Towards an operational model for estimating day and night instantaneous near-surface air temperature for urban heat island studies: outline and assessment

Nicola Colaninno [*], Eugenio Morello

Department of Architecture and Urban Studies (DAStU), Politecnico di Milano, Milan, Italy

ABSTRACT

Near-surface air temperature (NSAT) is key for assessing urban heat islands, human health, and well-being. However, a widely recognized and cost- and time-effective replicable approach for estimating hourly NSAT is still urgent. In this study, we outline and validate an easy-to-replicate, yet effective, operational model, for automating the estimation of high-resolution day and night instantaneous NSAT. The model is tested on a heat wave event and for a large geographical area. The model combines remotely sensed land surface temperature and digital elevation model, with air temperature from local fixed weather station networks. Achieved NSAT has daily and hourly frequency consistent with MODIS revisiting time. A geographically weighted regression method is employed, with exponential weighting found to be highly accurate for our purpose. A robust assessment of different methods, at different time slots, both day- and night-time, and during a heatwave event, is provided based on a cross-validation protocol. Four-time periods are modelled and tested, for two consecutive days, i.e. 31st of July 2020 at 10:40 and 21:50, and 1st of August 2020 at 02:00 and 13:10 local time. High R^2 was found for all time slots, ranging from 0.82 to 0.88, with a bias close to 0, RMSE ranging from 1.45 °C to 1.77 °C, and MAE from 1.15 °C to 1.36 °C. Normalized RMSE and MAE are roughly 0.05 to 0.08. Overall, if compared to other recognized regression models, higher effectiveness is allowed also in terms of spatial autocorrelation of residuals, as well as in terms of model sensitivity.

1. Introduction

Climate change and global warming are now at the center of the international debate and, in the last decades, great attention is paid by governments globally to address such concerns. As the urban environment affects the local-to-global climate and, in turn, suffers from global warming, cities are the key to mitigation and adaptation actions to cope with climate change. Local physic-morphological features of the built-up space critically affect urban air temperature, causing the urban environment to reach warmer temperatures than surrounding rural and peri-urban areas. This occurrence is typically known as urban heat island (UHI), an urban climate phenomenon caused by strong daytime absorption of solar radiation, released back at night time (Oke, 1981).

Despite the growing interest and studies in recent years, the UHI still represents an extensive area of research within urban climatology and planning (Rasul et al., 2017; Souch and Grimmond, 2006). Generally, the UHI phenomenon is assessed using air

temperature obtained through fixed weather stations and/or traverse measurements, or based on remotely sensed land surface temperature (Roth et al., 1989; Voogt, 2014). However, although they account for different indicators, both approaches show advantages and disadvantages.

Over the past few decades, special attention has been placed on combining ground stations temperature with satellite-derived land surface temperature (LST) to model and estimate NSAT for urban microclimate, also supported by field data of both surface and air temperatures (Florio et al., 2004; Nichol, 2005; Schwarz et al., 2012). Nonetheless, until the last decade, the approaches were rarely combined (Schwarz et al., 2012).

In fact, although data from ground weather stations is key for assessing local air temperature, spatial coverage is normally limited due to the low density and uneven distribution of the stations. Since temperature normally varies considerably both spatially and temporally, ground-based measurements are not effective for continuous spatial observations along with large and heterogeneous areas. Applying spatial interpolations, may not always reflect the real climatic conditions of an interpolated point (Ozelkan et al., 2015). As a consequence, the assessment of the UHI based on weather stations only is not practicable (Cristóbal et al., 2008; Kloog et al., 2014; Xu et al., 2014).

To overcome such limitations, different interpolation methods have been applied to fixed weather stations, and mobile traverse measurements, to estimate urban temperature, including Inverse Distance Weighting (IDW), Regularized Spline with Tension (RST), Ordinary Kriging (OK), Multiple Linear Regression (MLR), and Regression Kriging (RK). Nonetheless, while relevant progress has been made, models for simulating UHI advanced slowly (Ozelkan et al., 2015; Rasul et al., 2017; Yang et al., 2004).

An effective resource comes from satellite-derived information, namely optical and thermal remotely sensed data. While weather stations provide discrete point measurements, thermal infrareds have the advantage of continuous surface (gridded) observation. Hence, the spatial dependence and heterogeneity of urban thermal landscapes can be increasingly investigated based on thermal remote sensing (Xue et al., 2012). The increased availability and accessibility of satellite data, at different spatial, temporal, and spectral resolutions, as well as the advances in tools and processing techniques, have led remote sensing to be increasingly used in the field of urban climate and planning. Among the most explored remotely sensed data, either GOES data (Zhang and Du, 2022) or thermal-infrared-derived LST, in particular, from Landsat (Elmes et al., 2020) or the MODerate resolution Imaging Spectroradiometer (MODIS) on-board the Terra and Aqua satellites, have been extensively used for mapping daily surface air temperature as a key variable for assessing urban climate (Zhang and Du, 2022).

Concerning surface temperatures, mean LST from instantaneous LST products derived from MODIS has been explored based on multiple linear regressions of several MODIS LST observations taken at different moments, both day- and night-time (Xing et al., 2021). Support vector machine regression (SVR) has been used for estimating next-day night-time surface UHI based on factors such as land cover, solar radiation, temperature, relative humidity, precipitation, wind speed, aerosol optical depth, and soil moisture, which have been also found to account for daily SUHI variations (Lai et al., 2021). More recently, SVR and Random Forest algorithms have been employed for setting up GIS add-ins for automatic modelling of SUHI (Jato-Espino et al., 2022).

However, for assessing the UHI, near-surface air temperature is required, as it directly reflects thermal conditions above the ground. Indeed, it is reported the surface temperature not to be a suitable proxy to assess the air temperature, mostly day-time (Alvi et al., 2022; Pepin et al., 2016). Hence, because retrieving systematic, continuous, and high spatial-temporal resolution NSAT, in extensive areas, is still challenging, numerous recent studies have focused on advanced regression and machine learning techniques.

The use of MLR has been widely investigated, based on the combination of ground measurements and remotely sensed data, either to study the correlation and relevance of different variables or for modelling, and for mapping day and night NSAT. MLR was explored by including geographical features elaborated from DEM as independent variables (Ninyerola et al., 2000), by combining MODIS thermal imagery and site data, namely longitude, latitude, and altitude (Yan et al., 2009), or satellite-derived LST with meteorological data (Xu et al., 2012). Also, predicting daily 24-h NSAT from LST, on a 1-km grid from MODIS, has been tested to extend both the temporal and spatial coverage of air temperature data in large geographical areas (Kloog et al., 2014, 2017). Very-high 20 m spatial resolution air temperature has also been modelled by means of Landsat 8 LST, CORINE Land Cover and MODIS, and ASTER for emissivity (Alvi et al., 2022). However, MLR generates inaccurate results when modelling non-stationary variables (Florio et al., 2004; Rasul et al., 2017).

To compensate for such a concern, other approaches have been investigated, including Regression Kriging (RK) (Touati et al., 2020) and Geographically Weighted Regression (GWR) models. The latter, in particular, is increasingly experimented with, to account for spatial non-stationarity (Fotheringham et al., 1998). GWR enables site-specific investigation based on local statistics, thus providing enriched information regarding the local spatial variations than global regression models (Xue et al., 2012). Compared to MLR, or other conventional regression analyses such as Ordinary Kriging and Universal Kriging, GWR is better suited for predicting air temperature and modelling the spatial structure of UHI as it takes into account the non-stationarity nature of the phenomenon (Harris et al., 2010; Su et al., 2012; Szymanowski et al., 2013; Szymanowski and Kryza, 2009, 2011).

GWR with adaptive bi-square kernel function has been successfully employed for estimating maximum, minimum, and mean air temperature, at different temporal and spatial resolutions, using satellite-derived DEM data and both day- and night-time LST as auxiliary data. However, the approach is explored for retrieving monthly air temperature at 5-km resolution, and 8-day air temperature at 1-km resolution (Chen et al., 2015).

To improve the accuracy of air temperature modelling, further significant methods have been investigated, including the Temperature-Vegetation Index (TVX) algorithm (Zhu et al., 2013), successfully employed for estimating daily maximum air temperature based on Normalized Difference Vegetation Index (NDVI) and MODIS land products, or the Energy Balance Bowen Ratio model (Hou et al., 2013) used for NSAT retrieval based on Landsat Thematic Mapper images. Both methods significantly improved the accuracy, however, errors are reported to be higher than 2 °C, while the level of precision generally considered accurate is 1–2 °C

(Benali et al., 2012).

Recently, machine learning is investigated, with the capacity of lowering the error below 2 °C. The rule-based Cubist regression model was proposed to estimate all-sky daily average surface air temperature (SAT) using weather station data and corresponding satellite products (Rao et al., 2019). However, the daily average SAT was obtained at a spatial resolution of 0.05° × 0.05°. Also, deep learning, mainly based on space remote sensing and ground station observations, is investigated to obtain daily maximum air temperature at a spatial resolution of 0.01° (Shen et al., 2020).

Random forest (RF) has been successfully used for the capability of characterizing the complex temperature distribution, by establishing non-linear relationships among air temperature, LST, and auxiliary data. Monthly average temperatures were estimated using the RF approach and MODIS LST, NDVI, night-time light, and DEM data (Li and Zha, 2019). Daily maximum and minimum air temperatures were estimated using eight time-series MODIS LST data with other auxiliary variables such as elevation, solar radiation, NDVI, latitude and longitude, aspect, and the percentage of impervious area (Yoo et al., 2018). A stacking ensemble model was also explored to improve the accuracy of spatial interpolation of daily maximum air temperature based on MLR, support vector regression, and RF optimized by the SVR (Cho et al., 2020). However, the elevated number of variables and processing steps is often time-consuming and can reduce the capacity of replicability. Lastly, a machine learning-based ensemble model was investigated to estimate hourly T_a at a spatial resolution of 1 km based on the use of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) (Zhou et al., 2020). Though, to reach the spatial resolution of 1 km, SEVIRI data, at a lower spatial resolution, is combined with MODIS LST.

Generally, NSAT estimation focuses on maximum and minimum, and average monthly, 8-days, and/or daily average air temperature. Reconstruction of air temperature at very high temporal resolution such as sub-daily or hourly is carried out in very few studies (Zhang and Du, 2022; Zhou et al., 2020). With this research, the objective was to establish the first step toward the definition of an operational cost- and time-effective, and replicable method, for estimating high spatio-temporal resolution NSAT. The method, based on open data, is planned to be suitable for extensive geographical areas, worldwide. In particular, at this stage, we have outlined and extensively tested an original easy-to-replicate, yet effective model for automating the estimation of 1 km spatial resolution day- and night-time instantaneous NSAT, during heat wave events, and for four different time observations. The planned model relies on the use of geographically weighted regression (GWR), with exponential weighting, which combines on-site measured air temperature, from local fixed weather station networks, satellite-derived LST, and elevation data (DEM). The obtained NSAT is time instantaneous data with daily and hourly frequency consistent with MODIS revisiting time.

2. Case study and datasets

2.1. Study area and period under investigation

The case study is Lombardy region, north-western Italy, which is the fourth Italian region by extension, covering an area of about 23,860 km^2, and the first in terms of the resident population, with around 10 million inhabitants in 2020. In terms of climate, the region is framed as continental, characterized by cold winters and hot summers. According to the Köppen climate classification scheme, the area is classified as *Cfa* (humid subtropical), i.e. temperate, without dry season, and affected by hot summers (Belda et al., 2014; Peel et al., 2007). Over the last decades, the region has experimented with high environmental, social, and economic vulnerability due to climatic drift and increasing extreme weather events, even higher than the average of European countries. Since 1850, the average air temperature in Lombardy increased by about 2 °C. The global warming effect increased considerably in the last 30 years, with anomalies in the average air temperature of about +0.2–0.3 °C, compared to the average of the reference period 1968–1996 (Terrádez Mas et al., 2016).

