.ronchetti@polimi.it; Tel.: +39-02-23-996-543
† These authors contributed equally to this work.

check for
updates

Abstract: Climate issues are nowadays one of the most pressing societal challenges, with cities being identified among the landmarks for climate change. This study investigates the effect of urban land cover composition on a relevant climate-related variable, i.e., the air temperature. The analysis exploits different big geo-data sources, namely high-resolution satellite imagery and in-situ air temperature observations, using the city of Milan (Northern Italy) as a case study. Satellite imagery from the Landsat 8, Sentinel-2, and RapidEye missions are used to derive Local Climate Zone (LCZ) maps depicting land cover compositions across the study area. Correlation tests are run to investigate and measure the influence of land cover composition on air temperature. Results show an underlying connection between the two variables by detecting an average temperature offset of about 1.5 °C between heavily urbanized and vegetated urban areas. The approach looks promising in investigating urban climate at a local scale and explaining effects through maps and exploratory graphs, which are valuable tools for urban planners to implement climate change mitigation strategies. The availability of worldwide coverage datasets, as well as the exclusive use of Free and Open Source Software (FOSS), provide the analysis with a potential to be empowered, replicated, and improved.

Keywords: urban climate, geo-data; Free and Open Source Software; Geographic Information Systems; satellite imagery; local climate zones; air temperature

1. Introduction

Assessing and mitigating human-driven impacts on climate is nowadays a well-established societal concern, as well as a highly pressing challenge for a number of actors operating across different scales and sectors. Interest from the scientific community in climate change dates back to more than a century ago [1]. However, it was only over the last few decades that the growing participation of public and non-public institutions to international summits on climate has provided a clear proof of the increasing importance attributed to this phenomenon from both the economic and the civil society [2,3]. This has raised attention on a number of climate-related issues, recognized as priorities by the international community.

Among them, the increasing trend in urbanization has been identified as a landmark for climate change. Today, cities host 55% of the world's population and this number is projected to rise to 68% by 2050 [4]. Cities generate more than 80% of global Gross Domestic Product (GDP), but they also consume about 2/3 of the world's energy and account for more than 70% of global greenhouse

gas emissions [5]. Cities are also recognized as one of the most vulnerable places in terms of the future effects of climate change such as increased water shortages, negative health effects from Urban Heat Islands (UHI), and increased risk of flooding, among others [6]. This challenge is explicitly addressed by the United Nations (UN) Sustainable Development Goals (SDGs), in particular SDG 11 "Make cities inclusive, safe, resilient and sustainable" [7]. Nevertheless, the cities themselves can play a pivotal role in addressing climate change, since the way they are planned and governed may generate profound impacts on the minimization of emissions while also sustaining high living standards currently threatened by climate change [8]. This represents a unique opportunity for the academic and scientific community.

Climate change adaptation and mitigation strategies must certainly start from collecting, understanding, and extracting knowledge from the multitude of geo-data sources available for cities [9]. Such sources can be roughly classified into space sensors and in-situ sensors. The former category includes optical and radar imagery, captured by satellites orbiting the Earth, as well as their derived products. An increasing number of Earth Observation (EO) initiatives around the globe is providing free and open access to such datasets with an unprecedented coverage and resolution. The most comprehensive example is the Copernicus Programme, which is empowering the European Commission with remotely-sensed information for land, marine, and atmospheric monitoring, emergency mapping, and climate change analysis [10]. The second category includes instead the burgeoning and non-stop amount of information generated by physical sensors, e.g., environmental sensor networks, and Internet of Things (IoT) devices and applications. These are complemented by, and often intersected with, crowdsourced geographic information collected by citizens [11]. This information can be contributed passively, e.g., from social networks, or actively within so-called Volunteered Geographic Information (VGI) projects, the most popular of which is OpenStreetMap [12], or Citizen Science (CS) initiatives. In urban contexts, many CS projects were accomplished to collect measurements of air temperature, air pollution, urban noise, etc. [13–15].

With this background, the present work contributes to the investigation of urban climate at the local scale through an approach based on the combination of space and in situ observations, the latter including crowdsourced geographic information. More in detail, the objective of the study is to analyze the effect of urban land cover on one of the most relevant climate-related variables, i.e., air temperature. This aims to provide insights into the study of urban heating issues, which are typically linked to anthropogenic factors such as Urban Heat Islands (UHIs) [16]. UHIs are identified where temperature warming patterns for urban areas are significantly higher than the rural surroundings. UHIs are due to structural and land cover differences between urban and rural areas and are generally measured through a climate-based classification system called Local Climate Zone (LCZ) [17]. This classification defines 17 unique area types based on the physical and thermal properties of their surface. LCZ maps are widely used in urban climatology since they can explain the contribution of urban surface characteristics to heat fluxes [18,19]. These maps are usually derived from the supervised classification of satellite imagery, eventually in combination with other EO products such as Digital Elevation Models (DEMs) [20]. The World Urban Database and Access Portal Tools (WUDAPT) [21,22] offers detailed guidelines on how to perform LCZ classification and provides access to LCZ maps of different cities.

Literature has seen many authors developing studies on the correlation of land cover and temperature, e.g., [23,24]. However, these have mainly focused on the analysis of surface temperature rather than air temperature [25]. Despite air temperature and surface temperature are generally correlated measures, they have a different physical meaning. Surface temperature is a measure of the radiation emitted by the land surface, thus it can be acquired by satellite sensors. Air temperature is instead a ground-based measure which is strongly connected to atmospheric, seasonal, and geographical conditions [26]. Most important, it represents the characterizing factor of UHIs and, as such, it is considered to evaluate their intensity and persistence [27]. The present study focuses on the urban area of Milan (Northern Italy), where no similar studies have been carried

out so far. The proposed analysis aims to provide insights into the underlying effect of land surface features on the air temperature by leveraging geo-data exploratory procedures. This enables to answer questions such as how and how much each LCZ class is correlated to local air temperature observations as well as to highlight strengths and limits of the considered geo-data in achieving the analysis goal.

The remaining of the paper is structured as follows. Section 2 introduces the area chosen to analyze the relation between urban land cover and temperature, as well as the data sources and software technology used. In Section 3 the methodology undertaken to preprocess the datasets, compute the LCZ maps and correlate them with temperature observations are presented in detail. Section 4 discusses the results achieved. Section 5 concludes the paper by outlining the main findings of the study.

2. Materials

2.1. Study Area

The analysis framework developed within this study is tested on the city of Milan (Northern Italy). Milan is located in one of the most densely populated European regions, the Padana Plain (see Figure 1), affected by poor wind circulation and high urbanization which favour the stagnation of pollutants as well as the persistence of a historically well-known UHI [28]. This makes Milan a suitable test area to investigate the influence of land cover on local air temperature that is the primary goal of the proposed analysis.
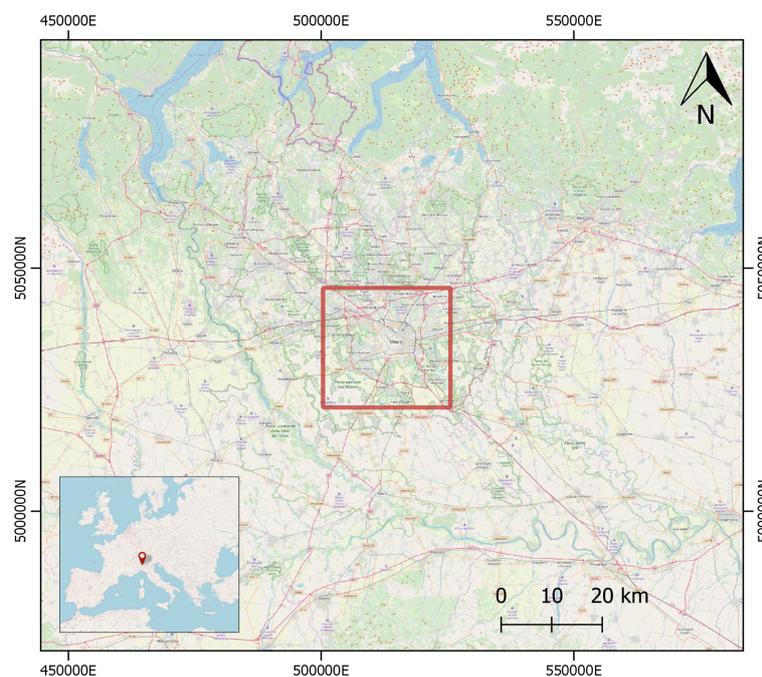
Figure 1. Location of the study area. Coordinate Reference System (CRS): WGS84/UTM zone 32 N. Map data: © OpenStreetMap contributors.

The study region is characterized by a flat terrain with an average altitude of 120 m above sea level. Minimum and maximum altitudes are 85 m and 175 m above sea level, respectively. Due to this small range, we did not account for any relevant variation of temperature produced by orographic phenomena. The region climate characteristics can be summarized according to the Köppen-Geiger Climate Classification [29] in warm temperate, fully humid with hot summer (Cfa zone). Such environmental peculiarities are also common to other cities in Italy and Southern Europe [30], thus the same work might be replicated or adapted in other sites.

