

Article

# Remote Sensing and Data Mining Techniques for Assessing the Urban Fabric Vulnerability to Heat Waves and UHI

Flavio Borfecchia <sup>1,\*</sup>, Vittorio Rosato <sup>2</sup>, Emanuela Caiaffa <sup>2</sup>, Maurizio Pollino <sup>2</sup>, Luigi De Cecco <sup>1</sup>, Luigi La Porta <sup>2</sup>, Simone Ombuen <sup>3</sup>, Lorenzo Barbieri <sup>3</sup>, Federica Benelli <sup>3</sup>, Flavio Camerata <sup>3</sup>, Valeria Pellegrini <sup>3</sup> and Andrea Filpa <sup>3</sup>

<sup>1</sup> ENEA-SSTP-PROTER-OAC (Sustainability of Productive and Territorial systems Dep., Protection and valorization of territory and natural heritage Div., Earth & Climate Observation and Analysis Lab.); Via Anguillarese, 301, Rome 00123, Italy; luigi.dececco@enea.it

<sup>2</sup> ENEA DTE-SEN-APIC (DTE-Energy Technologies Dep. SEN-Smart Energy Div. APIC-Analysis and Protection of Critical Infrastructures Lab.), ENEA (Italian National Agency for New Technologies, Energy and Sustainable Economic Development), C.R. Casaccia, via Anguillarese, 301, Rome 00123, Italy; vittorio.rosato@enea.it (V.R.); emanuela.caiaffa@enea.it (E.C.); maurizio.pollino@enea.it (M.P.); luigi.laporta@enea.it (L.L.P.)

<sup>3</sup> Architecture Department, Roma Tre University, via della Madonna dei Monti, 40, Rome 00154, Italy; simone.ombuen@uniroma3.it (S.O.); lorenzo.barbieri@uniroma3.it (L.B.); federica.benelli@uniroma3.it (F.B.); flavio.camerata@uniroma3.it (F.C.); valeria.pellegrini@uniroma3.it (V.P.); andrea.filpa@uniroma3.it (A.F.)

\* Correspondence: flavio.borfecchia@enea.it; Tel.: +39-6-3048-6042

**Abstract:** Densely urbanized areas, with a low percentage of green vegetation, are highly exposed to Heat Waves (HW) which nowadays are increasing in terms of frequency and intensity also in the middle-latitude regions, due to ongoing Climate Change (CC). Their negative effects may combine with those of the UHI (Urban Heat Island), a local phenomenon where air temperatures in the compact built up cores of towns increase more than those in the surrounding rural areas, with significant impact on the quality of urban environment, on citizens health and energy consumption and transport, as it has occurred in the summer of 2003 on France and Italian central-northern areas. In this context this work aims at designing and developing a methodology based on aero-spatial remote sensing (EO) at medium-high resolution and most recent GIS techniques, for the extensive characterization of the urban fabric response to these climatic impacts related to the temperature within the general framework of supporting local and national strategies and policies of adaptation to CC. Due to its extension and variety of built-up typologies, the municipality of Rome was selected as test area for the methodology development and validation. First of all, we started by operating through photointerpretation of cartography at detailed scale (CTR 1: 5000) on a reference area consisting of a transect of about 5x20 km, extending from the downtown to the suburbs and including all the built-up classes of interest. The reference built-up vulnerability classes found inside the transect were then exploited as training areas to classify the entire territory of Rome municipality. To this end, the satellite EO HR (High Resolution) multispectral data, provided by the Landsat sensors were used within a on purpose developed "supervised" classification procedure, based on data mining and "object-classification" techniques. The classification results were then exploited for implementing a calibration method, based on a typical UHI temperature distribution, derived from MODIS satellite sensor LST (Land Surface Temperature) data of the summer 2003, to obtain an analytical expression of the vulnerability model, previously introduced on a semi-empirical basis.

**Keywords:** HR satellite remote sensing; urban fabric vulnerability; UHI & heat waves; landsat & MODIS sensors; LST & urban heating; segmentation & objects classification; data mining; feature extraction & selection; stepwise regression & model calibration

---

## 1. Introduction

During the last decades, despite the slow increase in monthly average temperatures and precipitation, many regions around the world have experienced extreme climate-related events like HW, heavy rainfall, droughts and fires, as most noticeable phenomena linked to CC [1]. The climate simulations reinforce the long-term prediction of the trends increase in the number and intensity of these phenomena related to extremes of climate variability [2, 3]. On cities such phenomena have a bigger impact not only for the highest concentration of inhabitants but also for their accentuation arising from alterations of the local thermal response and of the hydrological cycle, typical of many conurbations with high population densities, lack of green space and concentration of built-up and sealed areas, also in synergy with other impact factors (pollution, anthropogenic heat inputs, ...). Here as vegetated and evaporating soil surfaces are replaced by impervious, low albedo paving and building materials, the reduction in the latent heat flux and parallel increase in the sensible heat leads to the formation of the typical UHI's related to the spatial distribution of thermal response [4], with local maxima of LST (Land Surface Temperature) of the most dense areas of urban fabric rising several degrees over that of the rural surroundings [5, 6, 7, 8]. These concomitant factors may negatively affect the urban environment quality increasing at same time the impact of these climate extrema phenomena on citizen's life level and health, particularly in coincidence with HW, which are currently considered an emerging environmental health concern due to growing number of the linked fatalities [9, 6]. In fact the recent HW phenomena (i.e. in 2003, 2010 and 2015) are estimated to have resulted in an increasing number of victims, especially among the elderly and those suffering from specific diseases in various parts of the world and also in the countries of Europe bordering the Mediterranean. Here during the HW event of the summer of 2003, warmer than the past 500 years, for several days the average daily temperatures have been higher by various degrees compared with those of previous years, and the values in the months of July and August grew up to 10 ° above the average, with a maximum in France and in Italian central and northern areas[10]. These factors have had a major impact on the health of citizens in Europe, with increase in deaths (over 50,000) attributable to these extreme meteo-climatic events (Figure 1), concentrated mainly in the cities[11, 2]. In order to prevent and minimize these potential negative consequences of these extreme events related to the CC, there are ongoing activities of analysis and research to support interventions and policies concerning mitigation and adaptation in urban areas[12], even in the context of Smart Cities nationals and EU research programs.

In this context, the present research work aimed at developing methods for the morphological and typological urban settlements characterization related to these climate aspects associated to the temperature and hydrology (HW, UHI, water cycle, radiative balance,...). In particular the focus was on the evaluation of contribution to the local LST increase from typological, geometric and structural aspects of the city urban fabric whose development is subject to decision-makers policies and administrative planning in the context of local and national mitigation strategies and adaptation to CC [13]. In this context, since innovative spatially explicit approaches are increasingly required for obtaining suitable information and for assessing the expected impacts of these meteo-climatic factors [14], here the recent aero-spatial remote sensing techniques at medium (i.e. NASA MODIS dual satellite system with daily and night acquisition capability) and high (HR) ground resolution (i.e. Landsat and Sentinel 2 operated respectively by NASA and ESA within the European Copernicus program) integrated by GIS (Geographical Information Systems) and object analysis techniques were proposed for methodology development, with the goal of improving the standardization and operative aspects, repeatability as well as validation capability of implemented models. In particular, taking into account that on Rome the typical UHI LST distributions usually occur during the nights, in summer, as in other towns of Mediterranean basin [15, 16], the basic idea was to use the related TIR frames detected by MODIS sensors during the 2003 summer, in coincidence with the well-established HW situation [11], for supporting a physically based evaluation of the specific contribution from urban fabric built-up to the temperature spatial patterns. In this perspective the 30 m. Landsat HR (High Resolution) multispectral data were preliminary processed through object classification and data mining procedures to assess the distribution of built-up morpho-typologies

over the entire territory of the Rome municipality on the basis of training classes found within the transect area and characterized for their temperature response in the preliminary photointerpretation step.

Although the VHR (Very High Resolution) satellite (i.e. WorldView, QuickBird, RapidEye,...) remote sensing techniques (multispectral spatial resolution between 1 and 5 m.), working on commercial basis and based on user specific acquisition request, have been frequently employed to map the fine scale spectral heterogeneity of materials usually present within urban settlements [17, 18], due to their favorable spatial/spectral and operative features, also the 30 m. HR multispectral data, systematically provided by Landsat family satellite platforms and made freely available to the user under form of historical series starting since the early 80's, have been widely exploited in many other remote sensing applications for characterizing the metropolitan areas. For instance the Landsat multispectral data have been used for retrieving the urban LUCC (Land Use Cover Changes) classes and the concentration of impervious surfaces [19, 20], by means of vegetation indices or aerial photos [21, 22], while others authors dealt with supervised classification algorithms testing on urban areas using different approach including the per pixel and object oriented ones [18, 23].

The Landsat family satellite HR sensors are conceived for systematic acquisitions of reflectance data in seven acquisition bands from 180x180 Km. tiles over entire Earth surface with typical GSR (Ground Sampling Resolution) of 30 m. Starting from the ETM+ on board of Landsat 7 satellite launched on 2000, the GSR was improved by adding a panchromatic channel at 15 m. Thus in the implemented methodology developed on the basis of previous results [24, 25], these panchromatic data having higher GSR were used in synergy with the related multispectral ones with the goal of improving the detection capability of the finer scale textural features of urban fabric.

In the subsequent calibration/validation step a statistical analysis by means of regression modeling was carried out through the exploitation of LST distributions corresponding to know HW periods in summer 2003 (Figure 1) and the local nocturnal UHI maxima, frequently occurring over Rome [11], as told before. In such a way we were able to assess the impact of both these phenomena often acting in synergy to deteriorate the urban microclimate conditions. The HW and UHI phenomena affect the LST distribution which may be derived from thermal infrared (TIR) remotely sensed images having an important role urban studies due to the potentially large-area and repetitive nature of their coverage. Besides the local topography and impervious soil surfaces concentration [8], numerous RS-based UHI studies have suggested the dependence of spatial patterns of intra-city LST, from various attributes of urban area of interest like population [26], land use/cover [27], vegetation coverage [28], anthropogenic heat and urban fabric features [26]. The TIR data are acquired from different polar and geostationary platforms e.g. Landsat, Terra/Aqua MODIS, Terra, ASTER, NOAA AVHRR and Meteosat MVIRI. Although the Landsat sensors are able to detect only the LST diurnal distribution through the TIR acquisition channel (~10-12  $\mu\text{m}$ ) at 60 m. of GSR, according to similar applications [26, 29], in this study the Terra MODIS sensor was selected since it is able to provide nighttime data for studying the nocturnal LST distribution, typical of UHI of Rome and others European cities. MODIS LST products (MYD11A1-MODIS/Aqua Land Surface Temperature and Emissivity Daily L3 Global 1 km Grid SIN) even with lower spatial resolution (1 Km.), are particularly suitable for the LST product due their easy availability, global coverage and high calibration accuracy based on multiple

thermal and spectral bands for properly taking into account the atmospheric effects (split window water-vapor atmospheric correction) and emissivity of earth surfaces; furthermore validation of version 3 standard products from Terra MODIS data shows that their accuracy is better than 1  $^{\circ}\text{C}$  in the range from -10 to +50  $^{\circ}\text{C}$  [30, 31].

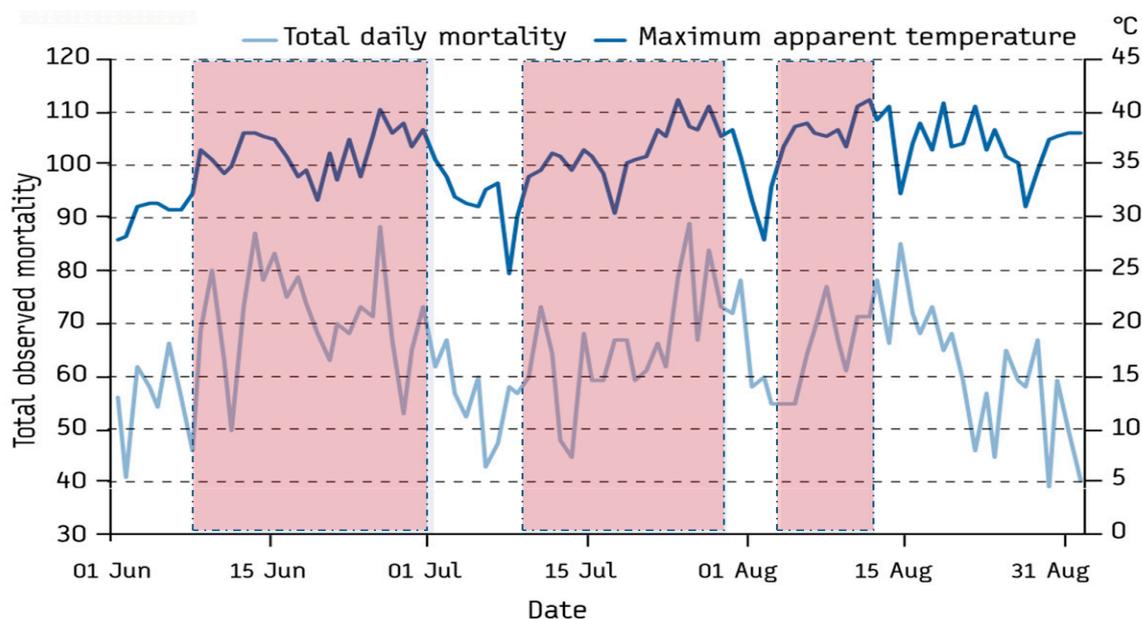


Figure 1 - Mortality and temperature trends in Rome area during the HW periods (reddish windows) of the summer 2003 (Michelozzi et al., 2005)

## 2. Materials and Methods

### 2.1. Area of interest

The Rome territory is located in the central Italy (Figure 2), characterized by Mediterranean climate and annual daily mean temperature of 20°, with its 1.287 Km<sup>2</sup> and about 2.875.000 inhabitants it represents the biggest municipality in terms of area and population at European level. Although the Rome municipality includes many protected and green areas, parks and rural landscapes, the huge concentration of historical monuments and cultural heritage coupled with pronounced urban sprawl of the recent decades and consequent widening of transportation network and infrastructures, often without an appropriate planning, increase the urbanized zones potentially subjected to impacts and pressures deriving from urban environment quality deterioration arising from ongoing CC and anthropogenic factors.