Because our focus is on estimating NSAT during heatwave events, we first identified heat waves that occurred during the summer of 2020 in Lombardy. Based on a period of 46 years, from 1973 to 2019, we compared the 90th percentile of daily maximum temperature over the reference period, with the daily maximum temperature of each day during July and August 2020, which are normally the months most likely to be inclined to heatwaves. Days with temperatures that exceed the 90th percentile of the reference period are recognized as heatwave events. Two events are identified, from the 29th of July to the 1st of August, and from the 8th to the 12th of August 2020, respectively. However, for this study, we selected the first heat wave because maximum temperatures reached the highest values, and we tested the methodology for the two most critical days, i.e. the 31st of July and the 1st of August 2020. Table 1 shows the days of the first heatwave, including one day before and one day after, and reports the daily average temperature, the daily minimum, and maximum temperatures, and the 90th percentile of maximum temperature.

Table 1
Heatwave occurred from the 29th of July to the 1st of August 2020, including one day before and one day after, and the two days selected for the experiment, namely the 31st of July and the 1st of August 2020.

DATE	Average Daily Value	Minimum Daily Value	Maximum Daily Value	90th of Maximum Daily Temperature
28–07-20	28.7	24.4	32.9	33.86
29–07-20	29.2	24.7	33.6	33.18
30–07-20	30.5	26.4	34.8	34
31–07-20	**30.9**	**26.1**	**35.1**	**33.51**
01–08-20	**31.3**	**26.3**	**36.4**	**33.7**
02–08-20	27	22	31.5	33.7

2.2. Employed data

2.2.1. MODIS sensor and derived products

MODIS instrument is on-board two satellites, named Terra and Aqua, part of NASA's Earth Observing System (EOS), a long-term mission that started in early 2000. Terra (EOS AM) was launched on December 18th, 1999, and Aqua (EOS PM) was launched on May 4th, 2002, and they are still operating. Both satellites follow a sun-synchronous near-polar circular orbit, at an altitude of 705 km, which allows passing over the same area at the same time every 12 h, thus acquiring data continuously and resulting in global coverage every one to two days.

Terra moves from north to south (descending) passing over the equator at around 10:30 a.m. and 10:30 p.m. each day, while Aqua moves from south to north (ascending) and passes over the equator daily, at approximately 1:30 p.m. and 1:30 a.m. Both satellites have a degree scanning of ±55 and work in tandem, approximately three hours apart from each other, to optimize cloud-free surface observation and allow a sub-daily time scale. MODIS instrument provides 36 bands ranging from wavelengths of 0.4 μm to 14.4 μm. While, spatial resolution is 250 m, for bands 1 and 2, 500 m for bands 3 to 7, and 1000 m for bands 8 to 36 (EarthData, 2020).

Besides raw data, several standard processed products are provided. For this research, we worked with MODIS LST and emissivity (MOD11 and MYD11), available at processing levels 2 and 3, and distributed in Hierarchical Data Format (HDF). LST and emissivity, at a daily temporal resolution, are obtained at two different spatial resolutions, i.e. roughly 1 km, obtained through a generalized split-window algorithm, and 6 km based on a physics-based day/night algorithm. Day and night LST and emissivity are calculated from pairs of MODIS observations using seven bands along with the thermal infrared (TIR). We employed 1-km MOD11A1 and MYD11A1 version 6 products, which allow daily per-pixel LST and Emissivity in a 1200 × 1200-km grid, with some improvements compared to previous collection 5 (EarthData, 2020; Metz et al., 2017). Table 2 summarizes the main information of MOD11A1 and MYD11A1 v6.

2.2.2. Shuttle radar topography mission DEM

The Shuttle Radar Topography Mission (SRTM) employed an active sensor, based on radar interferometry, which was aboard the Space Shuttle Endeavor, launched in 2000, and managed by the National Aeronautics and Space Administration (NASA) and the National Geospatial-Intelligence Agency (NGA). The endeavor was operating at an altitude of 233 km, with an inclination angle of 57°, and an orbiting period of 89.2 min. The shuttle revolved around the Earth about 16 times each day, thus completing 176 orbits during the 11-day mission. The SRTM successfully collected radar data over 80% of the Earth's land surface between 60° north and 56° south latitude with data points posted every 1 arc-second (approximately 30 m). Currently, SRTM-DEM data is available at two spatial resolutions of processed datasets, namely high-resolution elevation data at 1 arc-second (30-m) for the United States only, and medium-resolution elevation data at 3 arc-second (90-m), for global coverage. Additionally, a 30-m resolution DEM in GeoTIFF format at a radiometric resolution of 16 bits per pixel is available, at a global level, derived from resampling the high-resolution dataset using a cubic convolution interpolation.

2.2.3. ARPA weather data

In 2003 the Lombardy region assembled all regional meteorological skills into the ARPA Agency and established the Regional Meteorological Service (SMR). The SMR is active in ARPA Lombardia since January 1st, 2004, and currently has a meteorological monitoring network of about 250 automatic stations. Through the weather station system, ARPA elaborates and stores daily meteorological and climatological data, and provides data on the hydrometric level, snow height, precipitation, temperature, relative humidity, global radiation, wind speed and direction. The time of the datum is solar UTC + 1. We employ the air temperature data (°C), at 2 m from the ground, which is captured daily, every 10 min, and given as open data in a comma-separated value (CSV) format including points coordinates. Besides data and time, a station ID is given. Also, weather stations are provided as georeferenced point features, in a shapefile format, with the station ID. Points have been first projected to WGS84 UTM 32 Nord, then the ID is used as a

Table 2
Main technical features of MODIS Land Surface Temperature and emissivity products.

Short name:	MOD11A1 | MYD11A1
Platform:	Terra | Aqua
Instrument:	MODIS
Collection:	6
Processing Level:	Level-3
Temporal Resolution:	Daily
Spatial Resolution:	~1 km
Temporal Extent	March 2000 - Present
Coordinate System	Sinusoidal
Geographic Dimensions	1200 km × 1200 km
File Name Convention (sample):	MOD11A1.AYYYYDDD.hHHvVV.CCC.YYYYDDDHHMMSS.hdf

- ■ YYYYDDD = Year and Day of Year of acquisition
- ■ hHH = Horizontal tile number (0–35)
- ■ vVV = Vertical tile number (0–17)
- ■ CCC = Collection number
- ■ YYYYDDDHHMMSS = Production Date and Time

common field to join air temperature and spatial position.

3. Methodology

In this research, we designed an operational and replicable GWR model, which relies on the use of open data for effectively estimating pixel-based day- and night-time instantaneous near-surface air temperature. The model relies on the use of satellite-derived LST and DEM, and air temperature (T_a) from fixed weather stations. After a reconstruction phase of MODIS LST, a GWR model with an exponential weighting scheme is employed. MODIS LST and DEM are the predictors (independent variables), while T_a is the dependent variable. Instantaneous NSAT is estimated both day- and night-time consistent with MODIS spatial and temporal resolution. The methodology, as shown in Fig. 1, relies on four main steps: [1] MODIS LST re-projection and reconstruction of missing values due to cloud coverage (if needed); [2] DEM resampling to match MODIS LST spatial resolution; [3] Converting weather data to points, and projecting; [4] Model calibration and run.

3.1. MODIS LST processing and reconstruction

MODIS LST is obtained through the Earth Explorer portal, the one-stop shop managed by the US Geological Survey (USGS). MOD11A1 and MYD11A1 LST were downloaded for the 31st of July and the 1st of August 2020. As in Table 3, we used four images (dark grey), i.e. 10:42 local time (LT) and 21:48 LT for Terra on the 31st of July, and 13:06 LT and 2:00 LT for Aqua on the 1st of August 2020. Some images are severely limited by cloud coverage (light grey).

MODIS images were processed with ENVI software. Two steps are undertaken, namely a rigorous reprojecting and a reconstructing procedure to fill in small no-data areas caused by cloud coverage. For the reprojecting step, we used the MODIS Conversion Toolkit (MCTK) plugin, which ingests the MODIS data as HDF files and allows the selection of the dataset(s) as stored in the HDF file. A standalone open source version of the converter is also available, namely the HDF-EOS to GeoTIFF Conversion Tool (HEG). MCTK includes support for swath projection and grid reprojection. We reprojected all images to WGS84 UTM 32 North, based on rigorous georeferencing, and cubic convolution resampling. MODIS LST for selected days and times are shown in Fig. 2.

Images are given masked for cloud coverage, with cloudy pixels converted to no-data. Therefore, a reconstruction process is undertaken (if needed) to estimate missing values caused by cloud coverage. For operational reasons, and because of the low cloud coverage, we have followed a simple reconstruction approach that does not relies on a physical basis. Advanced surface energy balance models, or fusion methods, such as the spatial and temporal data fusion approach (STDFA) or the enhanced STARFM (Wang and Wang, 2020; Wu et al., 2012; Zhu et al., 2010), should be explored. Hence, we report this issue to be a major caveat of the proposed method.

Here, depending on the amount of missing data, one-to-two steps are used. For Aqua 214 Night (Fig. 2,d), because the area to be reconstructed is relatively small, in-painting-based approach is used, which relies on propagating the spectro-geometrical information of the same image, from surrounding uncontaminated data, to fill no-data (Lorenzi et al., 2011). Because in-painting methods are feasible for reconstructing small areas, for Terra 213 Day (Fig. 2a), Terra 213 Night (Fig. 2b), and Aqua 214 Day (Fig. 2c), a two-step procedure is employed. First, a multi-temporal-based method is applied to reconstruct most of the missing information. Then, in-painting is used to fill in small residual no-data areas. The multitemporal reconstruction of images 2a, b, and c, relies on the combination of two images, one main and one secondary (auxiliary) image, per each scene.
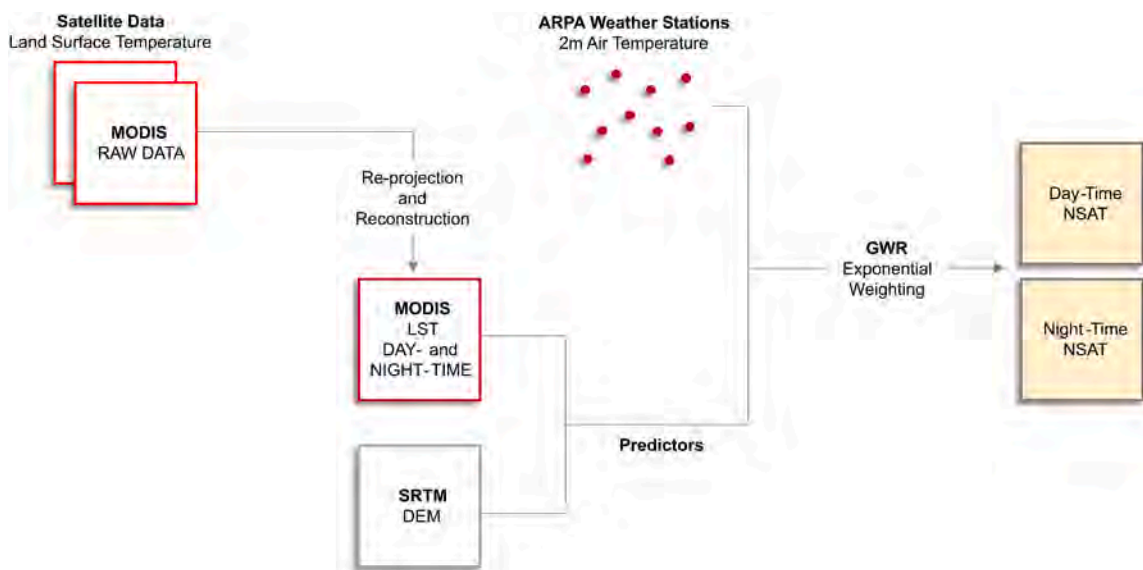

Fig. 1. Workflow of the methodology.