2.2. Data Collection

Collected data consists mainly of optical satellite imagery and temperature series from ground sensors. Data sources have been selected by considering data availability as well as licenses. The whole data used for this work is available under fully open licenses or at most distributed by the providers under custom open licenses enabling exploitation for research purposes. A detailed description of the two data sources and the strategies for their collection is reported in the following.

2.2.1. Satellite Imagery Collection

Three datasets are exploited for the assessment of the LCZ mapping performance. These consist of Landsat 8, Sentinel-2, and RapidEye satellite imagery available for Milan for the Year 2016.

- Landsat 8: Landsat 8 satellite collects images of the Earth every 16 days in 11 different spectral bands, through two main sensors, an optical sensor (the Operational Land Imager, OLI), and a thermal one (the Thermal Infrared Sensor, TIRS). With the aim of generating LCZ maps exclusively from optical data, thermal, panchromatic, aerosol, and cirrus bands were not considered in this study. Hence, only six bands were considered, which have a spatial resolution of 30 m. Landsat 8 data is freely available for download from GloVis, EarthExplorer, and also LandsatLook Viewer. In this study, data was downloaded through the EarthExplorer platform [31].
- Sentinel-2: The Sentinel-2 satellites acquire, approximately every 5 days, multispectral imagery with 13 bands, but only 10 bands are useful for classification purposes with a spatial resolution of 10 m and 20 m. Sentinel-2 imagery is distributed under a fully open license through the Copernicus Open Access Hub [32] interface. Data access through a dedicated Application Programming Interface (API) is also available.
- RapidEye: The RapidEye satellite constellation provides multispectral images every 5.5 days at nadir. Data are available at 5 m spatial resolution in 5 different spectral bands: Blue, green, red, red-edge, and NIR. RapidEye imagery is distributed under commercial fee by, among the others, Planet Labs Inc. through its Planet Explorer platform [33] API. Free of charge access to the RapidEye archive data API is kindly provided for research groups, with the download limit of 10,000 Km^2 per month.

According to the satellite revisiting time and weather conditions, two different images were downloaded for each satellite: The first representing a non-vegetative period (from October to March) and the second representing a vegetative period (from April to September). The two periods were defined according to seasonal features. Non-vegetative includes autumn and winter seasons, whereas vegetative refers to spring and summer seasons. Table 1 summarizes the acquired images for Landsat 8, Sentinel-2, and RapidEye. Descriptions of the reference synoptic conditions at the dates of acquisition of the satellites are reported in Table 2.

Table 1. Acquisition date and time for satellite images. Time is given in Greenwich Mean Time (GMT).

Satellite	Acquisition Date and Time	
	Non-Vegetative	**Vegetative**
Landsat 8	18-03-2016 10:10 a.m.	22-06-2016 10:10 a.m.
Sentinel-2	13-01-2016 10:20 a.m.	22-05-2016 10:20 a.m.
RapidEye	18-03-2016 11:14 a.m.	22-06-2016 10:46 a.m.

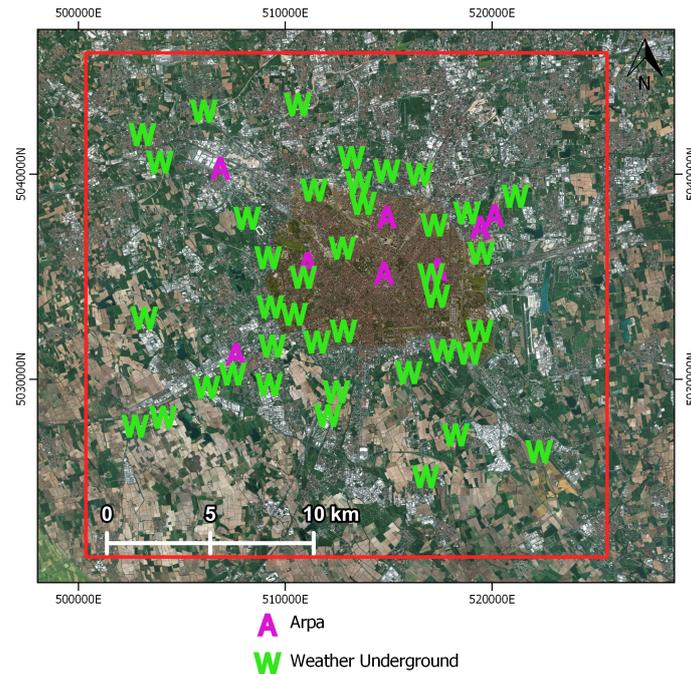### 2.2.2. Air Temperature Observations

Time series of environmental and weather observations are traditionally collected and distributed by public authorities with fewer restrictions. In Italy, official environmental monitoring is accomplished by the Regional Agency for Environmental Protection (ARPA, *Agenzia Regionale per la Protezione dell'Ambiente*) through its sensor networks located in all the 20 administrative regions. Time series

can be requested by any user through the ARPA data download platform. Data is provided under a custom open license with quality standard certification as requested by the national law.

Table 2. Synoptic conditions at the acquisition dates of of the satellites [34,35]. Arrows for Geopotential height indicate decreasing (↓), steady (↔), and increasing (↑) trend.

Acquisition Date	Synoptic Parameters
13-01-2016	Geopotential height 500 hPa: 5450 m (↓) Wind Direction 500 hPa: 315 deg Advection 850 hPa: Cold Thermodinamical Profile: Stable
18-03-2016	Geopotential height 500 hPa: 5560 m (↔) Wind Direction 500 hPa: 105 deg Advection 850 hPa: No advection Thermodinamical Profile: Stable
22-05-2016	Geopotential height 500 hPa: 5740 m (↑) Wind Direction 500 hPa: 200 deg Advection 850 hPa: Warm Thermodinamical Profile: Stable
22-06-2016	Geopotential height 500 hPa: 5890 m (↑) Wind Direction 500 hPa: 20 deg Advection 850 hPa: Warm Thermodinamical Profile: Stable

For this study, additional weather stations are considered in order to increase the spatial coverage of temperature observations over the study area. This is due to the need of investigating the correlation between temperature observations and LCZ at a very local scale, i.e., taking into account the likely point-wise nature of the two phenomena and their possible cause-effect relationship (see Section 3). The additional observations are collected from the Weather Underground (WU) platform [36]. This is an Internet service providing weather observations collected from both private and public weather stations from all over the world. Being based on a crowdsourced data collection strategy, information is delivered with no quality standard certification. Weather observations are publicly shared through the Weather Underground API, which allows requesting data at no cost for non-commercial purposes and within the limits expressed by the API provider. Because of new company policies, starting from June 2018 WU is no longer providing free API keys. All the data used in this study was acquired before this policy change. WU temperature time series for the stations of interest in Milan are downloaded in comma-separated values (CSV) format using a custom JavaScript collector module developed by the authors. The collected series refer to the period 1 January to 31 December 2016. Temperature measures from both ARPA and WU are available with a sub-hourly frequency. Stations locations are also retrieved and organized in a separate table linked with the time series CSV table for both data sources. The distribution of ARPA and WU stations providing temperature observations in Milan is depicted in Figure 2. The satellite basemap clearly shows that the stations in the South and South-West sides of the study area are mainly located in rural or not heavily urbanized areas.

Figure 2. Locations of Regional Agency for Environmental Protection (ARPA) and Weather Underground stations in the study area providing air temperature observations in Milan for the Year 2016. CRS: WGS84/UTM zone 32 N. Basemap data: Bing Satellite © Microsoft.

2.3. Software Tools

This work makes exclusive use of Free and Open Source Software (FOSS) as data analysis tools with the aim of favoring the reproducibility of the study. The premier FOSS *QGIS* [37] is exploited for general geodata management, layout design, and part of the raster and vector processing. The programming languages Python and R are also widely used for data analysis and processing. The most exploited Python library is the Python Data Analysis Library (*pandas*) [38], an open source library providing high-performance data structures and data analysis tools for Python. The geospatial extension of this library, i.e., *geopandas* [39], is also used. Statistical processing is performed by using the *NumPy* [40] and the *SciPy* [41] libraries that provide many built-in methods for numerical and statistical computations. The *rasterstat* library [42] is also adopted to enable raster layer processing. Regarding R, ad hoc scripts are used for processing and classification of satellite imagery, by exploiting the *raster* [43] and *randomForest* [44] packages. The former allows to read, write and manipulate gridded spatial data, while the latter performs data classification and regression based on a forest of trees using random inputs. Details about the use of software and libraries are included in the following sections.