Although with an average maximum temperature above 30 °C, the Roman summer was already very hot, in the last decades it has undergone further transformations with a demonstrable increase in the frequency and intensity of HW and due to temperature levels (day and night) rising, consequent uncomfortable conditions of malaise for human health. The center of Rome is, roughly, 25 km from the Tyrrhenian coast. In summer, the moderating influence of the Tyrrhenian Sea is more noticeable on the western slopes of the city, thanks to the local characteristic wind from west to east inhibiting the excessive heat of summer afternoons and relieving discomfort. The situation is different at the center, only partially achieved by this air flow due to heavy urbanization, with temperatures which can record up to 3-4 °C higher than the western side. Even warmer is the eastern districts, which happens almost entirely without a suitable air circulation.

The average annual precipitation was around 692 mm to the end of the 80s. Current value rose to about 750 mm. During the summer, the combination of humidity and high temperatures combined with a modest infiltration of fresh air from the north is the spark that frequently bursts storms of intensity almost unknown in the past.

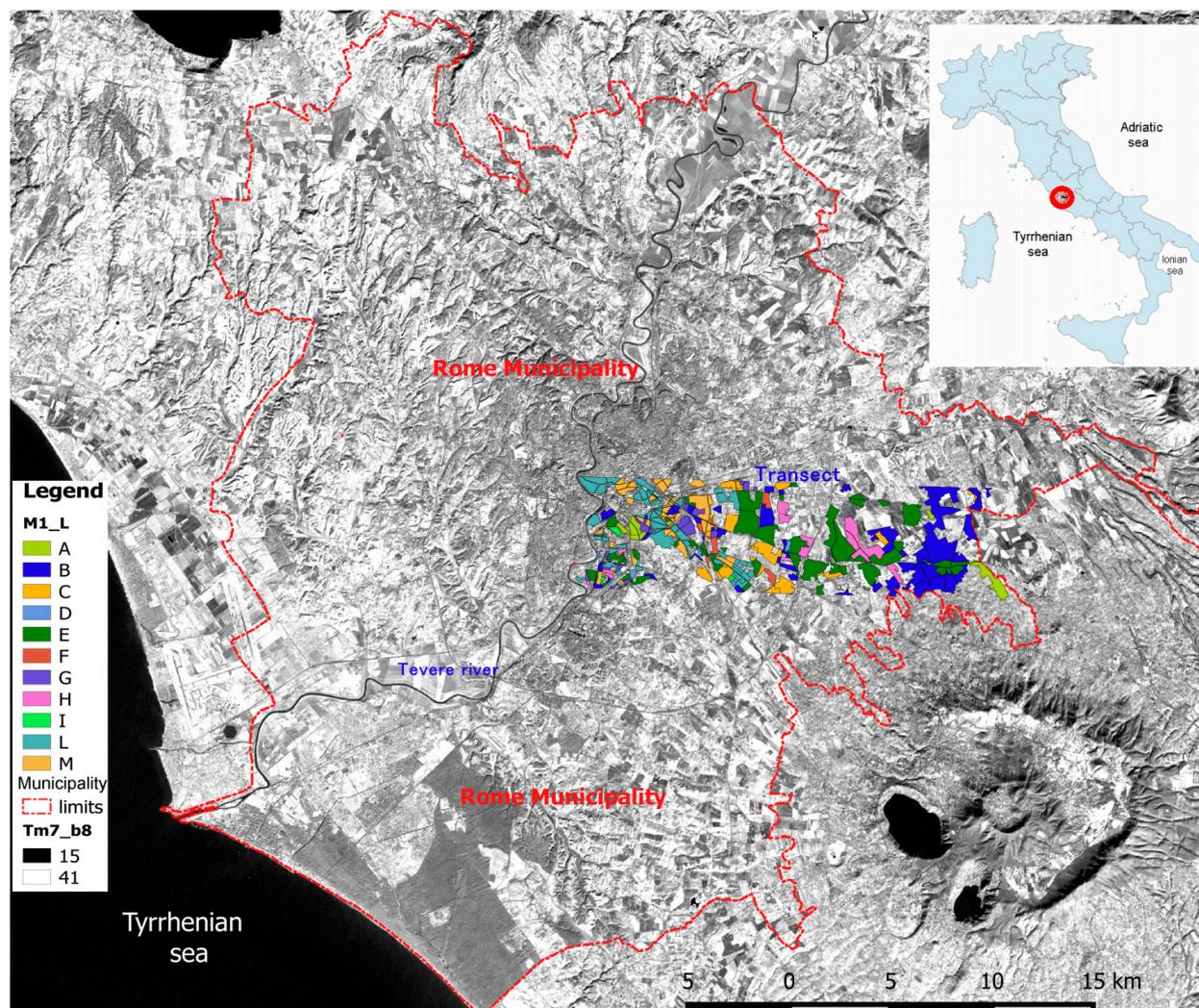


Figure 2 - Area of interest including the Rome municipality territory and M1-L built up classes (coloured polygonals) within the transect derived from photointerpretation methodology (see next chapter). The various vector features were superimposed to orto-rectified B/W panchromatic channel of the Landsat ETM+ frame exploited for object classification procedures.

## 2.2. Photointerpretation and training areas

The initial goal of the work was the photointerpretation-based characterization of the various areas of the transect, homogeneous from the urbanistic point of view and previously identified and delimited on the basis of the existing road network and cartography at appropriate scale (Figure 3). The work was accomplished by assigning to each polygonal area the quantitative attributes derived from the building type and density and from the vegetation/permeable surfaces presence, linked to their hydrological and thermal potential response to the HW and UHI. This approach aimed at the evaluation of a numerical index of vulnerability (NVI) for each of the areas, including the cumulative effect of these three main factors related to urban fabric features and estimated with their relative scores preliminarily introduced on a semi-empirical basis. Starting from preexisting knowledge and specific urbanistic studies on the urban areas of interest [13, 32, 33], three reference variables were defined: typology and compactness, related specifically to built-up portion of the area and the permeability of the surrounding surfaces. For each of them, on the basis of different relative weights, increasing degrees of partial vulnerability (scores) were assigned through photo interpretation methods. Finally, the NVI estimation for each area of the transect, was derived by

summarizing the partial contributions arising from the three variables scores. The 285 polygonal of urban residential areas found in the transect, previously identified using the road network, were thus characterized by photo-interpretation approach in terms of the prevalent buildings typology within 11 predefined types and related compactness level, as the number of floors and mean width/density of contiguous streets, in addition, their percentage of permeable surfaces, In order to assess the global thermal response as NVI (Numerical Vulnerability Index) of the urban fabric areas inside the transect the quantitative contribution of the three parameters was evaluated in term of intensity levels (score) within a fixed scale of a linear model defined by relative weights (Table 1) and empirically introduced to account for their specific impact in term of local temperature augmentation and related uncertainty:

1. predominant type of buildings within 11 types defined and characterized on the basis of previous studies and knowledge [32], grouped within 6 score classes as reported in Table 2;
2. compactness of the urban fabric, 4 classes: high, medium-high, medium-low, low, with relative scores as reported in Table 1;
3. presence of permeable surfaces, 3 classes: low, medium, high, with relative scores as shown in Table 1. vegetation included, was also visually estimated.

<b>Class n.</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>max score %</b>
<b>Typology</b>	2	4	6	8	10	20	20
<b>Compactness</b>	5	10	20	35			35
<b>Permeability</b>	1	20	45				45

Table 1- Urban fabric parameters and related score classes for NVI assessment

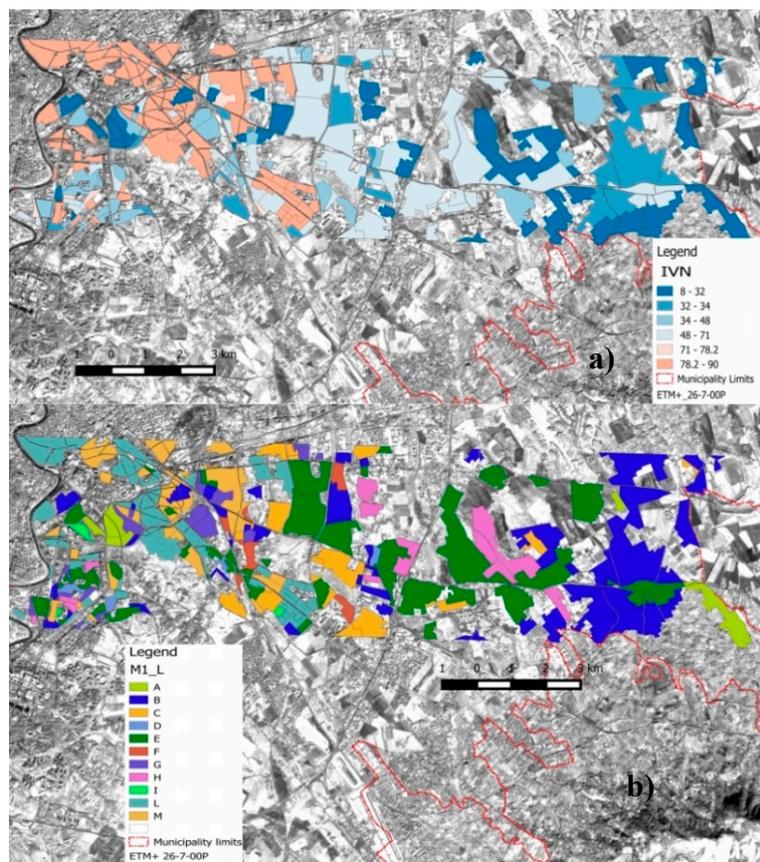
The 11 building typologies were associated with the relative photo-interpreted levels of compactness, for a total of 11 “morpho-types” (Table 3) classes (M1\_L) found in the transect (Figure 2 3), with a summed score in the range from 7 to 45, which synthesizes the specific partial thermal response of built-up and building fraction (Figure 3, b). Subsequently, for each area in the transect, the scores within the 11 morpho-type classes was added to the corresponding score referring to the permeability parameter, obtaining a distribution of NVI rightly increasing going from suburbs to the more vulnerable downtown zones (Figure 3, a). Thus the NVI was calculated according to a simple linear semi-empirical model, that is, by performing the arithmetic sum of the partial contributions of the class levels corresponding to each of the three photo-interpreted features as specific vulnerability factors appropriately scaled according to the relative weight as reported in Table 1 and 3.

<b>n</b>	<b>Typology</b>	<b>class</b>	<b>score</b>
1	palazzina	1	2
2	villino	1	2
3	linea	2	4
4	schiera	2	4
5	palazzona	3	6
6	spontaneo	3	6
7	blocco	4	8
8	semi-intensivo	5	10
9	intensivo	5	10
10	articolato	6	20
11	torre	6	20

Table 2 - Building typologies and related score classes

<b>M1_L</b>		
<b>class</b>	<b>Score</b>	<b>Pol. n.</b>
A	7	3
B	12	55
C	16	17
D	18	5
E	22	54
F	26	6
G	28	19
H	30	11
I	37	3
L	43	65
M	45	47

Table 3 - Urban fabric built-up morpho-typologies and related score classes.



**Figure 3** - Detailed transect distribution of the 11 classes of partial vulnerability of the urban fabric morpho-types M1\_L ( b ) and overall vulnerability NVI ( a ). In overlay panchromatic channel of Landsat ETM+ acquired on 7/27/2003.