Table 3
MODIS passing date and time for Terra and Aqua during the heatwave under investigation, including not usable images (light grey) and selected images for the research (dark grey).

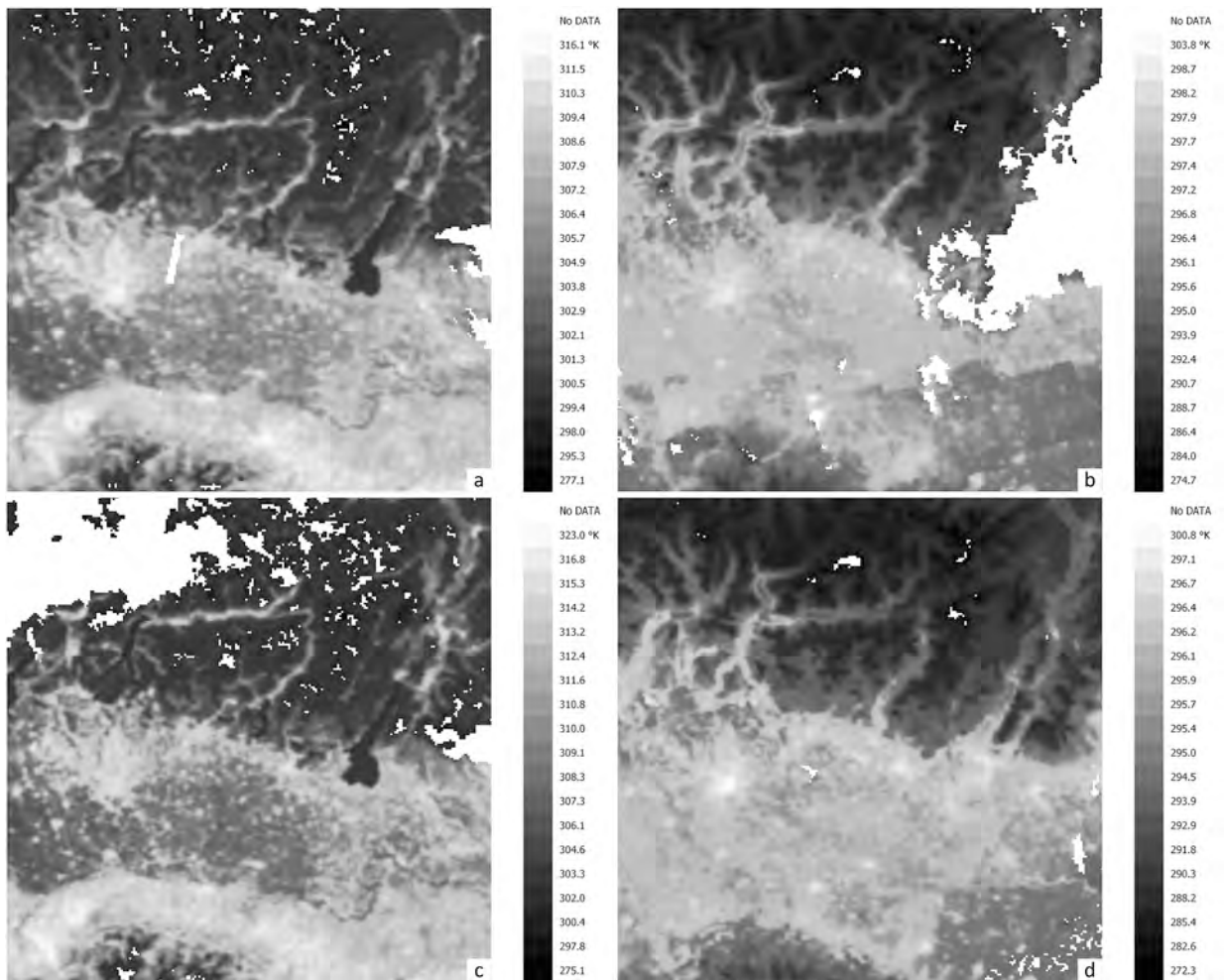| MODIS date and time		Day Number of the Year			
		211	212	213	214
Sensor	Date/Time	29.07.2020	30.07.2020	31.07.2020	01.08.2020
Aqua	Day	12:36	13:18	12:24	**13:06**
	Night	03:06	02:12	02:54	**02:00**
Terra	Day	10:54	11:36	**10:42**	11:24
	Night	22:00	21:06	**21:48**	20:54

Fig. 2. MODIS reprojected imagery over the Lombardy Region, Italy, i.e. the case-study. From top left to bottom right, respectively, a) Terra the 31st of July (213) 10:42 (day), and b) 21:48 (night); c) Aqua the 1st of August (214) 13:06 (day), and d) 02:00 (night). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As an example, we show the reconstruction process for Terra 213 Night (Fig. 2b). The image is combined with a Terra night image taken on the 9th of August 2020 (MOD11A1 20,200,809, day 222) as the secondary image. Passing time is almost the same, i.e. 21:48 and 21:42, respectively for the main and the secondary images. To minimize LST differences, the secondary image is adjusted to the main one, based on a linear regression. The correlation line among the images, built upon 44,851 pixels, shows a very high correlation, with Pearson (r) 0.965. Offset (a) and Gain (b) are respectively −4.374, and 1.022. The secondary image is adjusted and combined with

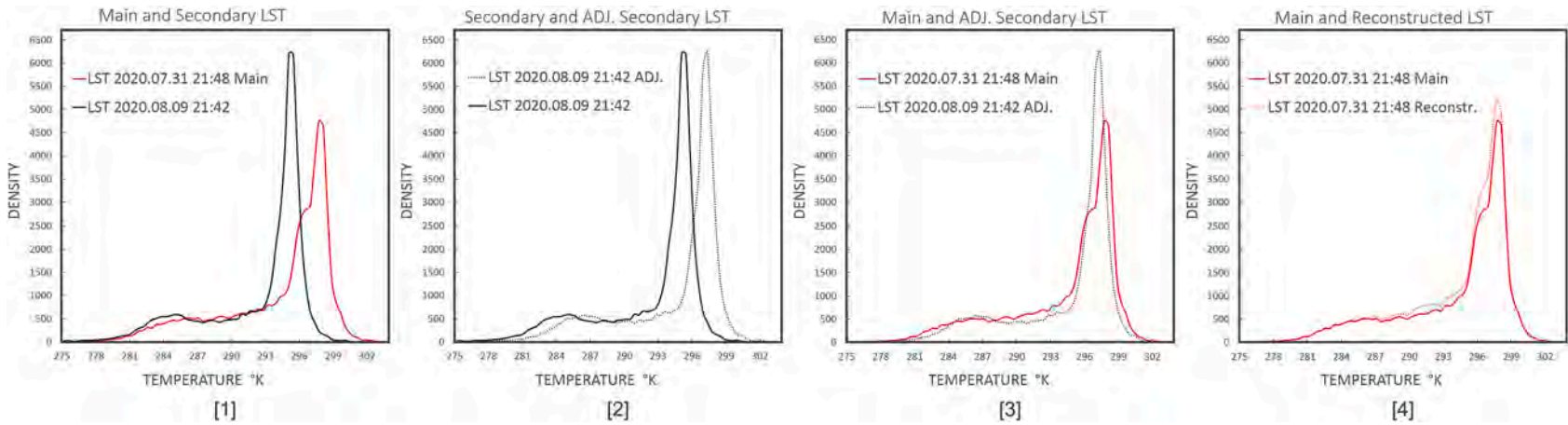**Fig. 3.** Histogram's comparison of the LST reconstruction process. [1]. Terra (213) 31st of July 21:48, the main image, and Terra (222) 9th of August 2020 21:42, secondary image; [2]. Terra (222) 9th of August 2020 21:42, before and after adjusting; [3]. Terra (213) 31st of July 21:48, and the adjusted Terra (222) 9th of August 2020 21:42; [4]. Terra (213) 31st of July 21:48 and reconstructed Terra (213) 31st of July 21:48.

the main image using the ENVI Mosaicking tool. Alternatively, an open-source tool for image mosaicking is available through the Monteverdi - Orfeo ToolBox (OTB).

A feathering distance of 5 pixels is used, and colour balancing based on stats from overlapping regions, between the two images, is further applied to the auxiliary image only. Finally, in-painting is applied to fill in small residual missing data. GRASS GIS r.fillnulls algorithm is used. Values of the main image are not changed. Fig. 3 shows the histograms of both images during the adjusting and reconstruction phases. From left to right, first, the histogram of the main image is compared with the secondary image [1], then the histograms of the secondary and adjusted secondary image are shown [2], then the histogram of the main image is compared with the adjusted secondary image [3]. Finally, the original (main) image is compared with the final reconstructed image, i.e. Terra 31st of July (213) 21:48 [4].

Reconstruction steps are shown in Fig. 4, including the main image, i.e. Terra (213) 31st of July 21:48 (Fig. 4a), and the secondary Terra (222) 9th of August 2020 21:42 (Fig. 4b). After mosaicking, missing areas are significantly reduced (Fig. 4c). In-painting is then applied to fill in residual missing data. Because the MCTK tool recognizes MODIS with a resolution of around 926.625 m per pixel, we have resampled the final image (Fig. 4d) to 900 m. Resampling is based on a Spline interpolation.

Similarly, Terra 213 Day (Fig. 2a), and Aqua 214 Day (Fig. 2c) are reconstructed. Terra 213 Day, which passes over the study area at 10:42, is combined with TERRA 222 Day on the 9th of August 2020 as the auxiliary image. It passes over the area at 10:36. The regression is made upon 56,235 pixels and gives a correlation (r) of 0.917, Offset (a) of -2.284, and Gain (b) of 1.011. Aqua 214 Day, which passes over the study area at 13:06, is combined with Aqua 213 on 31st of July 2020, which is the auxiliary image and passes over the area at 12:24. The regression is made upon 50,212 pixels, with a correlation (r) of 0.946, Offset (a) 0.406, and Gain (b) 1.0107. Image pairs are then mosaicked, filled in, and resampled. All four final scenes (Fig. 5) are converted from Kelvin to Celsius degrees (°C).
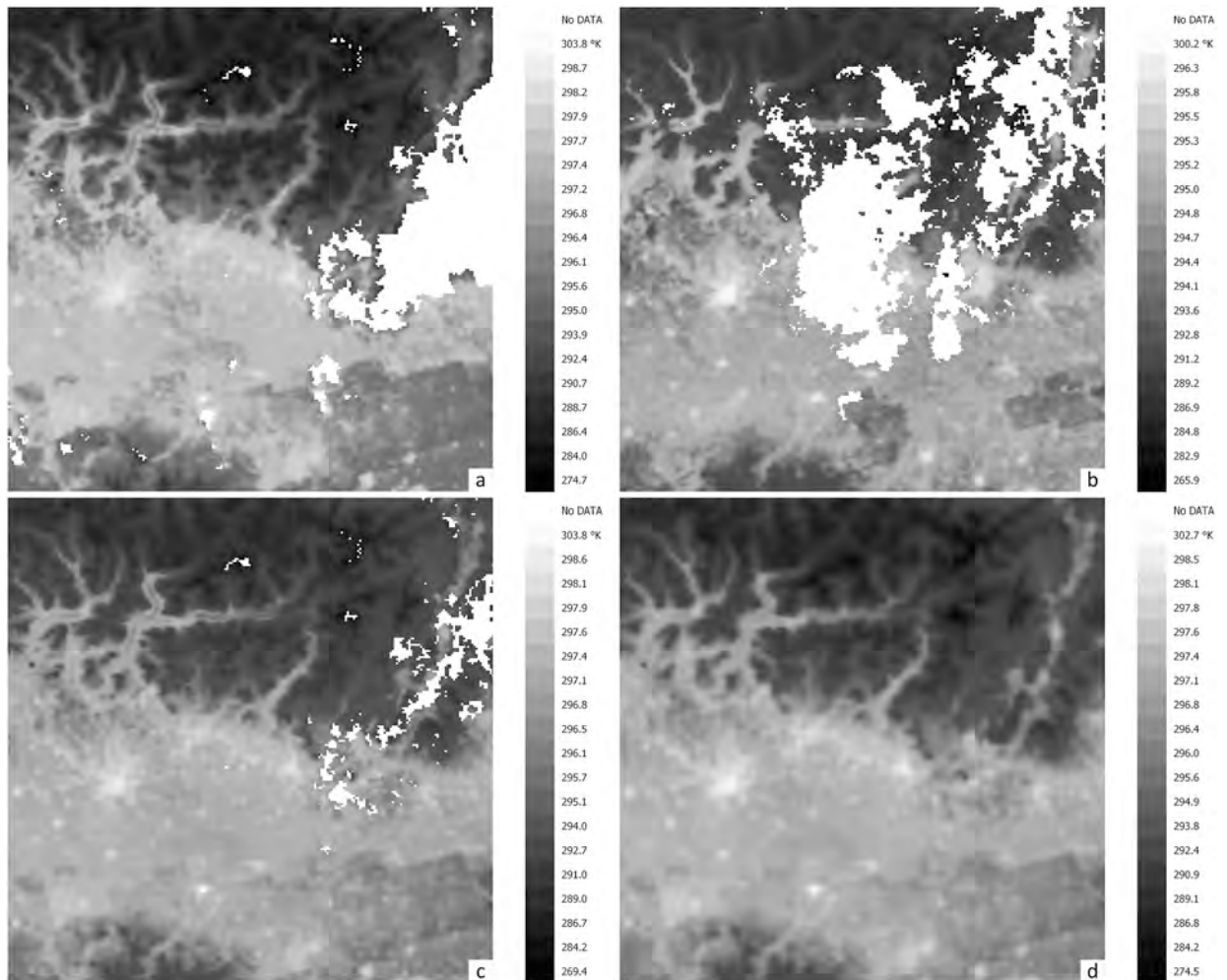

