A combined Remote Sensing and GIS-based method for Local Climate Zone mapping using PRISMA and Sentinel-2 imagery

Alberto Vavassori [a,*], Daniele Oxoli [a], Giovanna Venuti [a], Maria Antonia Brovelli [a], Mario Siciliani de Cumis [b], Patrizia Sacco [b], Deodato Tapete [b]

[a] *Department of Civil and Environmental Engineering, Politecnico di Milano, Milano 20133, Italy*
[b] *Italian Space Agency (ASI), Rome 00133, Italy*

A B S T R A C T

In the last decade, several methods have been developed for Local Climate Zone (LCZ) mapping, encompassing Remote Sensing and Geographic Information Systems (GIS) −based procedures. Combined approaches have also been proposed to compensate for intrinsic limitations that characterized their separate application. Recent work has disclosed the potential of hyperspectral satellite imagery for improving LCZ identification. However, the use of hyperspectral data for LCZ mapping is yet to be fully unfolded. A combined Remote Sensing and GIS-based method for LCZ mapping is proposed to exploit the integration of hyperspectral PRISMA and multispectral Sentinel-2 images with ancillary urban canopy parameter layers. Random Forest algorithm is applied to the feature sets to obtain the LCZ classification. The method is tested on the Metropolitan City of Milan (Italy), for the period from February to August 2023. A spectral separability analysis is carried out to investigate the improvement in LCZ identification using PRISMA in comparison to Sentinel-2 data, as well as improvements in LCZ spectral separability on PRISMA pan-sharpened images. The resulting maps' quality is evaluated by extracting accuracy metrics and performing inter-comparisons with maps computed from the LCZ Generator benchmark tool. Inter-comparisons yield promising results with a mean Overall Accuracy increase of 16% using PRISMA for each LCZ class. Furthermore, we find that PRISMA improves the detection of LCZs compared to Sentinel-2, with a mean Overall Accuracy increase of 5%, in line with the higher spectral separability of PRISMA spectral signatures computed on the training samples.

1. Introduction

United Nations Sustainable Development Goal (SDG) 11 points out the urgent need to make cities and urban areas safe, resilient, and sustainable. Cities are particularly sensitive to the negative effects of climate change since they gather people and infrastructure while hosting crucial economic activities (Kumar, 2021). The increasing frequency, intensity, and duration of heat waves, coupled with growing urbanization, are intensifying the urban heat island (UHI) effect globally (Ward et al., 2016; Almeida et al., 2021). The UHI occurs where temperature values are persistently higher in urban zones compared to the surrounding environment, due to physical reasons connected with building and asphalt heat-trapping, waste heat from factories, buildings, and vehicles, and reduced surface moisture of impervious surfaces. UHI affects human health and ecosystems while heightening the energy demand for air conditioning during summer (Vujovic et al., 2021).

In the last decades, the scientific community has developed physical and conceptual models to measure the intensity of this phenomenon, which is key for implementing evidence-based mitigation strategies. In this context, the Local Climate Zone (LCZ) model is a well-established classification system which divides the urban landscape into distinct area types based on the morphological characteristics and land cover composition of their surface, directly affecting air temperature at screen height (Stewart and Oke, 2012). The subdivision of urban landscapes into LCZs is well supported by in-situ, screen-height temperature observations and results of numerical modelling (Stewart et al., 2014). This explains why the LCZ system has been increasingly exploited in urban climatology studies (Huang et al., 2023).

Some standard methods are commonly leveraged to produce LCZ maps, namely Geographic Information System (GIS) and Remote Sensing (RS) based approaches, each one taking advantage of a wide range of geodata, e.g. satellite imagery, high-resolution land cover,

---

topographic databases, and meteorological sensor observations (Aslam and Rana, 2022). Regarding satellite images, optical multispectral and Synthetic-Aperture Radar (SAR) data, as well as a combination of the two, are largely exploited (Tavares et al., 2019). However, although a large interest has developed in novel technologies and hyperspectral satellite data for environment monitoring, the use of this kind of data for LCZ mapping is yet to be fully unfolded (Liang et al., 2023). Recent work has uncovered the potential of using hyperspectral satellite images for LCZ mapping, promising higher accuracy compared to multispectral data due to the more detailed description of the spectral properties of urban surfaces (Vavassori et al., 2023b). The combination of multispectral and hyperspectral images may improve the detection of urban surface features and thus the identification of LCZ. Nevertheless, managing heterogeneous satellite and geospatial data sources with different spatial and temporal resolutions, which is required for LCZ mapping, may not be a trivial task, especially for non-expert users. Furthermore, different approaches may be required to obtain maps with adequate spatial and temporal resolution, accessibility, and distribution format, depending on the different applications.

Given the above considerations, in this work we present the hybrid RS/GIS-based approach for LCZ mapping that was developed during the "Local Climate Zone & Open Data Cube" (LCZ-ODC) project, by integrating multispectral Sentinel-2 and hyperspectral PRISMA satellite data along with multiple geospatial layers for the description of the urban morphological characteristics (e.g., building height and density). The approach was designed and tested accounting for future implementation of output LCZ maps in urban applications by final users. This choice addressed the specific goal of "Innovation for Downstream Preparation for Science (I4DP_Science)" program, i.e. the program funded by the Italian Space Agency (ASI) to promote demonstration of novel scientific algorithms and satellite-based product workflows to support applications in the space downstream sector (Tapete and Coletta, 2022). The selected testbed is the Metropolitan City of Milan (Northern Italy), which is characterized by extensive urbanization and where territorial features (e.g., poor wind circulation) may foster the persistence of the UHI (Bacci et al., 1992; Pichierri et al., 2012). LCZ-ODC primarily aimed to produce multi-temporal LCZ maps and assess their correlation with air temperature observations, while addressing the user community needs by developing open and user-friendly application tools (e.g., software) and products. A key asset of the project was the use of free and open-source tools for data management, data processing, and software development, allowing for potential future improvements of the implemented tools.

2. State-of-art of LCZ mapping

Stewart and Oke (2012) conceived and formalized the LCZ concept to overcome the limitations of the basic urban–rural dichotomy, which is insufficient for investigating the UHI distribution and space–time patterns. Based on urban morphology and land cover composition, the LCZ classification system distinguishes 17 area types. Specifically, the urban landscape is classified into 10 built-up and 7 land cover categories based on surface structure (i.e., height and density of trees and buildings), and land cover (i.e., pervious and impervious materials). The physical properties of each zone essentially include geometric and surface cover features (e.g., Sky View Factor, Aspect Ratio, Building Surface Fraction, and Impervious Surface Fraction) as well as thermal properties (e.g., Surface Admittance, Albedo, and Anthropogenic Heat Output), whose values have been defined (Stewart and Oke, 2012) and can be measured or computed, thus providing a framework to determine LCZs based on physical and geometrical quantities.

The interest in LCZ mapping has significantly increased over the last decade. In particular, the development of the World Urban Database and Access Portal Tools (WUDAPT) protocol fostered the building of a large database, that can be freely exploited and enriched (Bechtel et al., 2015). In 2022, the first two global, freely available LCZ datasets with

100 m resolution were developed (Zhu et al., 2022; Demuzere et al., 2022), representing a benchmark for the user community (accessible from: https://mediatum.ub.tum.de/1633461; https://doi.org/10.5281/zenodo.6364594; last access 18/04/2024).

Several methodologies have been defined for LCZ mapping which can be grouped into three categories, namely RS-based, GIS-based, and combined methods. Between 2012 and 2021, most scientific publications dealt with RS-based approaches (66.2 %), followed by GIS-based (24.1 %) and combined strategies (9.6 %), as indicated in (Huang et al., 2023). RS-based methods take advantage of the supervised classification of satellite imagery (Bechtel et al., 2015), with classification being performed at pixel-level, object-level, or scene-level. In this context, the WUDAPT protocol defines a straightforward workflow relying on a pixel-based supervised classification of multispectral satellite imagery. The WUDAPT workflow suggests the selection of training samples with Google Earth, based on expert knowledge, and the classification of Landsat data through the Random Forest (RF) algorithm (Bechtel et al., 2015). The WUDAPT methodology may be easily applied by employing an online platform, the LCZ Generator, which simplifies the process by leveraging Google Earth Engine (GEE) cloud-based infrastructure (Demuzere et al., 2021), and it only requires the user to provide a training dataset. The classification leverages a fixed set of features extracted from Landsat 8, Sentinel-1, and Sentinel-2 data, and ancillary geospatial layers, as listed in (Demuzere et al., 2021), which cannot be changed by the user.

Landsat (8) (Puche et al., 2023) and Sentinel (1 and 2) data (Hu et al. 2018, Qiu et al. 2019) appear to be the most exploited. However, many other satellites have been used, including RapidEye (Oxoli et al., 2018), PALSAR-2 (Chen et al., 2021), and ASTER (Xu et al., 2017). Among the numerous classification algorithms employed for RS-based LCZ mapping, it is worthwhile mentioning the RF, naïve Bayes, support vector machine, and neural networks (Bechtel et al., 2016; Bechtel and Daneke, 2012). Two key features can be pointed out for RS-based methods. First, the spatial resolution of the map should not be too high. In fact, a pixel size around 100–150 m is an appropriate range considering typical urban block scales (Bechtel et al., 2015). Finally, training samples must be collected appropriately, avoiding class and spatial imbalances, that may negatively affect classification accuracy. Despite the good performance of RS-based methods, the need for ancillary geospatial data (e.g., the building database) for the description of urban morphology and the lack of a global high-quality training sample represent shortcomings to their application (Ma et al., 2021).

In this context, a few applications of hyperspectral data can also be found in the literature. Specifically, in (Liang et al., 2023), the combined use of Zhuhai-1 hyperspectral images, ALOS DEM data, and nighttime light Luojia-1 data has been tested, achieving an overall accuracy (OA) of 87 %. The LCZ classification performance of hyperspectral PRISMA and multispectral Sentinel-2 images has been compared in (Vavassori et al., 2023b), aiming to assess the improvements in accuracy using different combinations of bands and the building height layer. The results of this work showed that the two data types yield comparable values of accuracy, with a slight increase in the accuracy considering a proper number (to be evaluated for each case study) of PRISMA Principal Components (PCs), which represent uncorrelated, linear combinations of the original image bands, allowing us to retain a smaller set of variables, still preserving most of the information in the original band set.

On the other hand, GIS-based approaches aim to determine LCZs by computing their physical parameters from a wide range of geospatial data such as Digital Surface Models (DSMs), building data, and land cover. RS multispectral images are also used to compute surface material parameters, e.g. surface albedo (computed as the average of atmospherically corrected reflectance values over a certain spectral range) (Lelovics et al., 2014). In GIS-based methods, LCZ maps are typically produced at two basic spatial units, either parcel or grid units. The former refers to lot area polygons, urban islets, or urban blocks, with a

Table 1

Satellite imagery used for LCZ mapping. For PRISMA (cloud coverage always > 1.6%), the tile name is indicated. For Sentinel-2, T32TMR and T32TNR granules are used. Time is expressed in UTC.

PRISMA dataset	Sentinel-2 dataset
Acquisition Date and Time	Acquisition Date and Time
09–02-2023 10:24 a.m. [PRS_L2D_STD_20230209102412_20230209102416_0001]	10–02-2023 10:20 a.m.
22–03-2023 10:30 a.m. [PRS_L2D_STD_20230322103037_20230322103042_0001]	22–03-2023 10:16 a.m.
08–04-2023 10:24 a.m. [PRS_L2D_STD_20230408102359_20230408102403_0001]	26–04-2023 10:16 a.m.
17–06-2023 10:30 a.m. [PRS_L2D_STD_20230617103042_20230617103046_0001]	25–06-2023 10:16 a.m.
10–07-2023 10:27 a.m. [PRS_L2D_STD_20230710102719_20230710102724_0001]	10–07-2023 10:16 a.m.
08–08-2023 10:27 a.m. [PRS_L2D_STD_20230808102725_20230808102730_0001]	19–08-2023 10:16 a.m.

minimum size of 200–500 m (Stewart and Oke 2012). Conversely, the appropriate size of grid units is determined based on empirical knowledge (Geletič and Lehnert 2016), spatial autocorrelation of the building height (Zheng et al 2018), or by testing several possible sizes (Hu et al. 2019). Nonetheless, most studies proposed spatial sizes equal to or higher than 200 m (Cai et al., 2019; Hu et al., 2019; Zheng et al., 2018). The layers of LCZ parameters are finally combined using decision-making algorithms, such as standard or modified standard rule-based classifiers (Wang et al. 2018), or classification algorithms, e.g. the RF. GIS-based methods may appear as more suitable for LCZ mapping due to the direct computation of the physical properties of LCZs; however, as for the RS-based methods, the availability of necessary geospatial data, including building height, is not always guaranteed, which may prevent the application of these methods to several urban contexts.