### 2.3. Data mining and supervised classification

The objective of this stage was the assessment of the distribution of the urban morpho-typologies found within the transect over entire territory of the Rome municipality through supervised classification integrated methodologies on purpose implemented. The automatic land use/cover classification of urban areas using remotely sensed digital imagery is a particularly challenging task due to their high spectral heterogeneity arising from the mixing of natural and impervious artificial surfaces delimiting urban 3-D elements, like buildings, infrastructures, roads, and parking lots whose spectral responses may be generally too similar to be suitably separated using only the limited information derived from single pixels of the remotely sensed multispectral raster images. Therefore often additional information, are required for improving the discrimination of these land covers by the classifier, to this end the object-based schemas respects to those pixel-based, provide a more effective way to derive and integrate into the classification process other kind of data related to morphology and context of the contiguous pixels groups, suitably extracted from the same imagery, exploiting also their higher robustness against mis-registration between different input raster layers. In this context to increase the effectiveness of spectral classification semi-automatic approaches, many authors successfully exploited the object-based image analysis methods for properly describing and quantifying the typical spatial heterogeneity of urban land covers using the HR/VHR remotely sensed data [17, 34, 35]. These objects classification approaches in the supervised schemas allowed them to better discriminate multi-pixels ground

These NVI index values, calculated with a maximum of 100, were displayed with shades of color from green to red as shown in the upper part of Figure 2. Here a first consideration can be derived by observing that the NVI highest values concentrate mainly on the left side of the transect corresponding to the city center where there are still some exceptions arising in the green areas and city parks. The NVI distribution appears to have rightly a negative gradient going from the center to the periphery. Obviously this result is mostly determined by the parameters relating to impervious surfaces and compactness that have similar trends except for some peripheral areas subject to special or intensive urban deployment.

land-use patches containing a variable mix of different built materials, infrastructures and natural land cover features but pertaining to the same land-use (i.e. urbanistic class) category [36], on the basis of a subset of limited training object preliminarily characterized through visual interpretation or in situ data. These patches (objects) including a variable number of pixels having different spectral signatures work as pre-defined object boundaries, preliminarily obtained through segmentation algorithms based on predefined spectral and geometric parameters (i.e. scale/dimension, spectral homogeneity, convexity,...). Therefore in order to suitably utilize the increasing amount of the spectral and contextual information provided by the object approach and improved satellite sensors in terms of spatial resolution, number of acquisition channels and radiometry, in our case various data mining methods were exploited to properly handle the different spectral and textural components (features) extracted from the Landsat ETM+ multispectral data under form of spectral indices (SI) and used as independent variables (70 features) in the subsequent modelling and classification steps as detailed below. In our classification schema a pixel-based and object-based method were integrated with the objective to obtain more accurate result in the context of urban and suburban landscapes of Rome municipality using Landsat HR imageries under form of normalized spectral indices as described in the following paragraphs.

### *2.3.1. Data processing schema and object classification*

The classification procedure implemented with the goal of recognizing the M1\_L built up classes over the Rome municipality encompass various interconnected processing steps as described in this paragraphs. This classification has been carried out using a supervised approach on the basis of the results of photointerpretation at transects levels which have been extended to entire municipal territory using the Landsat ETM+ EO multispectral data as main input. The Figure 4 shows the logic schema of the entire processing and classification procedure where the processes and data are symbolically indicated and linked together. In particular, as reported here, first of all the EO data primary input (LDS8) are processed (PRO) to derive the three spectral normalized indices (SIL) suitable to minimize the atmospheric and illumination noise effects (see next paragraph) and used as inputs for the subsequent steps [37]. The three indices multilayer was exploited for a segmentation step [38], tuned according to the photo-interpretation constraints for obtaining a vector layer of 3478 polygonal (POL) covering the entire territory of Rome. For each of them, the set of 70 independent variables (spectral and textures from GLCM) were extracted from the three indices layer as polygonal attributes.

The layer containing the 285 polygonal of the transect (CAM), classified by means of photointerpretation approach was overlaid to that obtained from segmentation (POL). Subsequently various GIS tools and local majority analysis (TRA) were exploited to recognize 104 segmented polygonal sample (POLA), enough pure and complete to be used as training areas for morpho-types M1\_L classification of the remaining segmented urbanized polygonal areas. These formers were selected (URB) among all segmented polygonal on the basis of the presence at least of the 10 % of urban pixel (POLU) as threshold, using the results of a per pixel Maximum Likelihood (ML) supervised classification (CLS) of urban areas (RAS), previously obtained from the Landsat 7 ETM+ multispectral image. The rough evaluation of the accuracy of this result has been performed on the basis of comparison with the CLC 2006 (Corinne Land Cover) urban thematic layer with an agreement of about 85%. The final morpho-types M1\_L (POLC) classification (OBC) of the urbanized areas (POLU), including those to be used as training set, was performed using the previously selected algorithms and reduced spectral features input data set, the results including 1539 classified object polygonal are displayed in Figure 4. These algorithm and data reduction techniques were previously selected (DMA) among the most effective [39], on the basis of the results in terms of overall accuracy (ACC) and ER values obtained using the training set within the comprehensive selection method. In the next paragraphs firstly the features implementation and train data extraction were widely described then the selection processes referring to feature reduction and object classification by means of data mining algorithms were explained too.

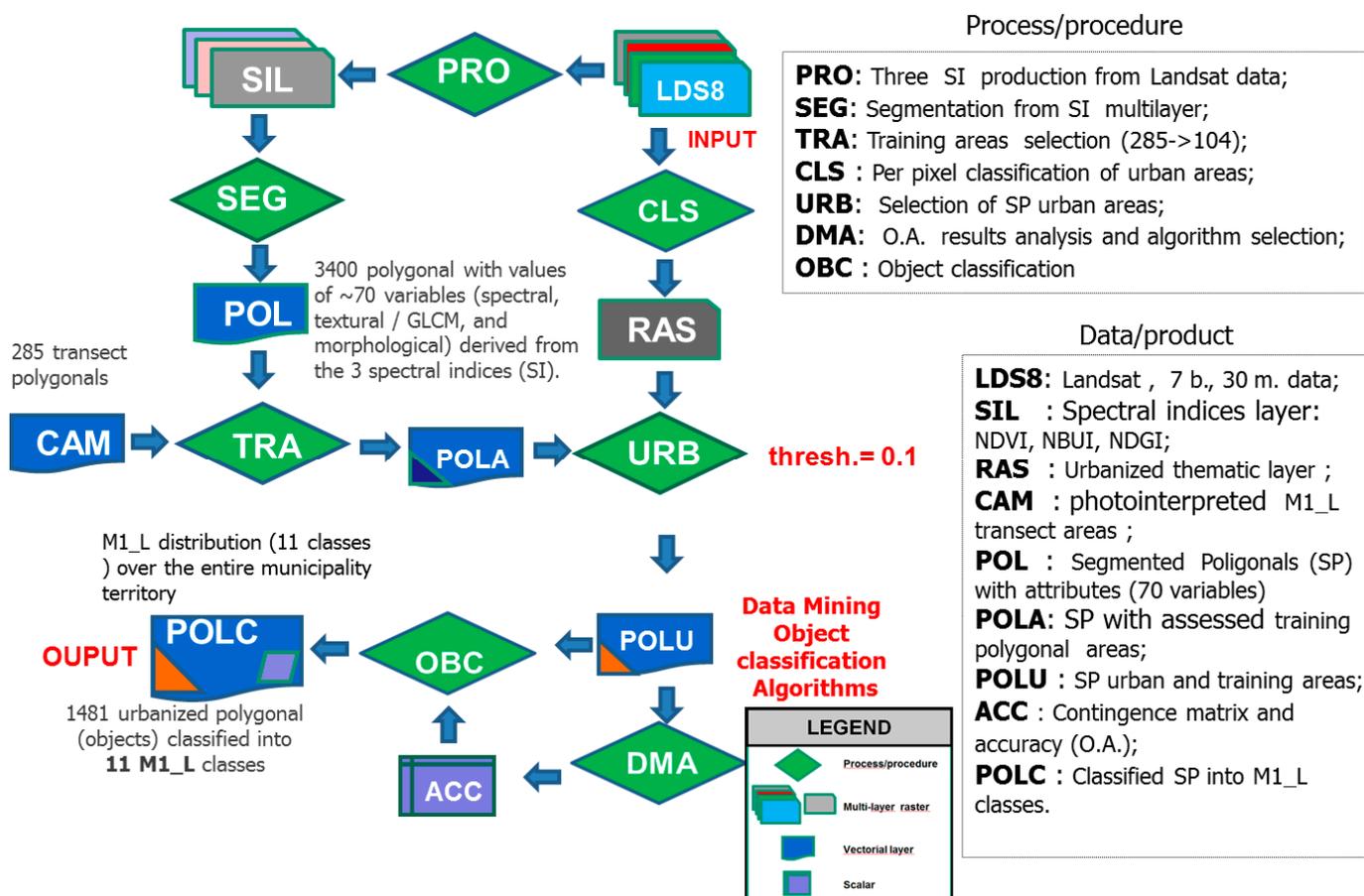


Figure 4 – Data processing and Object classification schema of urban fabric morpho-types sensible to HW and UHI effects in the municipality of Rome

### 2.3.2. Features and training data

The characterization of thermal response of the urban fabric in the urbanized areas of the entire municipality territory was therefore accomplished by means of a supervised object-classification procedure, on purpose developed, using the Landsat ETM+ multispectral data and the morfo-types classes distribution obtained in the transect by mean of photointerpretation methods. Considering that the Landsat 7 ETM+, starting from the middle of 2003, suffered from the SLC (Scan Line Corrector) subsystem failure which determined degradation of images acquired after the end of May, with introduction of some gaps without useful data [40, 41], more dense at borders, one of the last good frames acquired on 18 of March 2003, before the SLC damage was processed for the morfo-typologies distribution assessment through object-classification procedure. Such an approach allowed us to extend the knowledge about urban fabric thermal response gained by means of photointerpretation methods on polygonal (objects) of the transect to the other 3400 polygonal objects previously obtained by means of semiautomatic segmentation and covering the entire territory of interest. This was accomplished through advanced clustering data mining algorithms [18, 35], using the 70 spectral-spatial variables suitably extracted from Landsat reflectance data, mostly under form of devoted spectral indices and including those more linked to local textural (Zhang et al. 2014) content as proxy of built-up compactness, introduced in the previous paragraphs [37].

First of all, to reduce the basic variables number in view of the subsequent object-oriented/data mining approach trying to maintain their useful information contents, the 30 m. TOA (Top Of Atmosphere) reflectance signals acquired in the six bands were transformed into three normalized spectral indices (SI), more linked to spectral properties of the natural (i.e. vegetation,...) and artificial surfaces of interest [43] and, at same time, to reduce the noise contribution associated to the varying topography and atmosphere:

$$NDVI=(R_4-R_3)/(R_4+R_3) \quad (1)$$

$$NBUI=(R_6-R_5)/(R_5+R_6) \quad (2)$$

$$NDGI=(R_3-R_2)/(R_2+R_3) \quad (3)$$

where, according to Landsat ETM+ sensor spectral configuration,  $R_n$  state for reflectance in the  $n$ th acquisition band. Bands  $n. 1,2,3$  are respectively the RGB visible channels, while the remaining indicate those in NIR (4) and SWIR (5,6) ranges.

The NDVI (Normalized Difference Vegetation Index) derived from red and NIR (Near Infrared) reflectance signals characterizing the photosynthetic vegetation, was designed for more effectively capturing their spatial and temporal changes. It was widely used for vegetal ecosystems monitoring, in particular it was exploited for usefully assessing various important biophysical parameters of plants linked to their biomass, productivity and health and for studying grasslands and rangelands since the early applications of the satellite EO, starting from the late of the 1970 [44]. According to various authors who subsequently employed this index as proxy of impervious surface density [21, 45], and of surface emissivity for assessing the LST distribution at urban level [46, 29, Lin & Zhang, 2011), in our approach the NDVI was exploited directly as proxy of the previously introduced permeability variable and in synergy with the others variable/features for objects characterizing in terms of built up typology and density variables using also the related textural features. Although the NBUI index is less common than NDVI, it was successfully exploited for enhancing the specific spectral responses of building and artificial surfaces detected by remote sensors [37, 43], while the remaining index NDGI was introduced to recover the green spectral response fraction considering the high reflectance of urbanized infrastructures in the visible range and taking advantage from its normalized ratio formulation. The cover of polygonal objects spreading over entire municipality territory has been obtained from these three-indices raster layer through the M. Baatz segmentation algorithm [35], using an appropriate combination of the related input parameters (i.e. scale, color and compactness) preliminarily tuned to produce objects enough compatible in term of size with those previously found into the transect using the visual estimation.

