



# An urban energy balance-guided machine learning approach for synthetic nocturnal surface Urban Heat Island prediction: A heatwave event in Naples

Ana Oliveira <sup>a,\*</sup>, António Lopes <sup>b</sup>, Samuel Niza <sup>a</sup>, Amílcar Soares <sup>c</sup>

<sup>a</sup> IN+ Center for Innovation, Technology and Policy Research, Instituto Superior Técnico, Universidade de Lisboa, Portugal

<sup>b</sup> Centro de Estudos Geográficos, IGOT - Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, Portugal

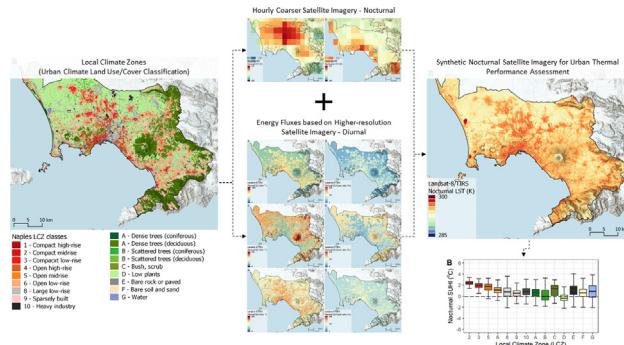
<sup>c</sup> CERENA, Instituto Superior Técnico, Universidade de Lisboa, Portugal



## HIGHLIGHTS

- Satellite thermal imagery offers insights into the urban climate performance.
- Those insights are constrained by long repeat cycles and diurnal overpass time.
- Land use/land cover (LULC) classes have different nocturnal thermal performances.
- Heat Fluxes (HF) disclose these nocturnal land surface temperature (LST) contrasts.
- Through machine learning, LULC and HF, synthetic nocturnal LST is predicted.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 4 July 2021

Received in revised form 19 August 2021

Accepted 31 August 2021

Available online 6 September 2021

Editor: Pavlos Kassomenos

### Keywords:

Urban climate adaptation  
Heatwave  
Urban Heat Island  
Land surface temperature  
Local climate zones  
Random forest  
Multisensor data fusion  
Satellite thermal imagery

## ABSTRACT

Southern European functional urban areas (FUAs) are increasingly subject to heatwave (HW) events, calling for anticipated climate adaptation measures. In the urban context, such adaptation strategies require a thorough understanding of the built-up response to the incoming solar radiation, i.e., the urban energy balance cycle and its implications for the Urban Heat Island (UHI) effect. Despite readily available, diurnal Land Surface Temperature (LST) data does not provide a meaningful picture of the UHI, in these midlatitudes FUAs. On the contrary, the mid-morning satellite overpass is characterized by the absence of a significant surface UHI (SUHI) signal, corresponding to the period of the day when the urban-rural air temperature difference is typically negative. Conversely, nocturnal high-resolution LST data is rarely available. In this study, an energy balance-based machine learning approach is explored, considering the Local Climate Zones (LCZ), to describe the daily cycle of the heat flux components and predict the nocturnal SUHI, during an HW event. While the urban and rural spatial outlines are not visible in the diurnal thermal image, they become apparent in the latent and storage heat flux maps – built-up infrastructures uptake heat during the day which is released back into the atmosphere, during the night, whereas vegetation land surfaces loose diurnal heat through evapotranspiration. For the LST prediction model, a random forest (RF) approach is implemented. RF results show that the model accurately predicts the LST, ensuring mean square errors inferior to 0.1 K. Both the latent and storage heat flux components, together with LCZ classification, are the most important explanatory variables for the nocturnal LST prediction, supporting the adoption of the energy balance approach. In future research, other locations and time-series data shall be trained and tested, providing an efficient local urban climate monitoring tool, where in-situ air temperature observations are not available.

© 2021 Elsevier B.V. All rights reserved.

\* Corresponding author.

E-mail addresses: [anappm oliveira@tecnico.ulisboa.pt](mailto:anappm oliveira@tecnico.ulisboa.pt) (A. Oliveira), [antonio.lopes@campus.ul.pt](mailto:antonio.lopes@campus.ul.pt) (A. Lopes), [samuel.niza@tecnico.ulisboa.pt](mailto:samuel.niza@tecnico.ulisboa.pt) (S. Niza), [asoares@tecnico.ulisboa.pt](mailto:asoares@tecnico.ulisboa.pt) (A. Soares).

## 1. Introduction

In the last decades, urban growth has become a major global trend – cities have been increasing in size, number, economic importance, and population density (United Nations, 2014). Accordingly, all aspects of urban sustainability have been gaining increasing attention, namely those related to climate change mitigation and adaptation responses (Major et al., 2011; World Bank, 2010). In fact, in the scope of regional climate, urban development is the main culprit of one of the most recognisable man-made alterations to the local near-surface air temperatures, the so-called Urban Heat Island (UHI) – a phenomenon firstly described for the city of London by Howard in the early 1800s (Howard, 1833), which is defined as the air temperature difference between the city centre and the surrounding non-urban/background areas (Oke, 1982, 1987, 1988; Oke et al., 2017a, 2017b). Two atmospheric vertical levels may be referred to in UHI studies: the urban boundary layer (UBL) UHI, corresponding to the air temperature difference in the fraction of the atmospheric boundary layer which is influenced by an urban presence; and the urban canopy layer (UCL) UHI, which corresponds to the air temperature difference between the ground surface and the buildings' mean height (Oke et al., 2017a, 2017b).

The majority of UHI studies measure its intensity and/or describe its patterns at the UCL level, not only due to the greater availability of air temperature observations near the surface but also because it is where people are exposed to its impacts (Oke, 1987; Oke et al., 2017a, 2017b). While the UCL positive air temperature anomalies are mostly found during the nocturnal period, the main cause for the UHI stems from the land-use/land cover (LULC) changes (and corresponding physical properties), as the urban artificial materials respond differently to the diurnal incoming solar radiation, compared to more natural/rural surfaces. As previously mentioned by several authors (e.g., Chrysoulakis et al., 2016, 2018; Feigenwinter et al., 2018; Kotthaus and Grimmond, 2014a, 2014b; Lopes, 2003; Nadeau et al., 2009; Oke, 1988; Rigo and Parlow, 2007; Wang et al., 2010), during the day, impervious/artificial urban surfaces are responsible for a greater absolute storage heat flux component (compared to rural/natural surfaces); in addition, at the city-level, urban areas are typically poorly ventilated (due to greater roughness length from the built infrastructures), while evapotranspiration is reduced, due to lower vegetation cover; during the night, the diurnal storage heat flux is progressively released back into the atmosphere, reducing the nocturnal cooling rate, compared to rural sites (Oke, 1982, 1988; Oke et al., 2017a, 2017b). An additional contribution to the UHI is the heat emitted due to anthropogenic activities, such as air conditioning systems and vehicles exhaust, although typically being a smaller contributor to the UHI intensity (Oke et al., 2017a, 2017b). In the last decades, the UHI effect has gained increasing attention, as both the climate change projections and the recent climate extremes trends have led to the prioritization of heat-related risk assessment actions, in urban settings (Baklanov et al., 2018; Mills, 2014; Oke et al., 2017a, 2017b; Stewart and Oke, 2012). Accordingly, urban climate case studies have multiplied in the last decades, covering all latitudes and longitudes (e.g., Anderson et al., 2018; Chrysoulakis et al., 2018; Emmanuel and Krüger, 2012; Freitas et al., 2007; Kotthaus and Grimmond, 2014b; Kourtidis et al., 2015; Lemonis et al., 2004; Liu and Zhang, 2011; Parlow, 2003; Shi et al., 2018; Wicki and Parlow, 2017).

Observational studies describing the processes, spatial patterns and/or magnitude of the atmospheric UHI is typically the aim of urban climate literature; however, air temperature intra-urban measurements are not always readily available with sufficient spatial coverage, temporal continuity and/or quality (Chapman et al., 2017; Meier et al., 2015; Muller et al., 2013; Napoly et al., 2018). Indeed, most cities do not have in-situ networks dense enough to provide air temperature datasets with adequate spatial resolution. On the other hand, freely available remote sensing satellite thermal imagery (such as those acquired by the Landsat, MODIS, or Sentinel-3 missions) has provided readily available thermal data, at much greater spatial resolution

(from 100 m to 1 km), ensuring global coverage and continuity, providing valuable inputs to many case studies (Bechtel et al., 2019; Cai et al., 2018; Liu and Zhang, 2011; Lopes, 2003; Parlow et al., 2014; Ramos and Cladera, 2016; Wicki et al., 2018; Wicki and Parlow, 2017). However, it should be noticed that this third UHI type, the surface urban heat island (SUHI), has quite specific spatio-temporal patterns and magnitudes, which should not be considered representative of the atmospheric ones (Lopes, 2003; Oke et al., 2017a, 2017b; Parlow, 2003; Parlow et al., 2014; Rigo et al., 2006). As the SUHI magnitude depicts the urban-rural differences of the land surface temperature (LST), a SUHI effect may be present also during the day, as satellite imagery provides a bird's eye view, and corresponding thermal imagery is greatly influenced by both the street-level and the rooftop surface materials.

The main issue in using LST for UHI assessment is related to the granularity versus spatial resolution dilemma (Anderson et al., 2021), and how it restricts the SUHI usage as an atmospheric UHI proxy. For urban studies, the sub-kilometre spatial resolution is of greatest interest, to allow the thorough recognition of the different urban fabric thermal performances. Such spatial resolution is usually provided by Low Earth Orbit (LEO) satellites, but readily available data correspond only to the thermal imagery acquired by the sun-synchronous Landsat mission satellites (e.g., Landsat 5, Landsat 7 and Landsat 8), but at the cost of large revisit times (less than 2 images per month) (USGS, 2016). In addition, only the scenes acquired during the diurnal descending orbit of the Landsat satellites are made readily available to the public. These correspond are acquired during the mid-morning hours of the day in midlatitude regions, which is not suitable for the atmospheric UHI assessment – at this time of the day, the atmospheric anomaly is typically absent, as the urban surfaces already emitted the previous day's stored heat, during the night, and have not yet received enough additional solar radiation to provide a clear 'urban signal' (Oke et al., 2017a, 2017b; Oliveira et al., 2020c).

Alternatively, nocturnal thermal imagery is available from other missions, such as those acquired by the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (Terra/Aqua MODIS) or the Sentinels 3 Sea Land Surface Temperature Radiometer (Sentinel 3 SLSTR). All these are LEO satellites as well, and they provide two images per day, but at the cost of a lower spatial resolution – circa 1 km-pixel size. Nonetheless, their nocturnal Land Surface Temperature (LST) products have been shown to provide good insights into the UHI effect in larger cities or metropolitan areas, even if with a lower level of detail (Shumilo et al., 2019; Ünal et al., 2020; Zhang et al., 2014). Finally, satellites in Geostationary Orbit (GEO), such as those from the GOES or Meteosat missions, provide as much as sub-hourly granularity, but with a 5 km pixel size, which is less suitable for intra-urban analysis.

Recently, several data fusion/data augmentation techniques have been developed to answer these and other satellite data acquisition gaps (Ghamisi et al., 2018; Zhang, 2010) – a data processing technique that aims to integrate multiple data sources into one combined data output. In satellite remote sensing, data fusion techniques are often used to merge imagery from different sensors, to fill in knowledge gaps and improve the original data information level (Ghamisi et al., 2018; Salcedo-Sanz et al., 2020; Zhang, 2010) – these can include fill in missing data, enhancement of the spatial level of detail, or improvement in the temporal granularity. Several techniques have been tested by multiple authors, including statistical downscaling, spatiotemporal data fusion, or combining satellite data with physical models to reconstruct the full daily cycle of the LST. Several examples of data augmentation workflows to retrieve synthetic thermal imagery have been reported in the literature, mostly focusing on data improvement (e.g., filling data gaps due to cloud cover, or retrieving higher granularity data). These are typically developed at either the global or regional scale, aiming for a kilometric pixel size (e.g., Hong et al., 2021; Jia et al., 2021; Long et al., 2020; Tan et al., 2021; Xiao et al., 2021; Xu et al., 2021a; Xu and Cheng, 2021; Zhang et al., 2020, 2021; Zhao and Duan, 2020).

Fewer cases are focused on the urban context and using sub-kilometric imagery - due to the above-mentioned sizable revisit times of the Landsat missions, these studies are often focused on improving the Landsat 8 temporal granularity, by using coarser LST to predict Landsat-like diurnal images (e.g., Cheval et al., 2020; Kesavan et al., 2021; Peng et al., 2019; Sharma et al., 2020; Shen et al., 2021; Wang et al., 2019, 2020, 2021; Zhao et al., 2020b). Conversely, studies using Landsat-8 nocturnal thermal imagery are quite rare, as these are scarcely available, globally (e.g., Wang et al., 2020).

Despite the diurnal acquisition-related constraints, Landsat-based diurnal surface heat fluxes have been shown to provide greater insights on the culprit of the atmospheric UHI signal – particularly, the diurnal storage heat flux has been noted for being much greater in impervious/artificial urban surfaces during the mid-morning overpass, which partially contributes to the greater availability of nocturnal sensible heat in the atmosphere (Lopes, 2003; Oke, 1982, 1988; Oke et al., 2017a, 2017b; Parlow, 2003; Parlow et al., 2014; Rigo and Parlow, 2007). The diurnal latent heat flux is also affected by the absence of extensive green coverage and is reduced in urban areas, which negatively affects natural cooling processes. Hence, the spatial patterns of the diurnal heat flux components are known to explain the corresponding nocturnal atmospheric temperature anomaly (Lopes, 2003; Oke, 1982, 1988; Oke et al., 2017a, 2017b; Parlow, 2003; Parlow et al., 2014; Rigo and Parlow, 2007).