3. Methods

3.1. Data Processing

LCZ maps can explain, in an intuitive way, the contribution of urban surface characteristics to heat fluxes. This is better achieved when LCZ maps show a good agreement with temperature observations, so that this coupled information can act as a valuable tool for urban climate management and local mitigation strategies. In addition to the classification accuracy, the resolution of LCZ maps is also fundamental for their adoption in urban planning and management. In the following, a FOSS-based methodology is described to exploit satellite imagery at a different spatial and spectral resolution, derive detailed LCZ maps, assess their classification accuracies, and perform a combined analysis with air temperature data.

3.1.1. Satellite Imagery Preprocessing

In order to perform LCZ classification of satellite imagery, besides standard procedures such as the clip and merge of tiles to reconstruct a complete scene of the study area at any time period, band pixels need first to be converted to their real reflectance values. This ensures that classifications are not negatively affected by atmospheric disturbances so that images acquired in different seasons can be compared. Specific processing for the different satellite imagery is described below.

- Landsat 8: Level-1 data available to users consists of radiometrically and geometrically corrected images. The Level-1 image is presented in units of DNs, which can be easily rescaled to spectral radiance or Top of Atmosphere (TOA) reflectance. In order to obtain real surface reflectance values, band pixels need to be further corrected. The Semi-Automatic Classification Plugin for *QGIS* [45] was used for such preprocessing.
- Sentinel-2: The Copernicus Open Access Hub provides Sentinel-2 imagery with different level products. In this study Level 1C imagery is used, which is ortho-corrected and with pixel radiometric measurements provided in Top of Atmosphere (ToA) reflectance. Moreover, with the aim of exploiting the complete band set of Sentinel-2 imagery at 10 m spatial resolution, resampling is required. As for Landsat 8 the preprocessing, including atmospheric correction and resampling, was performed through the Semi-Automatic Classification Plugin in *QGIS*.
- RapidEye: The Planet Explorer platform distributes only ortho-corrected imagery with bands at 5 m spatial resolution. Atmospheric disturbances can be adjusted by exploiting the band reflectance and geometric coefficients available in the metadata file of each tile, in either eXtensible Markup Language (XML) or Javascript Object Notation (JSON) formats, and by applying a Dark Object Subtraction (DOS) procedure [46]. To automate this process, an R script based on the open source Geospatial Data Abstraction Library (*GDAL*) [47] was created. The script makes also use of the *raster* R package.

3.1.2. Local Climate Zone (LCZ) mapping

LCZ classification system divides the urban area into different climate zones based on their physical and thermal characteristics. The classification consists of 17 classes, which are divided into "built types" and "land cover types" mainly based on the characteristics of the surface structure (height and density of buildings and trees) and surface cover (impervious and pervious) [17]. In the city of Milan, 8 different LCZ classes were found and considered out of the 17 original classes. A brief description of each class is reported in Table 3.

3.1.3. Random Forest (RF) Classification

To perform the LCZ classification, a number of steps were required which are summarized in the flowchart in Figure 3.
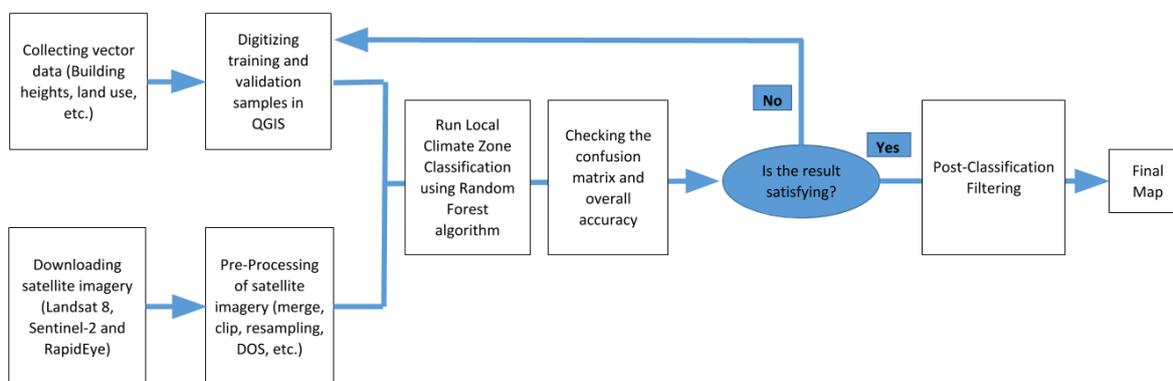

Figure 3. Flowchart adopted for Local Climate Zone (LCZ) classification.

Table 3. Local Climate Zone (LCZ) classes retained for the city of Milan: ID, Name, and Description [17].

Class ID	Class Name	Description
1	Compact midrise	Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction material.
2	Compact low-rise	Dense mix of low-rise building (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction material.
3	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.
4	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.
5	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.
6	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.
7	Low plants	Featureless landscape of glass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park.
8	Water	Large, open water bodies such as seas or lake, or small bodies such as rivers, reservoirs, and lagoons.

A supervised pixel-based classification was carried out using training areas digitized in *QGIS* and selected according to the LCZ classification system. Training polygons were defined by means of open geo-databases provided by the Lombardy Region. The most considered was the DUSAF (Destinazione d'Uso dei Suoli Agricoli e Forestali) [48], which includes detailed land use and land cover vector layers at scale 1:10,000, coupled with additional information such as the building heights from the regional topographic database [49].

The classifier adopted for this work was the Random Forest (RF) algorithm [50]. RF was chosen due to a number of reasons as follows. It is non-parametric, it provides high accuracy and fast computational performance on large databases, and since the classification unbiased error is estimated internally in the run, there is no need for cross-validation or for using a separate testing dataset [51]. RF classification was performed through the use of an R script, based on the RF implementation *randomforest* in R.

Accuracy of the RF classification outputs was assessed using the so-called "out-of-bag" (OOB) samples, not included in the training data. Common statistical measures were calculated from the confusion matrix [52], including Overall Accuracy (OA), Producer's Accuracy (PA) and User's Accuracy (UA). Since the pixel-based classification is typically noisy, a post-classification filtering was applied. In particular the Majority filter [53], available within the *raster* package in R, was applied on the resulting maps with a 3 × 3 moving window.

3.1.4. Air Temperature Time Series Adjustment

Time series of temperature observations were manipulated through Python Pandas scripting. Raw CSV files containing temperature observations from each station were collapsed in a single *pandas* DataFrame object. This is a two-dimensional tabular data structure with labeled axes, which uses the timestamps of collected observations as row indexes and the unique ids of the stations as column indexes. Sub-hourly observations are aggregated into hourly averages. A quick quality check of observations was performed by removing the stations for which data completeness—computed as the ratio between the number of hours in the study period with observations actually available and the total number of hours in the study period—was lower than 0.8. Hence, each considered station series contained less than 20% missing hourly observations. An additional check was performed

to remove stations having a minimum correlation coefficient—computed as the pairwise Pearson's correlation coefficient [54]—with all the other stations lower than 0.75. These two thresholds were arbitrarily selected as coarse indicators of the goodness of the time series. The values are aligned with those suggested by the Italian Institute for the Environmental Protection and Research (ISPRA) [55]. Finally, a rolling mean using a 5 h window was applied to smooth the series. This preliminary data cleaning step was useful to improve the reliability of the available data, especially the WU observations which—in contrast to the ones from ARPA—are provided without any prior quality check (Figure 4). The temperature time series resulting from these cleaning operations were considered as reference data for the correlation analysis described in the following sections.
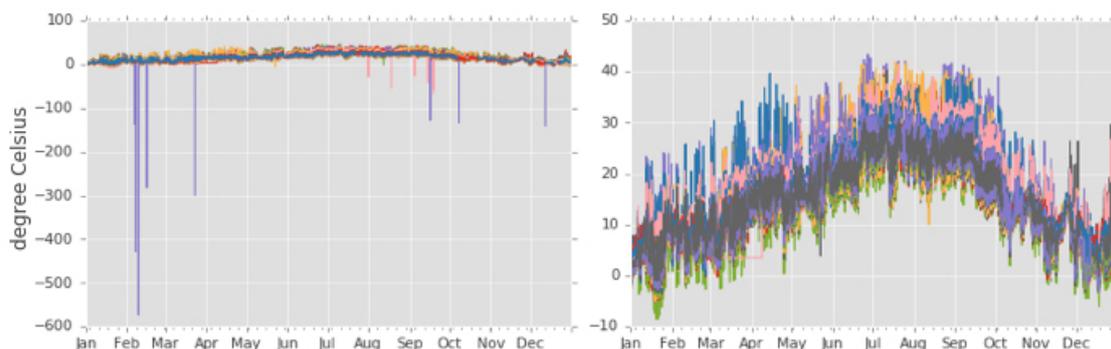
Figure 4. Example of temperature time series for the Year 2016 from Weather Underground stations in Milan before (**left**) and after (**right**) the cleaning and smoothing procedures. Coarse outliers (**left**) are clearly removed from the data set (**right**).