Fig. 4. MODIS LST reconstruction process: a). Terra 213, 31st of July 2020 at 21:48 (Night), the main image, b). Terra 222, 9th of August 2020 at 21:42 (Night), the auxiliary image, c). Mosaicking of main and auxiliary images, d) reconstructed Terra 213, 31st of July 21:48 after mosaicking, fill-in, and resampling.

3.2. Geographically weighted regression: An overview

Geographically weighted regression is a technique best fitted for exploring non-stationary phenomena and used in spatial analysis to investigate and/or predict heterogeneity in data relationships. GWR computes multiple local models, based on separate equations, whose parameters are a function of the observed locations. Accordingly, the model and estimated coefficients are calibrated locally, depending on a neighboring region, or bandwidth, around the target point. Both dependent and explanatory variables depend on such a bandwidth, and beyond this distance, the weight of the observations is assigned a value of 0. At a generic position i, where i is a vector of coordinates, either projected or geodetic coordinate system, the local regression, estimated with data whose influence decays with the distance, is defined by a general form according to Eq. (1) (Fotheringham et al., 1998, 2002; Lu et al., 2014).

$$Y_i = \beta_{i0} + \sum_{n=1}^{m} \beta_{in} X_{in} + \epsilon_i \tag{1}$$

Y_i is the dependent variable at location i (x_i, y_i), X_{in} is the nth independent (explanatory) variable at location i, m is the number of independent variables, β_{i0} is the intercept parameter at location i, β_{in} is the local regression coefficient for the nth independent variable at location i, and ε_i is the random error at location i.

At each regression point i, the parameters (or coefficients β) are estimated locally by the weighted least squares. The weights, expressed in a matrix form, depend on the observed location with respect to the other observations in the dataset (Leung et al., 2000a). The weights are computed using a weighting scheme based on a kernel or a distance function. Most common kernel schemes rely on Gaussian or Exponential functions, while kernel weights are defined as a function of the ratio between the distance from the
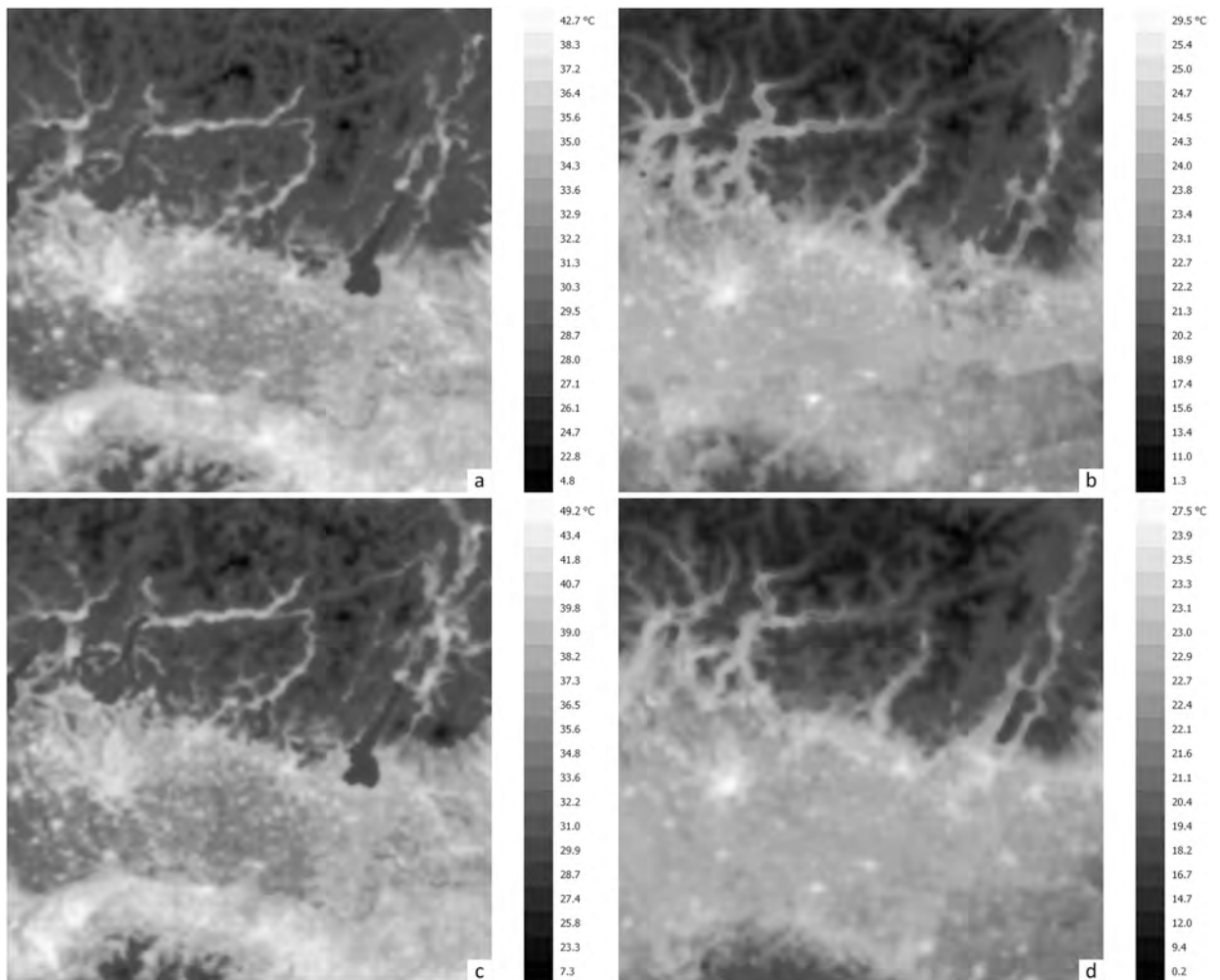
Fig. 5. MODIS LST final reconstructed image scenes for summer 2020: a). Terra 31st of July (213) 10:42 (Day), b). Terra 31st of July (213) 21:48 (Night), c). Aqua 1st of August (214) 13:06 (Day), d). Aqua 1st of August (214) 02:00 (Night). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. Spatial correlation and R^2 of measured T_a (°C) and Ws (m/s) for 112 weather stations. From left to right: 10:40 and 21:50 of the 31st of July 2020, 02:00 and 13:10 of the 1st of August 2020.

neighboring points to the target point, and the bandwidth. The bandwidth is the key controlling parameter and can be specified either by a fixed distance or by a fixed number of nearest neighbors, i.e. an adaptive bandwidth (A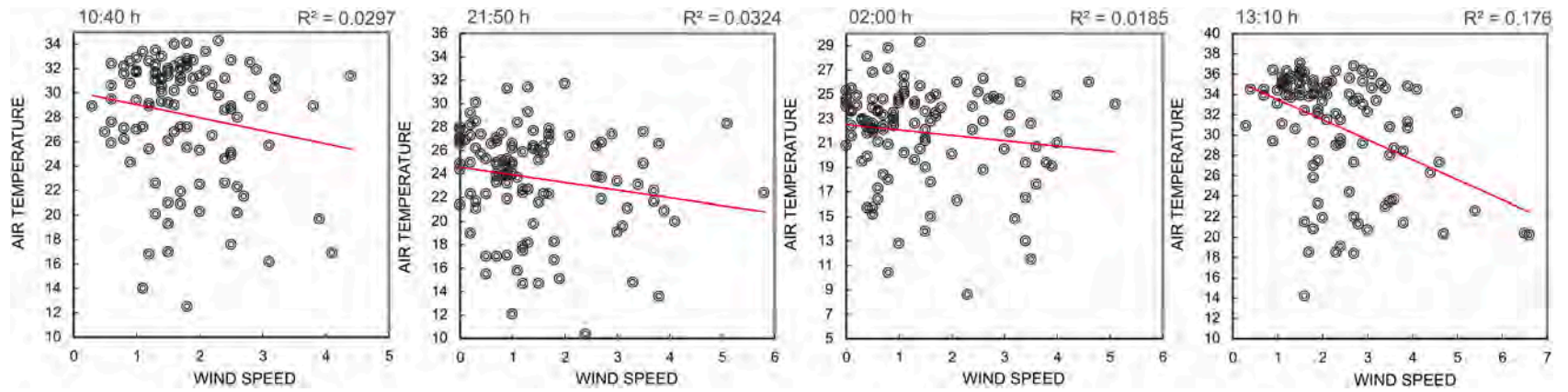nselin, 2018; Lu et al., 2014). Gaussian or Exponential weight w_i is computed according to a kernel function as represented by Eqs. (2), and (3), respectively (Czarnota et al., 2015; Lu et al., 2014). Also, the weighting scheme can be according to an inverse Euclidean distance function, as in Eq. (4) (Anselin, 2018).

$$w_i = exp\left[-\frac{1}{2}*\left(\frac{d_{ij}}{b}\right)^2\right]$$

(2)

$$w_i = exp\left(\frac{-d_{ij}}{b}\right)$$

(3)

$$w_i = \frac{1}{d_{ij}^\alpha}$$

(4)

Where d_{ij} is the distance between any observation j and regression point i, b is the kernel bandwidth, and α is the power value. For this study, a GWR model according to the Exponential weighting scheme has been preferred, as it shows a higher accuracy with respect to the Gaussian weighting, and more stability compared to the Inverse Distance (ID) weighting. Nevertheless, Gaussian and ID weighting schemes are tested and discussed.

3.3. GWR model setting and calibration

Surface temperature is proven to explain alone >80% of the observed variation of air temperature (Kawashima et al., 2000). Also, there is evidence of a strong correlation between temperature and elevation. Different studies focused on modelling daily and/or monthly air temperature or LST using elevation data to improve the performance of the model (Cristóbal et al., 2008; Malbéteau et al., 2017; Samanta et al., 2011a, 2011b).