Following previously published methodologies for satellite-based urban fluxes retrieval (e.g., Chrysoulakis et al., 2018; Feigenwinter et al., 2018; Parlow, 2003; Parlow et al., 2014; Rigo and Parlow, 2007; Wicki et al., 2018; Wicki and Parlow, 2017), in this study, the urban energy balance components are calculated for diurnal Landsat-8 imagery and Meteosat's Spinning Enhanced Visible and InfraRed Imager (MSG-SEVIRI), to investigate and disclose its relation with the nocturnal SUHI, during an HW event in Naples. In addition, a data fusion approach is tested to develop a sub-kilometric synthetic nocturnal LST prediction model, based solely on diurnal high-resolution heat flux components, and nocturnal kilometric thermal imagery.

While the retrieval of the urban energy balance components from satellite data is quite a mature technique, and several examples exist in the peer-reviewed literature (e.g., Chrysoulakis et al., 2016, 2018; Ezimand et al., 2021; Parlow, 2003; Parlow et al., 2014; Rigo and Parlow, 2007), its applicability for urban-relevant synthetic LST nocturnal imagery retrieval has not been tested. Here, a machine learning approach is tested, using a Random Forest (RF) algorithm to retrieve Landsat-equivalent nocturnal LST synthetic imagery, by combining geographic features, diurnal Landsat energy fluxes and the nocturnal LST from much the coarser MSG-SEVIRI satellite sensor. The aim is to develop and test an energy balance-based machine learning approach, to describe the daily cycle of the heat flux components and predict the nocturnal LST and SUHI, during an HW event. The study is focused on the Naples Functional Urban Area (FUA) region, during a specific heatwave (HW) day – the 9th of August 2017. The HW day selection follows a previously published work (Oliveira et al., 2020c), in which the Excess Heat Factor (EHF) index was used to identify extreme temperature events corresponding to Landsat-8 overpass days.

Previous work considered the interpretation of the diurnal LST patterns, per Local Climate Zones (LCZ) class, across five case study FUAs – the results highlighted the here-mentioned diurnal LST constraints for urban climate applications. This follow up study and the LST prediction model is now made possible due to the availability of a nocturnal (ascending orbit) cloud-free scene in the Landsat8 database, corresponding to the day of interest – such nocturnal imagery is rarely made available to the community, hence the need for alternatives. Hence, the nocturnal LST is used as the response variable, to train the RF regression model, guided by the physics-based energy balance components. The storage, sensible and latent heat fluxes are calculated from the visible and thermal bands of the equivalent diurnal scene (descending orbit), and used in the model, together with several geographical

and land cover factors, as the LST predictors. In addition, air temperature observations and coarser-resolution MSG/SEVIRI-based products (short and longwave radiation, albedo and LST) are used for preliminary exploratory data analysis on the full daily cycles of the storage heat flux, LST and air temperature, for comparison purposes.

## 2. Data and methods

### 2.1. Air temperature data and heatwave identification

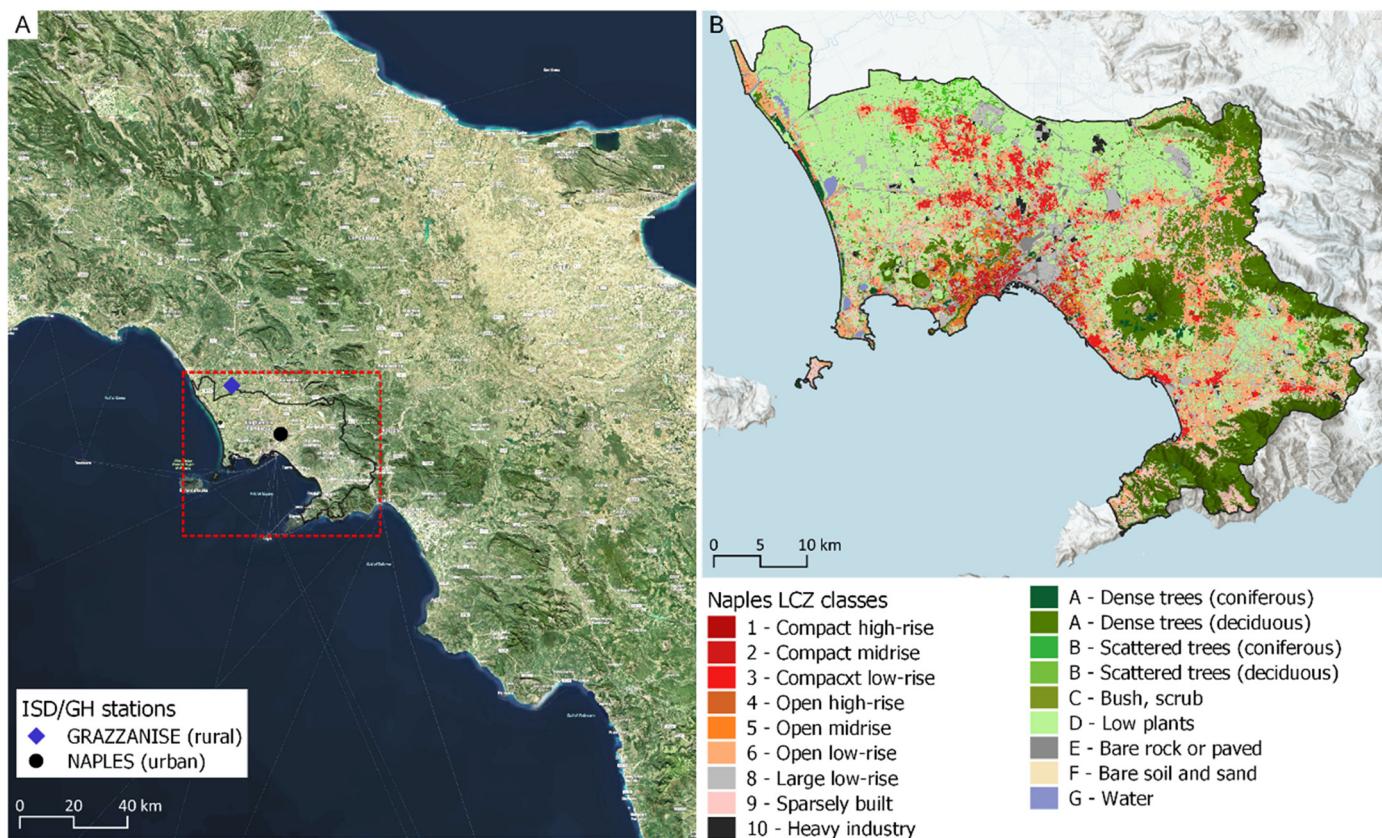
The Naples Functional Urban Area (FUA) is located in the western shoreline of the Italian Peninsula, at circa 40.90 N latitude and 14.10E longitude, being the capital of the Campania region and the third-largest Italian municipality (see Fig. 1). Its characterized by a hot dry-summer Mediterranean climate type (Mediterranean Csa) (Peel et al., 2007). While greater summer temperatures are typically in this European biogeographical region (compared to northern latitudes), recent studies have also shown that extreme temperature events are becoming increasingly more frequent. Such is the case of HW events detected through the Excess Heat Factor (EHF)-based – a percentile-based HW index that has been developed to provide a quadratic measure of the air temperature anomaly, relative to the long-term regional climate, accounting also for the short-term acclimatization (calculation details in Nairn et al., 2009; Nairn and Fawcett, 2013; Perkins, 2015). In addition, Naples demographic structure is characterized by an ageing population, which, together with the regions 20% annual probability of severe HW occurrence, puts this urban area in 13th place, amongst European FUA's (Oliveira et al., n.d.)

Here, air temperature data, from the Grazzanise station (name: GRAZZANISE, WMO code: 16253, rural site, 41.06 N latitude, 14.07E longitude, 9 m elevation) (1950 to 2018), was retrieved from the Integrated Surface Database – Global Hourly observations (ISD/GH), available at the National Centers for Environmental Information of the National Oceanic and Atmospheric Administration (NOAA) (Lott, 2004), and the 1961–1990 period used for the 90th threshold calculation. Days, when positive EHF occurred, were used to search for a set of diurnal and nocturnal imagery, in the Landsat 8 database – the 9th of August 2017 was chosen due to positive 3-days EHF values, elevated air temperatures (the maximum air temperature was 38 °C, at the Naples city station), and cloud-free diurnal and nocturnal imagery available. Hourly data from one additional urban station was also retrieved for the UHI magnitude analysis: the Naples Naval Air Station (name: NAPLES, WMO code: 16289, urban site, 40.09 N latitude, 14.30E longitude, 72 m elevation).

### 2.2. Remote sensing data

In this study, two cloud-free Landsat 8 scenes are retrieved from the United States Geological Survey EarthExplorer database (United States Geological Service, 2020; USGS, 2019), encompassing the study area – a mid-morning scene, corresponding to the satellite's descending orbit, and a nocturnal scene, corresponding to the satellite's ascending orbit. The scenes are obtained as Landsat 8 level 1 product, including 11 bands registered by the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments onboard the satellite: 8 bands in the visible (VIS), near-infrared (NIR) and shortwave infrared (SWIR) range, with 30 m resolution, a panchromatic band, with 15 m resolution, and 2 thermal infrared (TIRS) bands, with 100 m resolution (resampled to 30 m). Scene details are provided in Table 1.

The original level-1 data is delivered in 16-bit unsigned integer Digital Numbers (DN) format, without atmospheric correction. Hence, following the Data User's Manual (USGS, 2016), level 1 bands are translated into Top of Atmosphere (TOA) radiance and reflectance, and TOA Brightness Temperature (TOA-BT), using the rescaling coefficients available in each scene metadata files. The TOA reflectance is translated into surface reflectance (SR) using Dark Object Subtraction



**Fig. 1.** Naples Functional Urban Area (FUA): (A) location in the context of the central Italian Peninsula, identifying the available in-situ air temperature ( $T_{air}$ ) measurement sites, from the ISD/GH database; (B) Local Climate Zones classification as per (Oliveira et al., 2020a; Oliveira et al., 2020c), reclassified according to the average building height as per Italian General Population and Housing Censuses (ISTAT, 2018).

atmospheric correction (Chavez, 1996), implemented in the open-source geographic information system (GIS) QGIS software (version 3.10) Semi-Automatic Classification plugin (Congedo, 2019). Surface temperature (LST) is obtained from the TOA-BT, as per Eq. (1) – the physical-based atmospheric correction approach is conducted by the introduction of the atmospheric transmissivity, upwelling/atmospheric path radiance and downwelling/sky radiance parameters, corresponding to satellite overpass time – this information was retrieved from the web-based NASA Atmospheric Calculator (Barsi et al., 2004, 2005). The TOA-BT correction and conversion steps were also implemented in the QGIS software (version 3.10), through the raster calculator tool.

$$L_T = \tau \varepsilon L_\lambda + L_u + \tau(1 - \varepsilon) L_d \quad (1)$$

where  $L_T$  is the surface radiance,  $\tau$  is the atmospheric transmissivity,  $L_\lambda$  is the TOA radiance,  $L_u$  is the upwelling/atmospheric path radiance,  $L_d$  is the downwelling/sky radiance, and  $\varepsilon$  is the pixel-wise emissivity map. The  $\varepsilon$  parameter was calculated according to a normalized difference vegetation index (NDVI)-based approach (Sobrino et al., 2008), as modified by (Wicki and Parlow, 2017).

Accordingly, the modified methodology considers two additional Landsat 8-based indices, more appropriate for urbanized landscapes: the natural deviation water index (NDWI) and the natural deviation built-up index (NDBI). These indices are all based on normalized ratios between the VIS, NIR and SWIR bands, as per Eqs. (2), (3), and (4) (subscripts correspond to the centre wavelength of the corresponding Landsat 8 bands) and are used as the conditional arguments for pixel-wise reclassification into standard emissivity values. They are also considered as candidate predictors for the LST prediction model. The reclassification approach is described in detail in Wicki and Parlow (2017).

$$NDVI = \frac{NIR_{0.86} - VIS_{0.65}}{NIR_{0.86} + VIS_{0.65}} \quad (2)$$

$$NDWI = \frac{VIS_{0.56} - NIR_{0.86}}{VIS_{0.56} + NIR_{0.86}} \quad (3)$$

$$NDBI = \frac{SWIR_{1.61} - NIR_{0.86}}{SWIR_{01.61} + NIR_{0.86}} \quad (4)$$

**Table 1**

List of remote sensing data used in this study.

Acquisition date	Acquisition time	Satellite and sensor	Spatial Resolution	Scene/product Id	Source
09 Aug 2017	09:41 20:46	Landsat 8 OLI/TIRS	30 m (VIS, NIR, SWIR) 100 m (TIR)	LC81890322017221LGN00 LC80522122017221LGN00	USGS EarthExplorer, available at <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> (retrieved in August 2018)
09 Aug 2017	Every 30 min	MSG/SEVERI	5 km	MDSSF, LSA-201	EUMETSAT's Satellite Application Facility on Land Surface Analysis, available at <a href="https://landsaf.ipma.pt/">https://landsaf.ipma.pt/</a> (retrieved in May 2020)
09 Aug 2017	Every 30 min		5 km	MDSLF, LSA-204	
09 Aug 2017	5-days average		5 km	MDAL, LSA-101	
09 Aug 2017	Every 15 min		5 km	MLST, LSA-001/LSA-004	

Finally, LST can be calculated as per Eq. (5), in which  $k_1$  and  $k_2$  are constants provided in the Landsat 8 image metadata file, and  $LST$  is the surface radiance obtained from Eq. (3).

$$LST = \frac{k_2}{\ln\left(\frac{k_1}{LST} + 1\right)} \quad (5)$$

To complement the analysis, the downwelling radiation (Carrer et al., 2019a, 2019b; Josey et al., 2003; Prata, 1996), LST (Freitas et al., 2010; Götsche et al., 2016) and Albedo (Carrer et al., 2018) products from the MSG-SEVIRI were retrieved from the EUMETSAT's Satellite Application Facility on Land Surface Analysis (LSA SAF), for comparison purposes (see Table 1).