Combined approaches can help in overcoming the above limitations. In practice, LCZs are preliminarily classified using an RS-based approach, and the classification is finally refined through a GIS-based approach. Alternatively, LCZ parameters are directly integrated into the classification of satellite imagery (Zhou et al. 2020). Other combined methods perform the classification of built-up and natural classes separately, using an RS-based method for identifying land cover types and a GIS-based method for classifying built-up types (Du et al. 2020).

Considering the intrinsic limitations characterizing the separate application of RS and GIS-based methodologies, the present work proposes a combined RS/GIS-based method and a procedural workflow taking advantage of a fixed feature set, namely the original Sentinel-2 bands and the first 10 PRISMA PCs, as suggested in (Vavassori et al., 2023b), in combination with multiple UCP layers. UCPs are integrated into the spectral information to account for the urban morphological characteristics and thus improve the distinction of LCZs. Also, the present study aims to assess the possible increase in LCZ spectral separability leveraging PRISMA data, as well as through PRISMA pan-sharpening.

3. Case study and data

The case study selected to test the implemented methodology is the Metropolitan City of Milan, sited in the Po River valley (Northern Italy). According to the Köppen-Geiger Classification, the area climate can be defined as warm temperate, fully humid with hot summer (Cfa zone). Owing to the increasing frequency of heatwaves across Europe (Christidis et al., 2015), maximum summer temperatures higher than 35 °C are becoming more and more frequent in this area (ARPA Lombardia, 2023). Also, the poor wind circulation and consequent air stagnation characterizing the Po River valley fosters the persistence of humid heat conditions during summer as well as the UHI effect (Bacci and Maugeri, 1992). In addition, the study area is densely built-up (~44 % of urbanized areas in the Metropolitan City and ~81 % in the City of Milan, in 2022) (ISPRA, 2023). Such environmental and urban peculiarities are also common to other cities in Italy and Southern Europe, which makes Milan a suitable testbed for the analysis.

The datasets exploited in the present work encompass satellite imagery and other geospatial layers as follows. Regarding satellite imagery, Table 1 lists the hyperspectral images collected over the Metropolitan City of Milan by means of PRISMA (PRecursore IperSpettrale della Missione Applicativa) mission, i.e. the cutting-edge ASI's hyperspectral optical system, based on a single small class spacecraft, flying on a frozen Sun-Synchronous Low Earth Orbit at 615 km altitude and equipped with electro-optical devices (Loizzo et al., 2019; Caporusso et al., 2020). Each PRISMA scene integrates a panchromatic image at 5 m spatial resolution and 239 spectral layers from Visible-NearInfraRed to Short-Wave InfraRed (total VNIR-SWIR range: 400–2500 nm) with Ground Sampling Distance (GSD) of 30 m over a standard image size of 30 km × 30 km. This PRISMA dataset capitalized existing archive images and ad hoc tasking made during the LCZ-ODC project to ensure coverage of winter, spring, and summer seasons of the same year 2023. This dataset allows for annual monitoring of LCZ maps and seasonal comparisons. Table 1 also highlights the spatio-temporal match with the closest in time Sentinel-2 images with the least cloud coverage that were sourced from the Copernicus SciHub catalogue. In each epoch, two adjacent frames of Sentinel-2 were needed to overlap with the PRISMA footprint.

Multiple regional and global geospatial layers were also exploited to compute the Urban Canopy Parameters (UCPs). Table 2 reports the dataset name, source, resolution, reference date, and the list of UCPs derived from it. The latest release of each dataset was considered. Additional details about the computation of the UCPs are provided in Section 4. Regarding the regional geo-topographic database, the building layer was solely used in the present work, providing crucial data on building distribution and height. Apart from the building layer of the geo-topographic database, whose availability and characteristics depend on the specific regional and urban context, the other layers listed in Table 2 either cover the European region (Imperviousness Density) or have global coverage (Canopy Height and ALOS DEM). This allows replicability of the computation to other geographic regions.

Table 2

Datasets used to compute the Urban Canopy Parameters (UCPs).

Dataset name	Scale/spatial resolution	Reference date	Source	UCPs
Geo-topographic database (building layer)	1:2000 (across urban surfaces)	2023 (last revision)	Lombardy Region Geoportal (https://www.geoportale.regione.lombardia.it/)	Building Heights, Building Surface Fraction
Copernicus Imperviousness Density (10m)	10m	2018	Copernicus Land Monitoring Service (https://land.copernicus.eu/en/products/high-resolution-layer-imperviousness)	Impervious Surface Fraction
ETH Global Sentinel-2 (10m) Canopy Height	10m	2020	Google Earth Engine Catalog (https://gee-community-catalog.org/projects/canopy/)	Tree Canopy Height
ALOS DSM: Global (30m) v3.2	30m	2006	Google Earth Engine Catalog (https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V3_2)	Sky View Factor

4. Methodology

The combined RS/GIS-based procedure relies on the supervised classification of PRISMA and Sentinel-2 imagery through the RF algorithm. RF was chosen since it has been the most widely used in pixel-level LCZ classification, as indicated in (Huang et al., 2023), and it is suggested within the WUDAPT protocol (Bechtel et al., 2015). The UCP layers are integrated as additional feature sets within the classification step, to account for the urban morphological characteristics and reasonably improve the detection of LCZs. The main steps of the procedure are depicted in Figure 1 and can be summarized as follows.

Step #1: Pre-processing of satellite imagery, namely Level-2A Sentinel-2 and Level-2D PRISMA products (Bottom-of-Atmosphere reflectance). Given the known georeferencing accuracy of PRISMA data (nominally better than 200 m), PRISMA images are co-registered to the corresponding (closest in time) Sentinel-2 image using the free and open-source Python package GeFolki, which implements a Lucas-Kanade Iterative algorithm based on optical flow computation (Brigot et al. 2016). The displacement field is estimated on the 575.49 nm wavelength band of PRISMA, considering the 560 nm wavelength band of Sentinel-2 (resampled to 30 m using bilinear interpolation) as a reference, and then applied to the VNIR and SWIR PRISMA cubes. A Principal Component Analysis (PCA) (Gewers et al., 2022) is applied to the pre-processed PRISMA image by taking advantage of the dedicated scikit-learn Python library. The PCA is performed for PRISMA data dimensionality reduction to lower the computational time of the following image classification step while preserving the uncorrelated information contained in the original PRISMA bands. Regarding Sentinel-2, bands from B02 to B07, B8A, B11, and B12 (provided at 20 m spatial resolution by Copernicus) are exploited.

Step #2: Computation of the UCPs, namely Sky View Factor, Impervious Surface Fraction, Building Surface Fraction, Tree Canopy Height, and Building Heights (reported in the Appendix). These layers are computed in QGIS (Muhammad et al. 2022) at two spatial resolutions (i.e., 20 m and 30 m) and used to classify both Sentinel-2 (at 20 m) and PRISMA (at 30 m) images. The Sky View Factor is computed from the the ALOS DSM, using the dedicated module in SAGA GIS (Conrad et al. 2015). The Impervious Surface Fraction is computed from the Copernicus Imperviousness Density layer, while the building layer of the regional geo-topographic database is used to compute both the Building Surface Fraction and the Building Heights. Finally, the Global Sentinel-2 Canopy Height layer is exploited to derive the Tree Canopy Height. The layers are normalized to values in the [0–1] range to make them comparable with the reflectance data. Additional details on the methodology followed for computing the UCP layers are provided within the Appendix (see Table 6). Note that, the UCP layers are computed only once, considering the short time interval of the satellite acquisitions exploited within the project (i.e. 7 months).

Step #3: Collection of training and testing samples for performing the classification and assessing the result quality. The following ancillary data are used for this purpose: the 30 m RGB PRISMA image (central wavelengths: R, 664.89 nm; G, 575.49 nm; B, 482.55 nm) for the land cover detection through visual interpretation, the 5 m panchromatic PRISMA image for refining the polygon boundaries, and the building height layer for discriminating the built-up classes. Different samples have been collected for each PRISMA acquisition date both to account for the seasonal evolution of vegetated areas (that are included, e.g., in Scattered and Dense Trees and Low Plants LCZs) and to adapt the distribution of samples to the extent of the PRISMA image covering the study area. The samples are selected in such a way that they are as equally distributed as possible in space and within the LCZ classes, and possibly containing a similar number of pixels. The same polygons are used for the selection of training and testing samples in the corresponding Sentinel-2 images and provided as input to the LCZ Generator. It is worthwhile mentioning that the training sample is the only input required to the LCZ Generator which leverages a fixed set of images and ancillary layers for the classification.

Step #4: Assessment of the LCZ class spectral separability, by computing the Jeffries-Matusita (JM) distance (Richards, 2013). The JM distance is asymptotic to 2 when spectral signatures are completely separable and tends to 0 when signatures are identical. Spectral separability is evaluated on the signatures of training samples. However, owing to the high correlation between the PRISMA bands, the covariance matrices appearing in the JM distance formula turn out to be singular. To mitigate this problem, the spectral signatures computed from PRISMA are sampled with 10-band steps.
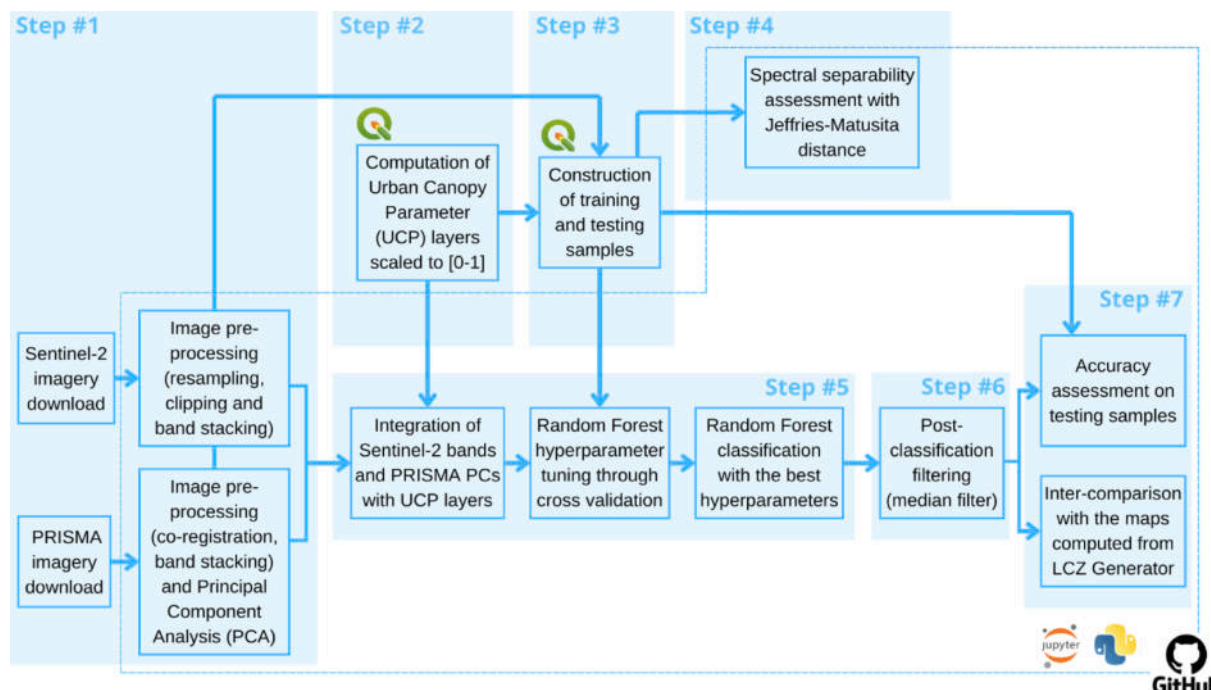