Thus, following the object oriented typical approach the features extraction has been accomplished using both the three indices layer (30 m. GSR) and the ETM+ panchromatic channel (15 m. GSR) for providing to each segmented polygonal object its own attributes synthesizing the spectral and textural features of the corresponding areas to be used as independent variables in the following supervised classification procedures. In addition to the usual spectral amplitude, mean, mode, standard deviation, sum and ratio also the three first moment GLCM (Gray Level Co-occurrence Matrix) dissimilarity, entropy and homogeneity [45] have been extracted from the four raster layers (three indices + panchromatic) at different GSR using a 3x3 kernel window and predefined offset vector. The ratio variable was calculated only for the three indices layer as pixels sum of fractions of the current component while the amplitude include the intensity difference

between maximum and minimum reflectance signal within every polygonal object. The table n. 4 shows the inventory of the nine features/variables extracted for every polygonal object using the available raster layers, including those referring to GLCM domain which have been introduced with the objective to better capture the local texture features of the urban fabric. In particular the three GLCM variables are dissimilarity (Dis) , entropy (Ent) and homogeneity (Hom) and the related pixel values were derived through the respective formulations:

$$Dis = \sum_{i=1}^{D-1} \sum_{j=1}^{D-1} p_{ij} (i - j) \quad (4)$$

$$Ent = - \sum_{i=1}^{D-1} \sum_{j=1}^{D-1} p_{ij} \log p_{ij} \quad (5)$$

$$Hom = \sum_{i=1}^{D-1} \sum_{j=1}^{D-1} \frac{p_{ij}}{1+(i-j)^2} \quad (6)$$

where the  $p_{ij}$  is the normalized number of intensity co-occurrences within the kernel,  $i$  and  $j$  indices indicate the gray level (intensity of the component) considered while  $D$  is their range corresponding to the dimension of the related GLCM square matrix derived from the number of intensity levels considered. The values of these GLCM variables have been firstly calculated at pixel level using the adopted kernel and offset vector while their mean over polygonal object area was then assumed for the related feature variable. Dis is a measurement of the intensity difference between the elements of the GLCM and it is high when the kernel region has a high contrast (i.e. repetitive high building strips and related shadows characterizing dense urban fabric areas). The Ent variable measures the local disorder in the distribution of pixel intensities, it increases as they don't happen in regular patterns and many GLCM elements have small values. Hom assumes higher values for smaller differences in the GLCM and little differences in local intensity pattern.

N.	Variables	Domain	Description
1	amplitude	Polygon	(Max-Min) Local contrast
2	dissimilarity	Poly(GLCM)	Linked to the intensity contrast of repetitive patterns in the region of interest
3	entropy	Poly(GLCM)	Linked to disordered intensity patterns in the region of interest
4	homogeneity	Poly(GLCM)	Linked to flattened intensity patterns in the region of interest.
5	mean	Polygon	average
6	mode	Polygon	mode
7	ratio	Polygon	normalized component
8	std-dev	Polygon	standard deviation
9	sum	Polygon	sum of components

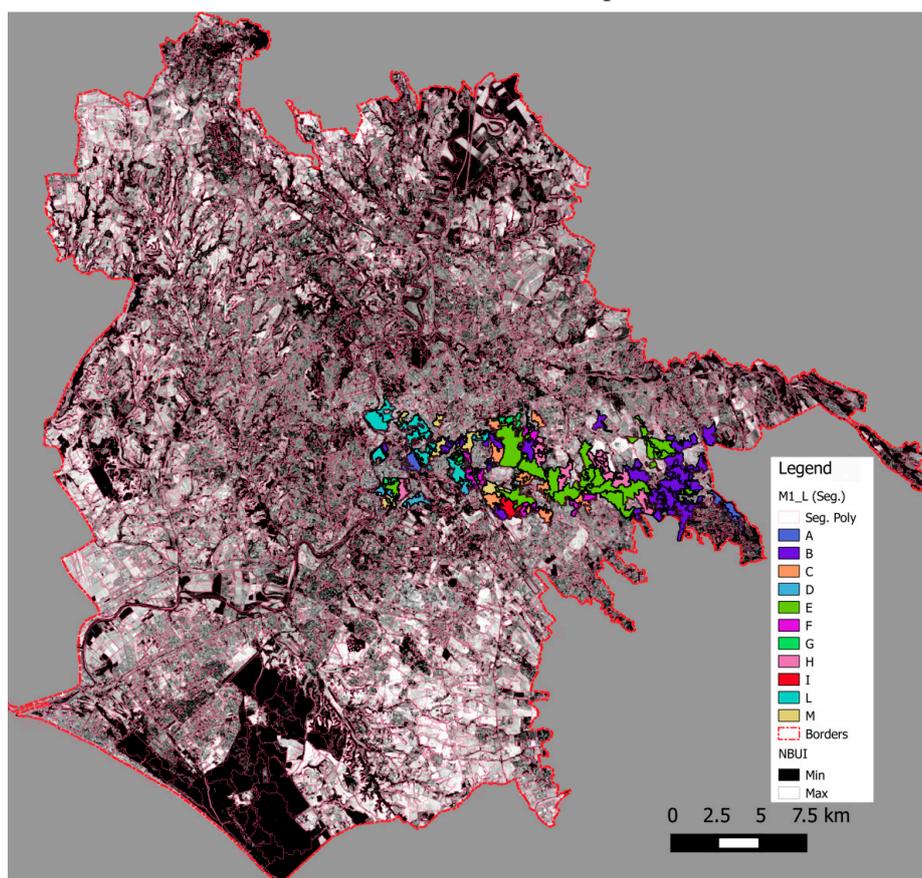
Table 4 - Feature variables extracted from the raster layers and associated to segmented polygonals objects

As you can see in table 4, the nine variables were derived from each of the three normalized indices stiched in the 30 m. GRS raster layer, while only eight (ratio excluded) have been calculated from the panchromatic channel.

These raster layers have been exploited for assessing the three raster GLCM features which were then averaged for every polygonal object. The obtained set of 35 variables has been doubled by

adding their normalized versions for a total of 70 feature variables assessed as attributes of polygonal objects. The objective of the subsequent object classification supervised process was to assess the distribution of the M1\_L classes over entire Rome municipality using suitable training data set derived the results obtained in the transect sub-area and the above EO-derived 70 feature independent variables. This former was a critical stage of the implemented methodology since the selection of the training areas in the polygonal cover obtained from segmentation on basis of those employed in the transect wasn't easy, in fact these formers, being delimited by road network they weren't generally exactly coincident with the results of automatic segmentation. Thus in order to overcome this difficulty first of all the segmented polygonal crossing those photo-interpreted in the transect were identified using the available GIS topological overlay tools, then for each of them a table of the related M1\_L classes area percentages in the transect was produced in order to allow the evaluation of the predominant cover class. Finally only the segmented polygonal having a majority class percentage higher than a predefined threshold (mainly up to 70%, 40% for two less populated classes) were retained as training areas.

The Figure 5 shows the cover of the selected training set as colored M1\_L classes obtained from the segmented polygonal for a total of 104 areas within or adjacent to the transect. These training areas were obtained by overlaying and spatial processing of the segmented polygonal and the 285 transect polygonal of the 11 morpho-types classes (table 3, Figure 2) assumed to be representative of the thermal response of the urban fabric in the Rome municipality. Considering the heterogeneity of residential areas which include different percentage of built-up mainly characterized by high reflectances in visible bands, in order to improve the characterization at municipality level of the



urban polygonal objects, a per pixel classification using standard algorithm (ML) was introduced in order to recognize the internal urbanized pixels and to allow a preliminary selection of the object polygonal based on a percentage threshold of their built-up surface (Figure 4 – URB process) for subsequent object classification step. The accuracy was assessed by means of the confusion matrix and photointerpretation results previously obtained for the transect areas, with a reached value bigger than 80 % in term of overall accuracy (OA).

Figure 5 - Urban fabric morpho-types distribution of training areas within the transect selected from segmented polygonal objects using GIS analysis in overlay to NBUI index distribution at municipality level.

### 2.3.3. Features and algorithm selection

First of all, five typical machine learning algorithms were preliminarily selected for supervised object classification taking into account their different capabilities and performance in various situation of noise, limited class samples and outliers presence in input data: C4.5 (decision tree), C-SVC (Support Vector Machine- discrete class, continuous input), Knn (K-nearest neighbors), ANN (Artificial Neural Network), RnTr (Random Forest Tree). Decision Tree allows to implement easy to interpret classification models by hierarchically splitting the data set [48, 49]. This algorithm selects the best subset of attributes based on an entropy measure and organizes the classes in a decision tree rule-based structure. Each node of the tree relates to a split in the feature space which is always orthogonal to its axes. Support Vector Machines, more often used in the per-pixel classification context, is a sophisticated non-parametric supervised statistical learning technique, robust against outliers, that estimates a hyperplane in the feature space that minimizes misclassifications [49, 50]. The k-nearest neighbor algorithm (Knn) is a method for classifying objects based on closest training examples in an n-dimensional features space. When given an unknown feature pattern the classifier searches the pattern space for the k training tuples that are closest to the unknown one [50]. The ANN non parametric algorithms are based on the neural network concepts and work without assumptions about input data distribution and independency. They learn from the training dataset and build relationships (networks) between input (features) and output nodes (classes) through hidden neurons layer connection weights modulation, a critical issue may be the amount of training occurrences which may be critical in our context (Wieland et al., 2014).

The Random Forest classifier consists of a group of decision trees induced with different sub-sets of the training data. Each tree of the forest casts a vote for the class to which a given analysis unit (in this case, a given segment) should be associated [50, 51]. The class with most votes is the one associated to the segment.

The high dimensionality of the extracted features used as EO-derived input data (i.e. our 70 variables) is potentially affected by a variable level of noise and mutual correlation of the components whose specific contribution may be irrelevant or even detrimental (introducing confusion) in object classification task by means of a particular algorithm. Thus an integrated approach for data reduction and algorithm selection was implemented for improving classes detection and processing efficiency [17], using the previously assessed polygonal training set. Initially the entire training data set containing the feature attributes and related class label for the 104 polygonal was processed in order to assess the accuracy in term of error-rate (ER) derived parameter (1-overall accuracy/100) for each of the algorithm above introduced. Subsequently only the randomly selected 70% of the data set was exploited for training phase while the test was carried out on remaining samples to evaluate the algorithm overtraining weakness. These last process was repeated three times and the results mean was retained as more representative ER on the basis of cases sampling increase (Table 5).

Trial	Train size	Test size	Error Rate	O.A.
1	72	32	0,6875	0,3125
2	72	32	0,7188	0,2812
3	72	32	0,6563	0,3437
			0,6875	0,3124

Table 5 - Train/test trials results and related mean (last line).

The following repetitive classification trials of the random selection of 30% of the training set by

In table 6 was reported the confusion matrix obtained from the classification result of C4.5 applied to the training set. The related overall accuracy (OA) as the fraction of number of correctly labeled objects (evidenced in bold along the principal diagonal) is 0,6538 while the corresponding error rate (ER) is 0,3462. The specific accuracy is varying for each class and drops to 0 for C and D (missed) classes.

means of the same C4.5 algorithm, suitably trained with the remaining input data, produced the ER results included in table 6, with related average in the last line. As you can observe, the 65% OA, obtained with the entire training data set, decreases to less than 30%, in case of repetitive train/test classification trials using respectively the 70% and 30% of the training data, while the number of undetected classes rises until 5 (Table 7).

<b>M1_L class</b>	<b>L</b>	<b>B</b>	<b>C</b>	<b>M</b>	<b>F</b>	<b>G</b>	<b>E</b>	<b>I</b>	<b>H</b>	<b>D</b>	<b>A</b>	<b>Sum</b>
<b>L</b>	6	0	0	2	2	0	1	1	0	0	0	<b>12</b>
<b>B</b>	0	20	0	0	2	0	1	0	1	0	0	<b>24</b>
<b>C</b>	0	0	0	0	3	0	1	0	4	0	0	<b>8</b>
<b>M</b>	0	1	0	4	0	2	0	0	0	0	0	<b>7</b>
<b>F</b>	0	0	0	0	6	1	1	0	0	0	0	<b>8</b>
<b>G</b>	0	0	0	0	0	3	1	0	0	0	0	<b>4</b>
<b>E</b>	0	1	0	0	0	0	15	2	1	0	0	<b>19</b>
<b>I</b>	0	0	0	1	0	0	0	3	0	0	0	<b>4</b>
<b>H</b>	0	0	0	0	0	1	0	0	8	0	0	<b>9</b>
<b>D</b>	0	2	0	0	1	0	0	0	0	0	2	<b>5</b>
<b>A</b>	0	1	0	0	0	0	0	0	0	0	3	<b>4</b>
<b>Sum</b>	<b>6</b>	<b>25</b>	<b>0</b>	<b>7</b>	<b>14</b>	<b>7</b>	<b>20</b>	<b>6</b>	<b>14</b>	<b>0</b>	<b>5</b>	<b>104</b>

Table 6 - Confusion matrix derived from object classification of training set using C4.5 decision tree algorithm.

The above described two phases tests were then carried out using the others algorithms above introduced. Then, according to general data-mining approach, two different methods for input data reduction were applied to the training set:

- the transformation of data to a lower dimensionality by means of the PCA (Principal Component Analysis) data transformation [52];
- the selection of most significant features, termed feature selection, using different statistical techniques: Fischer; Run Filt; Step.