### 2.3. Geographical predictors

Several open-source geospatial data layers were used in this study to depict the natural geographical factors, as well as the land use/land cover (LULC) characteristics. Altitude above mean sea level is retrieved from the ASTER Global Digital Elevation Model (DEM) (Team, 2009), with 30 m spatial resolution. The LULC layers include the Local Climate Zones classification (LCZ) based on Copernicus Land Monitoring Service's (CLMS) datasets (i.e., Urban Atlas, Corine Land Cover, Imperviousness Degree (IMD), CLMS Tree Cover Density (TCD) layer) (European Environment Agency (EEA), 2016a, 2016b, 2018a, 2018b) (Oliveira et al., 2020a; Oliveira et al., 2020c). The original LCZ classes were converted into LCZ-equivalent Bowen Ration ( $LCZ_{BR}$ ) values, as per (Oke et al., 2017a, 2017b; Oliveira et al., 2021; Stewart and Oke, 2012) (see Supplementary Materials V).

As the original LCZ classification does not contain building height information, one additional step was conducted, by retrieving the average number of building floors from the Italian General Population and Housing Censuses (ISTAT, 2018), per *Comuni* (per locality, in the rural areas outside Naples municipality) and per *Quartieri* (per city-block, within Naples municipality). LCZ reclassification was conducted by considering an average 3 m floor height, together with the original LCZ scheme thresholds, and implemented through the LCZ from Copernicus toolbox (Oliveira et al., 2020b), in ArcMap software (version10) – results of the improved LCZ classification are shown in Fig. 1.

### 2.4. Urban energy balance calculation

At the Urban Canopy Layer (UCL) level, the urban energy balance can be calculated as follows (Eq. (6)):

$$Q_{net} (+Q_F) = Q_E + Q_H + \delta Q_G (+\delta Q_A) \quad (6)$$

where  $Q_{net}$  is the net radiation,  $Q_F$  is the anthropogenic heat (released from human activities, such as from vehicles exhaustion or air condition systems),  $Q_H$  the turbulent sensible heat flux density,  $Q_E$  is the turbulent latent heat flux density,  $\delta Q_G$  is the net storage heat flux, and  $\delta Q_A$  is the net heat advection. Both  $Q_F$  and  $\delta Q_A$  terms are neglected in this study as (i) the former's remote sensing-based calculation is still under study, and its weight in the energy balance is considered small (Wicki et al., 2018), and (ii) the latter's effect during synoptically stable HW events can be neglected (Wicki et al., 2018). The  $Q_{net}$  term can be obtained from Eq. (7):

$$Q_{net} = SW\downarrow + SW\uparrow + LW\downarrow + LW\uparrow \quad (7)$$

where  $SW\downarrow$  and  $SW\uparrow$  are the upwelling and downwelling shortwave radiances, and  $LW\downarrow$  and  $LW\uparrow$  are the longwave atmospheric counter radiation and terrestrial emission, respectively. For the MSG-based analysis, the shortwave and longwave downwelling radiation terms are both available data (see Table 1). For the Landsat 8 analysis, the downwelling radiation components were calculated through a solar irradiance model

(Hofierka and Súri, 2002) implemented in the GRASS GIS (Neteler et al., 2012). The upwelling components can then be calculated as per Eqs. (8) and (9):

$$SW\uparrow = \alpha SW\downarrow \quad (8)$$

$$LW\uparrow = \varepsilon \sigma LST^4 + (1 - \varepsilon) LW\downarrow \quad (9)$$

where the  $\alpha$  is the broadband surface albedo,  $\varepsilon$  is the emissivity (previously mentioned NDVI-based classification),  $LST$  is the surface temperature, and  $\sigma$  is the Stefan-Boltzmann constant ( $5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ ). The MSG/SEVIRI albedo product was used for the corresponding MSG/SEVIRI analysis; the Landsat-based albedo  $\alpha$  was calculated from the weighted average sum of its bands, as developed by and employed by (see Eq. (10)).

$$\alpha = 0.356 \rho_2 + 0.130 \rho_4 + 0.373 \rho_5 + 0.085 \rho_6 + 0.072 \rho_7 - 0.0018 \quad (10)$$

where the  $\rho_{2-7}$  are the corresponding Landsat 8 surface reflectance from bands 2 to 7. The  $\delta Q_G$  component was calculated following the Objective Hysteresis Model (OHM), as per Eq. (11) (Grimmond and Oke, 1999; Rigo and Parlow, 2007), and the  $Q_E$  and  $Q_H$  components obtained from the Bowen Ratio (BR) approach (Lopes, 2003; Parlow, 2003; Wicki et al., 2018) as per Eqs. (12) and (13), considering the LCZ-based BR reclassification.

$$\delta Q_G = a_1 Q_{net} + a_2 \frac{\partial Q_{net}}{\partial t} + a_3 \quad (11)$$

$$Q_H = \frac{BR ( Q_{net} - \delta Q_G )}{(1 + BR)} \quad (12)$$

$$Q_E = \frac{( Q_{net} - \delta Q_G )}{(1 + BR)} \quad (13)$$

Recently, (Rigo and Parlow, 2007) have applied, for the first time, the OHM to remote sensing data, for the city of Basel, and successfully validated its superior performance, when compared to the previously used NDVI-approach. The  $a_1$ ,  $a_2$  and  $a_3$  parameters are standard parameters, for specific LULC classes, as per Table 2. The differential term is the  $Q_{net}$  change per  $t$  time interval – hence the OHM approach requires a minimum of two separate images available, which is not a problem for the MSG/SEVIRI analysis but is usually not possible for Landsat 8, except when a descending acquisition is available. Alternatively, diurnal  $Q_{net}$  can also be computed through the NDVI-based approach, with satisfactory results – see Rigo and Parlow (2007), for accuracy comparison.

### 2.5. Random forest LST fusion method

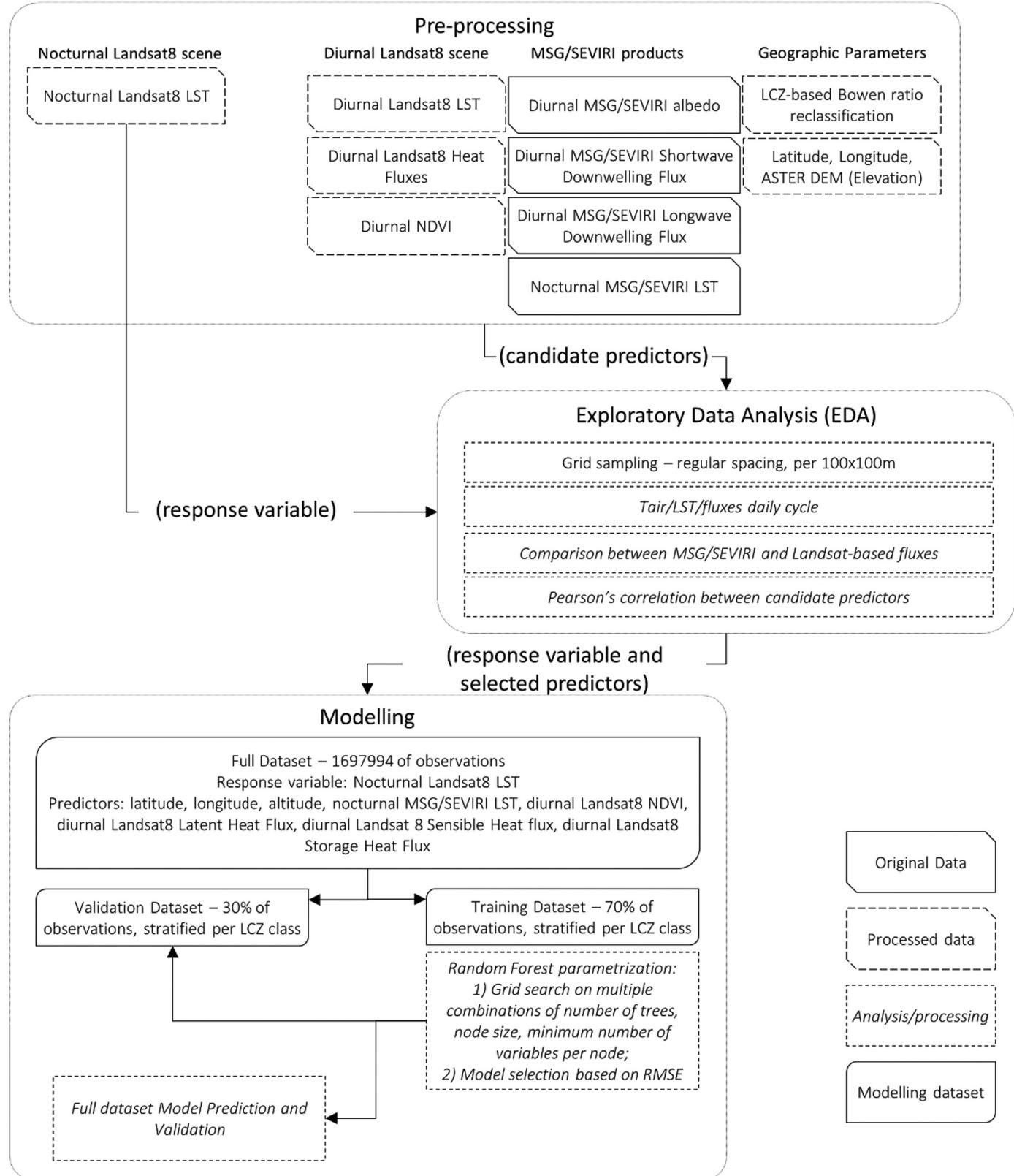
In Fig. 2, the workflow of the model implementation is presented. For the machine learning (ML) model development, all candidate predictors and the response variable (the nocturnal Landsat 8 land surface temperature, LSTn) were sampled in a regularly spaced point grid, set at 100x100m intervals. Given the complexity of factors that affect the spatial pattern of the LSTn, e preliminary exploratory data analysis (EDA)

**Table 2**  
Parameters used in the OHM calculation.

LULC	Corresponding LCZ classes	$a_1$	$a_2$	$a_3$
Forest <sup>a</sup>	A, B	-0.11	-0.11	12.30
High density/industrial <sup>b</sup>	1,2,3,8,10,E	-0.46	-0.16	49.00
Medium density <sup>b</sup>	4,5,6,F	-0.42	-0.27	36.00
Agriculture <sup>b</sup>	9,C	-0.21	-0.34	25.00
Grassland <sup>b</sup>	D	-0.16	-0.05	16.00

<sup>a</sup> Source: Grimmond and Oke (1999).

<sup>b</sup> Source: Rigo and Parlow (2007).



**Fig. 2.** Diagram of the model implementation workflow.

was conducted, where the relation between the Landsat8 and MSG/SEVIRI-based fluxes and LST were compared. In addition, the correlation between candidate predictors and the response variable was tested. Due to the existence of weak correlations between some candidate predictors (i.e., Pearson's coefficients inferior to 0.25), not all variables were

included in the final model, but their correlation matrixes are presented in [Section 3.1](#). The resulting multivariate spatial database was subject to data cleaning, by preserving only complete cases – the final database had a total of 1,697,994 observations, for 10 variables. The full database was split into training (30% cases) and validation (70% cases) subsets, by

randomly sampling each LCZ class, to ensure stratified LCZ representativity.

Random Forest (RF) is an ML technique based on randomly sampled decision trees (bootstrap sampling), often used for both classification and regression purposes (Breiman, 2001). Advantages of RF include being able to model non-linear and complex relationships between variables while requiring few parameters and providing high accuracy (Belgiu and Drăgu, 2016). As it is a so-called ‘black-box’ model, it does not provide a regression equation, but it is possible to retrieve the overall feature importance order – i.e., the relative weight of each predictor in the LTSn results. As referred to in Xu et al. (2021b), the RF algorithm does not account for spatial correlation. Hence, latitude and longitude are necessary to establish the corresponding spatial gradients in LSTn.

The random forest regression model (RF) was implemented in the R programming language (version 4.0.2) (R Development Core Team, 2011), using the ‘randomForest’ package (Liaw and Wiener, 2018). To optimize the RF hyperparameters, a smaller subset of 10% observations (due to computation limitation) was previously used – model performance was assessed through the mean square error (MSE) and mean absolute error (MAE), in the training set, for several alternative combinations of the number of trees and node size (per 10-fold cross-validation). The final RF model results are reported through the MSE and MAE performance in the data unseen by the model (validation