3.2. LCZ and Air Temperature Correlation Analysis

To investigate the correlation between the land cover type, expressed by the LCZ, and the air temperature, a 500 m buffer around each temperature station was considered. This relatively small buffer radius was introduced to account for the large variability of land cover features that characterize the urban environment. Moreover, temperature measurements are representative only in the spatial proximity of the sensors, where temperature changes can be assumed as negligible. By selecting a small buffer radius, also the point-wise nature of the temperature observations was taken into account for the correlation tests. Nevertheless, thanks to the abundance of weather stations across the Milan area, a satisfactory data coverage for the study area was achieved, especially within the city centre and the suburbs (Figure 2).

The fraction of each LCZ class coverage within each buffer was then calculated as the ratio between the number of pixels of a class and the total number of pixels within a buffer. A second measure of the LCZ class composition inside buffers was computed. This was achieved by assigning to each buffer its corresponding majority LCZ class id, computed as the class having the highest coverage inside the buffer out of the total number of pixels. Both measures were computed for each satellite at the two considered time periods, i.e., the non-vegetative and the vegetative (see Table 1).

These measures were used to run two separate correlation tests. The first one consisted of computing scatter-plots and connected regression lines by comparing the LCZ fractions at each station buffer with the temperature registered by the station at the exact acquisition time of satellites. This was accomplished using a custom Python script that combines *GeoPandas* and *Rasterstat* functionalities to process the station dataset and the LCZ raster maps. Scatter-plots were generated by using the *Matplotlip* library, while the linear regression was computed by the *SciPy* library. A positive slope of the regression line is interpreted as a positive contribution of the LCZ class to the temperature, while a negative slope as a mitigative effect. A scatter-plot was computed for each satellite at each time period and for each LCZ class. Results are reported and discussed in Section 4.

In the second experiment, the buffers were grouped according to their majority LCZ class id. Descriptive statistics such as the weekly temperature median and standard deviation were computed

for each group of buffers from the temperature series of the stations included in each group (see Table 5). The reference week is the one in which the satellite image used for LCZ classification was acquired. The majority LCZ class and the descriptive statistics were both computed using the *QGIS* Zonal Statistics Plugin as well as some Python scripts based on the *GeoPandas* library. This second experiment was designed with the goal of finding evidence of the persistence of higher temperatures within buffers characterized by a heavy urbanization and lower temperatures in buffers characterized by a more natural land cover composition. Results are reported and discussed in Section 4.

4. Results and Discussion

4.1. LCZ Maps

The LCZ maps obtained are shown in Figure 5, from which differences among different sensors as well as time periods can be outlined. Table 4 summarizes the OA for each classification output. OA is around 81% for the LCZ maps derived from Landsat 8, between 75% and 78% for the maps derived from Sentinel-2, and 71% for the maps derived from RapidEye. The slightly higher values registered for Landsat 8 can be mostly attributed to the pixel size and the choice of training datasets. In fact, in Landsat 8 imagery the pixel size allows smoothing reflectance information as well as noisy details such as buildings or tree shadows, which are instead more evident in Sentinel-2 and RapidEye imagery. Moreover, with the aim of comparing the classifications derived from the different satellite imagery, the same reference samples were used with a pixel size equal to the largest among the three, i.e., that of Landsat 8. A possible improvement of the OA values for the LCZ maps derived from Sentinel-2 and RapidEye might be achieved by selecting training samples more suited to the pixel size. In addition, the lower spectral resolution of RapidEye sensors negatively influences the classification results, as highlighted by the spectral signatures of the 8 LCZ classes reported in Figure 6. Except for the class water, which shows its peculiar spectral signature, the other classes have a similar spectral behavior. The main spectral differences are evident for infrared wavelengths, starting from 800 nm, represented by three bands in Landsat 8 imagery; four bands in Sentinel-2 dataset; and only one band in RapidEye collection. These considerably reduce OA values of the LCZ maps derived from RapidEye imagery. Comparing the non-vegetative and vegetative periods, OA values are almost the same for Landsat 8 and RapidEye, while for Sentinel-2 an improvement (more than 3%) is achieved for the map corresponding to the vegetative season. Major spectral differences among the classes are found for the summer period, with respect to the winter.

Table 4. Overall Accuracy (OA) for the Random Forest (RF)-classification outputs [%].

Satellite	Period	
	Non-Vegetative	**Vegetative**
Landsat 8	81.43	80.56
Sentinel-2	75.57	78.70
RapidEye	71.79	71.61

In terms of PA and UA of each LCZ class, the LCZ maps derived from all the three satellites show very high values for some classes such as water, low plants or large low-rise (higher than 80% and up to 99% for the water class). Other classes, e.g., compact low-rise and open midrise, show instead very low values of UA and PA, e.g., 11% PA for compact low-rise in the RapidEye-derived classification. These low values point out the difficulty of the classification algorithm in correctly detecting some LCZ classes. Among built-up classes, compact midrise and large low-rise score high values of UA (more than 65% in all the cases), while the remaining classes cannot be distinguished. This aspect has to be considered for further analysis. Finally, as expected, a significant improvement is registered for the class scattered trees in the vegetative season, since during the non-vegetative period the absence of leaves causes confusion with the low plants class [56].
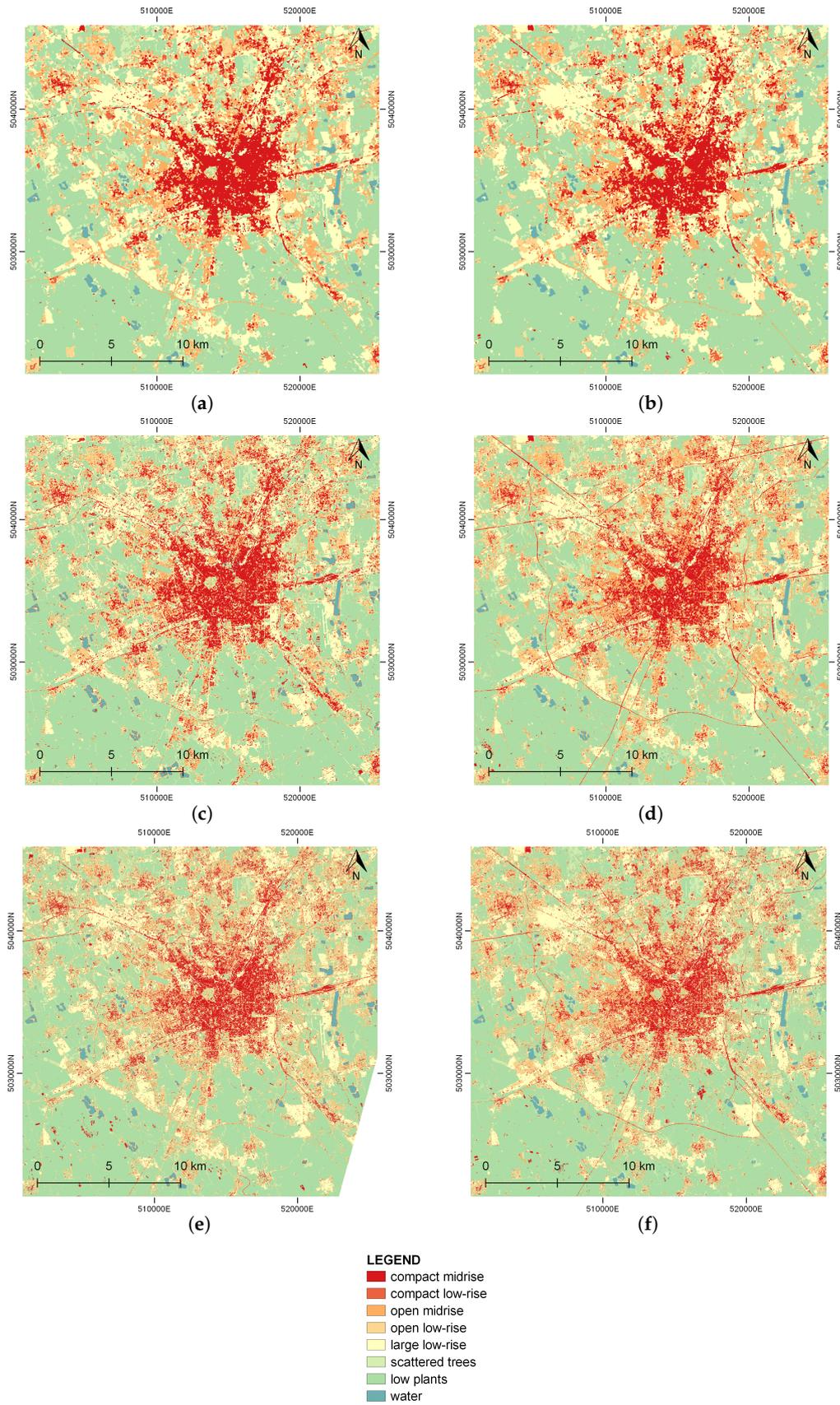
Figure 5. LCZ final maps: (**a**) Landsat 8, non-vegetative period; (**b**) Landsat 8, vegetative period; (**c**) Sentinel-2, non-vegetative period; (**d**) Sentinel-2, vegetative period; (**e**) RapidEye, non-vegetative period; (**f**) RapidEye, vegetative period. CRS: WGS84/UTM zone 32 N.
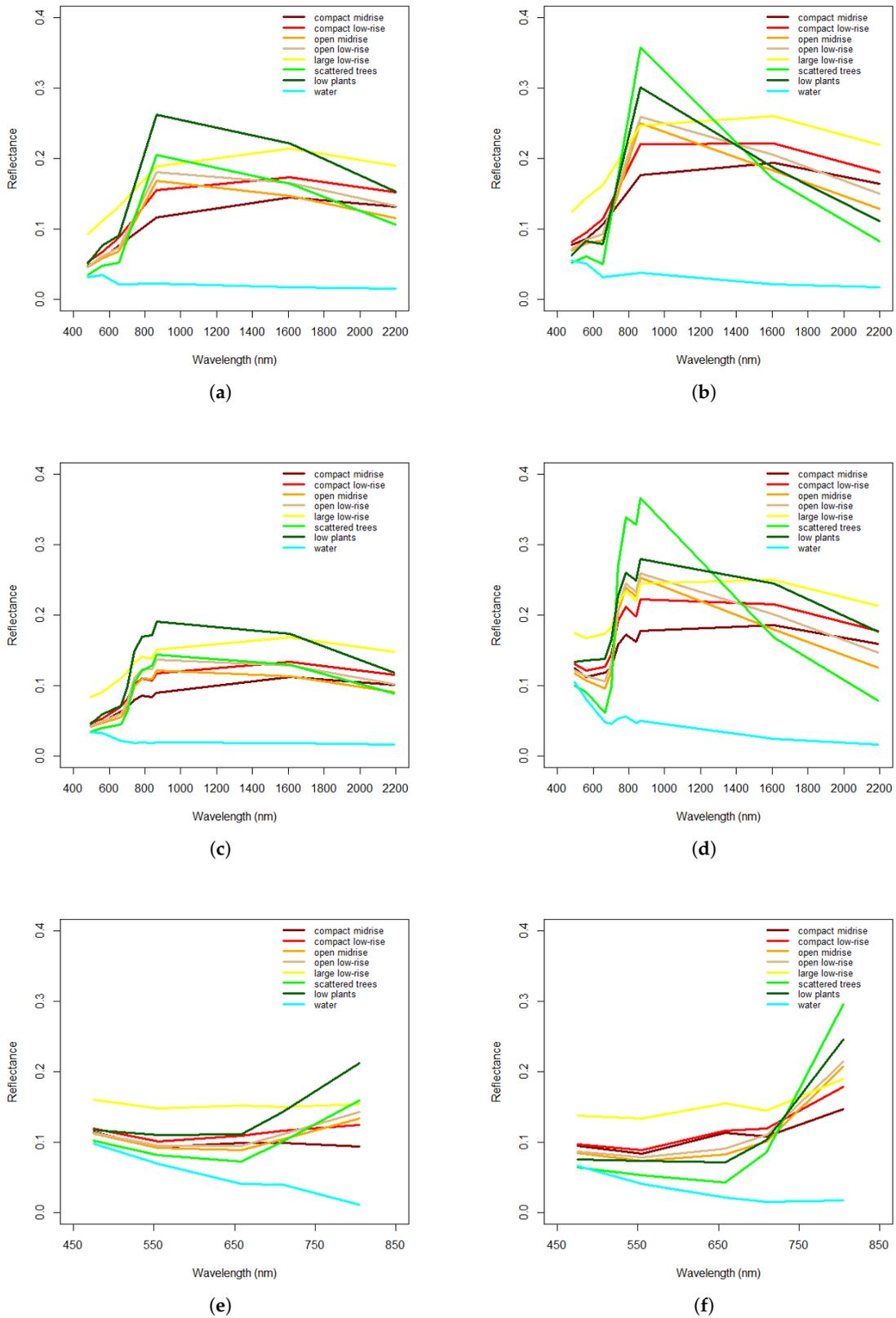
Figure 6. Spectral signatures of the 8 different LCZ classes: (**a**) Landsat 8, non-vegetative period; (**b**) Landsat 8, vegetative period; (**c**) Sentinel-2, non-vegetative period; (**d**) Sentinel-2, vegetative period; (**e**) RapidEye, non-vegetative period, (**f**) RapidEye, vegetative period.