The output of the PCA transformation of the original 70 features data set were uncorrelated components retaining a decreasing portion of the global variance, as depicted in figure 6. As you can see here, more than 90% of variance of the initial data set is preserved if we consider the first 21 PCA components instead of the original 70 variables. While the PCA transform input feature by mixing them on the basis of their mutual correlation, the feature selection methods allow to select a features subset from the original set without any transformation, and maintains the physical meanings of the original features. Even if, in this sense, feature selection is superior in terms of better readability and interpretability, here, in addition to the standard ones, we tried to apply in sequence these two approach (i.e. the feature selection was applied also to the PCA components). Starting from the feature attributes of the 104 polygonal in the training set whose dimensionality makes impossible the selection by means of the usual manual approach on the basis of visual comparison of distribution and multidimensional scatter plots, three techniques representative of the automatic continuous feature selection for for classification labelling were exploited. Thus the Fisher, Runfilt and Stepdisk [17,52, 53], were then introduced with the aim to select a smaller subset of input data minimizing their redundancy and maximizing relevance to the target scope, such as the suitable class labeling, in classification. The first two methods are based on the filter model which is independent from the subsequent learning algorithm and considers only the general characteristics

of the training data such as distance, consistency, dependency and information content for assessing the input feature quality. STEPDISC (Stepwise Discriminant Analysis) exploits a discriminant analysis criterion. Others approach (wrapper, embedded) include the specific classifier algorithm in input feature selection and weren't considered here since this step was in any case performed subsequently in our procedure. The feature selection step was applied also to the subset of the selected PCA components. Ultimately all combinations relating to five different algorithms, two training/test approach and feature selection techniques were tested for assessing their global accuracy in term of ER (O.A.) values which are arranged in the two following tables n. 7 and 8. In particular, the first table n. 7 includes the results obtained using the features of training set while the second one shows the ER's assessed for the subset of the first 21 PCA components, containing an amount higher than 90% of the original explained variance (Figure 5).

The first two lines of table n. 7 report the ER values obtained from classification of the training set using the five algorithms above introduced with all 70 features. The first one refers to the entire data set (all-70) while the second include the result for the related train/test approach (all-T). Similarly, in the following two row, the result of the feature selection using the Fisher filtering method are reported respectively for the selected data set including the first most relevant 35 components (35) and related train/test evaluations (35-T). The same schema is repeated for the remaining features selection methods applied here and in the subsequent table referring to PCA components. The "Feature n." columns (corresponding to different three selection techniques) of the two tables indicate the number of selected components while the T suffix state for train/test evaluation. All the resulting ER values are reported with a following number in brackets which indicates the missed classes (null accuracy) in classification, except those corresponding to detection of all the classes without brackets. Due to its structural overfitting aptitude some not significant results corresponding to RnTr algorithm aren't included. The most relevant ER values referenced in the following are shown in bold digit. In general the number of the selected descriptors/features with different methods is depending on their relevance/correlation with the training classes distribution assessed trough ANOVA analysis and F-test on the basis of predefined threshold.

In our case a number of selected features corresponding approximately to one half of the original one was adopted in order to provide reduction factor (50%) for enough effective

	Features n.	C4.5	C-SVC	Knn	ANN	RnTr
	all -70	0,3462 (2)	<b>0,1635</b>	0,5192 (1)	0,4519 (6)	
Sel. Tech.	all-T	0,7188 (5)	0,7292(5)	0,7500(5)	0,6875(5)	0,7917 (5)
<b>Fisher</b>	35	0,3365 (1)	0,3462	0,4615	0,5000(6)	
	35-T	0,7396 (6)	0,7396(5)	<b>0,6563 (3)</b>	0,7604 (7)	0,7188 (4)
<b>RunFilt</b>	35	0,3558 (2)	0,3173	0,4712	0,5192 (6)	
	35-T	0,7396 (6)	0,6563(4)	0,6979 (3)	0,7500 (8)	0,7396 (4)
<b>Step</b>	35	0,3365 (2)	0,3077	0,5192 (2)	0,5192 (5)	
	35-T	<b>0,6250(5)</b>	0,7292(4)	0,6667 (3)	0,6875 (8)	0,7604 (3)

Table 7 - ER values obtained from classification of the original and reduced training data sets

average) of missed classes.

The results of the C4.5 classifier improve with the feature subset selected using the Fisher and Step techniques even in the train/test trials, while those obtained by the Knn algorithm with Fisher

performance evaluation of the various combination of selection techniques and algorithms. As shown in the two first rows of table n. 7, although the ER obtained by C-SVC algorithm seems the best (lowest) in classifying the complete training set, its relative performance significantly decreases in test case (all-T second row), in which all the results happen poor, with a relevant number (5 on

selected data set might represent an optimal trade-off between the minor ER value and number of missed classes,

		C4.5	C-SVC	Knn	ANN	RnTr
	Pc21	0,4423 (4)	0,4808(1)	0,5288 (1)	0,5769 (7)	
Sel. Tech.	Pc21-T	0,7813 (6)	0,7477(6)	0,7917(7)	0,7292(7)	0,8333 (5)
Fisher	10	0,4519 (5)	0,5385(4)	0,5288(1)	0,6154(8)	
	10-T	0,8021 (6)	<b>0,6563(7)</b>	0,7708 (6)	0,7813 (8)	0,8021 (4)
RunFilt	10	0,4231 (2)	0,5577(4)	0,5481	0,6154 (7)	
	10-T	0,7604 (6)	<b>0,6354(7)</b>	0,7604 (3)	0,8021 (8)	<b>0,6979 (4)</b>
Step	10	0,4615 (2)	0,5577(5)	<b>0,5192</b>	0,6442 (8)	
	10-T	0,7604(7)	0,7396(7)	0,8125(7)	0,8125 (8)	0,7917 (7)

Table 8 - ER values obtained from classification of the PCA- compressed and reduced training data sets

related to Pc21 compressed dataset lowering at same time the missed classes number. Finally, considering the ER values/missed classes numbers obtained using the different selection techniques on the entire data set and the capability to be robust against overtraining by means of the train/test approach, firstly the 35-features data set derived through Fischer method and Knn classifier were exploited for the subsequent classification step of the global dataset of the 3478 polygonal objects, covering the entire municipal area. From the results reported in table 7, due to their performance in terms of ER / missed classes amount and/or robustness against overtraining, two

additional classifiers, namely C4.5 and C-SVC respectively with Fisher-selected and all features, were also considered. These three alternative M1\_L distributions indicated as KnnF (Knn classifier with features Fisher selection), C4.5S (C4.5 with Stepwise selection) and C-SVC (using all features), obtained from the above classification algorithms were preliminarily selected. The obtained thematic map referring to the former C-SVC is reported in Figure 7. The in depth evaluation for refining their selection was performed in the subsequent LST based calibration phase through their corresponding numerical sliced counterparts indicated respectively as KnnFN, C4.5SN and C-SVCN and obtained by assigning the relative class scores included in table 3.

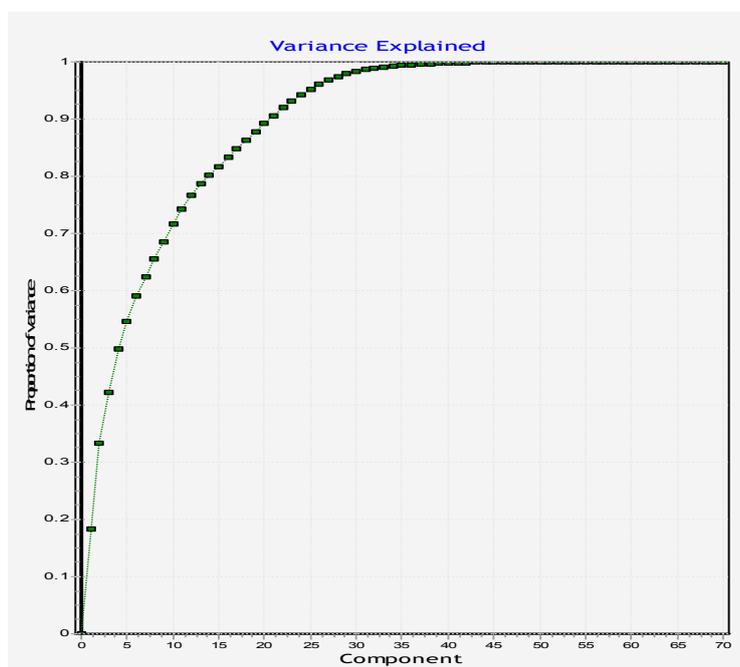


Figure 6 - PCA components and the related cumulative normalized amount of the explained variance

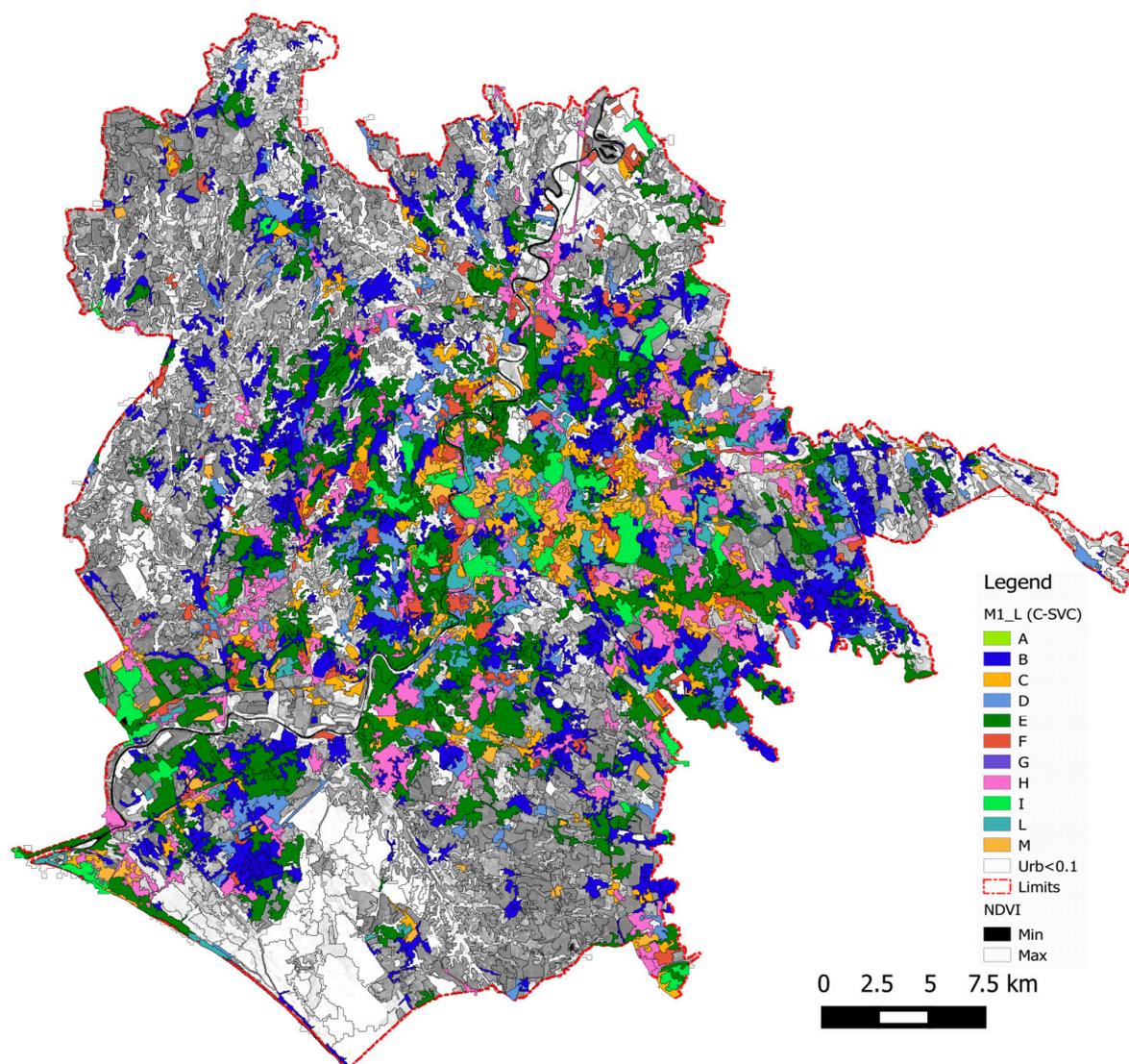


Figure 7 - M1\_L classes distribution on Rome municipality territory obtained by means of the implemented supervised object classification procedure through C-SVC algorithm using the transect training areas.

#### 2.4. HW and UHI Thermal calibration

Although a preliminary NVI distribution at municipality level might be obtained from the numerical distributions of M1\_L urban fabric partial vulnerability classes, provided by object classification procedure, and corresponding NDVI as permeability proxy, it will be affected by some uncertainties arising from the use of the best M1\_L distribution within the three previously selected, which optimize different aspects. Even if the photointerpretation approach allowed us to assess the NVI realistic distribution over transect area, based on the preceding identification of the 11 partial vulnerability M1\_L classes related to the urban fabric variables (typology and compactness) and permeability of the related ground surfaces, the relative weights between these different factors exploited in the related semi empirical model are assumed without a sufficiently robust physical basis.