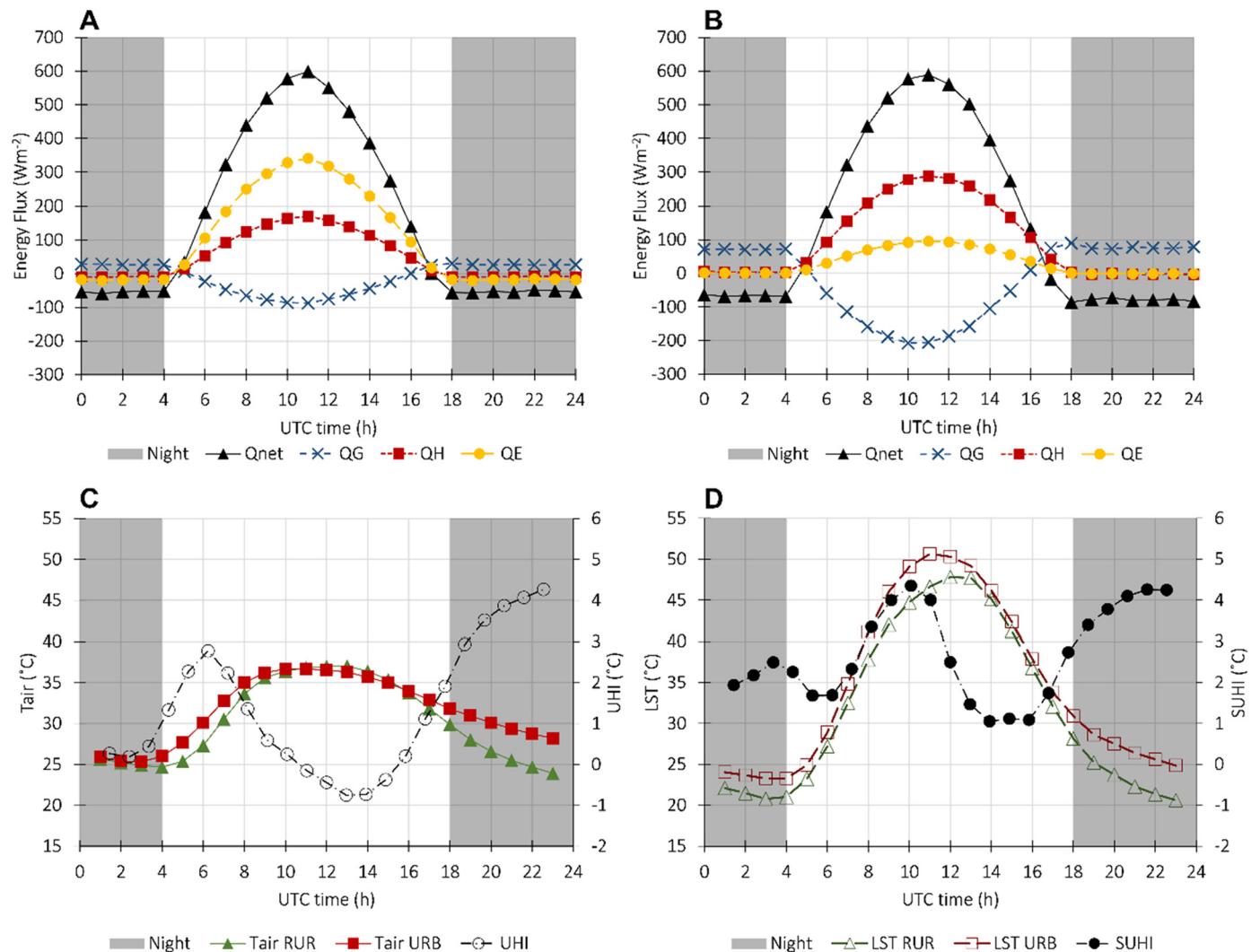
subset), together with the adjusted  $R^2$  coefficient of correlation between observed versus predicted values. All intermediate model results are reported in the Supplementary Material, Appendix A.

### 3. Results

#### 3.1. Exploratory data analysis

In this section, the higher granularity data (i.e., the MSG/SEVIRI radiation and LST imagery, and the air temperature observations) is used to describe the EDA insights on the following aspects: (i) the daily cycle of the urban thermal response in Naples, during the selected HW day; and (ii) a comparison between the MSG/SEVIRI and the Landsat 8 derived data. These are meant as a preliminary interpretation of the urban energy fluxes, and their relation with the nocturnal SUHI. Fig. 3 shows the daily cycle of the energy flux components, as obtained from the calculations based on MSG/SEVIRI products (Eqs. (7), (11), (12), and (13)), at the coarser  $5 \times 5$  km resolution – the rural and urban daily cycles (Fig. 3-A and B, respectively) correspond to the flux values measured at the air temperature measurement sites, as identified in Fig. 1-A.

Positive heat flux values depict an atmospheric heat gain (or surface heat loss), while negative values depict a surface heat gain (or atmospheric heat loss). The available energy balance describes a predictable



**Fig. 3.** Daily cycle of the urban thermal response, in the 9th of August 2017 (heatwave day): (A) urban energy balance components in the rural site (i.e., GRAZZANISE); (B) urban energy balance components in the urban site (i.e., NAPLES); (C) mean hourly air temperature ( $T_{air}$ ) observations in the urban and rural sites, and corresponding urban heat island intensities; and (C) equivalent mean hourly land surface temperature (LST) observations in the urban and rural sites and corresponding surface heat island intensities. A positive sign depicts an upward flux direction, i.e., an atmospheric heat gain, whereas a negative sign depicts a downward flux direction, i.e., a ground/surface heat gain (or atmospheric heat loss).

24 h-cycle: the maximum net heat flux ( $Q_{net}$ ) occurs approximately during the solar noon, with the diurnal curve following the solar radiation cycle, and negative  $Q_{net}$  values are observed during the night, when a net loss of heat is expected, in the atmosphere. No significant differences exist between the urban (Fig. 3-B) and rural (Fig. 3-A) sites, in this regard. Contrariwise, the sensible ( $Q_E$ ), storage ( $\delta Q_G$ ) and latent ( $Q_H$ ) heat flux components show contrasting profiles. In the rural site (Fig. 3-A), the evapotranspiration surface heat loss ( $Q_E$ ) accounts for more than half of the available  $Q_{net}$  energy, surpassing  $300 \text{ Wm}^{-2}$  at about 11 a.m. UTC (1 p.m., in local time), whereas the sensible heat flux ( $Q_H$ ) component does not reach  $200 \text{ Wm}^{-2}$ , in the same period. In the urban site, these proportions are reversed, due to the absence of vegetation - the daily maximum latent heat flux ( $Q_E$ ) does not surpass  $100 \text{ Wm}^{-2}$ , while the sensible heat flux ( $Q_H$ ) is three times greater ( $300 \text{ Wm}^{-2}$ ).

The storage/ground heat flux shows additional differences: in the rural site, the daily absolute maximum surface heat gain does not reach  $100 \text{ Wm}^{-2}$ , at solar noon, while in the urban site it is twice as large, surpassing an absolute value of  $200 \text{ Wm}^{-2}$ . This additional diurnal heat storage ( $\delta Q_G$ ) provides a larger amount of atmospheric heat gain during the night (circa  $70 \text{ Wm}^{-2}$ ), in the urban site. The impact of these flux exchanges in the near-surface air temperature ( $T_{air}$ ) is noticeable: during the day, the city has a negative  $T_{air}$  anomaly (urban cooling island effect (UCI)), due to increased ground heat flux gains (compared to the rural site), whereas during the night this previously-stored heat provides an added heat gain to the atmosphere, reducing the nocturnal cooling rate and contributing to the UHI effect. However, such a cycle does not correspond to the surface's thermal radiation emissions, which prove to be greater in the urban site, throughout the day (even though the surface anomaly, i.e., the SUHI is also greater during the night).

As only two  $T_{air}$  observation sites are available, they are too few to fully characterize the spatial patterns of these findings. In Fig. 4, the statistical distribution of the Surface Urban Heat Island (SUHI) intensities are depicted, per LCZ class. During the day, the most developed areas within the Naples FUA limits are not easily noticeable, in the LST map (see Fig. 7-A); areas corresponding to low vegetation types or bare soil/sand (i.e., LCZ C, D and F) have not only a greater diurnal LST variation than the compact urban fabric (i.e., LCZ 1-3) but have, in some cases, positive SUHI intensities (see Figs. 4-A and 7-A). The only LCZ consistently showing cooler surface temperatures is the 'Dense Trees' class (LCZ A) – also, it should be noted that satellite thermal imagery registers the top of the tree canopy. On the other hand, during the night, the most developed areas can be easily identified as having greater LST (Figs. 4-B and 7-B), which translates into a SUHI anomaly outlining artificial features such as built-up infrastructures or roads (LCZ 1-6,8,10, and E) (see Figs. 4-B and 7-B). In compact areas, this anomaly may surpass  $3^\circ\text{C}$ . The MSG/SEVIRI-based LST (Fig. 7-C and D) maps show the same spatial patterns and equivalent diurnal versus nocturnal SUHI magnitudes, despite the much coarser spatial resolution

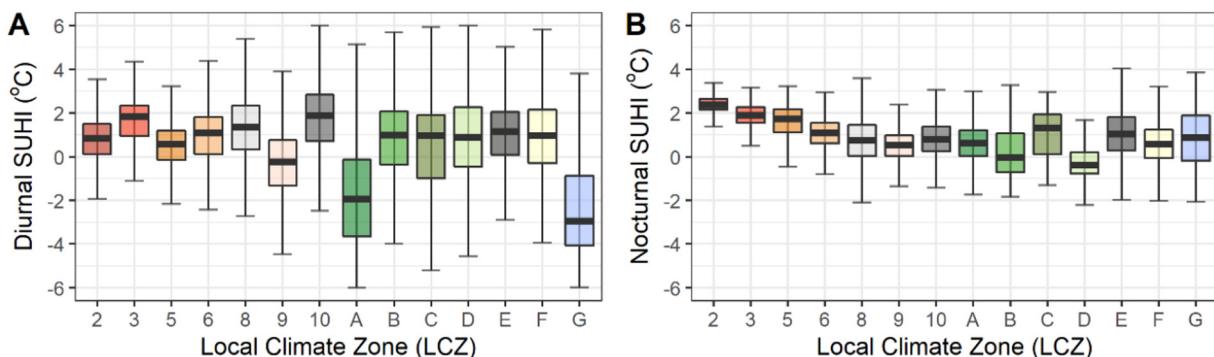
(5 km, versus 100 m in Lansat-8) – however, it should be noted that there is a 15 K difference in the diurnal LST, and a 5 K difference in the nocturnal LST, between them.

On the other hand, the diurnal Lansat-8/TIRS-based flux components disclose a clear urban pattern (see Figs. 5 and 8), particularly in the surface heat gains ( $\delta Q_G$ ) and evapotranspiration surface heat losses ( $Q_E$ ). From the available net heat flux ( $Q_{net}$  is, approximately,  $650 \text{ Wm}^{-2}$ , at the time of Lansat-8 overpass), the absolute heat storage ( $\delta Q_G$ ) component is largest in all artificial land cover classes (LCZ 2-8 and 10, reaching up to 50%) while becoming minimal (less than  $100 \text{ Wm}^{-2}$ ) in vegetated areas. Contrariwise, the latent heat component ( $Q_H$ ) is minimal in urbanized areas (less than 14%), and maximum in all LCZ vegetation land cover classes (LCZ A-D, reaching up to circa 60% of the available energy budget). These magnitudes and proportions of the energy balance components are quite similar between the two satellites, as depicted in Fig. 6. During the night, all heat components become close to zero, and the only significative values are found in the nocturnal storage heat flux  $\delta Q_G$ , which becomes positive, providing an atmospheric slight heat gain in urbanized areas.

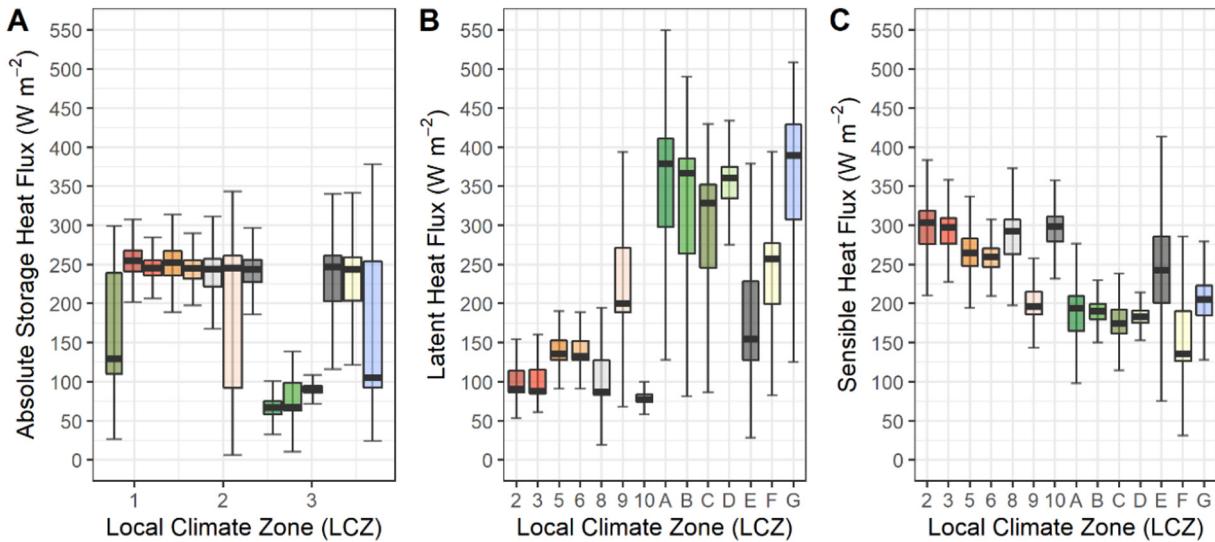
### 3.2. Model inputs and results

Since high-resolution (i.e., infra-kilometric) satellite thermal imagery is not readily available, globally, the RF algorithm was trained to predict the nocturnal LST ( $LST_{(n)}$ ), based on the nocturnal coarser LST (MSG/SEVIRI) and the higher resolution diurnal energy flux components, together with the elevation (DEM) and LCZ-based Bowen-ratio ( $LCZ_{BR}$ ) data. In Fig. 7, the LST data obtained from Eq. (5) (in the case of Lansat 8) and the MGS/SEVIRI database (in the case of LSA SAF products) are shown. The nocturnal LST from Lansat 8 (Fig. 7-B) is the response variable in the RF model, and the diurnal equivalent (Fig. 7-A) is used to calculate the upwelling radiation (Eq. (9)) and ultimately obtain the diurnal net energy and flux components (Eqs. (7), (11), (12), and (13)). The corresponding diurnal and nocturnal MSG/SEVIRI LST images are shown to highlight the resolution contrast (Fig. 7-C and D) – from these, only the nocturnal one is used as a predictor in the model (Fig. 7-D, Table 3).