Fig. 1. Workflow of the proposed RS and GIS-based method for LCZ mapping.
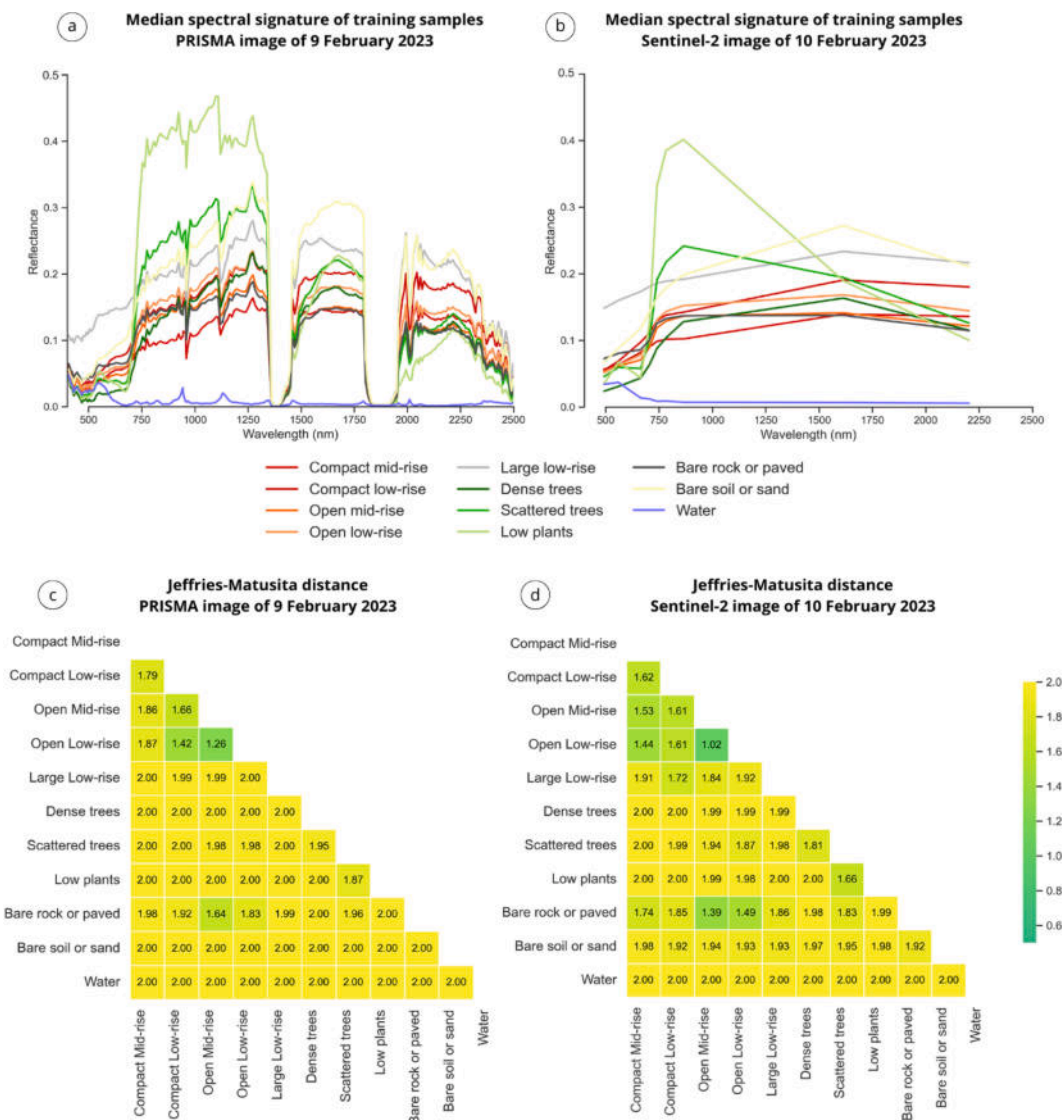
Fig. 2. Median spectral signature of the training samples and JM distance between the spectral signature of training samples, for each couple of LCZ classes. Spectral signatures and JM distance are computed on the 9th Feb PRISMA image (a and c) and on the 10th Feb Sentinel-2 image (b and d).

Step #5: Classification into LCZs through the RF algorithm, which is applied to each PRISMA and Sentinel-2 image using the same training samples. Regarding PRISMA, the algorithm is applied to the first 10 PCs, accounting for ~100 % of the variance in the original PRISMA bands, to speed up the computation and improve the classification performance. As for Sentinel-2, the algorithm is applied to all the selected bands listed above. The UCP layers are incorporated into the image bands (in the case of Sentinel-2) or PCs (in the case of PRISMA) as additional feature sets. Hyperparameter tuning is performed to select the combination of parameters yielding the highest accuracy score. To accomplish this task, several combinations of parameters are tested by splitting the training sample into training set (80 % of training sample) and validation set (20 % of training sample). Specifically, parameters are optimized through a repeated 5-fold cross-validation approach, based on the GridSearchCV class of the scikit-learn library. The parameters being tested are the number of estimators, the maximum number of features, and the criterion used to evaluate the quality of a split (between Gini impurity and entropy). The classification is finally carried out using the best combination of parameters.

Step #6: Application of post-processing median filter, with window size 3x3 pixels, to smooth the noisy output of the classification, by merging isolated LCZs into adjacent larger LCZ classes.

Step #7: Accuracy assessment of the LCZ map on the collected testing samples. Common accuracy metrics are derived from the confusion matrix, including OA, precision, recall, and F1-score. An inter-comparison is performed between the LCZ maps obtained leveraging PRISMA and Sentinel-2 images as well as the LCZ Generator (https://lcz-generator.rub.de) to assess the agreement of LCZ maps generated with different satellite sensors (i.e., PRISMA and Sentinel-2) and workflows (i.e., the LCZ-ODC project and LCZ Generator approaches). The maps are up-sampled to 10 m (common maximum resolution) by using a nearest-neighbour interpolator, and the LCZ Generator map is aligned to the common grid of PRISMA and Sentinel-2. The inter-comparison between each pair of maps is then performed by randomly picking pairs of corresponding (i.e., with the same coordinates) pixels on the two maps, following a stratified random sampling, thus selecting more pixels across more extended classes. The total number of pixels selected for each comparison is fixed to 1500 according to the Cochran's formula for large populations (Cochran, 1940). The maps being compared are obtained using the first 10 PRISMA PCs and the Sentinel-2 bands (along with the UCP layers), and the maps obtained with the LCZ Generator.

The procedure described above was applied to all the PRISMA and Sentinel-2 images listed in Table 1, resulting in multi-temporal (and multi-seasonal) LCZ maps, specifically 6 LCZ maps at 20 m resolution

(derived from Sentinel-2) and 6 LCZ maps at 30 m resolution (derived from PRISMA). Despite the different spatial extent of the resulting maps that depends on the footprint of the PRISMA acquisitions over the selected study area, the city of Milan is included in every LCZ map. For the sake of completeness, all LCZ maps are provided within the Appendix. In the following Section 5, a detailed description of the work outcomes is provided relatively to the LCZ maps computed for February and June 2023, which are representative of winter and summer time conditions, respectively.

A further experiment was carried out to investigate the effect of hyperspectral pan-sharpening on the spectral separability of LCZ classes. Specifically, the Adaptive Gram-Schmidt (GSA) decomposition method, based on Component Substitution (Loncan et al., 2015), was implemented and applied to the VNIR bands of one PRISMA image (acquired on 8th August 2023), encompassing the same spectral region as the panchromatic band (i.e. 400–700 nm). Although the image of 8th August was selected for running the test on pan-sharpening, the experiment can be easily replicated with the other PRISMA images listed in Table 1. The computation was restricted to the City of Milan area (~144 km^2 extent) and produced a super-resolved hyperspectral image, having 66 bands and 5 m resolution. The JM distance was computed on the spectral signature of training samples falling within the selected area,

before and after pan-sharpening. The spatial distribution of training and testing samples, relative to February, June, and August, is reported in the supplementary material.

The methodologies employed in this work were implemented through the creation of a collection of Python Jupyter notebooks (https://github.com/gisgeolab/LCZ-ODC/tree/Processing-Notebooks). These notebooks were designed to execute all computational steps described above. The procedures leverage exclusively free and open-source software and libraries by ensuring the reproducibility of outcomes. The entire codebase has been organized and made publicly accessible through GitHub (https://github.com/gisgeolab/LCZ-ODC) to ease future revision and replications behind the case study tackled in this paper.

To manage and integrate the diverse datasets, a unified endpoint built on the Open Data Cube (ODC) free and open-source data management platform was used (https://www.opendatacube.org). The ODC platform facilitated a cohesive approach to dataset handling. Satellite imagery, pre-processed and rendered analysis-ready, was registered within the ODC platform, together with the selected ancillary data. These datasets were programmatically accessed using the ODC API through the Jupyter notebooks.

Finally, to enhance the practical utility and accessibility of the
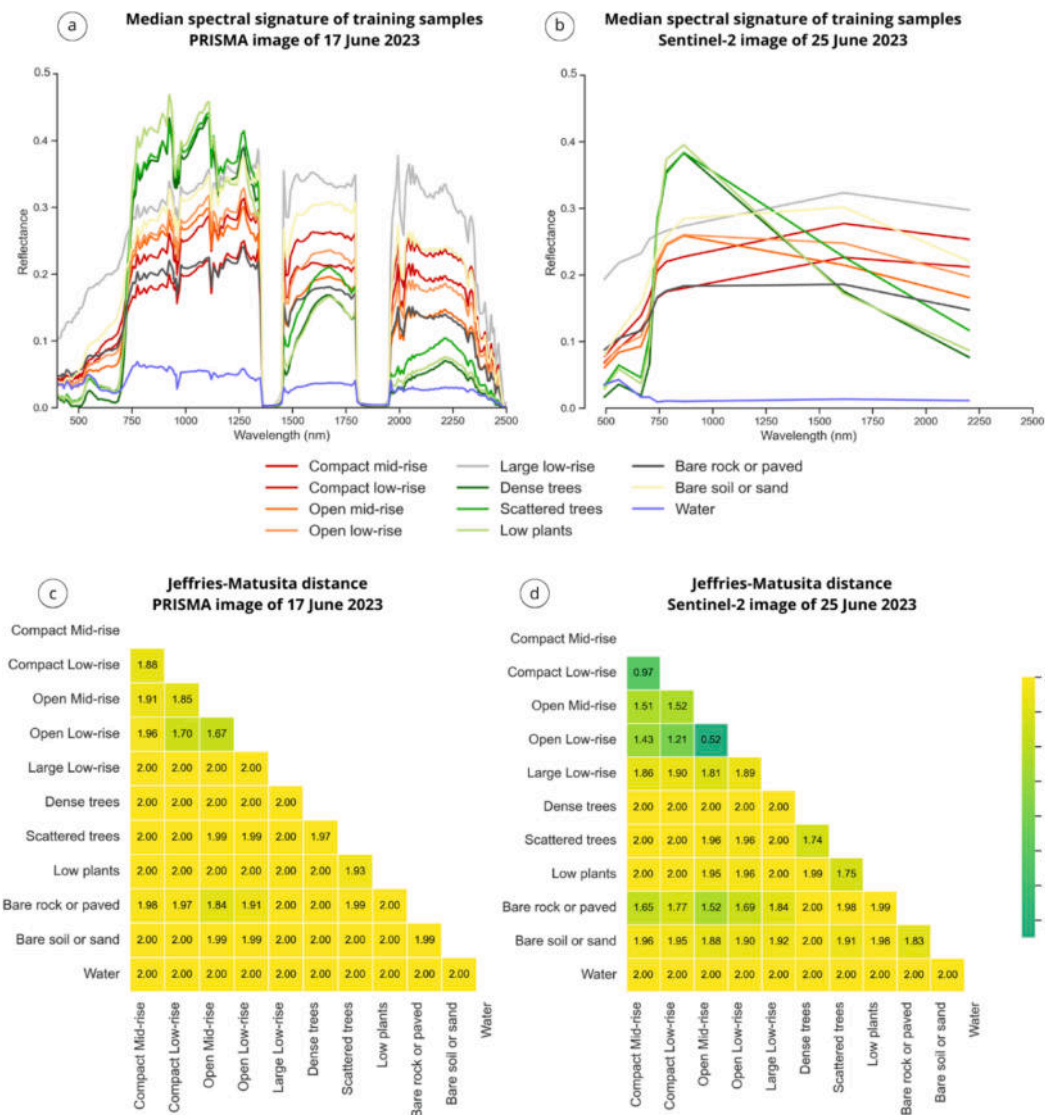
Fig. 3. Median spectral signature of the training samples and JM distance between the spectral signature of training samples, for each couple of LCZ classes. Spectral signatures and JM distance are computed on the 17th Jun PRISMA image (a and c) and on the 25th Jun Sentinel-2 image (b and d).

research findings, a demonstration version of the system was created and deployed using Docker technology. The source code was made available on GitHub (https://github.com/gisgeolab/LCZ-ODC/tree/Docker-ODC) for further replications of the system. This was instrumental to the dissemination and user engagement activities envisaged by the LCZ-ODC project. Such a project's activity is out of the scope of this paper. Further details were documented by the authors in separate publications (Vavassori et al. 2023a; Oxoli et al. 2023).