4.2. Air Temperature Patterns in Milan

Temperature maps were derived from the air temperature observations by interpolating the station measures using Kriging [57]. Ordinary Kriging was performed through an R script based on the *gstat* package [58], available from the *QGIS* Processing Toolbox. Interpolated maps provided a better understanding of the air temperature patterns in the city of Milan. These allowed to explore evidence and features of the Milan UHI. A qualitatively assess of the actual goodness for the temperature series adopted was also achieved. In fact, the series provided a reasonable feedback in describing both spatial and temporal temperature patterns in the city. The presence of a significant UHI in the North-Eastern area of the city is clearly outlined in Figure 7, especially during the non-vegetative period. Conversely, during the vegetative period the UHI is less clustered while remaining more intense around the city center.

(a)   (b)
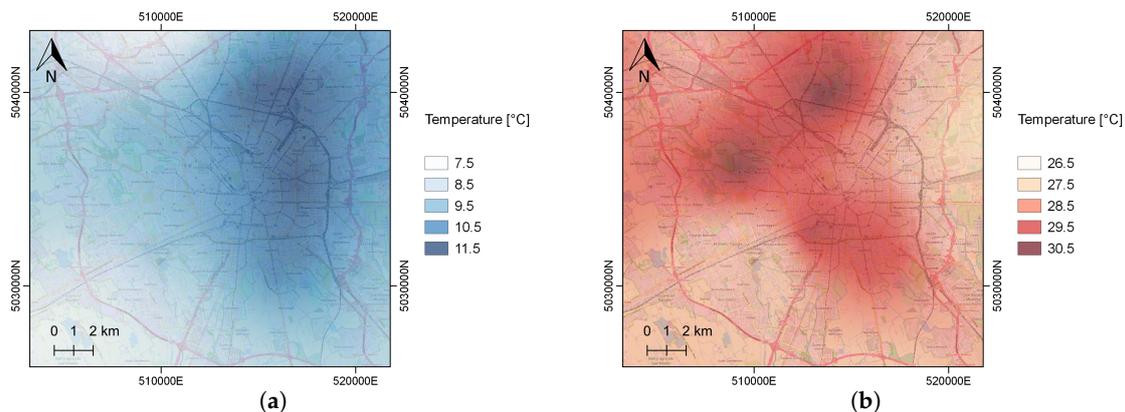
Figure 7. Temperature maps from the ordinary Kriging interpolation of the weekly average air temperature observations: Example of non-vegetative period (**a**) and vegetative period (**b**). Weekly averages were computed using the whole hourly observations for the week 11–17 January 2016 (**a**), and the week 20–26 June 2016 (**b**). CRS: WGS84/UTM zone 32 N.

4.3. Correlation Analysis

Results from the the correlation analysis between LCZ classes and temperature are included in Figures 8–10, and Table 5. For the sake of clarity, some LCZ class compositions for buffers were collapsed into a single class. These are classes 2, 3, 4, and 5, which represent slightly different land cover features of medium compact built-up areas. As a matter of fact, it was observed that the contribution of these classes to the correlation tests was comparable, and generally not interesting in terms of influence on air temperature. On the other hand, this might be also due to an intrinsically lower interclass accuracy of the LCZ classification procedure, that might be biased by the similar surface characteristics of these classes compared to more extreme classes such as 1, 6 and 7. Class 8, i.e., water, was not considered in the correlation analysis because no significant fractions (i.e., always < 0.04%) were detected in any of the station buffers.