In addition to the LST, the diurnal energy fluxes obtained from Eqs. (11), (12), and (13), are shown in Fig. 8 – contrary to the diurnal LST (Fig. 7-A), the urban presence can be easily noticed in all cases, particularly in the storage (Fig. 8-A) and latent (Fig. 8-C) heat flux components, which are greater/smaller in more compact areas, respectively. From available candidate predictors (see Table 3), Pearson's linear correlation coefficients are shown in Fig. 9, together with the corresponding significance levels. Due to non-meaningful correlations, both the diurnal LST and  $Q_{net}$  were excluded from the model. In the correlation matrix, strongest linear correlations (Pearson's coefficient  $> 0.75$ , p-values  $< 0.001$ ) are found between the LCZ-based Bowen-ratio ( $LCZ_{BR}$ ) and the diurnal sensible heat flux ( $Q_{H(d)}$ ), and between the diurnal storage ( $\delta Q_{G(d)}$ ) and latent ( $Q_{E(d)}$ ) heat flux components. The



**Fig. 4.** Lansat-8/TIRS-based Surface Urban Heat Island (SUHI), per Local Climate Zone (LCZ): (A) diurnal (B) nocturnal.



**Fig. 5.** Landsat-8/TIRS-based diurnal energy balance components, per Local Climate Zone (LCZ): (A) Storage Heat Flux ( $\delta Q_G$ ), in absolute units (for comparison purposes); (B) Latent Heat Flux Flux ( $Q_E$ ); and (C) Sensible Heat Flux Flux ( $Q_H$ ).

nocturnal LST ( $LST_{(n)}$ ) has the greatest correlations also with  $LCZ_{BR}$ , the nocturnal MSG/SEVIRI-base LST, and the diurnal latent heat flux ( $Q_{E(d)}$ ).

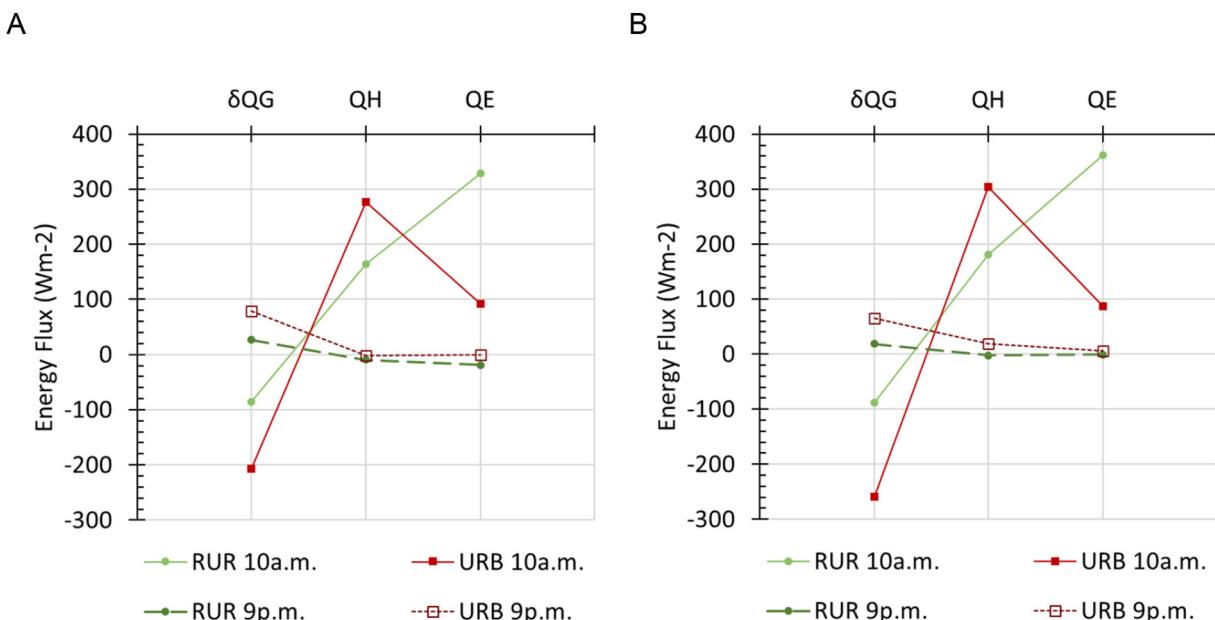
Despite highly correlated predictor variables, they are not equally correlated with the response variable ( $LST_{(n)}$ ). Hence, all (except  $Q_{net(d)}$  and  $LST_{(d)}$ ) were maintained in the model, as the random forest algorithms can handle multicollinearity (Liaw and Wiener, 2018). In fact, throughout the fine-tuning routine (reported in Supplementary Materials V), forcing the model to consider the maximum number of variables for each splitting ('mtry' parameter in the R-based function) (Liaw and Wiener, 2018), provided the most accurate results. In Fig. 10, observed versus predicted results from the best random forest model are presented. The model proved to predict the nocturnal LST ( $LST_{(n)}$ ) with a mean square error of inferior to  $0.1\text{ }^{\circ}\text{C}$ , and highly correlated predicted versus observed values (adjusted  $R^2 = 0.948$ , p-value  $< 0.001$ ).

From the predictor variables importance levels, several conclusions can also be drawn: (i) the three most important predictors are the spatial-related variables, i.e., altitude, longitude and latitude (which was expected due to the spatial nature of the data); (ii) these are followed

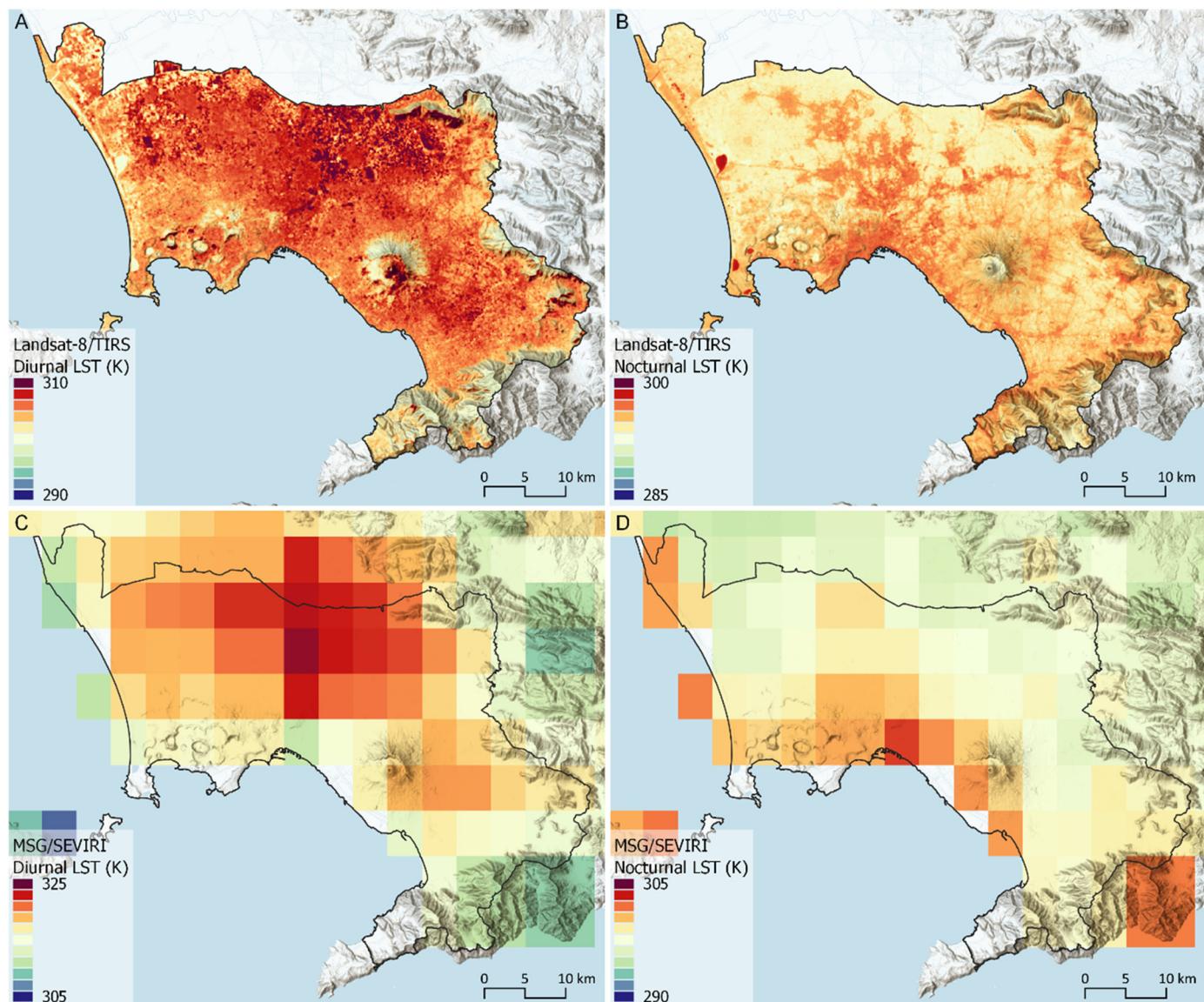
by MSG/SEVIRI-based nocturnal LST ( $LST_{MSG(n)}$ ) together with both the diurnal latent heat flux and storage heat flux components ( $Q_{E(d)}$  and  $\delta Q_{G(d)}$ , respectively), followed the LCZ-based Bowen-ratio. Not only do these findings agree with the spatial interpretation of the energy balance components, and their relation with diurnal and nocturnal LST, described in Section 3.2, but it proves the data fusion technique capabilities. The RF algorithm performance provides a data augmentation methodology to predict Lansat-8-equivalent higher-resolution LST, from coarser-resolution MSG/SEVIRI LST, by unmixing the latter according to the energy balance spatial signatures. While the model's accuracy was only tested for the 9th of August 2017 scene, the method shows promising results to pursue its application and validation in a future time-series analysis.

#### 4. Discussion and conclusions

While nocturnal infra-kilometric thermal imagery is still unavailable from open-source satellite platforms, data fusion techniques for data



**Fig. 6.** Comparison of Landsat-8/TIRS-based (A) and MSG-SEVERI-based (B) energy balance components, at the times of Landsat-8 overpass.



**Fig. 7.** Satellite-based land surface temperature (LST): (A) Landsat-8/TIRS diurnal LST (acquired at 9h41a.m.); (B) Landsat-8/TIRS nocturnal LST (acquired at 8h46p.m.); (C) MSG/SEVIRI diurnal LST (mean hourly, at 10 a.m.); and (D) MSG/SEVIRI nocturnal LST (mean hourly, at 9 p.m.).

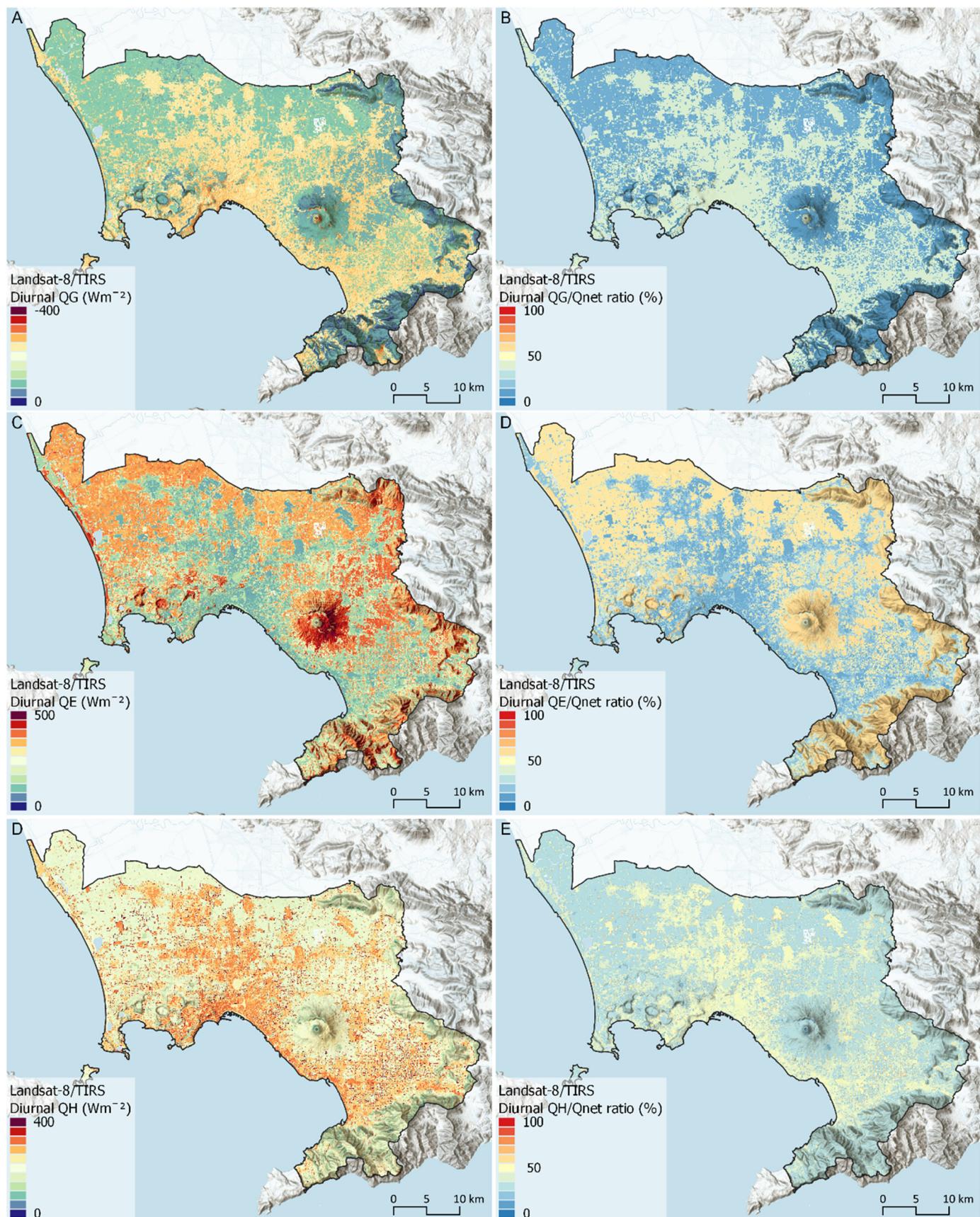
augmentation - i.e., to derive higher-resolution synthetic imagery from synchronous coarser-resolution data (Chen et al., 2014; Ghamsi et al., 2018; Zhang, 2010) - are providing alternative research pathways for research topics requiring higher resolution.