5. Results and discussion

5.1. Comparison of spectral separability and classification accuracy using PRISMA and Sentinel-2 bands

The median spectral signatures, computed from the training samples, and the spectral separability of LCZ classes are represented in Figure 2 (for February) and Figure 3 (for June). The values of JM distance (mostly close to 2) indicate a high spectral separability between the classes. However, two main considerations may be highlighted. Firstly, spectral separability is lower within the impervious/built-up LCZs. In fact, the distinction between built-up LCZ types is mostly linked to the urban canopy characteristics (e.g., building height and density) rather than the

spectral properties of their materials which are detected by optical satellite imagery. This is the reason why ancillary geospatial data describing urban morphology is needed to improve the distinction of built-up LCZs. Secondly, spectral separability between each pair of classes is higher for the spectral signatures computed from the PRISMA image, suggesting that the use of hyperspectral data can improve the distinction of LCZs.

Figure 4 represents the median spectral signatures (bounded to the VNIR range) of training samples for each LCZ class and the corresponding spectral separability, before and after the application of the GSA pan-sharpening algorithm. For all LCZs, the median spectral signatures are preserved after pan-sharpening, with very limited changes of reflectance (<0.02), meaning that a higher spatial resolution (5 m) was achieved without significant spectral distortion. This result points out the good performance of the GSA algorithm; however, a quantitative evaluation of the pan-sharpening quality has still to be carried out. Similar results (not reported in this paper) were obtained when applying the pan-sharpening algorithm to the February PRISMA image. The conservation of spectral signatures yields, in turn, similar values of JM distance (i.e. spectral separability of LCZs) when computed on the original and pan-sharpened PRISMA images.

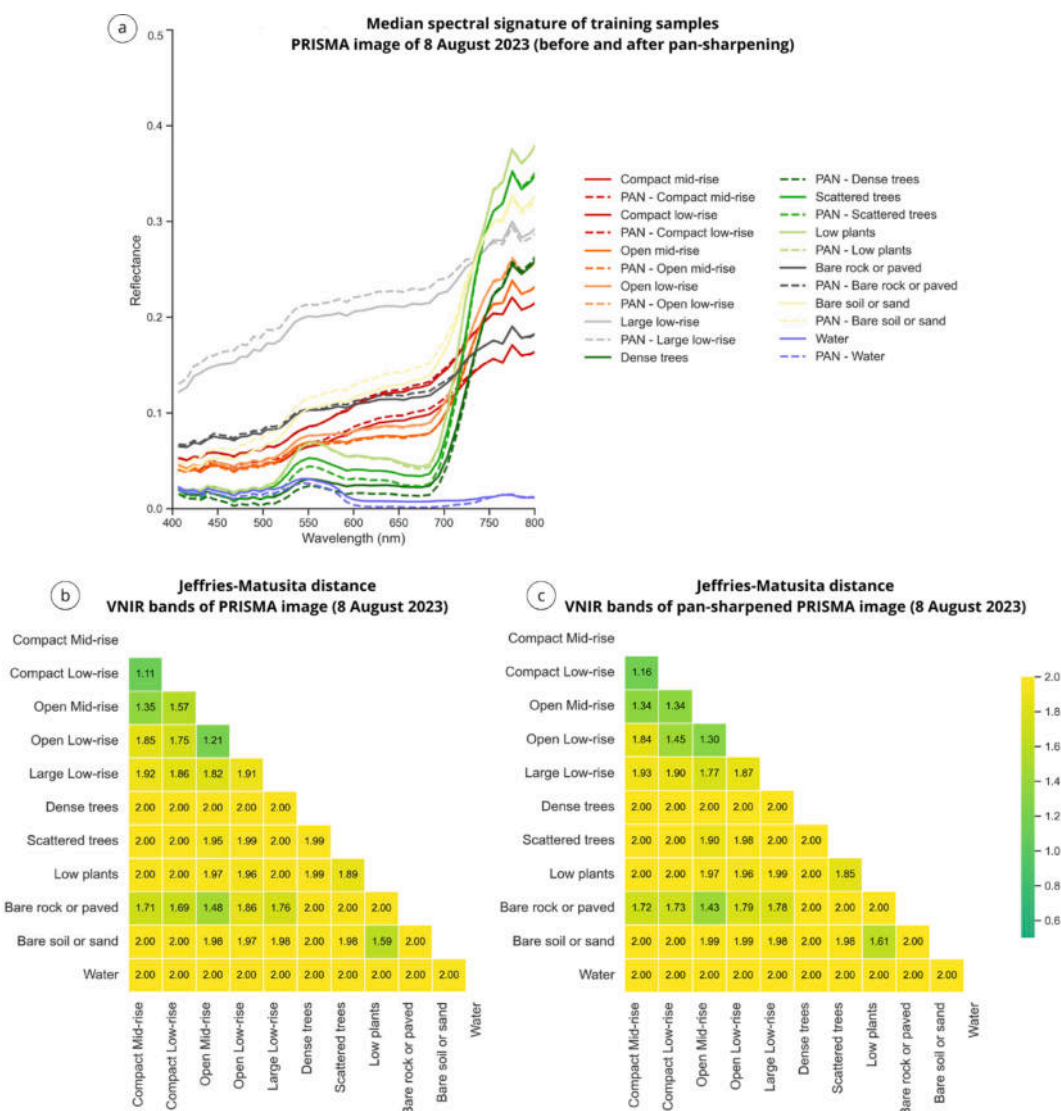Based on these findings, a first classification test was run on the 9th

Fig. 4. (a) Median spectral signature of the training samples and (b and c) JM distance between the spectral signature of training samples, for each couple of LCZ classes. Spectral signatures and JM distance are computed on the 8th Aug PRISMA image, (b) before and (c) after pan-sharpening.
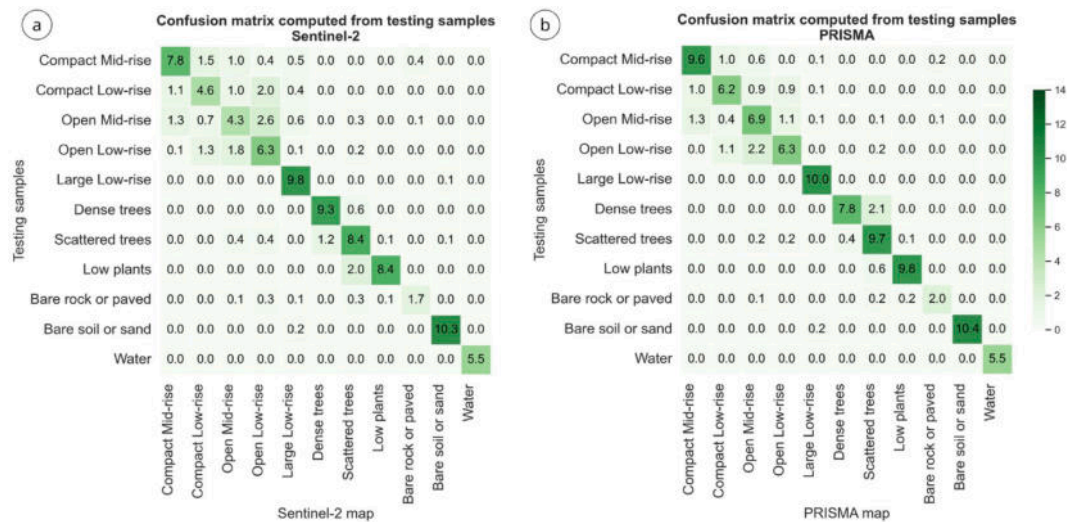
Fig. 5. Confusion matrices computed from the testing samples, relative to the LCZ maps computed from (a) the PRISMA image of 9th Feb, and (b) Sentinel-2 image of 10th Feb, without adding the UCPs to the spectral bands. Values are expressed as percentage of the total number of pixels in the testing sample, equal to 11,687 (computed on the map at 30 m).

Table 3
Precision, recall, F1-score, and Overall Accuracy (OA) of LCZ maps computed for June 2023, without adding the UCPs to the spectral bands.

February 2023	PRISMA			Sentinel-2		
	Precision	Recall	F1	Precision	Recall	F1
Compact Mid-rise	0.80	0.84	0.82	0.76	0.68	0.72
Compact Low-rise	0.71	0.68	0.70	0.56	0.51	0.53
Open Mid-rise	0.63	0.70	0.66	0.50	0.43	0.46
Open Low-rise	0.74	0.64	0.69	0.52	0.64	0.57
Large Low-rise	0.96	1.0	0.98	0.83	0.98	0.90
Dense Trees	0.95	0.79	0.86	0.88	0.94	0.91
Scattered Trees	0.75	0.92	0.83	0.72	0.80	0.76
Low Plants	0.97	0.94	0.96	0.98	0.81	0.89
Bare Rock or Paved	0.89	0.78	0.83	0.75	0.64	0.69
Bare Soil or Sand	1.0	0.98	0.99	0.98	0.97	0.97
Water	1.0	1.0	1.0	1.0	1.0	1.0
OA	0.84			0.76		

February PRISMA and 10th February Sentinel-2 images, without adding the UCP layers to the spectral information. The confusion matrices and corresponding accuracy metrics, computed on the testing samples, are provided in Figure 5 and Table 3, respectively. The comparison indicates the increase in accuracy using PRISMA, especially for the built-up LCZs and the Bare Rock or Paved class, as well as a lower confusion between the built-up classes, in line with the higher spectral separability of their signatures when computed on the hyperspectral data.

5.2. Comparison of classification accuracy using PRISMA and Sentinel-2 along with UCPs

The results described henceforth refer to the maps obtained by adding the UCP layers to the PRISMA PCs and Sentinel-2 spectral bands. Specifically, Figure 6 compares the LCZ maps computed from PRISMA (at 30 m), Sentinel-2 (at 20 m), and using the LCZ Generator (at 100 m resolution), for February and June. The quantitative evaluation of the product accuracy was performed by leveraging testing samples, collected upstream of the classification. The same testing sample was exploited to assess the quality of the LCZ maps computed with the two satellite sensors and the LCZ Generator. The confusion matrices are shown in Figure 7 whereas the corresponding accuracy metrics are reported in Table 4 (February) and Table 5 (June). Given the different resolutions of the maps, the values reported in the confusion matrices

are expressed as a percentage of the total number of pixels of the testing samples (i.e. relative frequencies).

The best combinations of hyperparameters, used for the RF classification, are reported in the supplementary material, along with the feature importance, expressed in terms of the mean decrease in impurity within each tree and mean decrease in accuracy. These indices suggest the UCP layers have a large importance in the classification, particularly the Imperviousness, Sky View Factor, and Canopy Height layers. It would be interesting to carry out a correlation analysis of the UCPs to interpret the results obtained out of the RF classification about the feature importance. This analysis is left for future work. Regarding the spectral information, the first 3 PRISMA PCs have higher importance than the other PCs, while the Sentinel-2 bands have a similar value of mean decrease in impurity and accuracy scores. These outcomes are corroborated by the comparison of accuracy metrics of LCZ maps obtained with (Table 4) and without (Table 3) the UCP layers, for February. Indeed, considering PRISMA, the UCPs yield a mean increase of F1-score equal to 0.08 for built-up classes (up to 0.16 for Open Mid-Rise) and 0.04 for land-cover classes (up to 0.13 for Bare Rock or Paved). The same happens for Sentinel-2, with a mean F1-score increase of 0.16 for built-up LCZs (up to 0.29 for Open Mid-Rise) and 0.05 for land-cover classes (up to 0.18 for Bare Rock or Paved). The smaller increase in accuracy recorded for the maps computed from PRISMA, after adding the UCP layers, is proof that the hyperspectral data partially compensates for the lack of ancillary layers describing urban morphological features, despite its lower spatial resolution.