By observing the scatter-plots in Figures 8–10, it can be argued that the contribution of highly urbanized areas, such as the one expressed by the Class 1 fractions, is always positive on the air temperature. This is true for any satellite imagery and time period considered. The contribution of the collapsed Classes 2, 3, 4, and 5 shows almost no correlation evidence with air temperature. Regarding classes 6 and 7, a sensible negative correlation is detected in almost all the cases. Besides a general low linear correlation between the two variables that is expressed by the coefficient of determination (R^2) computed for each graph, regression lines show comparable trends by considering the same LCZ class and time period for different satellites. The different spatial and spectral resolutions of the different satellite imagery seems not to have a significant influence on this exploratory experiment. The results show how temperature patterns are connected with the land cover type that is a well known effect, also reported in literature [59]. Nevertheless, the data used in this example, as well as the analysis

strategy adopted, would benefit from a further improvement in order to conceptualize robust models that can better explain this correlation.
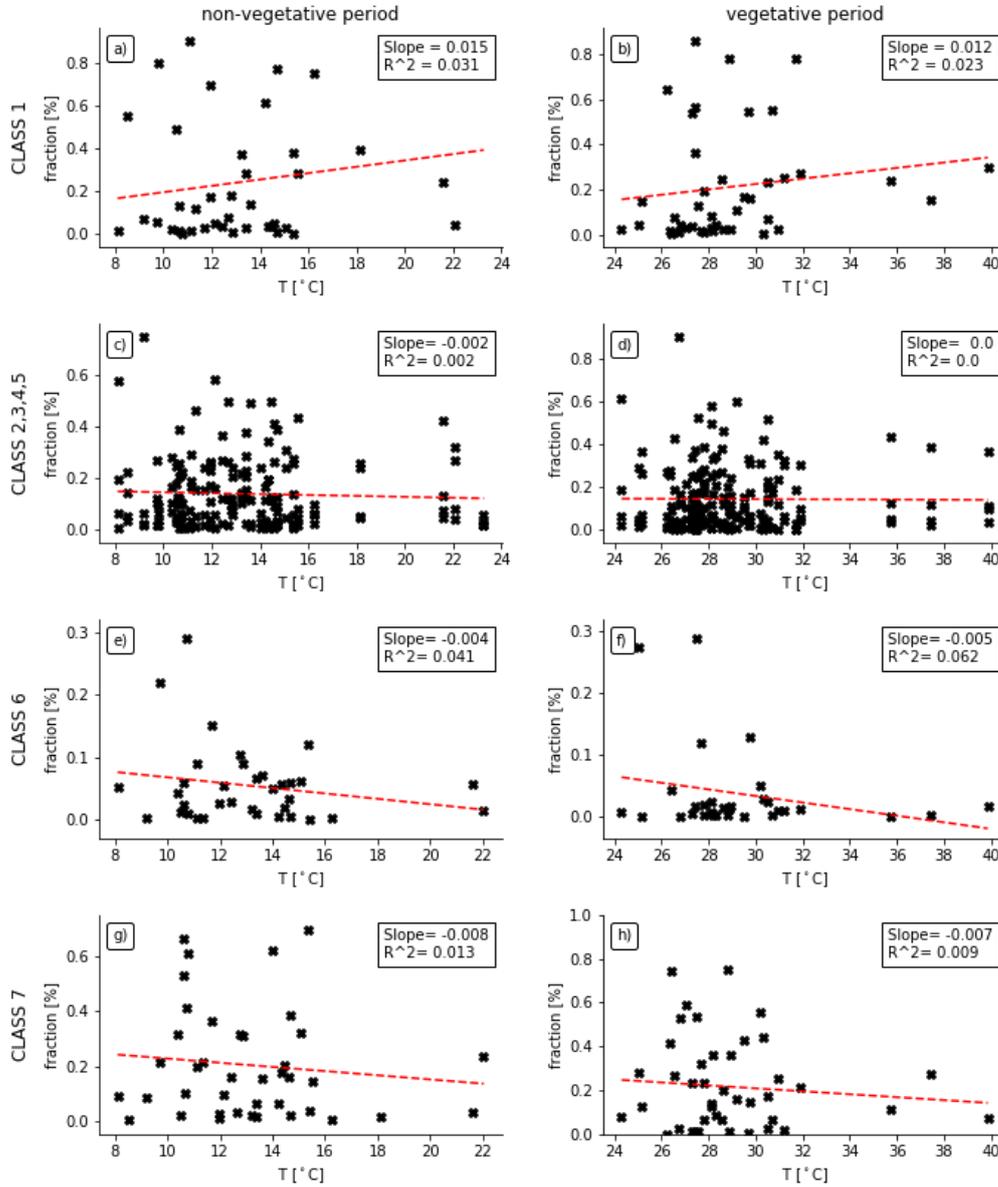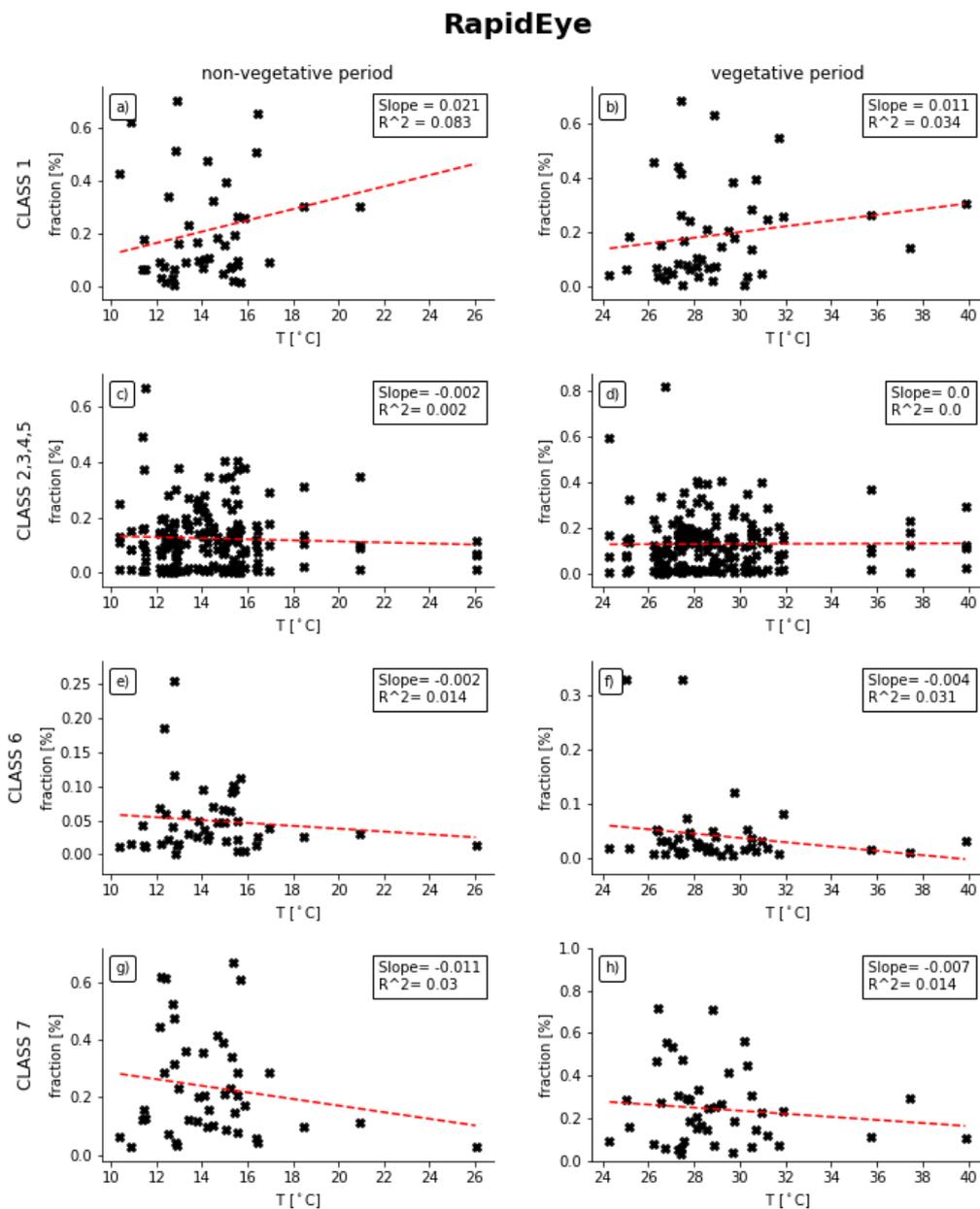