As an example, several authors have developed successful workflows based on either thermal sharpening or spectral unmixing techniques,

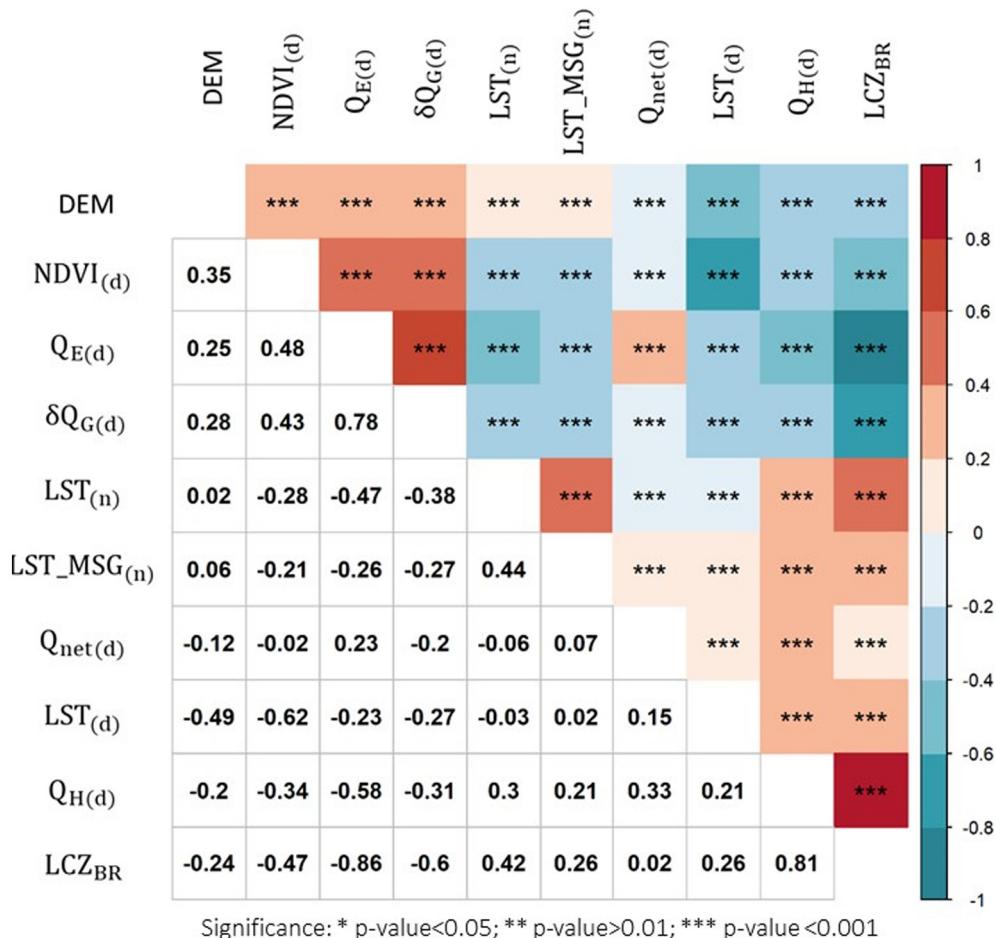
increasing pixel size by a factor of 10, or greater (Huryna et al., 2019; Jin and Han, 2017; Jung and Park, 2014; Zhao et al., 2020b). More specifically, in the context of the Urban Fluxes project (Chrysoulakis et al., 2016) developed a multiple linear regression unmixing approach was developed for the city of Basel, ensuring  $R^2$  values ranging from 0.63 to 0.83, for summer days (Wicki and Parlow, 2017).

**Table 3**  
Nocturnal LST candidate predictor variables.

Designation	Acronym	Source	Final model
Local Climate Zones-based Bowen Ratio	$LCZ_{BR}$	Calculated from (Oliveira et al., 2020a)	Yes
Digital Elevation Model	$DEM$	(Team, 2009)	Yes
Latitude	$y$	Sample grid geographical coordinates	Yes
Longitude	$x$	Sample grid geographical coordinates	Yes
MGS/SEVIRI Nocturnal Land Surface Temperature	$LST\_MSG_{(n)}$	(Freitas et al., 2010)	Yes
Normalized Difference Vegetation Index	$NDVI_{(d)}$	Calculated from Landsat 8 (United States Geological Service, 2020)	Yes
Landsat 8 Diurnal Land Surface Temperature	$LST_{(d)}$		No
Landsat 8 Diurnal Net Heat Flux	$Q_{net(d)}$		No
Landsat 8 Diurnal Sensible Heat Flux	$Q_{H(d)}$		Yes
Landsat 8 Diurnal Latent Heat Flux	$Q_{E(d)}$		Yes
Landsat 8 Diurnal Storage Heat Flux	$\delta Q_{G(d)}$		Yes



**Fig. 8.** Landsat-8/TIRS-based energy balance components: (A) diurnal storage heat flux ( $\delta Q_G$ ) and (B) corresponding ratio from available net heat flux ( $\delta Q_G/Q_{\text{net}}$ ); (C) diurnal latent heat flux ( $\delta Q_E$ ) and (D) corresponding ratio from available net heat flux ( $\delta Q_E/Q_{\text{net}}$ ); and (E) sensible heat flux ( $\delta Q_H$ ) and (F) corresponding ratio from available net heat flux ( $\delta Q_H/Q_{\text{net}}$ ). A positive sign depicts an upward flux direction, i.e., an atmospheric heat gain, whereas a negative sign depicts a downward flux direction, i.e., a ground/surface heat gain (or atmospheric heat loss).



**Fig. 9.** Correlation matrix between all candidate predictors and the response variable ( $LST_{(n)}$ ).

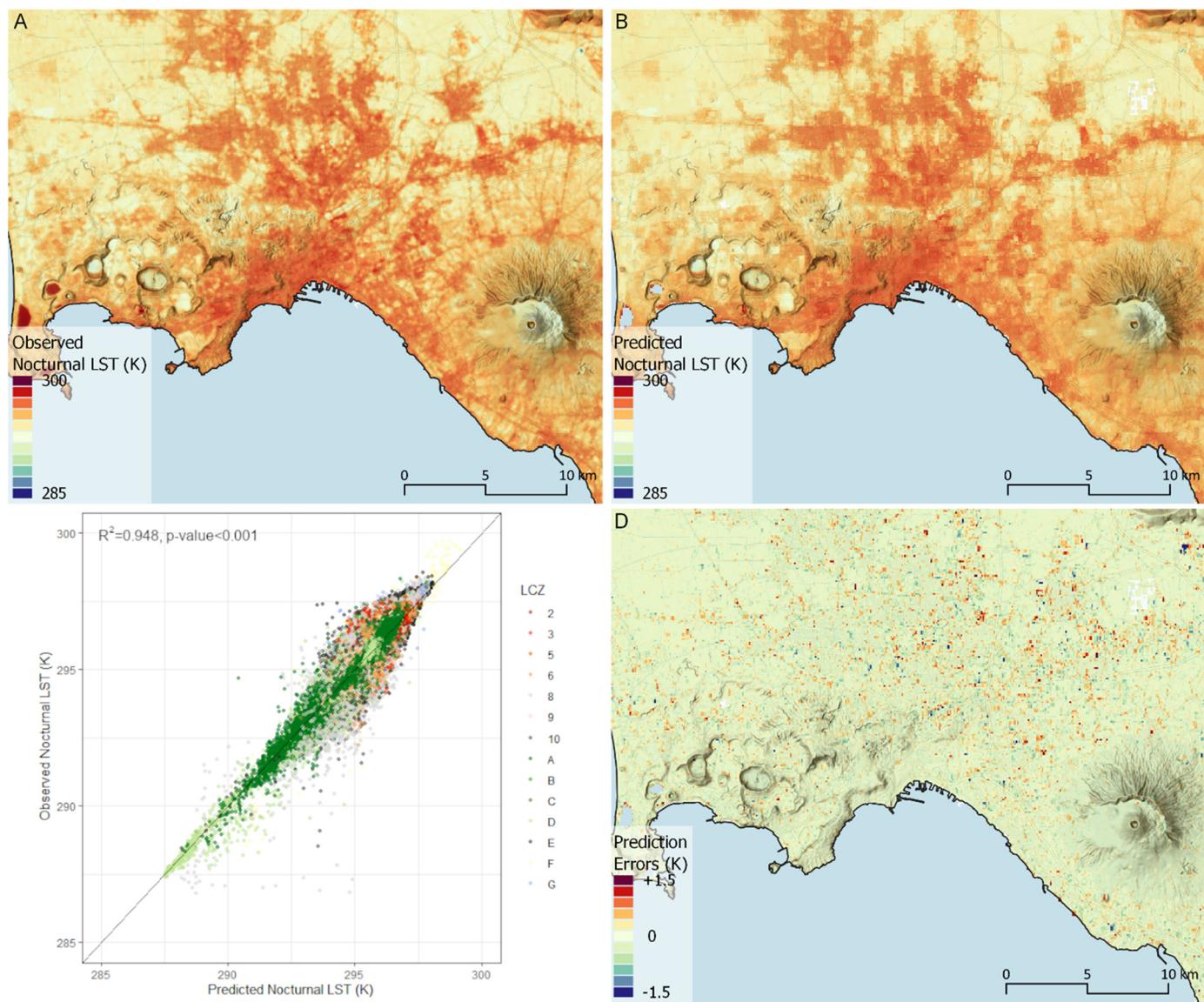
In this study, an energy balance-based machine learning approach is explored, considering the Local Climate Zones (LCZ) and geostationary satellite data (i.e., with sub-daily granularity, but supra-kilometric spatial resolution) to describe the daily cycle of the heat flux components and predict the nocturnal SUHI, during an HW event. The RF model approach, reaches a superior performance (greater correlation with observations and lower mean absolute errors), ensuring mean square errors inferior to 0.1 K, despite being case-specific of the region and the HW day of interest. Both the latent and storage heat flux components, together with LCZ classification, are the most important explanatory variables for the nocturnal LST prediction, supporting the adoption of the energy balance approach.

While in-situ measurements of the canopy urban heat island effect (UHI) are not readily available, globally, the urban climate research community has been increasingly focusing on either: (i) remote sensing-based mesoscale analysis; or (ii) microclimate numerical modeling approaches. Despite recent advances in computational capacity, at the metropolitan scale, the latter option does not provide a straightforward solution for metropolitan-scale urban climate studies – those are still mostly developed within academia, where such processing resources are not always accessible. On the other hand, satellite-based thermal imagery is readily available. The main advantage of remote sensing data is its ability to reveal the spatial patterns of urban thermal performance while ensuring global coverage and repeated observations.

However, given the constraints related to the time of overpass (detailed in Section 1), delving into the diurnal Land Surface Temperature (LST) patterns alone does not provide meaningful information on the local urban climate, specifically the UHI effect, or the Local Climate

Zones (LCZ) performance. Such limitations are especially significant during dry summer conditions, where low vegetation and bare soil areas abound, as is the case in the southern European Mediterranean region. A previous study (Oliveira et al., 2020c) has shown how the diurnal surface temperature did not respond to the most compact built-up occupation, nor a significative surface urban heat island effect (SUHI) was detected. In this study, those results are further developed, by showing that such limitations are possible to overcome when an integrated energy balance approach is considered, as highlighted by several authors (Lopes, 2003; Oke et al., 2017a, 2017b; Parlow, 2003; Parlow et al., 2014; Rigo et al., 2006; Wicki et al., 2018).

Indeed, during the day, the artificial land cover classes show the greatest surface energy gains, which can reach up to  $300 \text{ Wm}^{-2}$  in Compact areas of the metropolitan area (circa 50% of available net heat flux). Similar magnitudes were also found in Basel, during a summer day, as described and validated by (Rigo and Parlow, 2007). Despite being located in a distinct geographical region, Basel was also characterized by an equivalent diurnal surface heat gain (circa  $100 \text{ Wm}^{-2}$ ) in the rural site (Rigo and Parlow, 2007), while in a previous study (Parlow, 2003) the evapotranspiration component was described as reaching up to  $300 \text{ Wm}^{-2}$  as well. The urban-rural differences are mostly the results of these heat flux contrasts: during the day, artificial materials store up the heat, whereas the vegetation's evapotranspiration provides a natural cooling mechanism; during the night, this energy storage is released into the near-surface atmosphere, supplying heat to the development of the positive air temperature anomaly, the UHI effect (Grimmond and Oke, 1999; Lopes, 2003; Oke, 1982, 1988; Parlow, 2003; Rigo and Parlow, 2007).



**Fig. 10.** Comparison between Observed (A) and Random Forest LST predictions (B), Scatter plot of Observed versus predicted LST (C); and Absolute errors map (D).

Accordingly, the diurnal energy budget components are much more useful and meaningful inputs for mesoscale urban climate assessment. Still, heat flux information is more difficult to communicate to a general audience of urban stakeholders and/or the community at large, who are not familiar with their definitions, implications or aware of their measurement magnitudes. In that regard, satellite thermal imagery has one additional advantage of being easily understood by non-specialists, being temperature a familiar physical variable.

Limitations from the current model implementation arise from not being possible to use it to predict the nocturnal LST in other locations – i.e., empirical ML models are only suitable to be used within the scope of the training dataset used, and predictions outside of that context might be flawed, as the relation between the response variable and predictors might not be the same. While the data fusion model predictions cannot yet be generalizable to other locations or dates, the methodology itself is highly scalable to the other European FUAs (more than 700 cases) for which all the input variables are currently available, or easy to implement. While such scalability will be addressed in future steps of the research, the temporal component shall also be considered for re-training and re-testing the data fusion technique. Hence, in a future more complete time series analysis temporal-related variables related to seasonal or weather conditions (e.g., relative humidity, wind direction and

speed) will be added, depending on data and computational resources availability. Thus, this study's method and results are a precursor to upscale an energy balance-based data fusion approach for nocturnal LST/SUHI prediction, which would provide an efficient urban climate monitoring tool, whenever in-situ observations are not readily available.