The high values of the OA (>0.80) recorded for the maps computed from PRISMA and Sentinel-2, prove that the tested workflow provides very promising results compared to the LCZ Generator which achieves lower OA (0.70 and 0.75) on the same testing samples. This outcome holds for each LCZ class, for both February and June, indicating that the LCZ-ODC approach is generally better than the one implemented by the LCZ Generator in detecting both built-up and land cover LCZs. The most relevant improvement, in terms of F1-score, achieved using PRISMA (compared to the LCZ Generator) is for the classes Bare Rock or Paved (+0.48 for February and + 0.43 for June) and Bare Soil or Sand (+0.31 for February and + 0.52 for June). Nevertheless, an increase up to 0.21 is also recorded for the built-up classes (i.e. Large Low-rise, June). The relatively low F1-score of Bare Rock or Paved for the LCZ Generator maps, 0.48 and 0.52 respectively, may be attributed to the low spatial resolution of these maps which may negatively affect the detection of roads and railways.
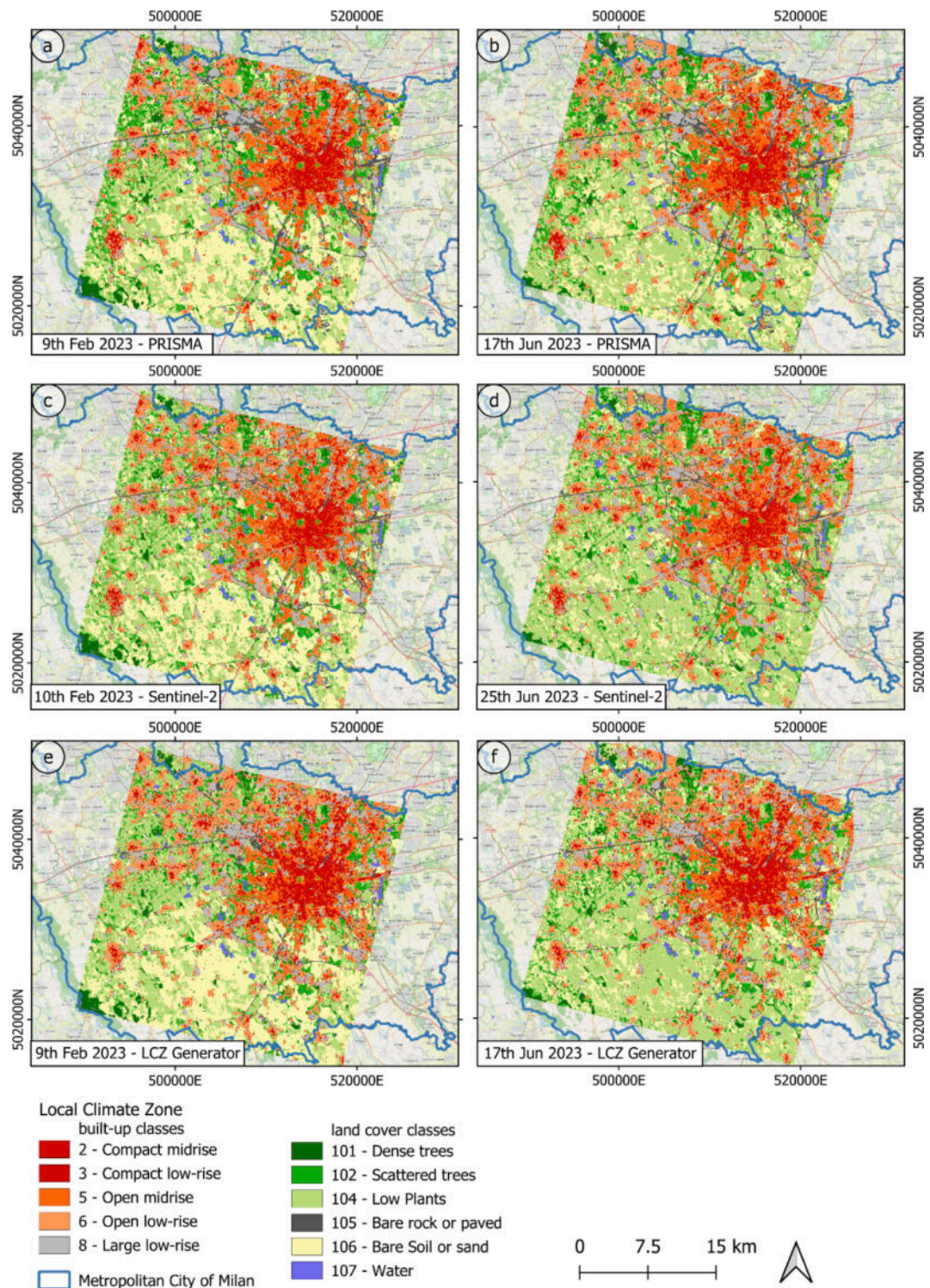
Fig. 6. LCZ maps obtained using the first 10 PRISMA PCs (30 m) and the Sentinel-2 bands (20 m) along with the UCP layers, and the LCZ Generator (100 m). LCZ maps are derived from (a) PRISMA image of 9th Feb, (b) PRISMA image of 17th Jun, (c) Sentinel-2 image of 10th Feb, (d) Sentinel-2 image of 25th Jun 2023, (e) LCZ Generator (training samples of 9th Feb), (f) LCZ Generator (training samples of 17th Jun).
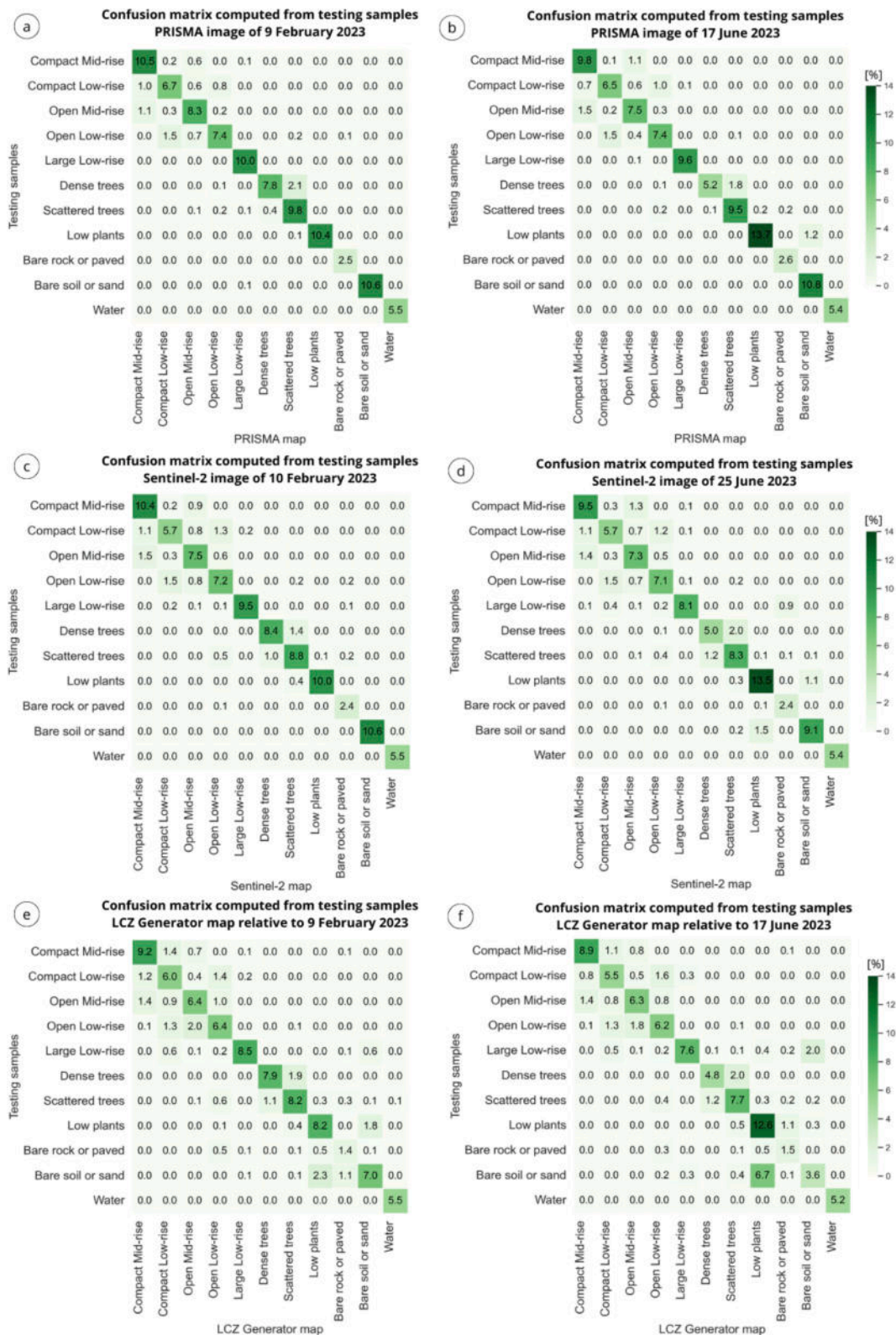
Fig. 7. Confusion matrices computed from the testing samples, relative to the LCZ maps computed from: (a) PRISMA image of 9th Feb, (b) PRISMA image of 17th Jun, (c) Sentinel-2 image of 10th Feb, (d) Sentinel-2 image of 25th Jun, (e) LCZ Generator (training samples of 9th Feb), (f) LCZ Generator (training samples of 17th Jun). Values are expressed as percentage of the total number of pixels in the testing sample, equal to 11,687 for February and 12,054 for June (computed on the map at 30 m).

Table 4
Precision, recall, F1-score, and Overall Accuracy (OA) of LCZ maps computed for February 2023.

February 2023	PRISMA			Sentinel-2			LCZ Generator		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Compact Mid-rise	0.84	0.92	0.88	0.80	0.90	0.85	0.77	0.80	0.79
Compact Low-rise	0.77	0.73	0.75	0.72	0.63	0.67	0.59	0.65	0.62
Open Mid-rise	0.80	0.84	0.82	0.74	0.76	0.75	0.66	0.65	0.66
Open Low-rise	0.86	0.75	0.80	0.73	0.72	0.73	0.63	0.65	0.64
Large Low-rise	0.98	1.00	0.99	0.96	0.95	0.96	0.95	0.85	0.90
Dense Trees	0.95	0.78	0.86	0.90	0.85	0.87	0.88	0.79	0.83
Scattered Trees	0.81	0.93	0.86	0.82	0.83	0.82	0.77	0.77	0.77
Low Plants	1.0	0.99	1.00	0.99	0.96	0.98	0.72	0.79	0.75
Bare Rock or Paved	0.94	0.98	0.96	0.82	0.92	0.87	0.45	0.52	0.48
Bare Soil or Sand	1.0	0.99	1.00	1.0	1.0	1.0	0.73	0.66	0.69
Water	1.0	1.00	1.00	0.99	1.0	1.0	0.99	0.99	0.99
OA	0.89			0.86			0.75		

Table 5
Precision, recall, F1-score, and Overall Accuracy (OA) of LCZ maps computed for June 2023.

June 2023	PRISMA			Sentinel-2			LCZ Generator		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Compact Mid-rise	0.82	0.89	0.85	0.78	0.85	0.81	0.79	0.82	0.80
Compact Low-rise	0.77	0.73	0.75	0.70	0.64	0.67	0.59	0.63	0.61
Open Mid-rise	0.77	0.78	0.78	0.72	0.76	0.74	0.67	0.67	0.67
Open Low-rise	0.82	0.78	0.80	0.74	0.73	0.74	0.64	0.66	0.65
Large Low-rise	0.98	0.99	0.99	0.96	0.83	0.89	0.92	0.68	0.78
Dense Trees	0.98	0.73	0.84	0.80	0.71	0.75	0.79	0.70	0.74
Scattered Trees	0.83	0.93	0.88	0.76	0.81	0.78	0.71	0.77	0.74
Low Plants	0.98	0.92	0.95	0.89	0.91	0.90	0.62	0.87	0.72
Bare Rock or Paved	0.93	0.97	0.95	0.69	0.94	0.80	0.45	0.60	0.52
Bare Soil or Sand	0.90	0.99	0.94	0.89	0.84	0.86	0.59	0.32	0.42
Water	1.0	1.0	1.0	0.99	1.0	1.0	1.0	0.99	0.99
OA	0.88			0.81			0.70		

Comparing the maps computed according to the LCZ-ODC workflow, higher accuracy is obtained using PRISMA (OA ~0.88), especially for June. This outcome is in good agreement with the results of the spectral separability analysis, which highlighted a higher separability between the spectral signatures computed with PRISMA (compared to Sentinel-2) for the June image (compared to the February image). Concerning this point, the following consideration can be pointed out. The lower accuracy of built-up LCZs is associated with a higher confusion between such classes (see the confusion matrix in Figure 7), which is in line with their lower spectral separability. Within the land cover types, a slightly higher confusion is found among the vegetated classes (Dense Trees, Scattered Trees, and Low Plants). Note that, the confusion is anyway reduced compared to the results obtained using the PRISMA and Sentinel-2 spectral bands only.