Figure 8. LCZ class fractions [%] computed into the 500 m buffer around the stations versus the measured air temperature at the stations. Air temperature measurements correspond to the hour of the Landsat 8 imagery acquisition. Acquisition times are 18 March 2016 (non-vegetative period) and 22 June 2016 (vegetative period) both at 10:10 am GMT. Red Lines and text boxes show the results of the linear regression for each plot.
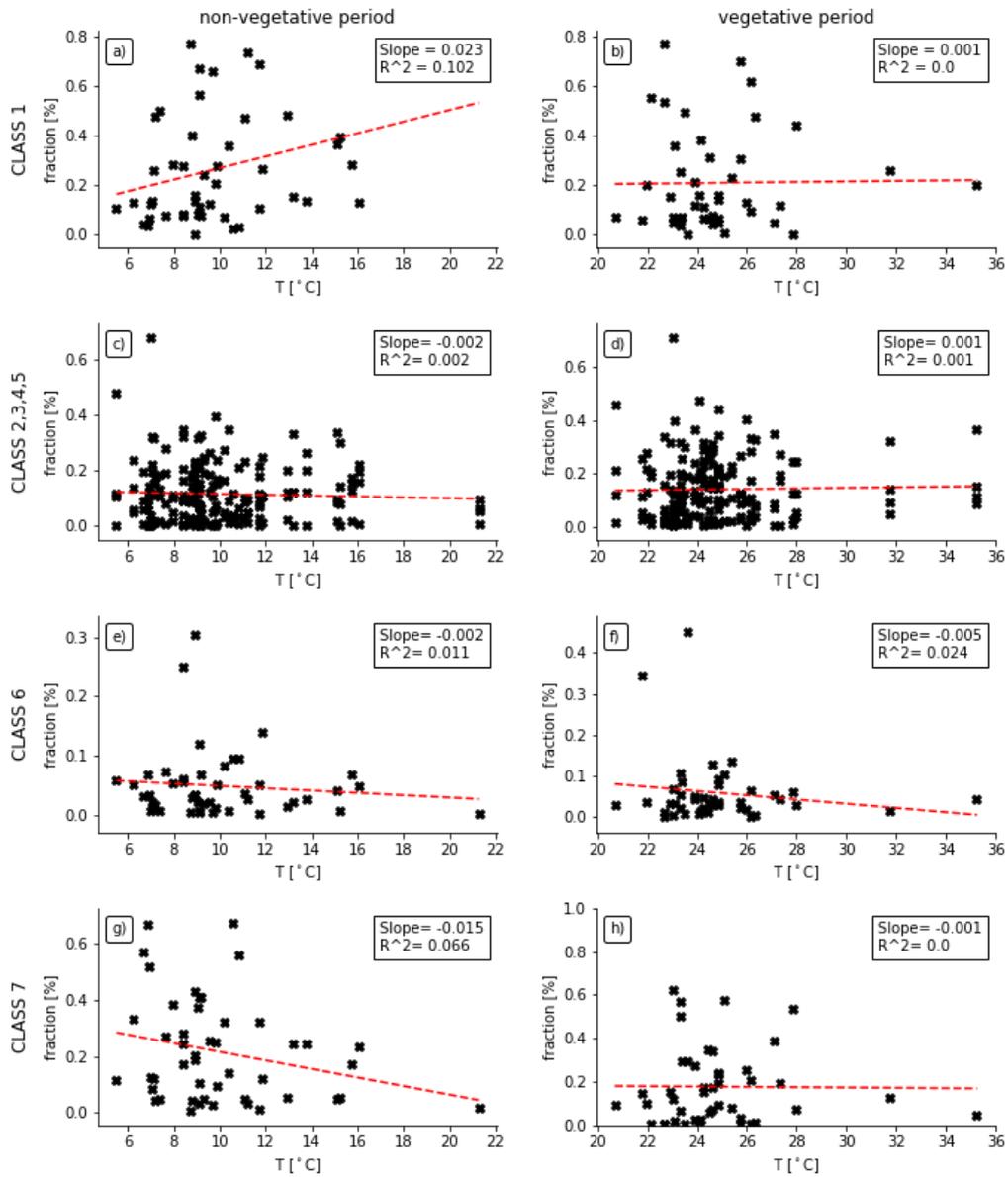
Figure 9. LCZ class fractions [%] computed into the 500 m buffer around the stations versus the measured air temperature at the stations. Air temperature measurements correspond to the hour of the Sentinel-2 imagery acquisition. Acquisition times are 13 January 2016 (non-vegetative period) and 22 May 2016 (vegetative period) both at 10:20 am GMT. Red Lines and text boxes show the results of the linear regression for each plot.

RapidEye

Figure 10. LCZ class fractions [%] computed into the 500 m buffer around the stations versus the measured air temperature at the stations. Air temperature measurements correspond to the hour of the RapidEye imagery acquisition. Acquisition times are 18 March 2016 at 11:14 a.m. GMT (non-vegetative period) and 22 June 2016 (vegetative period) at 10:46 a.m. GMT. Red Lines and text boxes show the results of the linear regression for each plot.

The most important outcome is, however, the demonstration of how heterogeneous geo-data can be combined and processed to explore urban climate patterns and get insight into these [9].

With the second experiment, we aim at providing some quantitative evidence of land cover influence on air temperature instead of qualitative results as we did in the previous experiment. Table 5 reports the median and standard deviation of weekly temperatures by LCZ majority buffers, for each time period and satellite considered. A marked difference of about 1.5 °C can be observed, which characterizes median temperatures of heavily urbanized areas, i.e., with Class 1 majority, and vegetated areas such as the ones with the majority in Class 7. The majority groups of the intermediate Classes 2, 3, 4, and 5 are always placed in between these two. Standard deviations are generally lower

than these differences. A different information can be retrieved by the buffers with the majority in Class 6, i.e., scattered trees. These few areas are detected in the Sentinel-2 and the RapidEye imagery in the vegetative period only. In Table 5, the Class 6 entry for the RapidEye represents the measurement and not the median because only a single buffer with this majority was detected. The absence of buffers with the majority in Class 6 for the Landsat 8 imagery can be explained by considering its lower spatial resolution (30 m). Hence, the spatial resolution of the imagery influences the classification of some LCZ classes by providing an approximated indication of the limit resolution to be used in this kind of analysis. During non-vegetative periods, part of Class 6 areas is instead classified with a different label, thus explaining its absence in the results achieved.

Table 5. Results of LCZ classes effects on air temperatures.

Satellite	Period	Majority Class ID	Number of Buffers	Median °C	Standard Deviation °C
Landsat 8	Non-Vegetative	1	13	10.59	1.2
		2,3,4,5	23	9.60	1.5
		6	0	/	/
		7	12	8.98	0.8
Landsat 8	Vegetative	1	11	21.97	0.9
		2,3,4,5	23	21.46	1.1
		6	0	/	/
		7	14	20.84	0.7
Sentinel-2	Non-Vegetative	1	18	5.46	1.2
		2,3,4,5	14	4.68	1.5
		6	0	/	/
		7	16	3.93	0.9
Sentinel-2	Vegetative	1	12	19.15	1
		2,3,4,5	21	18.58	1.1
		6	2	16.69	/
		7	13	18.08	0.6
RapidEye	Non-Vegetative	1	12	11.02	1.7
		2,3,4,5	18	9.97	0.9
		6	0	/	/
		7	18	9.02	1.1
RapidEye	Vegetative	1	13	21.97	0.9
		2,3,4,5	17	21.30	1.2
		6	1	19.53	/
		7	17	20.90	0.8

5. Conclusions

The experiments performed in this work showed an underlying relationship between the land cover composition and the air temperature for the chosen study area. The analysis of the average temperature patterns provided also a metric to quantify this relationship in terms of observed temperature offsets between different LCZs.

Data availability was commensurate to the expectation of the analysis. The combined use of open data with a worldwide coverage, such as Landsat 8 and Sentinel-2 imagery, with traditionally available datasets such as air temperature observations, makes it possible to replicate the analysis in almost all major cities on the Earth. Nevertheless, it must be said that an enormous amount of high-accuracy, high-frequency and high-resolution environmental data which is potentially key for studying, understanding and solving climate-related issues is only managed by private companies, with no or limited access to both the research and public community. At least at the European level, some efforts are however underway towards the preparation of policies aiming to promote and facilitate the open release of private data [60]. The use of FOSS technology was an asset for both data processing and analysis. This proves the maturity, completeness, and reliability of such technology for geo-data management [61], since flexibility and customization (e.g., in data modeling) are combined with the sustainability in terms of technology costs.

Departing from the urban environment, the establishment of good practices in environmental policy-making by leveraging the use of available geo-data represents the ultimate goal of the presented work. The proposed analysis proved to be promising for all those applications requiring insights into the human impact on relevant climate-related variables such as the air temperature. These include urban planning and design, land policies impact assessment, and urban climate monitoring, among others.

References

1. Arrhenius, S. On the influence of carbonic acid in the air upon the temperature of the ground. *Lond. Edinb. Dublin Phil. Mag. J. Sci.* **1896**, *41*, 237–276. [CrossRef]
2. Nordhaus, W.D. *Managing the Global Commons: The Economics of Climate Change*; MIT press: Cambridge, MA, USA, 1994; Volume 31.
3. Stern, N.; Peters, S.; Bakhshi, V.; Bowen, A.; Cameron, C.; Catovsky, S.; Crane, D.; Cruickshank, S.; Dietz, S.; Edmonson, N.; et al. *Stern Review: The Economics of Climate Change*; Volume 30, 2006. Available online: http://webarchive.nationalarchives.gov.uk/20100407163608/http://www.hm-treasury.gov.uk/d/Summary_of_Conclusions.pdf (accessed on 27 October 2018).
4. United Nations. World Urbanization Prospects: The 2018 Revision. Available online: https://population.un.org/wup (accessed on 16 September 2018).
5. World Bank. Urban Development. Available online: http://www.worldbank.org/en/topic/urbandevelopment/overview (accessed on 16 September 2018).
6. UN-Habitat. *Cities and Climate Change: Global Report on Human Settlements 2011*; UN-Habitat, United Nations Human Settlements Programme: Nairobi, Kenya, 2011.
7. United Nations. Sustainable Development Goals. Available online: https://www.un.org/sustainabledevelopment (accessed on 16 September 2018).
8. Nicholas, S. Growth, Climate and Collaboration: Towards Agreement in Paris 2015. 2014. Available online: http://eprints.lse.ac.uk/64538/1/Growth_Climate_and_Collaboration_Stern_2014.pdf (accessed on 27 October 2018).
9. Faghmous, J.H.; Kumar, V. A big data guide to understanding climate change: The case for theory-guided data science. *Big Data* **2014**, *2*, 155–163. [CrossRef] [PubMed]
10. European Commission. Copernicus. Available online: http://www.copernicus.eu (accessed on 22 September 2018).
11. Foody, G.; Fritz, S.; Fonte, C.C.; Bastin, L.; Olteanu-Raimond, A.M.; Mooney, P.; See, L.; Liu, H.Y.; Rodriguez, A.; Minghini, M.; et al. Mapping and the Citizen Sensor. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 1–12.
12. Mooney, P.; Minghini, M. A review of OpenStreetMap data. In *Mapping and the Citizen Sensor*; Foody, G., See, L.; Fritz, S., Mooney, P., Olteanu-Raimond, A.M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 37–57.