On the other hand, we did not find any significant evidence of a spatial correlation between air temperature and wind speed (Ws). Different analyses were undertaken. First, we have analyzed the spatial correlation between T_a and Ws measured at the time and days under investigation, namely 10:40 and 21:50 on July 31st, 2020, and 02:00 and 13:10 on August 1st, 2020. Since not all stations of the regional network have both sensors or, in some cases, one of the measures is not available for a temporal observation, or even sub-hourly data is not available, the analysis finally relies on 112 stations. Fig. 6 shows the correlation between T_a (°C) and Ws (m/s), including the R^2 for each observation. Although the correlation tends to be negative, i.e. increasing Ws decreases T_a, the R^2 is not significant.

Besides, for a selection of 20 sample stations, the sub-hourly correlation was explored during the two days, namely the 31st of July and the 1st of August, every 10 min through 48 h. 288 observations are examined for each selected station (Annex I, I.A). Finally, based on interpolations of hourly T_a and Ws, at a spatial resolution of 1.5 km, as provided by ARPA (Lussana et al., 2009; Uboldi et al., 2008), the correlation between the two variables was explored hourly, through 48 h, on 30,798 pixels (Annex I, I.B). In both analyses, no relevant evidence of correlation was found. Moreover, we report a great variability of Ws that results in a difficult identification of a well-defined behavior.
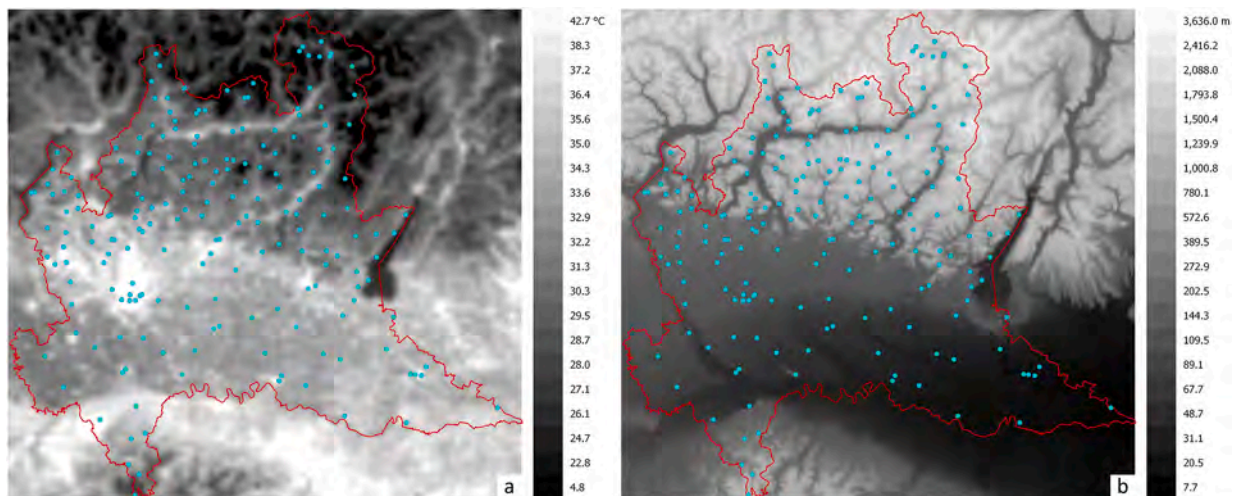

Fig. 7. ARPA weather stations (cyan dots) and Lombardy Region limit (red line) with a). MOD11A1 v6 daytime LST at 10.42 of the 31st of July 2020 and b). The SRTM DEM up-scaled to 900 m. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We also emphasize that, during the two days considered, Ws was mainly found to be barely 0 to 3 or 4 m/s. It has been observed that for wind speeds under three m/s, the temperature difference between urban and rural areas does not depend on Ws considerably, whereas, for greater Ws, the effect of wind on urban temperature increases significantly (Abbassi et al., 2022).

Once discarded wind speed, and under the assumption of spatial non-stationarity, we designed a GWR model based on the combined use of LST and DEM as predictors. The SRTM DEM at a spatial resolution of 90 m has been re-projected to WGS84 UTM 32 N, and up-scaled to 900 m to match MODIS LST resolution. Then, the predictive capacity of DEM and LST is explored concerning T_a. Because T_a has a temporal resolution of 10 min, a little difference occurs, in terms of time, between T_a and LST. To reduce the most such a difference, we selected T_a taken at the closest time as the MODIS local passing time. We used T_a at 10:40 and 21:50 for the 31st of July, and 02:00 and 13:10 for the 1st of August 2020.

Of the available 250 stations, only 205 provide T_a. Also, a few stations do not provide 10-min observation, but only hourly data. Hence, some stations at 10:40, 21:50, and 13:10 are omitted. Finally, 197 values are available for 13:10, 195 for both 10:40, and 21:50. In the latter two cases, we also excluded two more stations that provided abnormal values (assigned 9999). For the model at 02:00, we used all 205 stations since this is at the top of the hour.

Fig. 7 shows the spatial distribution of the ARPA stations for air temperature (cyan dots) in the Lombardy region (red line), with respect to the predictors, namely LST (Fig. 7a) and DEM (Fig. 7b). In particular, LST in Fig. 7 refers to Terra MOD11A1 v6 daytime product at 10:42 on the 31st of July 2020.

To calibrate the model, we analyzed the explanatory capacity of LST and DEM with respect to T_a (dependent variable). This allows us to figure out the different correlations among the predictors and T_a at the four-time slots, both day and night. Linear regression is computed to assess Pearson correlation (r) and F-test significance. As shown in Table 4, a very high correlation is found between both LST and DEM with T_a.

While DEM is inversely correlated to T_a, i.e. the higher the elevation the lower T_a, LST shows a very high positive correlation. However, DEM shows a higher correlation with T_a during daytime, being even higher correlated than LST. At the warmest time, i.e. 13:00, the correlation between T_a and LST is further reduced. On the other hand, the correlation between T_a and LST remarkably increases during night-time, when the two temperatures get closer to each other, with T_a more correlated to LST than the DEM. Because we aim at designing a model capable of estimating T_a at different hours, both LST and DEM are used as predictors. GWR, as previously defined by Eq. (1), is then computed according to Eq. (5), where Y_i is the air temperature T_{ai}, at location i (x_i, y_i), and X_{in} is LST and DEM.

$$T_{a_i} = \beta_{0i} + \beta_{1i}LST_{1i} + \beta_{2i}DEM_{2i} + \epsilon_i \tag{5}$$

The model is performed in SAGA GIS. We set four models, per each time observation, according to the different LST and T_a, namely 31st of July 2020 at 10:40, and 21:50, and 1st of August 2020 at 02:00 and 13:10. All available points (stations) are included in the analysis for weighting, i.e. not a preferred searching distance is established to select a limited number of observations around the target. While a bandwidth (b) is established.

Either for Gaussian or Exponential, defining the most effective bandwidth distance is challenging. Although time-consuming, a trial/error process would be required. However, automatic estimates of default bandwidth distance are allowed, based on input data and according to the map unit. This ensures each target point has at least one neighbor. A bandwidth of 60,000 m is automatically estimated for the input dataset, both for Gaussian and Exponential schemes. Yet, to assess the most appropriate bandwidth, we have tested different bandwidths. The testing is made on the model at 13:10 (daytime), which is the most challenging, due to the reduced correlation between LST and T_a.

The predictive capacity for different bandwidths is assessed through a Cross-Validation (CV) protocol, which normally results in less biased and less optimistic estimates. Generally, we refer to a k-fold CV because the available dataset for testing is split into k similar groups. One group is hold-out for validation, while the remaining group is used as the training dataset. The procedure is repeated k times, then an average error of k tests is calculated. If a specific value is given for k, such a value is used to identify the CV scheme. For a 2-fold CV, a random selection of 50% of all points is considered for training the model, while the remaining 50% is used as testing data. CV coefficient of determination R^2, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE) were used to test different bandwidths both for Gaussian and Exponential weighting, at 13:10, as shown in Table 5.

Table 4

Pearson correlation (r) and F-test significance among the dependent variable T_a and the independent variables (predictors) for each case study, i.e. 31st of July (day and night), and 1st of August (day and night) 2020.

Predictors		Air Temperature (T_a) ARPA Weather stations							
		2020.07.31				2020.08.01			
		10:40 T_a (195 pts.)		21:50 T_a (195 pts.)		02:00 T_a (205 pts.)		13:10 T_a (197 pts.)	
		r	F-test	r	F-test	r	F-test	r	F-test
DEM	*Elevation*	**−0.930**	**1299.80**	−0.899	844.34	−0.848	518.80	**−0.929**	**1285.66**
LST	*10:42 - Day*	0.801	364.48						
LST	*21:48 - Night*			**0.927**	**1221.48**				
LST	*02:00 - Night*					**0.909**	**967.34**		
LST	*13:06 - Day*							0.769	294.33

For Gaussian weighting, a bandwidth of 60,000, as automatically estimated, is performing better than a smaller or greater distance. The performance decreases either by decreasing or increasing the distance. For exponential weighting, we found that by decreasing the bandwidth distance to 40,000 the predictive capacity slightly improved. Once weighting scheme and bandwidth are established, the model performance is tested for three different combinations of predictors, namely both LST and DEM combined, LST only, and DEM only. A GWR with an exponential weighting scheme and bandwidth of 40,000 is employed. The predictive capacity is assessed through a 2-fold CV protocol.

In Table 6, beside the RMSE, MAE, and MBE, the Adjusted (Adj.) R^2 is computed. Although the Adj.R^2 assesses the variance explained by the model in the same way as the standard R^2, it is better suited for comparing models with a different number of predictors and is often the best indication of fit quality. It adjusts using the residual degree of freedom. Adj.R^2 is always less than or equal to R^2, but with the same 0 to 1 range, with values closer to 1 signifying a better match (Touati et al., 2020).

The model dramatically gets worse at daytime if we use LST only as a predictor. All errors increase, while Adj.R^2 critically decreases. The situation is significantly different at night. LST and T_a are much more correlated during night-time, making the model perform better. The worse performance is obtained at 13:10, which is the warmest time.

The situation is reversed using the DEM only. Although elevation is widely proven to be a fundamental variable to explain T_a (Good et al., 2017; Lussana et al., 2009; Mutiibwa et al., 2015; Uboldi et al., 2008), the impact of elevation on prediction is much more sensitive during daytime, being less correlated to T_a during night-time. The worse performance is obtained at 02:00, the coldest time.

Therefore, the final model is set upon the combined use of both LST and DEM to increase the predictive capacity at each time. The result of the modelled instantaneous NSAT, for all observations, is shown in Fig. 8.

4. Model performance assessment

The designed model is assessed, for a heat wave event, through different measures, and with respect to other methods. Since evaluating the goodness of the model based on known values only is not appropriate, both residuals at the observed (known) points, as well as the residuals for estimated (unknown) values, are considered. A 2-fold CV protocol is used. Although the 2-fold CV is a non-exhaustive method, as it does not consider all possible combinations, it provides significant advantages in terms of computation time. On the other hand, it is less optimistic because it trains the models with just half of the observations. Besides R^2, RMSE, MAE, and MBE, for the CV protocol the normalized RMSE and MAE, and *t*-Test are calculated.

While RMSE and MAE both estimate the predictive capacity by measuring the model error, in the same unit as the dependent variable, bias is the difference between the expected value by the estimator, and the true value (generally the mean) of the parameter to be estimated. MBE is not a measure of the model error, but it is mainly used for estimating the average bias of the prediction. It can indicate whether the model overestimates or underestimates the predicted variable, whit values equal to zero indicating unbiased results. According to Kambezidis (2012), both MBE and RMSE are subject to a drawback. For MBE, an overestimation in one observation can be cancelled by an underestimation in another, while for RMSE few large errors in the sum may produce a significant increase in the error. The use of the MAE can be complementary to the assessment because the residuals are not squared, while the absolute values avoid the effect of summing positive and negative errors.