#### Funding

This research was funded by national funds through FCT – Fundação para a Ciência e a Tecnologia [Ph.D. Grant NO. PD/BD/52304/2013].

#### Declaration of competing interest

The funding source had no involvement in study design, data collection/analysis/interpretation, writing the report or in the decision to submit the article for publication.

#### Acknowledgements

AO acknowledges “Fundação para a Ciência e a Tecnologia (FCT) – Portugal” for providing financial support to the research, through the individual PhD grant number PD/BD/52304/2013.

The Copernicus Land Monitoring Service datasets used in this study were produced "with funding by the European Union".

Map data copyrighted OpenStreetMap contributors and available from <https://www.openstreetmap.org>.

We acknowledge the use of imagery provided by services from the Global Imagery Browse Services (GIBS), operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS, <https://earthdata.nasa.gov>) with funding provided by NASA/HQ.

We thank the three anonymous reviewers for critically reading the manuscript whose comments/suggestions helped improve and clarify this manuscript.

We also thank the support provided through the editorial process.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2021.150130>.

## References

- Anderson, C.I., Gough, W.A., Mohsin, T., 2018. Characterization of the urban heat island at Toronto: revisiting the choice of rural sites using a measure of day-to-day variation. *Urban Clim.* <https://doi.org/10.1016/j.ulclim.2018.07.002>.
- Anderson, V., Leung, A.C.W., Mehdipoor, H., Jänicke, B., Milošević, D., Oliveira, A., Manavvi, S., Kabano, P., Dzyuban, Y., Aguilar, R., Agan, P.N., Kunda, J.J., García-Chapeton, G., de França Carvalho Fonsêca, V., Nascimento, S.T., Zurita-Milla, R., 2021. Technological opportunities for sensing of the health effects of weather and climate change: a state-of-the-art-review. *Int. J. Biometeorol.* <https://doi.org/10.1007/s00484-020-02063-z>.
- Baklanov, A., Grimmond, C.S.B., Carlson, D., Terblanche, D., Tang, X., Bouchet, V., Lee, B., Langendijk, G., Kolli, R.K., Hovsepyan, A., 2018. From urban meteorology, climate and environment research to integrated city services. *Urban Clim.* <https://doi.org/10.1016/j.ulclim.2017.05.004>.
- Barsi, J.A., Barker, J.L., Schott, J.R., 2004. An Atmospheric Correction Parameter Calculator for a Single Thermal Band Earth-sensing Instrument. <https://doi.org/10.1109/igars.2003.1294665>.
- Barsi, J.A., Schott, J.R., Palluconi, F.D., Hook, S.J., 2005. Validation of a web-based atmospheric correction tool for single thermal band instruments. *Earth Observing Systems X* <https://doi.org/10.1117/12.619990>.
- Bechtel, B., Demuzere, M., Mills, G., Zhan, W., Sismanidis, P., Small, C., Voogt, J., 2019. SUHI analysis using local climate zones—a comparison of 50 cities. *Urban Clim.* 28, 100451. <https://doi.org/10.1016/j.ulclim.2019.01.005>.
- Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Cai, M., Ren, C., Xu, Y., Lau, K.K.L., Wang, R., 2018. Investigating the relationship between local climate zone and land surface temperature using an improved WUDAPT methodology – a case study of Yangtze River Delta, China. *Urban Clim.* <https://doi.org/10.1016/j.ulclim.2017.05.010>.
- Carrer, D., Moparthy, S., Lellouch, G., Ceamanos, X., Pinault, F., Freitas, S.C., Trigo, I.F., 2018. Land surface albedo derived on a ten daily basis from Meteosat Second Generation Observations: the NRT and climate data record collections from the EUMETSAT LSA SAF. *Remote Sens.* 10. <https://doi.org/10.3390/rs10081262>.
- Carrer, D., Ceamanos, X., Moparthy, S., Vincent, C., Freitas, S.C., Trigo, I.F., 2019a. Satellite retrieval of downwelling shortwave surface flux and diffuse fraction under all sky conditions in the framework of the LSA SAF program (part 1: methodology). *Remote Sens.* 11. <https://doi.org/10.3390/rs11212532>.
- Carrer, D., Moparthy, S., Vincent, C., Ceamanos, X., Freitas, S.C., Trigo, I.F., 2019b. Satellite retrieval of downwelling shortwave surface flux and diffuse fraction under all sky conditions in the framework of the LSA SAF program (part 2: evaluation). *Remote Sens.* 11. <https://doi.org/10.3390/rs11222630>.
- Chapman, L., Bell, C., Bell, S., 2017. Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using netatmo weather stations. *Int. J. Climatol.* <https://doi.org/10.1002/joc.4940>.
- Chavez, P.S., 1996. Image-based atmospheric corrections - revisited and improved. *Photogramm. Eng. Remote. Sens.* 62 (9), 1025–1036.
- Chen, Y., Zhan, W., Quan, J., Zhou, J., Zhu, X., Sun, H., 2014. Disaggregation of remotely sensed land surface temperature: a generalized paradigm. *IEEE Trans. Geosci. Remote Sens.* 52. <https://doi.org/10.1109/TGRS.2013.2294031>.
- Cheval, S., Dumitrescu, A., Amihaezi, V.-A., 2020. Exploratory analysis of urban climate using a gap-filled Landsat 8 land surface temperature data set. *Sensors* 20, 5336. <https://doi.org/10.3390/S20185336>.
- Chrysoulakis, N., Heldens, W., Gastellu-Etchegorry, J.P., Grimmond, S., Feigenwinter, C., Lindberg, F., Frate, F.del, Klostermann, J., Mitraka, Z., Esch, T., Albitar, A., Gabey, A., Parlow, E., Olofson, F., 2016. A novel approach for anthropogenic heat flux estimation from space. *International Geoscience and Remote Sensing Symposium (IGARSS)* <https://doi.org/10.1109/IGARSS.2016.7730768>.
- Chrysoulakis, N., Grimmond, S., Feigenwinter, C., Lindberg, F., Gastellu-Etchegorry, J.P., Marconcini, M., Mitraka, Z., Stagakis, S., Crawford, B., Olofson, F., Landier, L., Morrison, W., Parlow, E., 2018. Urban energy exchanges monitoring from space. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-29873-x>.
- Congedo, L., 2019. Semi-Automatic Classification Plugin Documentation Release 6.2.0.1. Release.
- Emmanuel, R., Krüger, E., 2012. Urban heat island and its impact on climate change resilience in a shrinking city: the case of Glasgow, UK. *Build. Environ.* <https://doi.org/10.1016/j.buildenv.2012.01.020>.
- European Environment Agency (EEA), 2016a. Urban atlas 2012 [WWW document]. <https://land.copernicus.eu/local/urban-atlas/urban-atlas-2012>.
- European Environment Agency (EEA), 2016b. Corine land cover 2012 [WWW document]. <https://www.eea.europa.eu/data-and-maps/data/external/corine-land-cover-2012>.
- European Environment Agency (EEA), 2018a. Imperviousness density [WWW document]. <https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps/2015>.
- European Environment Agency (EEA), 2018b. Tree cover density [WWW document]. <https://land.copernicus.eu/pan-european/high-resolution-layers/forests/tree-cover-density>.
- Ezimand, K., Chahardoli, M., Azadbakht, M., Matkan, A.A., 2021. Spatiotemporal analysis of land surface temperature using multi-temporal and multi-sensor image fusion techniques. *Sustain. Cities Soc.* 64, 102508. <https://doi.org/10.1016/j.scs.2020.102508>.
- Feigenwinter, C., Vogt, R., Parlow, E., Lindberg, F., Marconcini, M., del Frate, F., Chrysoulakis, N., 2018. Spatial distribution of sensible and latent heat flux in the City of Basel (Switzerland). *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* <https://doi.org/10.1109/JSTARS.2018.2807815>.
- Freitas, E.D., Rozoff, C.M., Cotton, W.R., Silva Dias, P.L., 2007. Interactions of an urban heat island and sea-breeze circulations during winter over the metropolitan area of São Paulo, Brazil. *Bound.-Layer Meteorol.* <https://doi.org/10.1007/s10546-006-9091-3>.
- Freitas, S.C., Trigo, I.F., Bioucas-Dias, J.M., Götsche, F.M., 2010. Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat. *IEEE Trans. Geosci. Remote Sens.* 48. <https://doi.org/10.1109/TGRS.2009.2027697>.
- Ghamisi, P., Rasti, B., Yokoya, N., Gloaguen, R., Wang, Q., Höfle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Atkinson, P.M., Benediktsson, J.A., 2018. *Multisource and Multitemporal Data Fusion in Remote Sensing* arXiv.
- Götsche, F.M., Olesen, F.S., Trigo, I.F., Bork-Unkelbach, A., Martin, M.A., 2016. Long term validation of land surface temperature retrieved from MSG/SEVIRI with continuous in-situ measurements in Africa. *Remote Sens. Environ.* 8. <https://doi.org/10.3390/rs8050410>.
- Grimmond, C.S.B., Oke, T.R., 1999. Heat storage in urban areas: local-scale observations and evaluation of a simple model. *J. Appl. Meteorol.* 38. [https://doi.org/10.1175/1520-0450\(1999\)038<0922:HSIUAL>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<0922:HSIUAL>2.0.CO;2).
- Hofierka, J., Súri, M., 2002. The solar radiation model for Open Source GIS: implementation and applications. *Open Source GIS - GRASS Users Conference*.
- Hong, F., Zhan, W., Götsche, F.-M., Lai, J., Liu, Z., Hu, L., Fu, P., Huang, F., Li, J., Li, H., Wu, H., 2021. A simple yet robust framework to estimate accurate daily mean land surface temperature from thermal observations of tandem polar orbiters. *Remote Sens. Environ.* 264, 112612. <https://doi.org/10.1016/j.rse.2021.112612>.
- Howard, L., 1833. The Climate of London: Deduced from Meteorological Observations, Made at Different Places in the Neighbourhood of the Metropolis. Vol.1. IAUC edition. <https://doi.org/10.1177/030913309339794>.
- Huryna, H., Cohen, Y., Karniel, A., Panov, N., Kustas, W.P., Agam, N., 2019. Evaluation of TsHARP utility for thermal sharpening of Sentinel-3 satellite images using Sentinel-2 visual imagery. *Remote Sens.* 11. <https://doi.org/10.3390/rs11192304>.
- ISTAT, 2018. *Italian census. Permanent Census of Population and Housing*.
- Jia, A., Ma, H., Liang, S., Wang, D., 2021. Cloudy-sky land surface temperature from VIIRS and MODIS satellite data using a surface energy balance-based method. *Remote Sens. Environ.* 263, 112566. <https://doi.org/10.1016/j.rse.2021.112566>.
- Jin, H.S., Han, D., 2017. Multisensor fusion of landsat images for high-resolution thermal infrared images using sparse representations. *Math. Probl. Eng.* 2017. <https://doi.org/10.1155/2017/2048098>.
- Josey, S.A., Pascal, R.W., Taylor, P.K., Yelland, M.J., 2003. A new formula for determining the atmospheric longwave flux at the ocean surface at mid-high latitudes. *J. Geophys. Res. Oceans* 108. <https://doi.org/10.1029/2002jc001418>.
- Jung, H.S., Park, S.W., 2014. Multi-sensor fusion of landsat 8 thermal infrared (TIR) and panchromatic (PAN) images. *Sensors (Switzerland)* 14. <https://doi.org/10.3390/s141224425>.
- Kesavan, R., Muthian, M., Sudalaimuthu, K., Sundarsingh, S., Krishnan, S., 2021. ARIMA modeling for forecasting land surface temperature and determination of urban heat island using remote sensing techniques for Chennai city, India. *Arab. J. Geosci.* (11), 1–14. <https://doi.org/10.1007/S12517-021-07351-5>.
- Kotthaus, S., Grimmond, C.S.B., 2014a. Energy exchange in a dense urban environment - part II: impact of spatial heterogeneity of the surface. *Urban Clim.* <https://doi.org/10.1016/j.ulclim.2013.10.001>.
- Kotthaus, S., Grimmond, C.S.B., 2014b. Energy exchange in a dense urban environment - part I: temporal variability of long-term observations in Central London. *Urban Clim.* <https://doi.org/10.1016/j.ulclim.2013.10.002>.
- Kourtidis, K., Georgoulas, A.K., Rapsomanikis, S., Amiridis, V., Keramitsoglou, I., Hooyberghs, H., Maiheu, B., Melas, D., 2015. A study of the hourly variability of the urban heat island effect in the greater Athens area during summer. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2015.02.062>.
- Lemonsu, A., Grimmond, C.S.B., Masson, V., 2004. Modeling the surface energy balance of the Core of an old Mediterranean City: Marseille. *J. Appl. Meteorol.* [https://doi.org/10.1175/1520-0450\(2004\)043<0312:mtsebo>2.0.co;2](https://doi.org/10.1175/1520-0450(2004)043<0312:mtsebo>2.0.co;2).
- Liaw, A., Wiener, M., 2018. *Classification and regression by randomForest*. R News 2 (3), 18–22. 2002. See Also `selectFeatures`, `selectFeaturesSlimPLS`, `getClassification`, `readExpMat` Examples# reads an expression matrix with class labels into exp\_