13. Dutta, P.; Aoki, P.M.; Kumar, N.; Mainwaring, A.; Myers, C.; Willett, W.; Woodruff, A. Common sense: Participatory urban sensing using a network of handheld air quality monitors. In Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, Berkeley, CA, USA, 4–6 November 2009; pp. 349–350.

14. Maisonneuve, N.; Stevens, M.; Niessen, M.E.; Steels, L. NoiseTube: Measuring and mapping noise pollution with mobile phones. In *Information Technologies in Environmental Engineering*; Springer: Berlin, Germany, 2009; pp. 215–228.

15. Snik, F.; Rietjens, J.H.; Apituley, A.; Volten, H.; Mijling, B.; Di Noia, A.; Heikamp, S.; Heinsbroek, R.C.; Hasekamp, O.P.; Smit, J.M.; et al. Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters. *Geophys. Res. Lett.* **2014**, *41*, 7351–7358. [CrossRef]

16. Oke, T.R. City size and the urban heat island. *Atmos. Environ.* **1973**, *7*, 769–779. [CrossRef]

17. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900. [CrossRef]

18. Stewart, I.D.; Oke, T.R.; Krayenhoff, E.S. Evaluation of the 'local climate zone'scheme using temperature observations and model simulations. *Int. J. Climatol.* **2014**, *34*, 1062–1080. [CrossRef]

19. Long, N.; Gardes, T.; Hidalgo, J.; Masson, V.; Schoetter, R. Influence of the urban morphology on the urban heat island intensity: An approach based on the Local Climate Zone classification. *PeerJ* **2018**, *6*, 1–9.

20. Bechtel, B.; Daneke, C. Classification of local climate zones based on multiple earth observation data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1191. [CrossRef]

21. WUDAPT. World Urban Database. Available online: http://www.wudapt.org (accessed on 22 September 2018).

22. Bechtel, B.; Alexander, P.; Böhner, J.; Ching, J.; Conrad, O.; Feddema, J.; Mills, G.; See, L.; Stewart, I. Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 199–219. [CrossRef]

23. Dousset, B.; Gourmelon, F. Satellite multi-sensor data analysis of urban surface temperatures and landcover. *ISPRS J. Photogramm. Remote Sens.* **2003**, *58*, 43–54. [CrossRef]

24. Weng, Q.; Lu, D.; Schubring, J. Estimation of land surface temperature—Vegetation abundance relationship for urban heat island studies. *Remote Sens. Environ.* **2004**, *89*, 467–483. [CrossRef]

25. Kawashima, S.; Ishida, T.; Minomura, M.; Miwa, T. Relations between surface temperature and air temperature on a local scale during winter nights. *J. Appl. Meteorol.* **2000**, *39*, 1570–1579. [CrossRef]

26. Mutiibwa, D.; Strachan, S.; Albright, T. Land surface temperature and surface air temperature in complex terrain. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4762–4774. [CrossRef]

27. Nichol, J.E.; Fung, W.Y.; Lam, K.s.; Wong, M.S. Urban heat island diagnosis using ASTER satellite images and 'in situ'air temperature. *Atmos. Res.* **2009**, *94*, 276–284. [CrossRef]

28. Bacci, P.; Maugeri, M. The urban heat island of Milan. *Il Nuovo Cimento C* **1992**, *15*, 417–424. [CrossRef]

29. Köppen, W. Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet. *Meteorol. Zeitsch.* **1884**, *1*, 5–226.

30. Kopf, S.; Ha-Duong, M.; Hallegatte, S. Using maps of city analogues to display and interpret climate change scenarios and their uncertainty. *Nat. Hazards Earth Syst. Sci.* **2008**, *8*, 905–918. [CrossRef]

31. U.S. Geological Survey. EarthExplorer. Available online: https://earthexplorer.usgs.gov (accessed on 14 September 2018).

32. European Commission. Copernicus Open Access Hub. Available online: https://scihub.copernicus.eu (accessed on 14 September 2018).

33. Planet Labs Inc. Planet Explorer. Available online: https://www.planet.com/explorer (accessed on 14 September 2018).

34. The World in Weather Charts. Available online: http://www1.wetter3.de (accessed on 18 October 2018).

35. University of Wyoming—Atmospheric Soundings. Available online: http://weather.uwyo.edu/upperair/sounding (accessed on 18 October 2018).

36. TWC Product and Technology LLC. Weather Underground. Available online: https://www.wunderground.com (accessed on 14 September 2018).

37. Open Source Geospatial Foundation. QGIS Geographic Information System. Available online: http://qgis.osgeo.org (accessed on 14 September 2018).

38. McKinney, W. Pandas: A foundational Python library for data analysis and statistics. In Proceedings of the Python for High Performance and Scientific Computing, Seattle, WA, USA, 18 November 2011; pp. 1–9.

39. GeoPandas Developers. GeoPandas 0.4.0. Available online: http://geopandas.org (accessed on 14 September 2018).

40. Walt, S.v.d.; Colbert, S.C.; Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [CrossRef]

41. Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open Source Scientific Tools for Python. Available online: http://www.scipy.org (accessed on 14 September 2018).

42. Perry, M.T. Rasterstats. Available online: https://pythonhosted.org/rasterstats (accessed on 14 September 2018).

43. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling. Available online: https://CRAN.R-project.org/package=raster (accessed on 10 July 2018).

44. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.

45. Congedo, L. Semi-Automatic Classification Plugin Documentation. Available online: https://semiautomaticclassificationmanual.readthedocs.io (accessed on 14 September 2018).

46. Chavez, P.S., Jr. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.* **1988**, *24*, 459–479. [CrossRef]

47. GDAL/OGR contributors. GDAL/OGR Geospatial Data Abstraction Software Library. Available online: http://gdal.org (accessed on 14 September 2018).

48. Credali, M.; Fasolini, D.; Minnella, L.; Pedrazzini, L.; Peggion, M.; Pezzoli, S. Tools for territorial knowledge and government. In *Land Cover Changes in Lombardy Over the Last 50 Years*; ERSAF: Milan, Italy, 2011; pp. 17–19.

49. Regione Lombardia. Database Topografico Regionale (DBTR). Available online: http://www.geoportale.regione.lombardia.it/download-dati (accessed on 14 September 2018).

50. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

51. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogram. Remote Sens.* **2012**, *67*, 93–104. [CrossRef]

52. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles And Practices*; CRC Press: Boca Raton, FL, USA, 2008.

53. Rozenstein, O.; Karnieli, A. Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Appli. Geogr.* **2011**, *31*, 533–544. [CrossRef]

54. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

55. ISPRA. Variazioni E Tendenze Degli Estremi Di Temperatura E Precipitazione in Italia. 2013. Available online: https://tinyurl.com/ycay4jaj (accessed on 14 September 2018).

56. Sapkota, B.B.; Liang, L. A multistep approach to classify full canopy and leafless trees in bottomland hardwoods using very high-resolution imagery. *J. Sustain. For.* **2018**, *37*, 339–356. [CrossRef]

57. Cressie, N. Spatial prediction and ordinary kriging. *Math. Geol.* **1988**, *20*, 405–421. [CrossRef]

58. Gräler, B.; Pebesma, E.; Heuvelink, G. Spatio-Temporal Interpolation using gstat. *R J.* **2016**, *8*, 204–218.

59. Arnfield, A.J. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **2003**, *23*, 1–26. [CrossRef]

60. De Groof, H. Remarks at the Closing Plenary of the INSPIRE 2018 Conference. In Proceedings of the INSPIRE 2018 Conference, Antwerp, Belgium, 18–21 September 2018.

61. Brovelli, M.A.; Minghini, M.; Moreno-Sanchez, R.; Oliveira, R. Free and open source software for geospatial applications (FOSS4G) to support Future Earth. *Int. J. Digit. Earth* **2017**, *10*, 386–404. [CrossRef]