An inter-comparison among the LCZ maps was also carried out (Figure 8), providing further pieces of information about the result quality. In this case, the accuracy metrics that can be extracted from the confusion matrices, is a proxy of the agreement (or consistency) between the products.

Not surprisingly, a higher consistency was found between the maps generated with PRISMA and Sentinel-2 (OA equal to 0.79 and 0.74 for February and June, respectively) which were computed according to the same workflow and using the same ancillary geodata. In this case, lower F1-score is obtained for Dense Trees (0.62 for February and 0.50 for June) and Scattered Trees (0.64 and 0.65). The comparison between the PRISMA (or, equivalently, Sentinel-2) and LCZ Generator products results in a lower consistency (OA equal to 0.60 and 0.56 for February and June). This outcome is expected, given that the LCZ Generator exploits a different workflow, satellite images, and ancillary datasets for computing the LCZ maps. On the contrary, the same methodology and UCPs were leveraged for the classification of PRISMA and Sentinel-2 imagery into LCZs. Focusing on the inter-comparison between the

PRISMA and LCZ Generator maps computed for February, F1-scores lower than 0.50 were calculated for the classes Bare Rock or Paved (0.39) and Compact Low-Rise (0.47). Nonetheless, the maps computed for June showed low F1-score for Bare Rock or Paved (0.33) and Bare Soil or Sand (0.36). To understand these results, the differences between the maps generated with PRISMA, Sentinel-2, and LCZ Generator were computed (and reported in the supplementary material) for the classes Bare Rock or Paved (both February and June), Compact Low-Rise (February), and Bare Soil or Sand (June), providing us with the following insights.

Regarding Bare Rock or Paved, it appears quite clearly that the differences between the maps are related to their spatial resolutions, thus confirming the hypothesis described above. In fact, the small differences between PRISMA and Sentinel-2 maps mostly appear at the boundaries of roads and railways. The LCZ Generator does not properly detect them due to the low spatial resolution or, in some cases, it confuses this class with Compact Low-Rise and Bare Soil or Sand. As for the class Compact Low-Rise, the main differences are around the City of Milan, where the dense urban texture is gradually substituted by lower, more sparse buildings, and a mixture of urban classes is present. Finally, large portions of the rural area south of Milan show disagreement between the PRISMA (or Sentinel-2) and LCZ Generator maps relatively to the class Bare Soil or Sand. The misalignment mostly appears with the class Low Plants, and the photointerpretation of the PRISMA image suggests that the PRISMA map is more reliable than the LCZ Generator map, which is also proved by the higher accuracy of the former on the testing samples. Furthermore, the differences between PRISMA and Sentinel-2 classifications for the class Bare Soil or Sand can be attributed to the temporal misalignment between the satellite images (8 days), during which crop fields have partially undergone grass cutting or harvesting.

A final remark is noteworthy. The visual interpretation of the spatial differences between the maps revealed that the disagreements mainly

Fig. 8. Confusion matrices representing the inter-comparison between the LCZ maps obtained with PRISMA, Sentinel-2, and LCZ Generator: (a) PRISMA/Sentinel-2 Feb, (b) PRISMA/Sentinel-2 Jun, (c) PRISMA/LCZ Generator Feb, (d) PRISMA/LCZ Generator Jun, (e) Sentinel-2/LCZ Generator Feb, (f) Sentinel-2/LCZ Generator Jun. Values are expressed as percentage of the total number of pixels (i.e., 1500).

occurred where a unique LCZ type is difficult to define even through an expert visual inspection. This primarily happens in zones constituted by a mixture of regularly distributed vegetation (both grass and trees) and concrete or asphalt paths (occurring, e.g., in parking lots or graveyards), mainly due to the different sizes of these objects and thus impact on the capability of being resolved and spectrally separated from satellites. As a result, PRISMA classifies these areas as Bare Rock or Paved, despite the presence of vegetation, whereas Sentinel-2 tends to assign them to the class Open Low-Rise, even though buildings are generally not present. Therefore, misclassification errors can be attributed to a lack of information rather than to the classification algorithm. This is especially true for pixel-based classification algorithms (like the RF) that classify each pixel independently of the surrounding landscape features. Nevertheless, the accuracy assessment presented in this paper was carried out on testing samples corresponding to areas where the interpretation of the LCZ type was straightforward, allowing for a reliable evaluation of the map quality.

6. Conclusions

In this paper, we explore the contribution and possible benefits of hyperspectral PRISMA images to the generation of LCZ maps. This required the definition of a complete workflow, from the pre-processing steps to the validation of the obtained results, which was implemented through free and open-source software. The main conclusions can be outlined as follows: (i) the hyperspectral PRISMA data improves the accuracy of the resulting LCZ map compared to multispectral Sentinel-2 data thanks to a better separability of the LCZ classes' spectral signatures; (ii) further improvements can still be obtained by integrating the UCP layers to the spectral information through a hybrid RS/GIS-based approach; (iii) PRISMA pan-sharpening does not significantly affect the spectral separability of LCZ classes, thus opening the possibility to exploit the resulting higher spatial resolution for LCZ mapping applications; (iv) the proposed hybrid approach proves to be more accurate than the state-of-art workflow implemented within the LCZ Generator tool.

More specifically, the proposed hybrid approach is based on the combined exploitation of hyperspectral PRISMA or multispectral Sentinel-2 imagery and geospatial layers for the computation of the UCPs, which are normalized and integrated into the PCs (in the case of PRISMA) or satellite bands (in the case of Sentinel-2) as additional feature bands.

The presented experiments primarily focus on the LCZ maps computed for February and June 2023 (representative of winter and summer time conditions, respectively) in the Metropolitan City of Milan. We carried out a preliminary spectral separability analysis by computing the JM distance on the median spectral signatures (of the training samples) for each pair of classes. This experiment disclosed the general improvement of spectral separability provided by PRISMA with respect to Sentinel-2 imagery which is reflected in the better class-specific accuracy scores obtained from the classification of PRISMA bands. The lower spectral separability and the higher confusion of built-up LCZs justify the integration of the UCP layers into the classification step to improve their distinction. A further test was performed with the aim of evaluating the effect of PRISMA image pan-sharpening (through the GSA algorithm) on the spectral separability, revealing that spectral signatures are preserved after pan-sharpening, thus qualitatively proving the good performance of the algorithm. The quantitative assessment of pan-sharpening accuracy is left as a future development of this work.

We also provide a thorough evaluation of the result quality by assessing the map's accuracy on a set of testing samples and through an inter-comparison with the maps computed using the LCZ Generator tool. The results clearly indicate that the LCZ-ODC approach yields more accurate maps compared to the LCZ Generator, both in terms of overall and class-specific accuracies, thus suggesting that the proposed

workflow is better in detecting both built-up and land cover LCZ types. Focusing on the maps computed with the LCZ-ODC workflow, higher accuracy is obtained using PRISMA (with respect to Sentinel-2). The UCP layers contribute to reducing the confusion among the built-up LCZs, especially when integrated into the Sentinel-2 bands, indicating that the hyperspectral data partially compensates for the lack of ancillary layers describing the urban morphological features. Finally, the differences between the maps can be attributed to either actual misclassifications (and, in some cases, to the different spatial resolution) or the presence of "mixed" classes which are difficult to interpret even through an expert visual inspection.

In line with the purposes of the LCZ-ODC project, a relevant outcome of the work consists in the development of completely free and open-source software that can be directly exploited, allowing any potential user to process raw PRISMA and Sentinel-2 images, analyse spectral signatures (in terms of separability and statistics), apply a PCA, perform an LCZ classification and assess the accuracy of the generated maps. Therefore, the user can improve each step of the workflow, and tailor the processing to different case studies.

Besides the obtained results, some open issues may be highlighted and put off for future work. These include (i) the investigation of possible benefits of object-based classification algorithms for dealing with hard-to-define LCZ types, and (ii) experiments on the scalability of the proposed procedure, by applying it to other case studies, e.g., cities with significantly different urban structures compared to Milan.

CRediT authorship contribution statement

Alberto Vavassori: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Daniele Oxoli:** Writing – original draft, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Giovanna Venuti:** Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Maria Antonia Brovelli:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Mario Siciliani de Cumis:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Patrizia Sacco:** Writing – review & editing, Supervision, Project administration, Data curation, Conceptualization. **Deodato Tapete:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

Data availability

The research presented in this paper is founded entirely on open data, except for PRISMA data, which is subject to a license from ASI Original PRISMA Product - © Italian Space Agency (ASI) - 2023.

Appendix

Table 6
Input data and processing steps for computing the UCP layers.

UCP	Source of the dataset used to compute the UCP	Processing steps in QGIS
Building Height	• Geo-topographic databas; layer "UN_VOL"; attribute "UN_VOL_AV". • Provider: Lombardy Region. • Data type: vector (geopackage). • Scale: 1:2000 (across urban areas).	• Conversion to raster (5m) (*). • Down-sampling to 20/30m using the function "r.resamp.stats" (aggregation method: average). • Values scaled to the range [0–1] with the raster calculator.
Tree Canopy Height	• ETH Global Sentinle-2 10m Canopy Height (2020); tile relative to the Milan area: N45E006. • Provider: Google Earth Engine (GEE). • Data type: raster. • Resolution: 10m (global coverage).	• Down-sampling to 20/30m using the function "r.resamp.stats" (aggregation method: average). • Values scaled to the range [0–1] with the raster calculator.
Sky View Factor	• ALOS DSM (Global 30m v3.2) • Provider: GEE. • Data type: raster. • Resolution: 30m (global coverage)	• ALOS DMS (30m) used as an input to the SAGA-GIS module "Sky View Factor". Input parameters (default): maximum search radius: 100; number of sectors: 16. • Values already returned in the range [0–1] by SAGA-GIS.
Impervious Surface Fraction	• Imperviousness Density 2018 (raster 10m), Europe, 3-yearly; tile relative to the Milan area: IMD_2018_010m_E42N24_03035_v020. • Provider: Copernicus Land Monitoring Service (CLMS). • Data type: raster (GeoTIFF). • Resolution: 10m (Europe).	• Down-sampling to 20/30m using the function "r.resamp.stats" (aggregation method: average). • Values divided by 100 with the raster calculator, to scale the range to [0–1].
Building Surface Fraction	• Building Height raster layer at 5m resolution (*).	• Reclassification to: "1″ (=building is present) for the pixels with height value >0; "0" (=building is absent) for the pixels with height values =0. • Down-sampling to 20/30m using the function "r.resamp.stats" (aggregation method: average). Note that, since the reclassified layer has values "0"/"1", computing the mean is equivalent to computing the percentage of "1" values, and thus the area percentage covered with buildings.
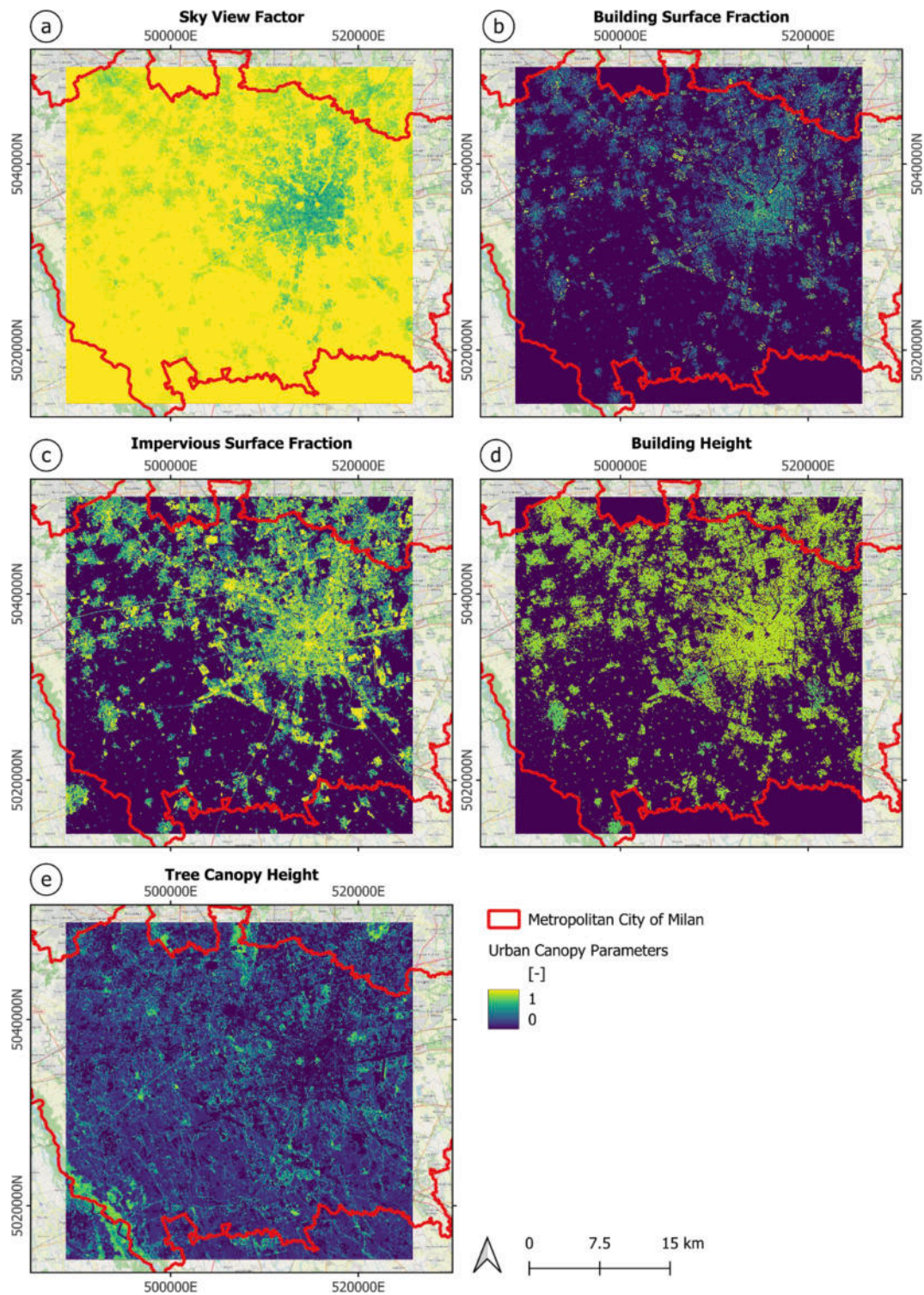
Fig. 9. Urban Canopy Parameters: (a) Sky View Factor, (b) Building Surface Fraction, (c) Impervious Surface Fraction, (d) Building Height, (e) Tree Canopy Height. Layers are clipped to the extent of the PRISMA image of 9th Feb. All parameters are normalized to values in the range [0–1].