- mat2## Not run: exp\_mat2<-readExpMat (" golub\_leukemia\_data\_with\_classes\_training.csv", TRUE)## End (Not run) tr 15.
- Li, L., Zhang, Y., 2011. Urban heat island analysis using the landsat TM data and ASTER data: a case study in Hong Kong. *Remote Sens.* <https://doi.org/10.3390/rs3071535>.
- Long, D., Yan, L., Bai, L., Zhang, C., Li, X., Lei, H., Yang, H., Tian, F., Zeng, C., Meng, X., Shi, C., 2020. Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach. *Remote Sens. Environ.* 246, 111863. <https://doi.org/10.1016/j.rse.2020.111863>.
- Lopes, A., 2003. Changes in Lisbon's urban climate as a consequence of urban growth. *Wind, Surface UHI and Energy Budget*.
- Lott, J.N., 2004. The quality control of the integrated surface hourly database, American Meteorological Society Paper 71929, 14th Conference on Applied Climatology, Seattle, WA. available at: <https://ams.confex.com/ams/84Annual/webprogram/Paper71929.html> (last access: September 2016).
- Major, D.C., Omojola, A., Dettinger, M., Hanson, R.T., Sanchez-Rodriguez, R., 2011. Climate Change and Cities: First Assessment Report of the Urban Climate Change Research Network. Climate Change and Cities: First Assessment Report of the Urban Climate Change Research Network. <https://doi.org/10.1017/CBO9780511783142>.
- Meier, F., Fenner, D., Grassmann, T., Jänicke, B., Otto, M., Scherer, D., 2015. Challenges and benefits from crowdsourced atmospheric data for urban climate research using Berlin, Germany, as testbed. ICUC9 - 9th International Conference on Urban Climate Jointly With 12th Symposium on the Urban Environment Challenges.
- Mills, G., 2014. Urban climatology: history, status and prospects. *Urban Clim.* <https://doi.org/10.1016/j.ulclim.2014.06.004>.
- Muller, C.L., Chapman, L., Grimmond, C.S.B., Young, D.T., Cai, X., 2013. Sensors and the city: a review of urban meteorological networks. *Int. J. Climatol.* <https://doi.org/10.1002/joc.3678>.
- Nadeau, D.F., Brutsaert, W., Parlange, M.B., Bou-Zeid, E., Barrenetxea, G., Couach, O., Boldi, M.O., Selker, J.S., Vetterli, M., 2009. Estimation of urban sensible heat flux using a dense wireless network of observations. *Environ. Fluid Mech.* <https://doi.org/10.1007/s10652-009-9150-7>.
- Nairn, J., Fawcett, R., 2013. Defining heatwaves: heatwave defined as a heat-impact event servicing all community and business sectors in Australia. CAWCR Technical Report.
- Nairn, J., Fawcett, R., Ray, D., 2009. Defining and predicting excessive heat events, a national system. Proceedings of the CAWCR Modelling ... <https://doi.org/10.1002/psc.518>.
- Napoly, A., Grassmann, T., Meier, F., Fenner, D., 2018. Development and application of a statistically-based quality control for crowdsourced air temperature data. *Front. Earth Sci.* <https://doi.org/10.3389/feart.2018.00118>.
- Neteler, M., Bowman, M.H., Land, M., Metz, M., 2012. GRASS GIS: a multi-purpose open source GIS. *Environ. Model. Softw.* 31. <https://doi.org/10.1016/j.envsoft.2011.11.014>.
- Oke, T.R., 1982. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* <https://doi.org/10.1002/qj.491010845502>.
- Oke, T.R., 1987. Boundary Layer Climates. Second edition. Routledge <https://doi.org/10.1017/CBO9781107415324.004>.
- Oke, T.R., 1988. The urban energy balance. *Prog. Phys. Geogr.* <https://doi.org/10.1177/03091338801200401>.
- Oke, T.R., Mills, G., Christen, A., Voogt, J.A., 2017. Urban heat island. *Urban Climates*. Cambridge University Press, Cambridge, pp. 197–237. <https://doi.org/10.1017/978139016476.008>.
- Oke, T.R., Mills, G., Christen, A., Voogt, J.A., 2017. *Urban Climates*. Cambridge University Press.
- Oliveira, Ana, Lopes, A., Niza, S., 2020a. Local climate zones datasets from five Southern European cities: Copernicus based classification maps of Athens, Barcelona, Lisbon, Marseille and Naples. *Data Brief* 31.
- Oliveira, Ana, Lopes, A., Niza, S., 2020b. Local climate zones classification method from copernicus land monitoring service datasets: an ArcGIS-based toolbox. *MethodsX* 7, 101150. <https://doi.org/10.1016/j.mex.2020.101150>.
- Oliveira, A., Lopes, A., Niza, S., 2020c. Local climate zones in five southern european cities: an improved GIS-based classification method based on copernicus data. *Urban Clim.* 33. <https://doi.org/10.1016/j.ulclim.2020.100631>.
- A Oliveira A. Lopes A. Soares , n.d. (in press) Excess heat factor climatology, trends and ranking functional urban areas exposure in Europe. *Glob. Environ. Chang.*
- Oliveira, A., Lopes, A., Correia, E., Niza, S., Soares, A., 2021. Heatwaves and summer urban heat islands: a daily cycle approach to unveil the urban thermal signal changes in Lisbon, Portugal. *Atmosphere* 12, 292. <https://doi.org/10.3390/atmos12030292>.
- Parlow, E., 2003. The urban heat budget derived from satellite data. *Geogr. Helv.* <https://doi.org/10.5194/gh-58-99-2003>.
- Parlow, E., Vogt, R., Feigenwinter, C., 2014. The urban heat island of Basel - seen from different perspectives. *Erde* <https://doi.org/10.12854/erde-145-8>.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* <https://doi.org/10.5194/hess-11-1633-2007>.
- Peng, Y., Li, W., Luo, X., Li, H., 2019. A geographically and temporally weighted regression model for spatial downscaling of MODIS land surface temperatures over urban heterogeneous regions. *IEEE Trans. Geosci. Remote Sens.* 57, 5012–5027. <https://doi.org/10.1109/TGRS.2019.2895351>.
- Perkins, S.E., 2015. A review on the scientific understanding of heatwaves-their measurement, driving mechanisms, and changes at the global scale. *Atmos. Res.* <https://doi.org/10.1016/j.atmosres.2015.05.014>.
- Prata, A.J., 1996. A new long-wave formula for estimating downward clear-sky radiation at the surface. *Q. J. R. Meteorol. Soc.* 122. <https://doi.org/10.1256/smsqj.53305>.
- R Development Core Team, R, 2011. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. <https://doi.org/10.1007/978-3-540-74686-7>.
- Ramos, B.A., Cladera, J.R., 2016. Identifying urban heat island: the Barcelona case. *11th Congress Virtual City and Territory*.
- Rigo, G., Parlow, E., 2007. Modelling the ground heat flux of an urban area using remote sensing data. *Theor. Appl. Climatol.* <https://doi.org/10.1007/s00704-006-0279-8>.
- Rigo, G., Parlow, E., Oesch, D., 2006. Validation of satellite observed thermal emission with in-situ measurements over an urban surface. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2006.04.018>.
- Salcedo-Sanz, S., Ghamisi, P., Piles, M., Werner, M., Cuadra, L., Moreno-Martínez, A., Izquierdo-Verdiguier, E., Muñoz-Marí, J., Mosavi, A., Camps-Valls, G., 2020. Machine learning information fusion in earth observation: a comprehensive review of methods, applications and data sources. *Inf. Fusion* 63, 256–272. <https://doi.org/10.1016/j.inffus.2020.07.004>.
- Sharma, K.V., Khandelwal, S., Kaul, N., 2020. Principal component based fusion of land surface temperature (LST) and panchromatic (PAN) images. *Spat. Inf. Res.* 29 (1), 31–42. <https://doi.org/10.1007/S41324-020-0033-X>.
- Shen, Y., Shen, H., Cheng, Q., Zhang, L., 2021. Generating comparable and fine-scale time series of summer land surface temperature for thermal environment monitoring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 2136–2147. <https://doi.org/10.1109/JSTARS.2020.3046755>.
- Shi, Y., Katschner, L., Ng, E., 2018. Modelling the fine-scale spatiotemporal pattern of urban heat island effect using land use regression approach in a megacity. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2017.08.252>.
- Shumilo, I., Kussul, N., Shelestov, A., Korsunskaya, Y., Yailymov, B., 2019. Sentinel-3 urban heat island monitoring and analysis for Kyiv based on vector data. *Conference Proceedings of 2019 10th International Conference on Dependable Systems, Services and Technologies, DESSERT 2019*.
- Sobrino, J.A., Jiménez-Muñoz, J.C., Soria, G., Romaguera, M., Guanter, L., Moreno, J., Plaza, A., Martínez, P., 2008. Land surface emissivity retrieval from different VNIR and TIR sensors. *IEEE Trans. Geosci. Remote Sens.* <https://doi.org/10.1109/TGRS.2007.904834>.
- Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* <https://doi.org/10.1175/BAMS-D-11-00019.1>.
- Tan, J., Che, T., Wang, J., Liang, J., Zhang, Y., Ren, Z., 2021. Reconstruction of the daily MODIS land surface temperature product using the two-step improved similar pixels method. *Remote Sens.* 13, 1671. <https://doi.org/10.3390/RS13091671>.
- Team, A., 2009. ASTER Global DEM Validation. <https://doi.org/10.1002/ar.a.20400> lpdaac usgs gov4 accessed 28 July 2009.
- Ünal, Y.S., Sonuç, C.Y., Inceci, S., Topcu, H.S., Diren-Üstün, D.H., Temizöz, H.P., 2020. Investigating urban heat island intensity in İstanbul. *Theor. Appl. Climatol.* 139. <https://doi.org/10.1007/s00704-019-02953-2>.
- United Nations, 2014. World Urbanization Prospects: 2014 Revision, New York, United. <https://doi.org/10.4054/DemRes.2005.12.9>.
- United States Geological Service, 2020. USGS EROS Archive - Landsat Archives - Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) [WWW Document]. USGS.
- USGS, 2016. Landsat 8 (L8) Data Users Handbook (LSDS-1574 version 2.0). USGS Landsat User Services.
- USGS, 2019. EarthExplorer - Home. U.S. Geological Survey.
- Wang, Z.-H., Bou-Zeid, E., Smith, J.A., 2010. Application of a sensor network to study the energy budget in urban canopies. *Proceedings of 15th Symposium on Meteorological Observation and Instrumentation, Atlanta*.
- Wang, J.W., Chow, W.T.L., Wang, Y.-C., 2019. A Global Regression Method for Thermal Sharpening of Urban Land Surface Temperatures From MODIS and Landsat. 41, pp. 2986–3009. <https://doi.org/10.1080/01431161.2019.1697009>.
- Wang, J., Schmitz, O., Lu, M., Karssenberg, D., 2020. Thermal unmixing based downscaling for fine resolution diurnal land surface temperature analysis. *ISPRS J. Photogramm. Remote Sens.* 161, 76–89. <https://doi.org/10.1016/j.isprsjprs.2020.01.014>.
- Wang, S., Luo, Y., Li, Xia, Yang, K., Liu, Q., Luo, X., Li, Xiuhong, 2021. Downscaling land surface temperature based on non-linear geographically weighted regressive model over urban areas. *Remote Sens.* 13, 1580. <https://doi.org/10.3390/RS13081580>.
- Wicki, A., Parlow, E., 2017. Multiple regression analysis for unmixing of surface temperature data in an urban environment. *Remote Sens.* <https://doi.org/10.3390/rs9070684>.
- Wicki, A., Parlow, E., Feigenwinter, C., 2018. Evaluation and modeling of urban heat island intensity in Basel, Switzerland. *Climate* <https://doi.org/10.3390/clif6030055>.
- World Bank, 2010. Cities and Climate Change an Urgent Agenda. Urban Development and Local Government <https://doi.org/10.1787/9789264091375-en>.
- Xiao, Y., Zhao, W., Ma, M., He, K., 2021. Gap-free LST generation for MODIS/Terra LST product using a random forest-based reconstruction method. *Remote Sens.* 13, 2828. <https://doi.org/10.3390/RS13142828>.
- Xu, S., Cheng, J., 2021. A new land surface temperature fusion strategy based on cumulative distribution function matching and multiresolution Kalman filtering. *Remote Sens. Environ.* 254, 112256. <https://doi.org/10.1016/j.rse.2020.112256>.
- Xu, S., Cheng, J., Zhang, Q., 2021a. A random forest-based data fusion method for obtaining all-weather land surface temperature with high spatial resolution. *Remote Sens.* 13, 2211. <https://doi.org/10.3390/RS13112211>.
- Xu, S., Cheng, J., Zhang, Q., 2021b. A random forest-based data fusion method for obtaining all-weather land surface temperature with high spatial resolution. *Remote Sens.* 13, 2211. <https://doi.org/10.3390/RS13112211>.
- Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. *Int. J. Image Data Fusion* <https://doi.org/10.1080/19479830903561035>.
- Zhang, P., Bounoua, L., Imhoff, M.L., Wolfe, R.E., Thome, K., 2014. Comparison of MODIS land surface temperature and air temperature over the continental USA meteorological stations. *Can. J. Remote. Sens.* 40. <https://doi.org/10.1080/07038992.2014.935934>.
- Zhang, X., Zhou, J., Liang, S., Chai, L., Wang, D., Liu, J., 2020. Estimation of 1-km all-weather remotely sensed land surface temperature based on reconstructed spatial-seamless

- satellite passive microwave brightness temperature and thermal infrared data. ISPRS J. Photogramm. Remote Sens. 167, 321–344. <https://doi.org/10.1016/J.ISPRSPRS.2020.07.014>.
- Zhang, X., Zhou, J., Liang, S., Wang, D., 2021. A practical reanalysis data and thermal infrared remote sensing data merging (RTM) method for reconstruction of a 1-km all-weather land surface temperature. Remote Sens. Environ. 260, 112437. <https://doi.org/10.1016/J.RSE.2021.112437>.
- Zhao, W., Duan, S.B., 2020. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. Remote Sens. Environ. 247, 111931. <https://doi.org/10.1016/J.RSE.2020.111931>.
- Zhao, G., Zhang, Y., Tan, J., Li, C., Ren, Y., 2020b. A data fusion modeling framework for retrieval of land surface temperature from landsat-8 and modis data. Sensors (Switzerland) 20. <https://doi.org/10.3390/s20154337>.