In order to better address these weaknesses of the photointerpretation based approach, the basic idea was to exploit the LST distributions measured in HW and UHI condition for a physically based model calibration. Given that the HW and UHI synergy reflects on the LST distribution as the main measurable effect, it is assumed as dependent variable of a multivariate regression model whose independent variables have to be selected within the most significant of all the previously

extracted object features. The three M1\_L numerical distributions, previously obtained from the classification procedure were added with the local altimetry to the independent variables data set before the selection for a total of 74 independent variables of a multivariate model to be optimized by selecting the more significant variables using step-wise regression statistical methods. All the (dependent and independent) model variables at polygonal level were arranged to provide a unique data set for the 1539 classified urban objects which was then exploited for variable selection and model calibration using the stepwise regression methods as described in the following. In this perspective, two typical LST distributions of Rome nocturnal UHI, captured by MODIS polar satellite sensor in the summer 2003 (Figure 8), were selected as input to the physically based calibration method of the urban fabric vulnerability M1\_L model, previously introduced on empirical basis. The LST distributions refer to the HW period of July 2003 (Figure 1) and are reported in the figure 8 (left and right pictures). Here you can see how the night LST distribution in the leftmost and rightmost images exhibit local UHI maxima in correspondence of the town center while the central diurnal image of the 16-7-2003 is characterized by a diffuse cloud cover which decrease the detected radiant temperature over land. Although the cloud presence has decreased in coincidence of the night acquisition (right image) some mist residuals have remained on the right borders of the municipality corresponding to mountains areas and coinciding with colder patch.

In the following graph of figure 9 the bi-spectral plot of the two night LST distribution of 12 and 16 July 2003 within the polygonal objects corresponding to different M1\_L classes are reported. Here you can see how the unreliable lowest temperatures corresponding mainly to these residual cloudiness affect the distributions, in particular that of the 16-07 which required a preventive selection of the cloud free and most reliable polygonal areas to be used in variable selection.

The objective of the stepwise methods introduction here was to assess a regression model by selecting only the most relevant variables within those available (74) and possibly including as few variables as possible because each irrelevant regressors decreases the precision of the estimated model coefficients and predicted values. In such a way both the selection of the most appropriate

	LST-12-07-2003		LST-16-07-2003	
	Bck	Frw	Bck	Frw
<b>Sig. Lev.</b>	0,1	0,0001	0,1	0,0001
<b>N° var.</b>	16	8	16	8
<b>C4.5SN</b>		0,013969	0,010265	0,006966
<b>C-SVCN</b>	0,011381	0,021581		0,014668
<b>KnnFN</b>				
<b>NDVI</b>	-0,08302	-0,05773		-0,03563
<b>NzNDVI</b>			-7,964	
<b>NzPan_mode</b>		0,0000001		0,0000001
<b>R<sup>2</sup></b>	0,477	0,38	0,362	0,28
<b>R<sup>2</sup><sub>adj</sub></b>	<b>0,472</b>	<b>0,377</b>	<b>0,354</b>	<b>0,276</b>
<b>n. samp.</b>	1534	1534	1262	1262
<b>F</b>	869,912	1176,056	442,863	611,608

Table 9 - Backward (Bck) and forward (Frw) stepwise results for multivariate regression models of the two night LST distribution on Rome territory during the HW of 2003

variables including those related to the previous NVI transect model and estimate of the related weights (coefficients of regression model) were achieved on the basis of the LST distributions. In the stepwise method the variables selection for the regression model is carried out by successively adding (forward) or removing (backward) variables based solely on the t-statistics of their estimated regression coefficients [54]. The stepwise option lets you either begin with no variables in the model and proceed forward (adding one variable at a time), or start with all potential variables in the model and proceed backward (removing one variable at a time) based on the previously selected thresholds for the coefficients. In the table n. 9 a synthesis of the results obtained through backward (Bck) and forward (Frw) stepwise regression application with the

two above described LST distributions is reported. The number of selected variables (N° var.) with

related significance level (Sig. lev.) of t- statistics threshold used for inclusion/exclusion and the estimated coefficients for the most significant of them in the assessed multivariate models are included in the table. The polygonal samples have been selected trying to exclude the clouds artifacts and according to a temperature threshold (285 T°) compatible with reliable summer (see x-LST value of graph on Figure 8) land values. As reported in Table 9, both the C4.5SN and C-SVCN independent variables derived from M1\_L are relevant in explaining the LST spatial distribution especially in models obtained from the forward stepwise method, which include the NDVI index and a minor number of independent variables (8) for the LST distribution detected in two different dates during the HW period. Finally the R<sup>2</sup> correlation levels ranging from 0.472 to 0,27, and F/p-value parameters are also included to highlight the models significance. Although the four models contain different independent variables here firstly we focused on those previously introduced for estimating NVI at transect level related to urban fabric features and soil permeability. Since these formers, namely the M1\_L numeric distributions and NDVI (or its derived normalized versions), as soil permeability proxy, are selected in all models their importance in determining the thermal response of urban fabric was confirmed. First of all it should be evidenced that the both C4.5SN and C-SVCN numerical urban fabric partial vulnerability distribution have been retained in the estimated models while the KnnF variable, despite its good statistical performance with variables selection, didn't agree enough with the considered HW-UHI LST patches. The increasing contribution of urban fabric M1\_L distributions to temperature rise is evidenced by the related coefficient sign which maintain positive in all models while the opposite sign corresponding to NDVI is related to cooling effects arising from the permeability and vegetation increase. In agreement with the general indications in our case although the estimated models using the backward method exhibit higher correlation coefficients (Table 9) they include more numerous significant independent variables which make difficult the physical explanation of their connection with the LST distribution. Thus given that the forward procedure is generally retained more suitable to provide an initial screening of the candidate variables within a large group and also for its robustness against the multicollinearity and outliers problems (Miller, 2002) the corresponding results are considered for the M1\_L distribution selection. In Table 10 and 11 all the independent variables of the complete models selected via forward stepwise for the Rome night LST distribution of the 12 and 16 July 2003 are reported.

Variable	Coef.	std	t(1529)	p-value
Intercept	297.430.196	0.529670	561.538.321	0.000000
C4.5SN	0.013969	0.005345	2.613.432	0.009052
rp_ratio_2	12.252.723	0.608251	20.144.190	0.000000
rp_mean_0	-0.057728	0.003120	-18.503.385	0.000000
nz_ratio_0	5.478.326	0.748758	7.316.554	0.000000
C-SVCN	0.021581	0.003839	5.622.098	0.000000
pnz_rode	0.000000	0.000000	-5.186.027	0.000000
rp_aude_1	0.004156	0.000760	5.465.880	0.000000
nz_rode_1	0.000000	0.000000	4.634.898	0.000004

Table 10 - Regression model obtained by means of the forward stepwise variable selection for the LST night distribution of 16-07-2003.

variables rp\_ratio\_2 and rp\_ratio\_0 state respectively for the normalized responses of NDGI and NDVI indices, pnz\_rode and nz\_rode\_1 refer correspondingly to the normalized mode related to

In addition to M1\_L numeric distributions previously introduced, the others independent variables included are referenced using the internal names with standard suffixes and prefixes assigned by the segmentation and data mining software packages previously exploited in features extraction and object classification. In particular the numeric suffixes 0, 1, 2 indicates the three previously introduced spectral indices( respectively NDVI, NBUI and NDGI), while the prefixes rp, nz, pnz state correspondingly for polygonal average, normalized polygonal average and panchromatic normalized polygonal average. The

panchromatic band and NBUI. The NBUI polygonal amplitude (rp\_aude\_1) was included also in this model.

Variable	Coef.	std	t(1253)	p-value
Intercept	303.454.833	0.740549	409.770.309	0.000000
C4.5-SN	0.006966	0.005457	1.276.643	0.201965
rp_mean_1	-0.020506	0.001413	-14.514.235	0.000000
rp_mean_0	-0.035626	0.002516	-14.160.640	0.000000
pnz_rote	0.000000	0.000000	-7.614.581	0.000000
nz_rtude_2	0.706314	0.143912	4.907.956	0.000001
prp_mode	-0.007164	0.001757	-4.077.290	0.000048
rp_eopy_1	0.000900	0.000231	3.904.598	0.000099
C-SVCN	0.014668	0.003932	3.730.728	0.000199

Table 11 - Regression model obtained by means of the forward stepwise variable selection for the LST night distribution of 12-07-2003.

GLCM entropy (rp\_eopy\_1). In both the models the forward selection included the two M1\_L C4.5-SN and C-SVCN numeric distributions referring to urban fabric features previously assessed by means of photointerpretation methods. In addition also the NDVI related rp\_mean\_0 and pnz\_rote polygonal variables were selected in both the LST models.

The selected variables for the model of the night LST distribution captured by MODIS sensor on the 12-07-2003, as the previous one, comprise the same two M1\_L numeric distributions, polygonal means of the NDVI and NBUI indices, and panchromatic channel normalized mode (pnz\_rote). Differently from the model referring to the LST distribution of the more cloudy night of 16-07-2003, other variables were included namely the normalized amplitude of NDGI (nz\_rtude\_2), the polygonal mode of panchromatic channel (prp\_mode) and the NDVI

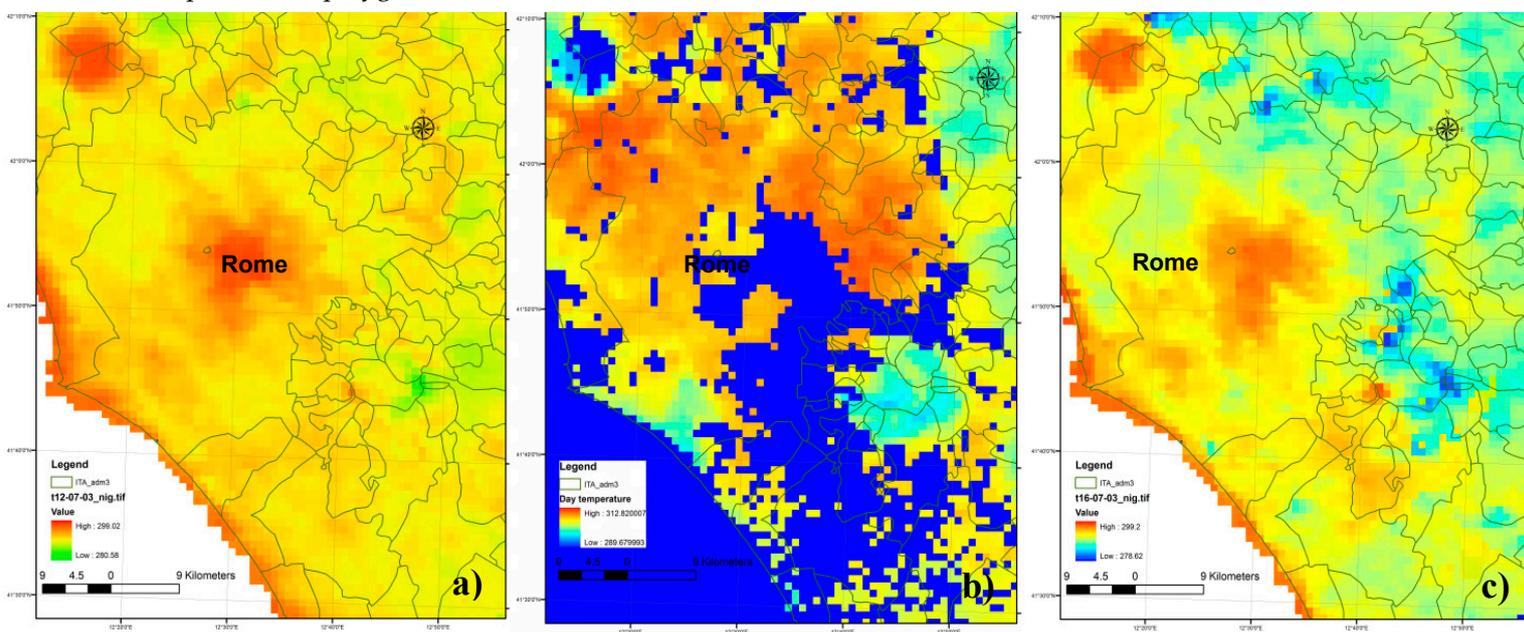


Figure 8 - LST distribution on Rome municipality acquired by MODIS sensor in 12-7-2003 at 21.30 ( a ) and in 16-7-2003 at 09.30 ( b ) and 21.30 ( c ).