Table 7

Optimum values of the Random Forest parameters: 'criterion' ('entropy', 'gini'), 'max_features' ('auto', 'log2', 'sqrt'), 'n_estimators' ('150', '200', '500'); accuracy resulting from the cross-validation: 'best_score'; accuracy resulting from the validation set: 'accuracy'.

| Satellite	PRISMA	Sentinel-2	PRISMA	Sentinel-2	PRISMA	Sentinel-2
Date	**9 Feb 2023**	**10 Feb 2023**	**22 Mar 2023**	**22 Mar 2023**	**8 Apr 2023**	**26 Apr 2023**
criterion	entropy	entropy	entropy	entropy	entropy	gini
max_features	auto	log2	auto	auto	auto	auto
n_estimators	100	500	500	500	500	200
best_score	0.938	0.884	0.944	0.900	0.942	0.889
accuracy	0.938	0.889	0.952	0.905	0.946	0.897
Date	**17 Jun 2023**	**25 Jun 2023**	**10 Jul 2023**	**10 Jul 2023**	**8 Aug 2023**	**19 Aug 2023**
criterion	entropy	gini	entropy	entropy	entropy	entropy
max_features	sqrt	sqrt	sqrt	log2	auto	sqrt
n_estimators	200	500	500	200	150	500
best_score	0.946	0.949	0.949	0.904	0.951	0.898
accuracy	0.946	0.953	0.953	0.901	0.958	0.905
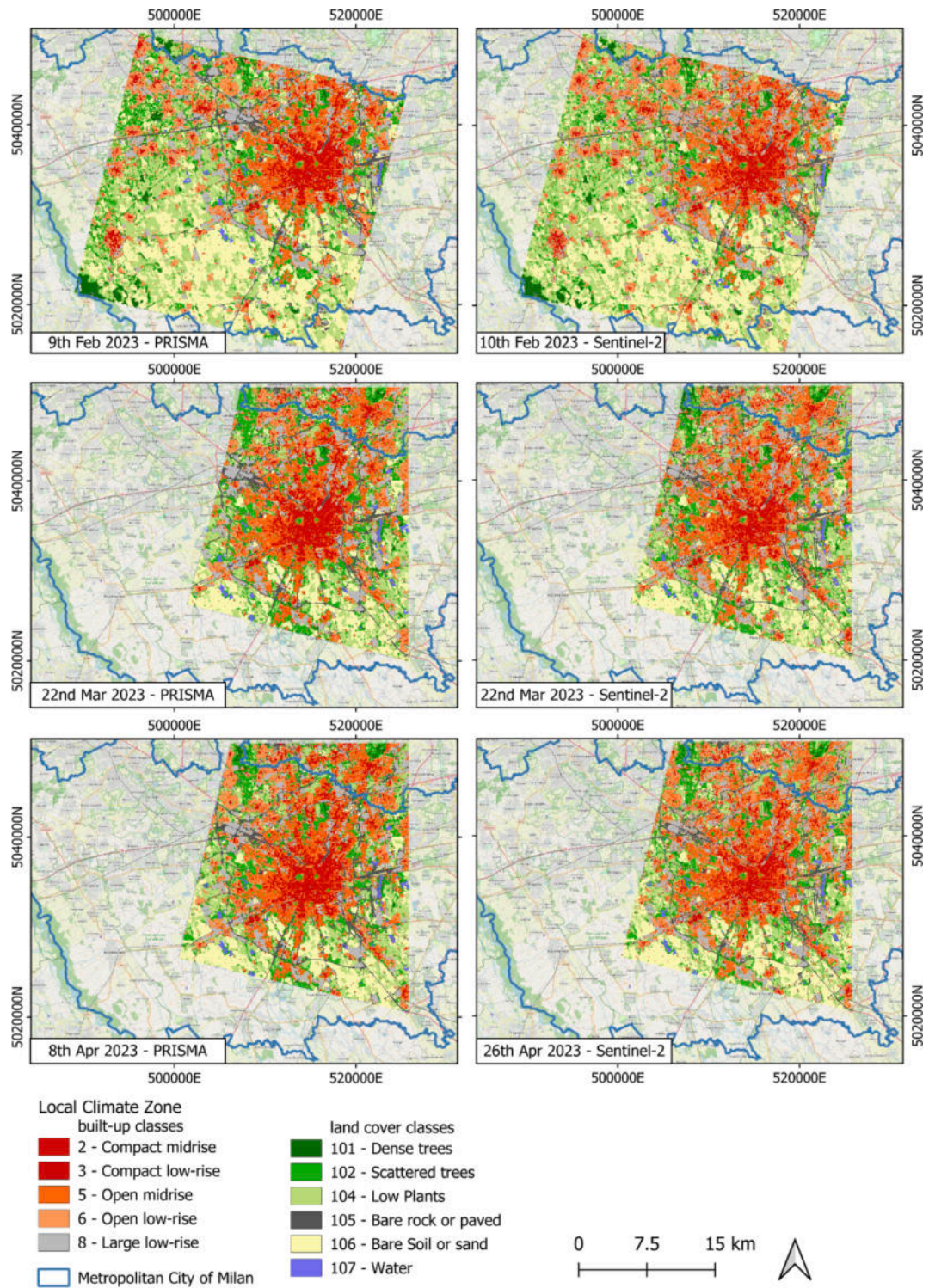
Fig. 10. LCZ maps obtained following the combined RS/GIS based approach of the LCZ-ODC project, using the first 10 PRISMA PCs (left panels) and the Sentinel-2 bands (right panels) along with the UCP layers.

Fig. 11. LCZ maps obtained following the combined RS/GIS based approach of the LCZ-ODC project, using the first 10 PRISMA PCs (left panels) and the Sentinel-2 bands (right panels), along with the UCP layers.
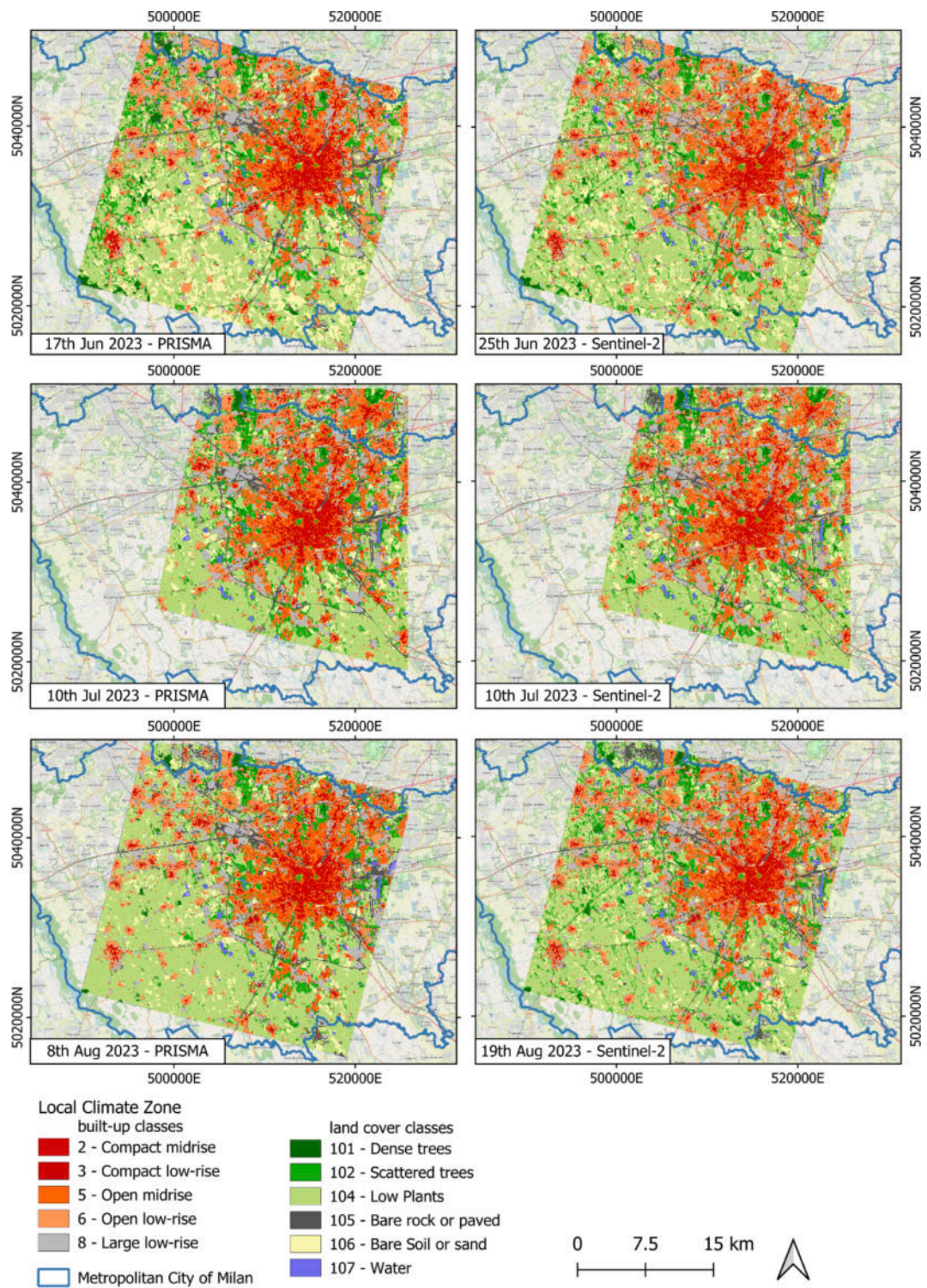
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2024.103944.

References

Almeida, C.R.D., Teodoro, A.C., Gonçalves, A., 2021. Study of the Urban Heat Island (UHI) Using Remote Sensing Data/Techniques: A Systematic Review. Environments 8 (10), 105. https://doi.org/10.3390/environments8100105.

ARPA Lombardia (2023) – Il clima in Lombardia. Retrieved January 29, 2024, from https://www.arpalombardia.it/temi-ambientali/meteo-e-clima/clima/il-clima-in-lombardia/.