Also, a t-Test can be used to overcome such drawbacks. The t-Test explains the overestimation or underestimation of the data at a certain significance level, usually 95%. For a population of $N > 120$ and the absolute value of t being less than, or equal to 1.96, there is no statistically significant difference between the measured and estimated data at the confidence level of 95%. Then, the more the t-value approaches 0, the higher the accuracy. Positive t-values mean that measured values are not statistically greater than the estimated values, while negative t-values mean that calculated values are not significantly smaller than the measured, at the confidence level of 95% (Kambezidis, 2012).

Moreover, because for RMSE and MAE it is difficult to evaluate the magnitude of the errors without comparing among different models, the normalized n-RMSE and n-MAE are given to allow direct evaluation. They are obtained by dividing RMSE and MAE by the range of the observed temperatures. The average error with the best value is equal to 0, and the worst equals 1.

All indicators are computed for three weighting schemes of the GWR, namely Exponential, Gaussian, and Inverse Distance, and for

Table 5
Cross-validation R^2, RMSE, MAE, and MBE for different bandwidths both for Gaussian and Exponential weighting.

GWR – 13:10 LT Daytime										
Gaussian weighting										
Bandwidth - meters	10,000	20,000	30,000	40,000	50,000	**60,000**	70,000	80,000	90,000	100,000
R^2	0.75	0.84	0.85	0.85	0.85	**0.86**	0.85	0.85	0.85	0.85
RMSE	2.45	1.94	1.89	1.89	1.83	**1.82**	1.83	1.84	1.85	1.85
MAE	1.76	1.40	1.36	1.39	1.39	**1.39**	1.40	1.41	1.42	1.42
MBE	0.33	0.19	0.14	0.11	0.09	**0.09**	0.09	0.10	0.10	0.10
Exponential weighting										
Bandwidth - meters	10,000	20,000	30,000	**40,000**	50,000	60,000	70,000	80,000	90,000	100,000
R^2	0.82	0.85	0.86	**0.86**	0.86	0.86	0.86	0.86	0.85	0.85
RMSE	2.05	1.85	1.77	**1.77**	1.79	1.80	1.81	1.82	1.83	1.85
MAE	1.44	1.36	1.35	**1.36**	1.37	1.38	1.39	1.40	1.40	1.43
MBE	0.24	0.15	0.10	**0.09**	0.09	0.09	0.09	0.09	0.10	0.10

Table 6

GWR exponential model performance for different predictor combinations, i.e. LST and DEM combined, LST only, and DEM only. A 2-fold CV protocol is employed.

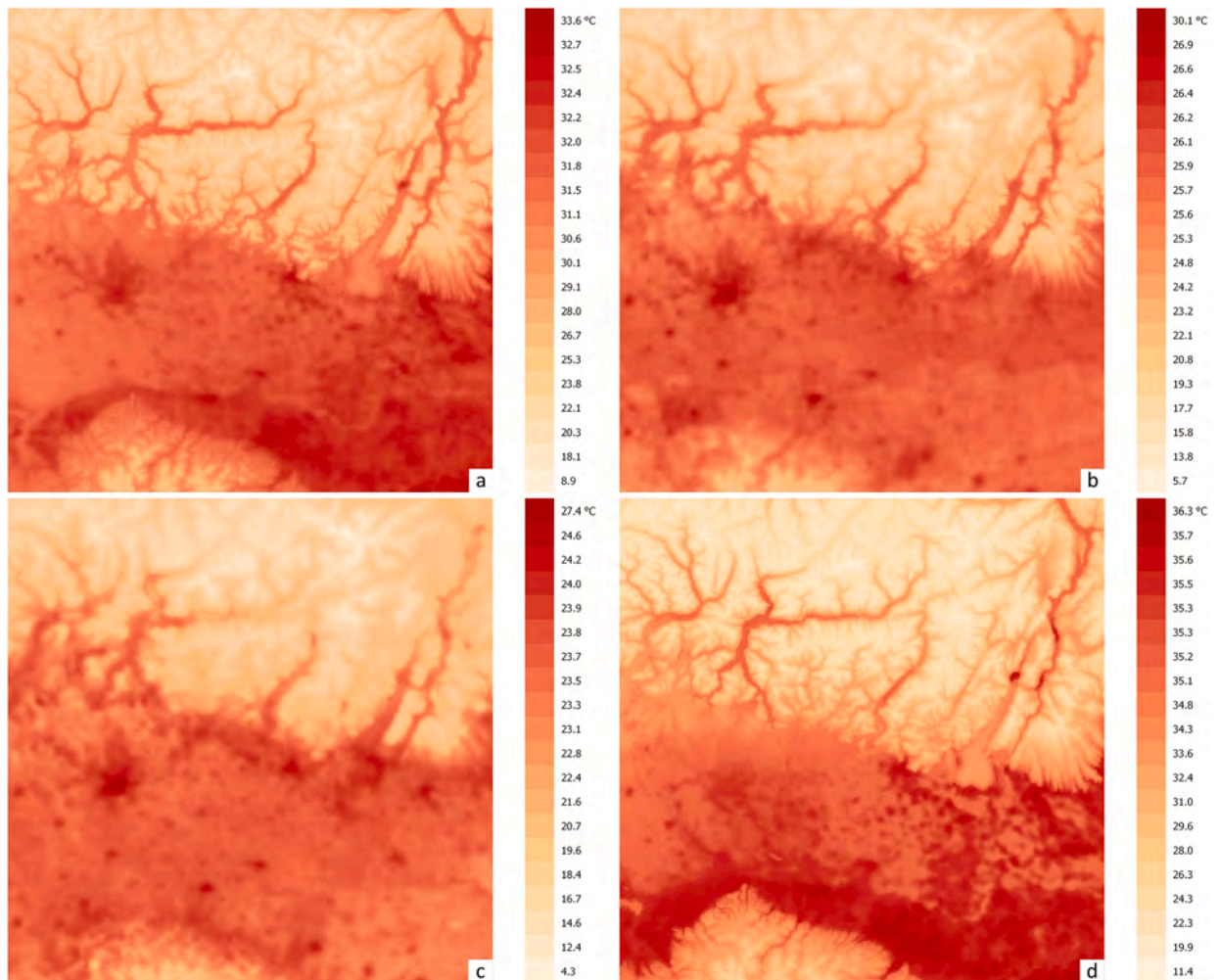| GWR Exponential													
2-fold Cross Validation		*LST and DEM*				*LST*				*DEM*			
Time	n. Stations	*Adj. R^2*	*RMSE*	*MAE*	*MBE*	*Adj. R^2*	*RMSE*	*MAE*	*MBE*	*Adj. R^2*	*RMSE*	*MAE*	*MBE*
10:40 - Day	195	0.88	1.50	**1.15**	−0.01	0.70	2.37	**1.85**	−0.22	0.87	1.54	**1.18**	0.07
21:50 - Night	195	0.87	1.45	**1.15**	0.05	0.83	1.93	**1.54**	−0.89	0.83	1.75	**1.39**	0.12
02:00 - Night	205	0.82	1.56	**1.23**	0.01	0.82	1.57	**1.22**	0.00	0.73	1.93	**1.52**	0.08
13:10 - Day	197	0.86	1.77	**1.36**	0.09	0.64	2.87	**2.28**	−0.16	0.86	1.79	**1.37**	0.09

Fig. 8. Predicted Near-Surface Air Temperature, for all time observations, based on GWR model with exponential weighting scheme, and a bandwidth of 40,000 m: a). 31st of July 2020 at 10:40, b). 31st of July 2020 at 21:50, c). 1st of August 2020 at 02:00, and d). 1st of August 2020 at 13:10.

global Multiple Linear Regression (MLR) and for Regression Kriging. Finally, because ARPA provides interpolations of weather stations' data, including T_a, on a raster basis, with a resolution of 1.5 km, we have assessed the accuracy of the models with respect to such data as a reference of the spatial distribution of T_a.

4.1. Validation of the GWR model with Exponential weighting scheme

For GWR Exponential model, with LST and DEM as predictors, Table 7 reports all metrics for each time observation. The assessment is given for both residuals on observed (known) points, and residuals for a 2-fold CV protocol. For CV, besides R^2, RMSE, MAE, and

MBE, normalized RMSE and MAE, and t-Test are also given.

For the 2-fold CV, i.e. for residuals at unknown points, the model is also assessed by using scatterplots, and Quantile-Quantile (Q-Q) plots for all time observations (Fig. 9).

Scatterplots refer to the correlation between measured and estimated T_a. The more the values scatter around a 45-degree line (reference line) the higher the similarity among estimated and measured values. In the case of the ideal predictive capacity of the model, the reference and regression line would completely be overlapping, while the Slope of the regression line and R^2 of the equation would tend to 1. The Q-Q plot considers the quantiles of the residuals with respect to the quantiles of a standard normal distribution. If residuals are normally distributed, the points on the graph should approximate the reference 45-degree line. Approaching the normal distribution means that most of the estimated values are in an interval of low error (negative or positive). In other words, the more the distribution approaches a normal distribution the higher the predictive capacity. Outliers are also emphasized by the point's distribution. Slope and R^2 of the Q-Q plot concerning the reference line approach 1 as the point's distribution approach the normality.

Finally, a sensitivity analysis is undertaken to test the impact that changing the input data of the model can have on the stability of outcomes. When the objective is the replicability of a model in further research, sensitivity analysis is relevant for establishing confidence in the findings of the initial study. To address such an issue, we have created some secondary models by splitting twice the input data, i.e. the observed T_a, into two datasets, hence we have run the model four times for different combinations of half of the points (50%), randomly selected, and for each period under investigation. All metrics are reported in Table 8.

If we plot all the models, at each time, made upon the original (whole) dataset, against the models calculated with the smaller datasets as in Table 8, the fit is very consistent, with a regression line with values of the slope that range from 0.96 to 1.02, and an R^2 that varies from 99.4 to 99.9.

Fig. 10 shows how much the R^2 varies, spatially, between the different models made with the subsets (50%) of the data. In particular, once plotted the spatialized R^2, we calculated the range among the four subsets (50%-1, −2, −3, and − 4) at each time. As it can be observed, overall, maximum values of the range do not exceed 0.1 at 10:40, 21:50, and 13:10, with values that reach 0.19 only for a small part of the image at 02:00. While, the average range of R^2 is 0.03 at 10:40 and 13:10, 0.04 at 21:50, and 0.08 at 02:00.

4.1.1. Model assessment by means of spatial autocorrelation of residuals

Generally, model assessment is made by using measures such as RMSE or MAE that average the errors for the entire scene, while little attention is paid to the spatial heterogeneity of the model performance. However, the latter is relevant for evaluating the predictive capacity. Such a concern can be addressed by assessing the spatial autocorrelation of residuals and, although it is reported the capability of GWR models to reduce spatial autocorrelation (Leung et al., 2000b; Propastin et al., 2008; Zhang and Gove, 2005), this should always be investigated.

For this research, we used two measures, namely the univariate Moran's I, for global spatial autocorrelation, and the Local Geary's C. Both analyses are undertaken using GeoDA. Moran's I is an effective measure, and appropriate test statistic, to explore the spatial distributions of the model errors (residuals) and the global predictive capacity. Small values indicate a weak autocorrelation structure in the residuals, which, in turn, suggests a good model fit (Lu et al., 2014; Propastin et al., 2008). Moran's I ranges from −1 to 1, with positive values indicating increasing aggregation, and negative values indicating dispersion. The closer the values to zero, the higher the spatial randomness patterns, with higher randomness indicating a better model (Wang et al., 2020).