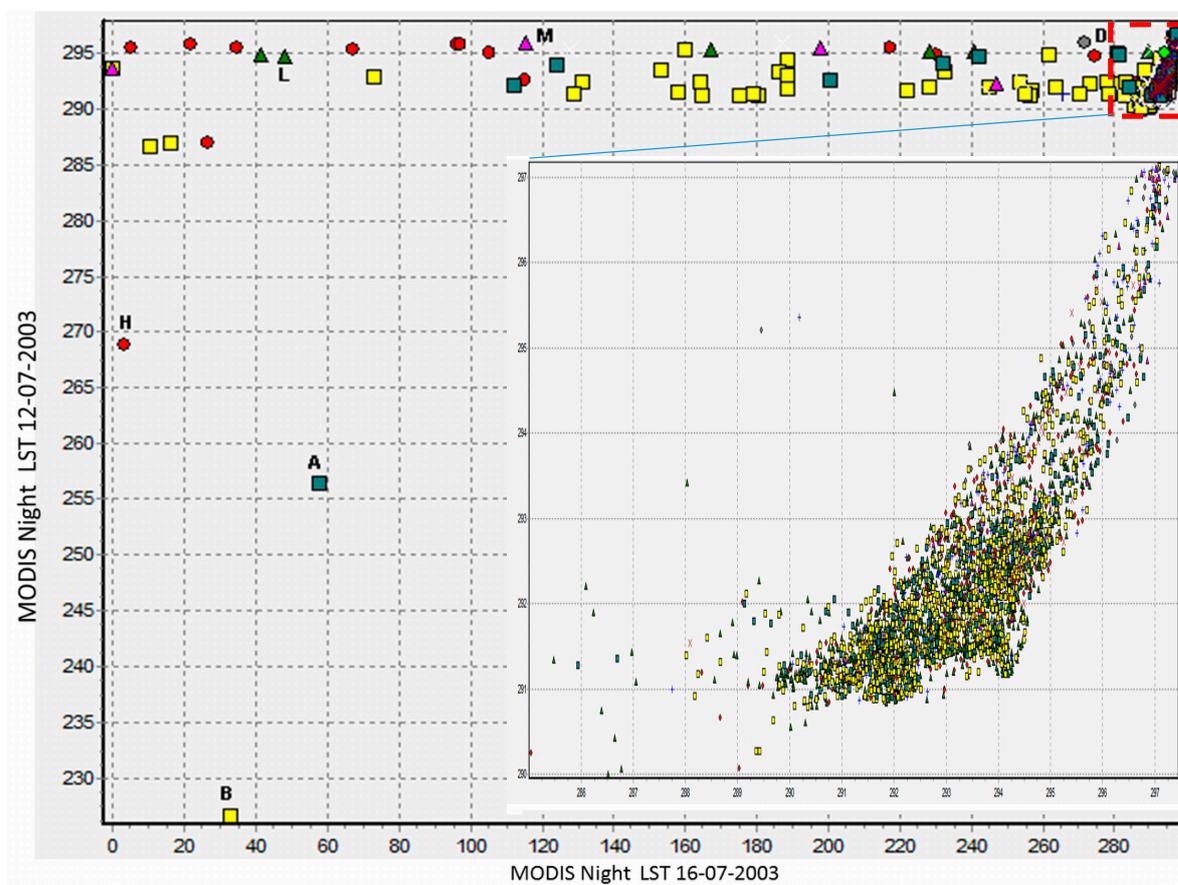


Figure 9 - Bi-spectral plot of 16-07-2003 vs 12-07-03 LST ( $T^{\circ}$ ) inside the segmented polygonal objects whose M1\_L class obtained using the C-SVC classifier is indicated using different point symbols. The detailed graph on the lower right side refer to the qualitative distribution in the interval of the 16-07-2003 LST, selected to avoid cloud noise effects.

### 3. Results and discussion

A semi-empirical linear model of thermal vulnerability of urban areas, based on built up characteristics and presence of permeable surfaces/vegetation was firstly implemented in order to quantitatively estimate the contribution of the urban fabric features related to temperature extrema regional meteo-climatic phenomena even in synergy with the well-known UHI local anomalies. The built up features focused for this purpose have been the building/infrastructure typology and compactness which were further characterized in term of classes and related numerical grades/scores on the basis of preexisting knowledge and studies on the various urban areas of Rome. This semi-empirical model was applied to a 5x20 km transect test area through its preliminarily road-delimited residential urbanistic homogeneous areas, using mainly an on purpose developed photointerpretation method of large scale detailed cartography. In such a way the preliminary distribution of thermal vulnerability M1\_L classes (Figure 2) and related scores coming only from urban fabric was produced. The association of the numerical scores assigned to M1\_L distribution with those derived from photo-interpreted permeable surface percentages for each of the previously introduced residential areas allowed us to obtain a preliminary and qualitatively reliable, thermal vulnerability distribution under form of NVI (Numerical Vulnerability Index) at transect level (Figure 2). Trying to extend these preliminaries results obtained in the transect area to entire municipality of Rome a semi-automatic methodology based on the remotely sensed data exploitation was conceived with the goal to detect the different

M1\_L classes on the basis of their spectral and textural features increasing at same time the efficiency, reliability, generality and applicability of the method. The implemented methodology consists in a original procedure based on integration between per pixel and object supervised classification steps working on polygonal objects preliminarily obtained for the entire municipal territory using a widely used multispectral segmentation algorithm. In addition to panchromatic channel three normalized spectral indices derived from Landsat ETM+ multispectral images were exploited as basic input data, with the objective to enhance the vegetation and built up typical reflectance responses, minimizing at same time the noise contribution from atmosphere and illumination sharp variations from cast shadows of buildings. In particular the usual NDVI linked to vegetation density was introduced as robust proxy of area permeability while NBUI was retained more linked to built-up typology and compactness. To improve the discrimination of these different fundamental characteristics of urban fabric, in particular those linked to built-up typology/compactness, also the spectral texture related indices derived from GLCM obtained from multispectral and panchromatic data were also included. In such a way a total of 70 independent variables were assessed for each of the 3400 polygonal areas obtained from segmentation and covering the entire municipality territory. A training set of 104 areas was then extracted from these polygonal by overlaying them with those exploited in the photointerpretation of transect features and using spatial GIS analysis to evaluate the majority M1\_L univocal class of each of them. This step was very problematic, especially for some M1\_L classes represented by too small and fragmented areas at transect level which often spread over different segmented polygonal without any prevalent class area percentage. It should be underlined also that some of these difficult classes are less representative, due their typologies (i.e. tower) including too few examples, important only for their architectural particularity. The training set of these segmented labeled polygonal and related attribute values of 70 variables was exploited to select both the best data-mining algorithms and related feature selection/compression strategies on the basis of their performance in term different accuracy and train suitability parameters. Following the general data-mining approach to deal with a big amount of input information, three classification algorithms and related input data handling strategies were selected according to their capabilities to optimize different aspects, let's say the classification accuracy in term of O.A. / ER, the number of missed M1\_L classes and the robustness against overtraining. The subsequent object classification by means of the selected combination of algorithms and related preprocessed features has been carried out on those segmented polygonal including an urban pixel percentage greater than 10%. The urban pixel distribution was previously assessed by means of an usual per pixel supervised classification using transect training areas, with a general agreement respect to comparable official thematic maps of urban areas (i.e. Urban Atlas, CLC 2006) provided by EEA(European Environment Agency) . According to the NVI semi-empirical linear model used at transect level, to quantitatively assess the thermal vulnerability for the entire territory of interest, the obtained distributions of M1\_L built up classes were transformed into their numerical counterparts using the table 3 and associated with NDVI index polygonal average, as effective proxy of permeable surface percentages. Conversely, to rightly address the remaining ambiguities about the best M1\_L distributions and the subjectivity of model weights which have been previously introduced on the basis of rough knowledge of building material used for different typologies in town districts, a more physically based approach, based on the real UHI night LST distribution captured by MODIS satellite sensors during a HW situation was implemented. Thus four models for 12-07-2003 and 16-07-2003 LST distributions were assessed using the three obtained M1\_L numerical distribution, altimetry and the 70 polygonal attributes as independent variables input which were then selected by means of backward and forward stepwise regression procedures. From the resulting statistical parameters reported in table 9, it happens that all the estimated models give a noticeable contribution to variance of the LST spatial distribution with adjusted correlation ( $R^2_{adj}$ ) going from 0,47 to 0,28 and confidence level higher than 99%. The results of backward stepwise procedures include models with a greater number of independent variables and higher correlation than those obtained by means of forward ones. Due perhaps to less cloud cover the number of cloud-free polygonal referring to LST

distribution of 12 July is increased respect to that of 16 July with consequent improvement of correlation and confidence level of the related models. Both the C4.5SN and C-SVCN distributions, as positive contributions to temperature rise (positive coefficients), were included in the models assessed through stepwise procedures while the KnnFN was excluded like altimetry. The NDVI index (always with rightly negative coefficients) and normalized mode referring to panchromatic channel (NzPan\_mode) with very small positive coefficient contributed significantly to the LST spatial variance. In the forward model of the 12 July (table 11) the NDVI and NBUI polygonal mode (rp\_mean\_0, rp\_mean\_1) and panchromatic mode (prp\_mode) provide negative contribution to the LST rise while the positive ones come from M1\_L distribution, C4.5-SN and C-SVCN (more statistically significant), normalized panchromatic mode (pnz\_rode) and finally from the GLCM entropy derived from NBUI (rp\_eopy\_1). The NBUI negative contribution could be associated to refrigerating effect from dense street shadows linked to photo-interpreted compactness level while the GLCM entropy could arise from the disordered built up concentration, especially in older city districts lacking of wind and air circulation cooling effects.

Various commercial and open sources packages and software were suitably exploited in integrated way for data processing following the above reported general schema. In particular the ERDAS-Imagine and ENVI commercial suites were used for EO and raster data processing while spatial analysis based mainly on vector coverages have been performed by means of ARC-GIS distributed by ESRI inc. and Q-GIS open source tools. Finally TerraView-Geodma and Tanagra freely available software platforms and the commercial E-Cognition package, supported the segmentation and object classification steps development.

#### 4. Conclusion

The urban heating and the formation of the UHI are typical features of the urban land transformation that are of interest across various science disciplines since they involve a broad suite of important biophysical changes of land surface, linked to urban sprawl and impacting on human health, ecosystem function, environment quality, local weather and, possibly, on the global CC characterized by also the rise of HW phenomena, as one of the most relevant effect at regional scales. The cities may be considered sensible areas to HW exposure as one of the CC impacts, their vulnerability assessment in terms of social and physical factor should be spatially assessed and possibly addressed to improve their resilience in the framework of local and national adaptation strategies. In this context the suitable spatial assessment of the thermal behavior of the different urban fabrics is an indispensable knowledge for designing the effective urban regeneration, recover, restoration and development. These activities are widely pursued by local administrations and city architects within their planning and management duties in the perspective of the environmental quality improving and energy saving within the general sustainability policies and EU framework of mitigation and adaptation strategies against the CC. Thus this work was focused on the implementation of an innovative methodology based on the integration of satellite multispectral and multiplatform remote sensing techniques in order to suitably support the extensive and operative characterization of the urban fabric thermal responses to UHI and HW, on the basis of its geometric and typological parameters, coupled with the presence of permeable soil/vegetation. Eleven typologies/compactness different classes of thermal response related to Rome urban fabric, were previously defined and recognized within a road-delimited blocks of 5X20 Km transect test area, using photointerpretation methods. They were then successfully detected over the entire territory of Rome using an on purpose developed automatic procedure, based on the ETM+ remotely sensed HR multispectral data and supervised object classification approach, optimized by means of data mining methods. This procedure includes innovative and original solutions for both classification algorithm selection and optimization of features selection/compression, in addition it includes the synergy of per pixel and objects supervised classification approaches to improve the detection and characterization of urbanized areas based on spectral and texture features.

Finally the physically-based extensive calibration/validation of the thermal vulnerability model of urban fabric, previously introduced on a semi-empirical basis at transect level, was originally accomplished by means of the LST night distributions detected by MODIS polar satellites sensors during the HW phenomena of the summer 2003 and documented by the historical series of meteorological and satellite data available, in conjunction with UHI condition, clearly evidenced in the processed frames. In such a way the effective contribution of typology/compactness built-up classes to temperature rise was confirmed and quantified like the mitigation effect of vegetation as proxy of permeable soil through NDVI. From the stepwise regression analysis also the NBUI spectral index, exploited by various authors as proxy of urban density, seemed to be able to capture the probably cooling effect provided by tall buildings shadow. Finally also various spectral and textural features derived from the panchromatic channel at 15 m. of GRS, happened significantly correlated with LST through opposite contributions to be further analyzed in term of their physical meaning and link with urban fabric parameters in a future work. Despite its established importance in LST distribution, the altimetry was non included within the more significant independent variables selected by stepwise algorithms, maybe for its too low variation range and perhaps also for uncertainties introduced in the GIS based selection of the training areas using the road-delimited transect stock areas. This former remains a critical passage in the developed procedure which will require further refinement to be better addressed using, for instance, a preliminary segmentation before the photointerpretation step.

Considering the dramatically augmented availability of the HR multispectral and SAR data provided by increasing number of operative satellite remote sensing missions and more and more suitable for characterizing the increasingly heterogeneous, complexes and wide urban areas [55], the preliminary results obtained in the present work may provide a robust contribution for developing advanced applications to face the growing needs of decision support knowledge about the different urban processes which more and more impact on citizen life and threatened urban environment, exploiting the big amount of geo-information and data provided by the remote sensing techniques, continuously improved not only in terms of spatial/spectral resolution and radiometry of sensors, but also in temporal coverage capability, innovative platforms (i.e. microsatellites, drones,...) and biophysical parameters detected, in conjunction with data mining and clustering algorithms advances.

This work was developed in the frame of the Project "RoMA", funded by Italian MIUR (Ministry of the University and Research), SCN\_00064 Call "Smart Cities and Communities".

We would like to thank U.S. Department of the Interior, NASA and U.S. Geological Survey for providing the publically accessible Landsat and MODIS data and product via respectively <http://earthexplorer.usgs.gov/> and <http://modis.gsfc.nasa.gov/>.

Author Contributions: Ombuen S., Barbieri L., Benelli F., Camerata F., Pellegrini V., Filpa A. designed and developed the vulnerability semi-empirical model for providing the built up categorization and photointerpretation of the transect area. Borfecchia F. conceived the basic ideas and overall schema for this study, supervised all the processing steps and wrote the first version of this manuscript. Caiaffa E., Rosato V., Pollino M. implemented all the procedures for spatial analyses based on GIS tools and methods, in addition they accomplished the English language reviews of the paper. De Cecco L., Martini S., La Porta L., performed processing and homogenization of MODIS and Landsat remotely sensed data.