Aslam, A., Rana, I.A., 2022. The use of local climate zones in the urban environment: A systematic review of data sources, methods, and themes. Urban Clim. 42, 101120 https://doi.org/10.1016/j.uclim.2022.101120.

Bacci, P., Maugeri, M., 1992. The urban heat island of Milan. Il Nuovo Cimento C 15 (4), 417–424. https://doi.org/10.1007/BF02511742.

Bechtel, B., Alexander, P., Böhner, J., Ching, J., Conrad, O., Feddema, J., Mills, G., See, L., Stewart, I., 2015. Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. ISPRS Int. J. Geo Inf. 4 (1), 199–219. https://doi.org/10.3390/ijgi4010199.

Bechtel, B., Daneke, C., 2012. Classification of Local Climate Zones Based on Multiple Earth Observation Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (4), 1191–1202. https://doi.org/10.1109/JSTARS.2012.2189873.

Bechtel, B., See, L., Mills, G., Foley, M., 2016. Classification of Local Climate Zones Using SAR and Multispectral Data in an Arid Environment. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (7), 3097–3105. https://doi.org/10.1109/JSTARS.2016.2531420.

Brigot, G., Colin-Koeniguer, E., Plyer, A., Janez, F., 2016. Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (7), 2923–2939. https://doi.org/10.1109/JSTARS.2016.2578362.

Cai, Z., Tang, Y., Chen, K., Han, G., 2019. Assessing the Heat Vulnerability of Different Local Climate Zones in the Old Areas of a Chinese Megacity. Sustainability 11 (7), 2032. https://doi.org/10.3390/su11072032.

Caporusso, G., Lopinto, E., Lorusso, R., Loizzo, R., Guarini, R., Darario, M.G., Sacco, P.. The Hyperspectral Prisma Mission in Operations. IGARSS 2020. https://doi.org/10.1109/IGARSS39084.2020.9323301.

Chen, C., Bagan, H., Xie, X., La, Y., Yamagata, Y., 2021. Combination of Sentinel-2 and PALSAR-2 for Local Climate Zone Classification: A Case Study of Nanchang. China. Remote Sensing 13 (10), 1902. https://doi.org/10.3390/rs13101902.

Christidis, N., Jones, G.S., Stott, P.A., 2015. Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. Nat. Clim. Chang. 5 (1), 46–50. https://doi.org/10.1038/nclimate2468.

Cochran, W. G. (1940). Note on an Approximate Formula for the Significance Levels of Z. In The Annals of Mathematical Statistics (Vol. 11, pp. 93–95). http://www.jstor.org/stable/2235972.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8 (7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Demuzere, M., Kittner, J., Bechtel, B., 2021. LCZ Generator: A Web Application to Create Local Climate Zone Maps. Front. Environ. Sci. 9, 637455 https://doi.org/10.3389/fenvs.2021.637455.

Demuzere, M., Kittner, J., Martilli, A., Mills, G., Moede, C., Stewart, I.D., Van Vliet, J., Bechtel, B., 2022. A global map of local climate zones to support earth system modelling and urban-scale environmental science. Earth Syst. Sci. Data 14 (8), 3835–3873. https://doi.org/10.5194/essd-14-3835-2022.

Du, P., Chen, J., Bai, X., Han, W., 2020. Understanding the seasonal variations of land surface temperature in Nanjing urban area based on local climate zone. Urban Clim. 33, 100657 https://doi.org/10.1016/j.uclim.2020.100657.

Geletič, J., Lehnert, M., 2016. GIS-based delineation of local climate zones: The case of medium-sized Central European cities. Moravian Geographical Reports 24 (3), 2–12. https://doi.org/10.1515/mgr-2016-0012.

Gewers, F.L., Ferreira, G.R., Arruda, H.F.D., Silva, F.N., Comin, C.H., Amancio, D.R., Costa, L.D.F., 2022. Principal Component Analysis: A Natural Approach to Data Exploration. ACM Comput. Surv. 54 (4), 1–34. https://doi.org/10.1145/3447755.

Hu, J., Ghamisi, P., Zhu, X., 2018. Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification. ISPRS Int. J. Geo Inf. 7 (9), 379. https://doi.org/10.3390/ijgi7090379.

Hu, J., Yang, Y., Pan, X., Zhu, X., Zhan, W., Wang, Y., Ma, W., Su, W., 2019. Analysis of the Spatial and Temporal Variations of Land Surface Temperature Based on Local Climate Zones: A Case Study in Nanjing, China. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (11), 4213–4223. https://doi.org/10.1109/JSTARS.2019.2926502.

Huang, F., Jiang, S., Zhan, W., Bechtel, B., Liu, Z., Demuzere, M., Huang, Y., Xu, Y., Ma, L., Xia, W., Quan, J., Jiang, L., Lai, J., Wang, C., Kong, F., Du, H., Miao, S., Chen, Y., Chen, J., 2023. Mapping local climate zones for cities: A large review. Remote Sens. Environ. 292, 113573 https://doi.org/10.1016/j.rse.2023.113573.

ISPRA (2023) – I dati sul consumo di suolo. Retrieved January 25, 2024, from https://www.isprambiente.gov.it/it/attivita/suolo-e-territorio/suolo/il-consumo-di-suolo/i-dati-sul-consumo-di-suolo.

Kumar, P., 2021. Climate Change and Cities: Challenges Ahead. Frontiers in Sustainable Cities 3, 645613. https://doi.org/10.3389/frsc.2021.645613.

Lelovics, E., Unger, J., Gál, C., Gál, C.V., 2014. Design of an urban monitoring network based on Local Climate Zone mapping and temperature pattern modelling. Climate Res. 60, 51–62. https://doi.org/10.3354/cr01220.

Liang, Y., Song, W., Cao, S., Du, M., 2023. Local Climate Zone Classification Using Daytime Zhuhai-1 Hyperspectral Imagery and Nighttime Light Data. Remote Sens. (Basel) 15 (13), 3351. https://doi.org/10.3390/rs15133351.

Loizzo, R., Daraio, M., Guarini, R., Longo, F., Lorusso, R., Dini, L., & Lopinto, E. (2019). Prisma Mission Status and Perspective. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 4503–4506. https://doi.org/10.1109/IGARSS.2019.8899272.

Loncan, L., De Almeida, L.B., Bioucas-Dias, J.M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G.A., Simoes, M., Tourneret, J.-Y., Veganzones, M.A., Vivone, G., Wei, Q., Yokoya, N., 2015. Hyperspectral Pansharpening: A Review. IEEE Geosci. Remote Sens. Mag. 3 (3), 27–46. https://doi.org/10.1109/MGRS.2015.2440094.

Ma, L., Zhu, X., Qiu, C., Blaschke, T., Li, M., 2021. Advances of Local Climate Zone Mapping and Its Practice Using Object-Based Image Analysis. Atmos. 12 (9), 1146. https://doi.org/10.3390/atmos12091146.

Muhammad, F., Xie, C., Vogel, J., Afshari, A., 2022. Inference of Local Climate Zones from GIS Data, and Comparison to WUDAPT Classification and Custom-Fit Clusters. Land 11 (5), 747. https://doi.org/10.3390/land11050747.

Oxoli, D., Ronchetti, G., Minghini, M., Molinari, M., Lotfian, M., Sona, G., Brovelli, M., 2018. Measuring Urban Land Cover Influence on Air Temperature through Multiple Geo-Data—The Case of Milan, Italy. ISPRS Int. J. Geo-Inform. 7 (11), 421. https://doi.org/10.3390/ijgi7110421.

Oxoli, D., Cedeno Jimenez, J.R., Capizzi, E., Brovelli, M.A., Siciliani De Cumis, M., Sacco, P., Tapete, D., 2023. QGIS AND OPEN DATA CUBE APPLICATIONS FOR LOCAL CLIMATE ZONES ANALYSIS LEVERAGING PRISMA HYPERSPECTRAL SATELLITE DATA. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-111-2023.

Pichierri, M., Bonafoni, S., Biondi, R., 2012. Satellite air temperature estimation for monitoring the canopy layer heat island of Milan. Remote Sens. Environ. 127, 130–138. https://doi.org/10.1016/j.rse.2012.08.025.

Puche, M., Vavassori, A., Brovelli, M., 2023. Insights into the Effect of Urban Morphology and Land Cover on Land Surface and Air Temperatures in the Metropolitan City of Milan (Italy) using Satellite Imagery and In Situ Measurements. Remote Sens. (Basel) 15 (3), 733. https://doi.org/10.3390/rs15030733.

Qiu, C., Mou, L., Schmitt, M., Zhu, X.X., 2019. Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network. ISPRS J. Photogramm. Remote Sens. 154, 151–162. https://doi.org/10.1016/j.isprsjprs.2019.05.004.

Richards, J.A., 2013. Remote Sensing Digital Image Analysis: An Introduction. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30062-2.

Stewart, I.D., Oke, T.R., 2012. Local Climate Zones for Urban Temperature Studies. Bull. Am. Meteorol. Soc. 93 (12), 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1.

Stewart, I.D., Oke, T.R., Krayenhoff, E.S., 2014. Evaluation of the 'local climate zone' scheme using temperature observations and model simulations. Int. J. Climatol. 34 (4), 1062–1080. https://doi.org/10.1002/joc.3746.

Tapete, D., Coletta, A. (2022). ASI's roadmap towards scientific downstream applications of satellite data. EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-5643. https://doi.org/10.5194/egusphere-egu22-5643.

Tavares, P., Beltrão, N., Guimarães, U., Teodoro, A., 2019. Integration of Sentinel-1 and Sentinel-2 for Classification and LULC Mapping in the Urban Area of Belém. Eastern Brazilian Amazon. Sensors 19 (5), 1140. https://doi.org/10.3390/s19051140.

Vavassori, A., Brovelli, M.A., Capizzi, E., Venuti, G., Betti, B., Siciliani De Cumis, M., Sacco, P., Tapete, D., 2023a. MAPPING LOCAL CLIMATE ZONES WITH MULTIPLE GEODATA AND THE OPEN DATA CUBE: INSIGHTS OF DOMAIN USER REQUIREMENTS AND OUTLOOKS OF THE LCZ-ODC PROJECT. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-581-2023.

Vavassori, A., Giuliani, G., Brovelli, M.A., 2023b. Mapping Local Climate Zones in Lausanne (Switzerland) with Sentinel-2 and PRISMA imagery: comparison of classification performance using different band combinations and building height data. Int. J. Digital Earth 16 (2), 4790–4810. https://doi.org/10.1080/17538947.2023.2283485.

Vujovic, S., Haddad, B., Karaky, H., Sebaibi, N., Boutouil, M., 2021. Urban Heat Island: Causes, Consequences, and Mitigation Measures with Emphasis on Reflective and Permeable Pavements. CivilEng 2 (2), 459–484. https://doi.org/10.3390/civileng2020026.

Wang, R., Ren, C., Xu, Y., Lau, K.-K.-L., Shi, Y., 2018. Mapping the local climate zones of urban areas by GIS-based and WUDAPT methods: A case study of Hong Kong. Urban Clim. 24, 567–576. https://doi.org/10.1016/j.uclim.2017.10.001.

Ward, K., Lauf, S., Kleinschmit, B., Endlicher, W., 2016. Heat waves and urban heat islands in Europe: A review of relevant drivers. Sci. Total Environ. 569–570, 527–539. https://doi.org/10.1016/j.scitotenv.2016.06.119.

Xu, Y., Ren, C., Cai, M., Edward, N.Y.Y., Wu, T., 2017. Classification of Local Climate Zones Using ASTER and Landsat Data for High-Density Cities. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (7), 3397–3405. https://doi.org/10.1109/JSTARS.2017.2683484.

Zheng, Y., Ren, C., Xu, Y., Wang, R., Ho, J., Lau, K., Ng, E., 2018. GIS-based mapping of Local Climate Zone in the high-density city of Hong Kong. Urban Clim. 24, 419–448. https://doi.org/10.1016/j.uclim.2017.05.008.

Zhou, X., Okaze, T., Ren, C., Cai, M., Ishida, Y., Mochida, A., 2020. Mapping local climate zones for a Japanese large city by an extended workflow of WUDAPT Level 0 method. Urban Clim. 33, 100660 https://doi.org/10.1016/j.uclim.2020.100660.

Zhu, X.X., Qiu, C., Hu, J., Shi, Y., Wang, Y., Schmitt, M., Taubenböck, H., 2022. The urban morphology on our planet – Global perspectives from space. Remote Sens. Environ. 269, 112794 https://doi.org/10.1016/j.rse.2021.112794.