The Local Geary's C (Anselin, 2019) is a Local Indicator of Spatial Association (LISA) that focuses on squared differences, or dissimilarity. Values lie between 0, and >1. Under the null hypothesis of spatial randomization, Geary approaches 1. Positive spatial autocorrelation is shown by values <1, whereas negative spatial autocorrelation is indicated by values >1.

While Moran's I assesses spatial autocorrelation on a global scale, whereas Geary's C is more sensitive to spatial autocorrelation on a local scale. In any case, determining a neighborhood search threshold distance is required for assessing autocorrelation. Both statistics are reliant on a geographic structural specification, such as a spatial weights matrix or a distance-related decline function. Either for Moran or Geary, we used a spatial weighting scheme based on a uniform Kernel, and an adaptive bandwidth by including 8 neighbors.

Table 9 shows both Moran's I, with corresponding z-score (Standard Deviations) and *p*-value (Probability) to account for the confidence level, and local significance of spatial autocorrelation as according to the Local Geary's C. In the latter case, it provides the number of no significant points (No-Sig.), the number of points significant at 0.05, 0.01, and 0.001, respectively, and the total percentage of significant points upon the total amount of observations (% Sig. pts.).

Moran values are mainly very close to 0, either positive or negative, which means no spatial autocorrelation is found globally. Also, values of z-score between −1.65 and 1.65, are considered not significant. For Local Geary, very few points are found to have significant spatial autocorrelation, mostly at 0.05. The only case that provides an increased significance is the model at 13:10 h, with a z-score of

Table 7

GWR Exponential model assessment for bandwidth of 40,000. Both observed (known) and unknown points are considered.

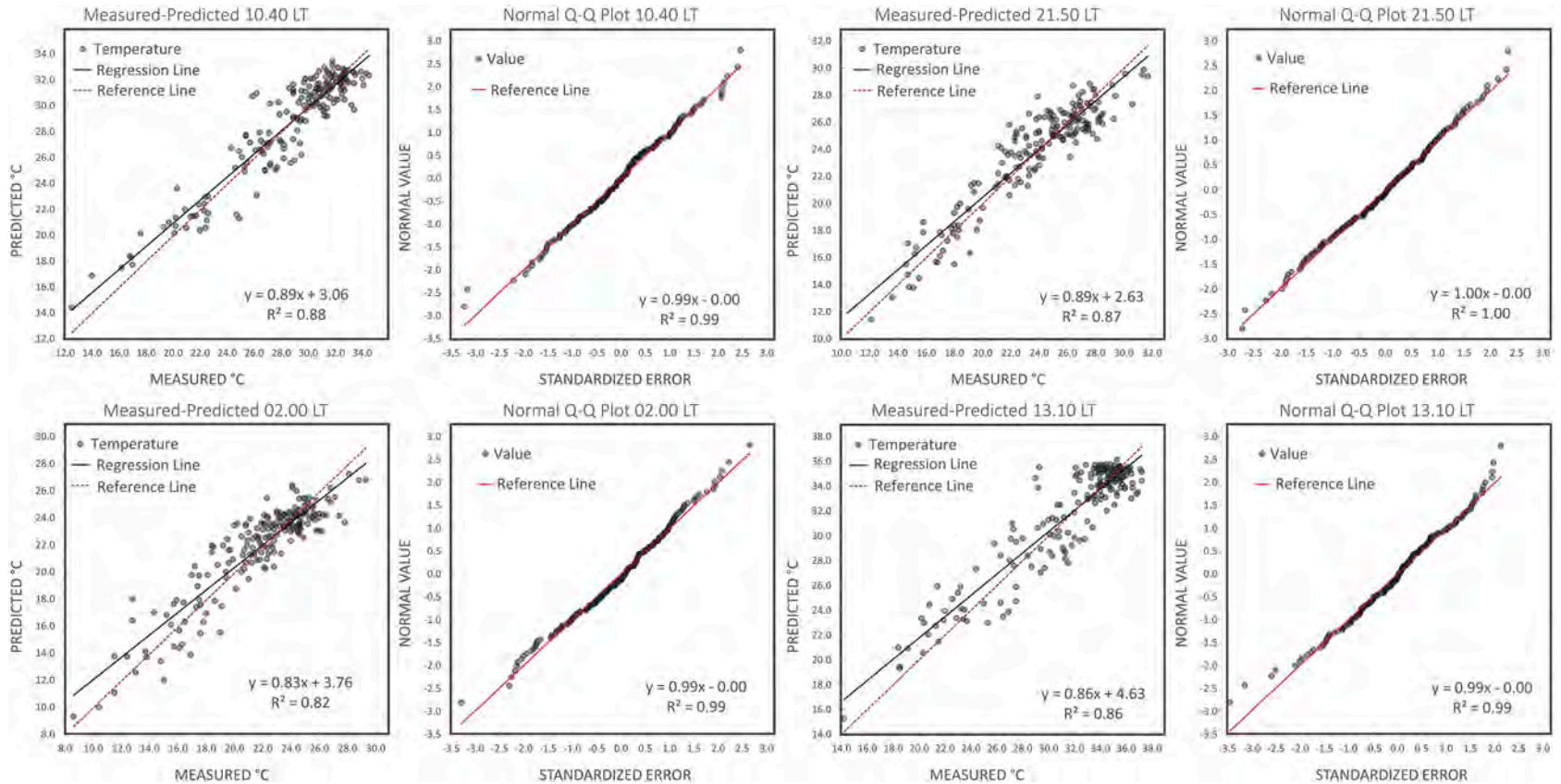| *GWR Exponential*		*Observed (Known) values*				*2-Fold Cross-Validation*						
Time	n. Stations	R^2	*RMSE*	**MAE**	*MBE*	R^2	*RMSE*	*n-RMSE*	**MAE**	*n-MAE*	*MBE*	*t-Test*
10:40 - Day	195	0.90	1.37	**1.08**	0.04	0.88	1.50	0.07	**1.15**	0.05	−0.01	0.00
21:50 - Night	195	0.89	1.36	**1.08**	0.04	0.87	1.45	0.07	**1.15**	0.05	0.05	−0.03
02:00 - Night	205	0.85	1.44	**1.13**	0.00	0.82	1.56	0.08	**1.23**	0.06	0.01	−0.01
13:10 - Day	197	0.89	1.55	**1.20**	0.06	0.86	1.77	0.08	**1.36**	0.06	0.09	−0.05

Fig. 9. Scatterplots and Q-Q plots for measured versus estimated temperatures according to a 2-fold CV protocol.

Table 8

Model sensitivity for each period under investigation, and four different combinations, randomly selected, of half of the input T_a, i.e. 50% of the points.

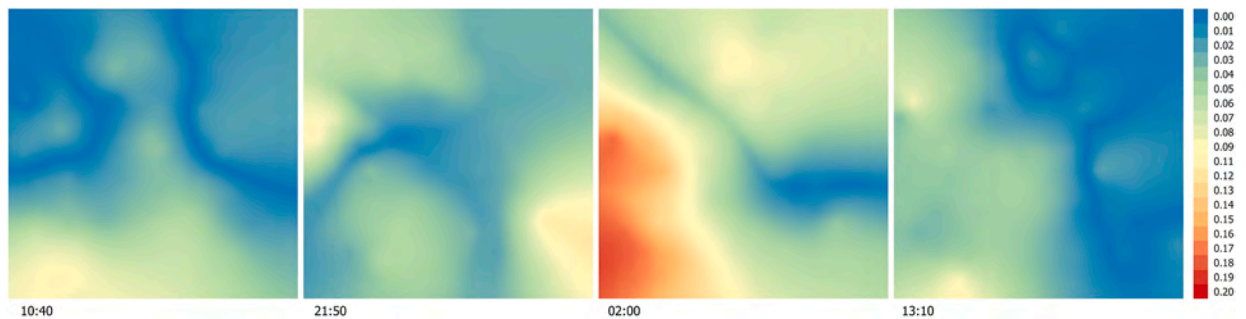| *GWR Exponential*										
Time	**n. Stations**	**50% - 1**	R^2	*RMSE*	*MAE*	*MBE*	**50% - 2**	R^2	*RMSE*	*MAE*	*MBE*
10:40 - Day	**195**	**98**	0.89	1.39	1.09	−0.19	**97**	0.87	1.59	1.21	0.18
21:50 - Night	**195**	**98**	0.87	1.48	1.16	−0.12	**97**	0.89	1.41	1.14	0.23
02:00 - Night	**205**	**103**	0.82	1.64	1.27	0.44	**102**	0.85	1.48	1.18	−0.42
13:10 - Day	**197**	**99**	0.87	1.64	1.26	−0.38	**98**	0.88	1.90	1.46	0.56
		50% - 3	R^2	*RMSE*	*MAE*	*MBE*	**50% - 4**	R^2	*RMSE*	*MAE*	*MBE*
10:40 - Day	**195**	**98**	0.89	1.59	1.21	0.06	**97**	0.89	1.31	1.06	−0.21
21:50 - Night	**195**	**98**	0.89	1.31	1.00	0.04	**97**	0.90	1.33	1.06	−0.04
02:00 - Night	**205**	**103**	0.85	1.41	1.09	−0.02	**102**	0.88	1.32	1.05	−0.01
13:10 - Day	**197**	**99**	0.86	1.72	1.29	0.22	**98**	0.88	1.76	1.34	−0.21

Fig. 10. Range of variation of R^2, spatialized, between different models made with an input subset of 50% of observed T_a.

Table 9

Moran's I, with z-score and p-value, and Local Geary's C significance analysis, for the GWR exponential model.

GWR Exponential		*Global Moran's I*						*Local Geary's C*				
		Observed (Known) values			*2-Fold Cross-Validation*			*2-Fold Cross-Validation*				
Time	n. Stations	*I*	*z-score*	*p-value*	*I*	*z-score*	*p-value*	*No-Sig.*	*0.05*	*0.01*	*0.001*	*% Sig. pts*
10:40 - Day	195	0.009	0.400	0.332	0.012	0.518	0.290	184	9	2	0	5.1
21:50 - Night	195	−0.014	−0.261	0.417	−0.012	−0.241	0.410	173	18	4	0	11.3
02:00 - Night	205	−0.001	0.138	0.434	0.004	0.237	0.388	180	23	2	0	12.2
13:10 - Day	197	0.099	3.069	0.003	0.074	2.407	0.007	155	31	10	1	21.3

2.407 and a *p*-value of 0.007, for a CV protocol. In this case, the null hypothesis that the attribute being analyzed is randomly distributed can be rejected. An increased percentage of significant points is also found locally (21.3%). Nonetheless, although significant, the spatial clustering of the residuals, either for known or unknown points, is very close to 0, namely 0.099 and 0.074.

The model at 13:10 results slightly less effective because the correlation between LST and T_a is reduced at the hottest time. To address such a limitation, the future investigation will focus on improving the model performance based on the use of either the QQ plots or Local Geary's C, to remove outliers' points.

4.2. Exponential versus Gaussian weighting

While the Gaussian scheme performs very similarly to the Exponential, the Inverse Distance weighting shows different behavior and results, depending on the distance and the value of the power. As the power increases, the weights decrease exponentially. To

Table 10

The predictive capacity of the GWR model with Gaussian weighting and bandwidth of 60,000.