Conflicts of Interest: The authors declare no conflict of interest.

## References

1. Giordano, L.; Giordano, F.; Grauso S.; Iannetta, M.; Sciortino, M.; Bonnati, G.; Borfecchia, F. Desertification vulnerability in Sicily. Proc. Of the 2nd Int. Conf. On New Trend in Water and

- Environmental Engineering for Safety and Life: Eco-Compatible Solution for Aquatic Environments, 2002.
2. Barriopedro, D.; Fischer, E. M.; Luterbacher, J.; Trigo, R. M.; Garcia-Herrera, R. The hot summer of 2010: Redrawing the temperature record map of Europe', *Science*, 2011, 332(6026), 220–224.
  3. EEA Report No 2/2012. Urban adaptation to climate change in Europe, Challenges and opportunities for cities together with supportive national and European policies. SBN 978-92-9213-308-5. doi:10.2800/41895.
  4. Borfecchia, F.; De Cecco, L.; Pollino, M.; Martini, S.; La Porta, L.; Zinzi, M.; Carnielo, E. Telerilevamento HR Multispettrale/Lidar e Modellistica SVAT 3d per la stima del bilancio Energetico in Ambiente Urbano. Proc. 15th National Conference ASITA , 2011, 449- 460, ISBN 978-88-903132-6-4. 15-18, Novembre Reggio di Colorno, Parma – Italy.
  5. Oke, T.R. The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 1982,108(455), 1-24. <http://onlinelibrary.wiley.com/doi/10.1002/qj.49710845502/pdf>.
  6. 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.
  7. Rajasekar, U.; Weng, Q. Urban heat island monitoring and analysis using a non-parametric model: A case study of Indianapolis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2009, 64 (1), 86-96.
  8. Imhoff, M. L.; Zhang, P.; Wolfe, R. E.; Bounoua, L. Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sensing of Environment*, 2010, 114, 504–513.
  9. Gabriel, KM.; Endlicher WR. Urban and rural mortality rates during heat waves in Berlin and Brandenburg, Germany. *Environmental Pollution*, 2011, 159(8-9), 2044-2050.
  10. IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 582.
  11. Michelozzi, P.; De’Donato, F.; Bisanti, L.; Russo, A.; Cadum, E.; De Maria, M.; D’Ovidio, M.; Costa, G.; Perucci, C. A. The impact of the summer 2003 heat waves on mortality in four Italian cities. *Eurosurveillance, Surveillance report*, 2005, 10 (7), n. 6.
  12. Mackey, C.W.; Lee, X.; Smith, R.B. Remotely sensing the cooling effects of city scale efforts to reduce urban heat island. *Building and Environment*, 2012, 49, 348-358.
  13. Borfecchia, F. Impatto dei cambiamenti climatici sulle città: ondate di calore e UHI. Il clima cambia le città. In *Strategie di adattamento e mitigazione nella pianificazione urbanistica*. Musco, F.; Zanchini, E. Franco Angeli/Urbanistica, 2014, pp.416 Codice ISBN: 9788820487232.
  14. Inostroza, L.; Baur, R.; Csaplovics, E. Urban sprawl and fragmentation in Latin America : A dynamic quantification and characterization of spatial patterns. *Journal of Environmental Management*, 2013, 115, 87-97.
  15. Fabrizi, R.; Bonafoni, S.; Biondi, R. Satellite and Ground-Based Sensors for the Urban Heat Island Analysis in the City of Rome. *Remote Sens.* 2010, 2(5), 1400-1415; doi:10.3390/rs2051400.
  16. Santamouris M. Advances in Building Energy Research, Volume 1, Earthscan. Copyright. London-Sterling – VA, 2007, ISBN 978-1-84407-389-4.
  17. Novack, T.; Esch, T.; Kux, H.; Stilla, U. Machine Learning Comparison between WorldView-2 and QuickBird-2-Simulated Imagery Regarding Object-Based Urban Land Cover Classification. *Remote Sens.* 2011, 3, 2263-2282; doi:10.3390/rs3102263.
  18. Li, C.; Wang, J.; Wang, L.; Hu, L.; Gong, P. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. *Remote Sens.* 2014, 6, 964-983; doi:10.3390/rs6020964.
  19. Michishita, R.; Jiang, Z.; Xu, B. Monitoring two decades of urbanization in the Poyang Lake area, China through spectral unmixing. *Remote Sensing of Environment*, 2012, 117, 3–18.
  20. Weng, Q. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, 2012, 117, 34–49
  21. Kaspersen, P. S.; Fensholt, R.; Drews, M. Using Landsat Vegetation Indices to Estimate Impervious Surface Fractions for European Cities. *Remote Sens.* 2015, 7, 8224-8249; doi:10.3390/rs70608224.

22. Parece, T. E.; Campbell J. B. Comparing Urban Impervious Surface Identification Using Landsat and High Resolution Aerial Photography. *Remote Sens.* 2013, 5, 4942-4960; doi:10.3390/rs5104942.
23. Scott, D.; Petropoulos, G. P.; Moxley, J.; Malcolm, H. Quantifying the Physical Composition of Urban Morphology throughout Wales Based on the Time Series (1989–2011) Analysis of Landsat TM/ETM+ Images and Supporting GIS Data. *Remote Sens.* 2014, 6, 11731-11752; doi:10.3390/rs61211731.
24. Borfecchia, F.; Caiaffa, E.; Pollino, M.; De Cecco, L.; La Porta, L.; Ombuen, S.; Barbieri, L.; Benelli, F.; Camerata, F.; Pellegrini, V.; Filpa, A. Assessment della vulnerabilità del tessuto urbano a heat waves ed UHI tramite tecniche di Remote Sensing ed object classification,. Proc.18th ASITA Conference, 14-16 Oct. Florence,. 2014, pp. 187–194, ISBN 978-88-903132-9-5.
25. Borfecchia, F.; Rosato, V.; Caiaffa, E.; Pollino, M.; De Cecco, L.; La Porta, L.; Ombuen, S.; Barbieri, L.; Benelli, F.; Camerata, F.; Pellegrin, V.; Filpa, A. Assessing the Urban Fabric Vulnerability To Heat Waves And UHI using Remote Sensing and Object Classification. IEEE International Geoscience & Remote Sensing Symposium, 2015, Catalog Number: CFP15IGA-USB, ISBN: 978-1-4799-7928-8.
26. Wang, J.; Huang, B.; Fu, D.; Atkinson P. M. Spatiotemporal Variation in Surface Urban Heat Island Intensity and Associated Determinants across Major Chinese Cities. *Remote Sens.* 2015, 7, 3670-3689; doi:10.3390/rs70403670.
27. Stathopoulou, M.; Cartalis, C. Daytime urban heat islands from Landsat ETM+ and Corine land cover data: An application to major cities in Greece. *Solar Energy*, 2007, 81, 358–368.
28. Odindi, J. O.; Bangamwabo, V. ; Mutanga, O. Assessing the Value of Urban Green Spaces in Mitigating Multi-Seasonal Urban Heat using MODIS Land SurfaceTemperature (LST) and Landsat 8 data. *Int. J. Environ. Res.* 2015, 9(1):9-18, ISSN: 1735-6865.
29. Keramitsoglou, I.; Kiranoudis, C. T.; Ceriola, G.; Weng, Q.; Rajasekar, U. Identification and analysis of urban surface temperature patterns in Greater Athens, Greece, using MODIS imagery. *Remote Sensing of Environment*, 2011, 115(12), 3080–3090.
30. Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z. L. Quality assessment and validation of the MODIS global land surface temperature. *Int. J. Remote sensing* , 2004, 25(1), 261–274.
31. Wan Z. New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sensing of Environment*, 2008, 112, 59–74.
32. D’Onofrio Rosalba, Talia Michele . La rigenerazione urbana alla prova. Franco Angeli Editore. Collana: Urbanistica: Studi urbani e regionali. Marzo 2014. ISBN-13: 978-88-917-0960-8.
33. Filpa, A.; Ombuen, S. Comprendere i cambiamenti climatici. Pianificare per l’adattamento. Understanding climate change. Planning for adaptation,. *Urbanistica*, 2014, 5 (2); ISSN: 1973-9702.
34. Ceccarelli, T.; Smiraglia D.; Bajocco S.; Rinaldo S.; De Angelis A.; Salvati L.; Perini L. Land cover data from Landsat single-date imagery: an approach integrating pixel-based and object-based classifiers. *European Journal of Remote Sensing* , 2013, 46, 699-717.
35. Hussain , M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing* , 2013, 80,91–106.
36. Berger, C.; Voltersen, M.; Hese, S.; Walde, I.; Schullius, C. Using Geographic Object-Based Image Analysis (Geobia) for Urban Land Cover Mapping and Settlement Density Assessment. Proceedings of the 4th GEOBIA, May 7-9, 2012 - Rio de Janeiro – Brazil, 2012, .503-508.
37. Liu, L.; Zhang, Y. Urban Heat Island Analysis Using the Landsat TM Data and ASTER Data: A Case Study in Hong Kong. *Remote Sens.* 2011, 3, 1535-1552; doi:10.3390/rs3071535.
38. Baatz, M.; Schäpe, A. Multiresolution Segmentation - an optimisation approach for high quality multi-scale image segmentation” Strobl, J., et al. (Eds.): *Angewandte Geographische Informationsverarbeitung*. Herbert Wichman Karlsruhe 2000, Vol XII, 12-23.
39. Jungo, I.; Zhenyu, L.; Jinyoung, R.; Lindi, J. Q. Impervious surface quantification using a synthesis of artificial immune networks and decision/regression trees from multi-sensor data, *Remote Sensing of Environment*. 2012, 117, 102–113.
40. Zhang, C.; Li, W.; Travis, D. Gaps-fill of SLC-off Landsat ETM+ satellite image using a geostatistical approach. *International Journal of Remote Sensing*, 2007, 28, 5103–5122
41. Chen, J.; Zhu, X.; Vogelmann, J. E.; Gao, F.; Jin, S. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. *Remote Sensing of Environment*, 2011, 115, 1053-1064.

42. Zhang, C.; Li, P.; Wang J. Urban Built-Up Area Extraction from Landsat TM/ETM+ Images Using Spectral Information and Multivariate Texture. *Remote Sens.* 2014, 6, 7339-7359; doi:10.3390/rs6087339.
43. Xu, H. Extraction of Urban Built-up Land Features from Landsat Imagery Using a Thematic oriented Index Combination Technique. *Photogrammetric Engineering & Remote Sensing*, 2007, 73(12), 1380–1391.
44. Todd, S.W.; Hoffer, R.M.; Milchunas D.G. Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International Journal of Remote Sensing*, 1998, 19, 427–438.
45. Salehi, B.; Zhang, Y.; Zhong, M.; Dey, V. Object-Based Classification of Urban Areas Using VHR Imagery and Height Points Ancillary Data. *Remote Sens.* 2012, 4, 2256-2276; doi:10.3390/rs4082256.
46. Zhang, H.; Jing, X. M.; Chen, J. Y.; Li, J. J.; Schwegler, B. Characterizing Urban Fabric Properties and Their Thermal Effect Using QuickBird Image and Landsat 8 Thermal Infrared (TIR) Data: The Case of Downtown Shanghai, China. *Remote Sens.* 2016, 8(7), 541; doi:10.3390/rs8070541.
47. Keramitsoglou, I.; Kiranoudis, C. T.; Ceriola, G.; Weng, Q.; Rajasekar, U. Identification and analysis of urban surface temperature patterns in Greater Athens, Greece, using MODIS imagery. *Remote Sensing of Environment*, 2011, 115(12), 3080–3090.
48. Quinlan, R. C4.5: Programs for Machine Learning; Morgan Kaufmann Publishers: San Mateo, CA, USA, 1993; p. 302.
49. Pinho, C. M. D.; Silva, F. C.; Fonseca, L. M. C.; Monteiro, A. M. V.. Intra-Urban Land Cover Classification From High-Resolution Images Using The C4.5 Algorithm. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B7. Beijing ,2008.
50. Wieland, M.; Pittore M. Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images. *Remote Sens.* 2014, 6, 2912-2939; doi:10.3390/rs6042912.
51. Breiman, L. Random Forests. *Machine Learning* . 2001, 45 (1), 5–32. doi:10.1023/A:1010933404324.
52. Rakotomalala , R. TANAGRA: a free data mining software for research and academic purposes. Proceedings of European Grid Conference, 2005, RNTI-E-3, 2, 697-702, Amsterdam, February 2005.
53. Webb, A. R. Statistical Pattern Recognition. Second Edition, QinetiQ Ltd., Malvern, UK, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England., 2002.
54. Miller, A. Subset Selection in Regression. 2002, Chapman and Hall/CRC, ISBN 9781584881711.
55. Poursanidis, D.; Chrysoulakis, N.; Mitraka, Z. Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urban land cover mapping. *International Journal of Applied Earth Observation and Geoinformation*, 2015, 35, 259–269.



© 2016 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).