GWR Gaussian		*Observed (Known) values*				*2-Fold Cross-Validation*						
Time	n. Stations	R2	*RMSE*	**MAE**	*MBE*	R2	*RMSE*	*n-RMSE*	**MAE**	*n-MAE*	*MBE*	*t-Test*
10:40 - Day	195	0.89	1.44	**1.13**	0.05	0.88	1.51	0.07	**1.16**	0.05	0.01	−0.01
21:50 - Night	195	0.88	1.43	**1.13**	0.04	0.87	1.47	0.07	**1.18**	0.06	0.05	−0.03
02:00 - Night	205	0.84	1.51	**1.18**	0.01	0.82	1.57	0.08	**1.23**	0.06	0.02	−0.01
13:10 - Day	197	0.88	1.66	**1.28**	0.06	0.86	1.82	0.08	**1.39**	0.06	0.09	−0.05

assess the goodness of the Exponential weighting scheme, the latter is compared with both the Gaussian and the ID. LST and DEM are the predictors for all models.

For Gaussian weighting, a bandwidth of 60,000 is used to maximize the predictive capacity (see Table 5). All performance indicators are provided for the Gaussian weighting model in Table 10, both for residuals at the observed (known) points, as well as for the 2-fold CV protocol, while Table 11 reports on the model sensitivity. Finally, in Table 12 Moran's I, and Local Geary's C are given for assessing spatial autocorrelation of 2-Fold CV residuals.

Although the predictive capacity of the Gaussian scheme is very similar to the Exponential, the latter shows an improved performance either for the values at the observed points or the residuals of the CV protocol. Both errors, namely RMSE and MAE, for the Gaussian scheme slightly increase for each observation, with a more significant difference for the model at 13:10. Likewise, regarding the sensitivity of the Gaussian scheme, though the stability of the model, overall we report a slight worsening in the model. This also applies to spatial autocorrelation, with an increased percentage of significant points, according to the Local Geary, mostly during daytime.

4.3. Exponential and inverse distance weighting

Regarding ID weighting, we tested different values of the power. Commonly, values used for the power (α) are greater or equal to 1.0. However, there is no theoretical justification for predefined values. The effect of the power should be assessed by Cross-Validation statistics. In practice, the power is seldom estimated but typically set to a fixed value. If α is very high, only the immediate surrounding points will influence the estimation. Conversely, if α equals zero, there is no decrease in the distance (Anselin, 2018). We tested two powers, namely 1.0 and 2.0, as shown in Table 13.

Although increasing the power from 1.0 to 2.0 significantly increases the accuracy of the estimated T_a at the observed (known) points, the predictive capacity for unknown values is dramatically reduced. At observed points, all measures improve, and a relevant reduction of the errors occurs when increasing the power from 1.0 to 2.0. However, the situation is completely different when considering the predictive capacity according to a CV protocol. Although for power 1.0 the values are similar to Exponential and Gaussian, for power 2.0 the errors slightly increase with respect to the Exponential model. Also, MBE and *t*-Test get slightly worse.

Such a significant difference in the predictive capacity, between observed and unknown points, is critical, it underlines a problem of model over-fitting, and generates an unfavorable heterogeneity of the results. Abnormal peaks emerge around the observed points. As an example, we provide the spatialized R^2 for the ID model at 10.40, which provides key information about the predictive capacity of the model, locally.

In Fig. 11, we compare the spatial distribution of R^2 for the Exponential model (11a), with respect to Gaussian (11b), and ID with power 1.0 (11c) and 2.0 (11d). First, we emphasize that, although the average R^2 of the whole image is similar for all weighting schemes, minimum and maximum values show a smaller range for Exponential and Gaussian (around 0.75 to 0.94), with respect to ID weighting 1.0 (around 0.61 to 0.98) and 2.0 (around 0.42 to 0.99). This increased heterogeneity of the distribution of R^2, in the case of ID models, and the significant difference between observed and unknown points, is then reflected in the spatial distribution of the estimated T_a, which will result in anomalous peaks of temperature.

If we consider a sample transect (North-South), as in previous Fig. 11c (E-E^1), and compare the behavior of the R^2 along it, for each time observation (Fig. 12), the graphs clearly show the heterogeneity of results in the case of ID weighting. Such heterogeneity critically increases by increasing the power. Locally, the R^2 shows a strong variability, and improbably high values, both positive and negative. On the other hand, the Gaussian scheme flattens out the R^2 distribution, with respect to the Exponential. According to this study, we discarded the ID weighting scheme for our purposes.

4.4. GWR versus MLR and regression kriging

The GWR performance is also compared with two widely recognized regression techniques, namely Multiple Linear Regression (MLR) and Regression Kriging (RK). Such methods have been extensively explored, employed, and reported in the literature. MLR uses the linear correlation between a dependent variable, or response, and the predictors, or explanatory variables. Because the model is based on a global correlation, in MLR it is assumed that residuals among observed and estimated values do not show significant

Table 11

Sensitivity analysis of the Gaussian weighting model for different combinations of half of the input T_a, i.e. 50% of the points, randomly selected.

GWR Gaussian										
Time	*n. Stations*	*50% - 1*	*R^2*	*RMSE*	*MAE*	*MBE*	*50% - 2*	*R^2*	*RMSE*	*MAE*	*MBE*
10:40 - Day	**195**	**98**	0.89	1.40	1.11	−0.17	**97**	0.87	1.61	1.21	0.20
21:50 - Night	**195**	**98**	0.87	1.51	1.19	−0.13	**97**	0.88	1.44	1.16	0.22
02:00 - Night	**205**	**103**	0.82	1.65	1.27	0.45	**102**	0.85	1.49	1.18	−0.41
13:10 - Day	**197**	**99**	0.87	1.69	1.29	−0.39	**98**	0.87	1.95	1.49	0.57
		50% - 3	*R^2*	*RMSE*	*MAE*	*MBE*	*50% - 4*	*R^2*	*RMSE*	*MAE*	*MBE*
10:40 - Day	**195**	**98**	0.88	1.62	1.26	−0.03	**97**	0.88	1.31	1.06	−0.01
21:50 - Night	**195**	**98**	0.87	1.38	1.06	0.03	**97**	0.89	1.42	1.14	−0.03
02:00 - Night	**205**	**103**	0.84	1.49	1.16	−0.02	**102**	0.85	1.43	1.14	−0.01
13:10 - Day	**197**	**99**	0.86	1.67	1.27	−0.04	**98**	0.87	1.78	1.38	−0.02

Table 12

Moran's I, with z-score and *p*-value, and Local Geary's C significance analysis, for the GWR Gaussian model.

GWR Gaussian		Global Moran's I						Local Geary's C				
		Observed (Known) values			2-Fold Cross-Validation			2-Fold Cross-Validation				
Time	n. St.	I	z-score	p-value	I	z-score	p-value	No-Sig.	0.05	0.01	0.001	% Sig. pts
10:40 - Day	195	0.031	1.073	0.146	0.032	1.303	0.100	182	11	2	0	6.7
21:50 - Night	195	0.022	0.814	0.206	0.017	0.691	0.242	173	18	4	0	11.3
02:00 - Night	205	0.028	1.055	0.148	0.032	1.141	0.134	180	21	4	0	12.2
13:10 - Day	197	0.132	4.042	0.001	0.107	3.435	0.001	149	33	14	1	24.4

Table 13

Performance of the GWR with ID weighting scheme, for two powers, namely 1.0 and 2.0.

GWR - ID Weighting		Observed (Known) values				2-Fold Cross-Validation					
Power $\alpha = 1.0$											
Time	n. Stations	R2	RMSE	**MAE**	MBE	R2	RMSE	n-RMSE	**MAE**	n-MAE	MBE	t-Test
10:40 - Day	195	0.97	0.74	**0.56**	0.00	0.88	1.49	0.07	**1.14**	0.05	−0.02	0.01
21:50 - Night	195	0.96	0.77	**0.60**	0.04	0.88	1.43	0.07	**1.14**	0.05	0.06	−0.04
02:00 - Night	205	0.95	0.80	**0.61**	0.01	0.82	1.60	0.08	**1.25**	0.06	−0.01	0.00
13:10 - Day	197	0.97	0.82	**0.62**	0.01	0.87	1.76	0.08	**1.36**	0.06	0.11	−0.06
Power $\alpha = 2.0$											
Time	n. Stations	R2	RMSE	**MAE**	MBE	R2	RMSE	n-RMSE	**MAE**	n-MAE	MBE	t-Test
10:40 - Day	195	0.99	0.39	**0.26**	0.08	0.87	1.53	0.07	**1.18**	0.05	−0.03	0.02
21:50 - Night	195	1.00	0.25	**0.17**	0.02	0.88	1.44	0.07	**1.15**	0.05	0.07	−0.05
02:00 - Night	205	1.00	0.19	**0.14**	0.00	0.81	1.63	0.08	**1.27**	0.06	0.02	−0.01
13:10 - Day	197	0.99	0.47	**0.31**	0.10	0.86	1.82	0.08	**1.41**	0.06	0.18	−0.10

variability.

An MLR model is designed to estimate T_a based on day- and night-time LST, and DEM as predictors. MLR is run in SAGA GIS, using the Nearest Neighbor resampling method, for each time observation. All the performance indicators, as well as the spatial autocorrelation of residuals, are reported in Table 14 and Table 15, respectively. Due to the non-stationarity of T_a, the predictive performance of MLR is dramatically reduced both day- and night-time. Because the difference between T_a and LST is much higher during daytime, MLR performance is significantly reduced mostly during daytime.

The Kriging is a family of interpolation methods that rely on the semi-variance analysis for characterizing the spatial autocorrelation of geo-located observations. Although different kriging methods are available, in general, they presume the existence of intrinsic stationarity of the phenomena (Corona et al., 2014). However, improved linear unbiased prediction for unknown data is achieved by a moving-average-based approach.

The Regression Kriging (RK) combines regression and kriging into two separate consecutive steps. After the regression, a kriging procedure is applied to residuals. Hence, the kriged residual is added to the regression result. In practice, RK is an improved MLR that, by incorporating residual kriging into regression, enhances the prediction accuracy (Bostan et al., 2012). RK uses a variogram that, based on different functions, estimates the variance between pairs of points at various distances. Variogram is then used to compute weights that minimize variance and improve predictive capacity.

Kriging employed in regression can be Simple Kriging (SK) and Ordinary Kriging (OK). However, while SK assumes the stationarity of the first moment (a global mean is given) and covariance, over the entire area, OK employs a constant mean over limited areas by a moving search radius. Hence, OK assumes a quasi-stationarity condition by varying the mean but keeping constant the covariance. This improves the estimation even with non-stationarity (Lloyd, 2005).

In this study, RK-OK is run in SAGA GIS. Different variogram functions are experimented with, including linear, exponential, logarithmic, Gaussian, cubic, and spherical. Lag distance, fitting range, and samples in the range are automatically estimated depending on the points' distribution. Among the tested functions, linear and logarithmic are those that provide higher R^2, with higher values at night-time for the linear function, and higher values during daytime for the logarithmic one. In particular, R^2 (%) for linear and logarithmic variogram functions are, respectively, 59.1 (linear) and 78.3 (logarithmic) at 10:40 (day), 65.6 and 85.7 at 13:10 (day), 83.3 and 57.4 at 21:50 (night), and 38.8 and 35.6 at 02:00 (night). The latter, i.e. the coldest hour, proves the less significant difference between the two functions.

Accuracy assessment for RK-OK with linear and logarithmic functions, and for both observed (known) and unknown points (2-Fold CV) is provided in Table 16. While, Table 17 reports the spatial autocorrelation of the residuals.

Although compared with MLR, the RK is estimating better at observed (known) points, the predictive capacity of unknown points based on the 2-Fold CV protocol is only slightly improved. However, regarding spatial autocorrelation, for a logarithmic function, overall, results improve mostly locally. On the other hand, if we compare the performance of the RK-OK for known and unknown points, while the linear function maintains certain stability (despite the overall reduced performance), for the logarithmic function the performance, for the CV indicators, dramatically worsens. In particular, we report a critical reduction of the performance for the 2-Fold CV at 02:00. Therefore, in Table 18 we further provide the sensitivity analysis of the RK-OK Logarithmic, which effectively accounts for
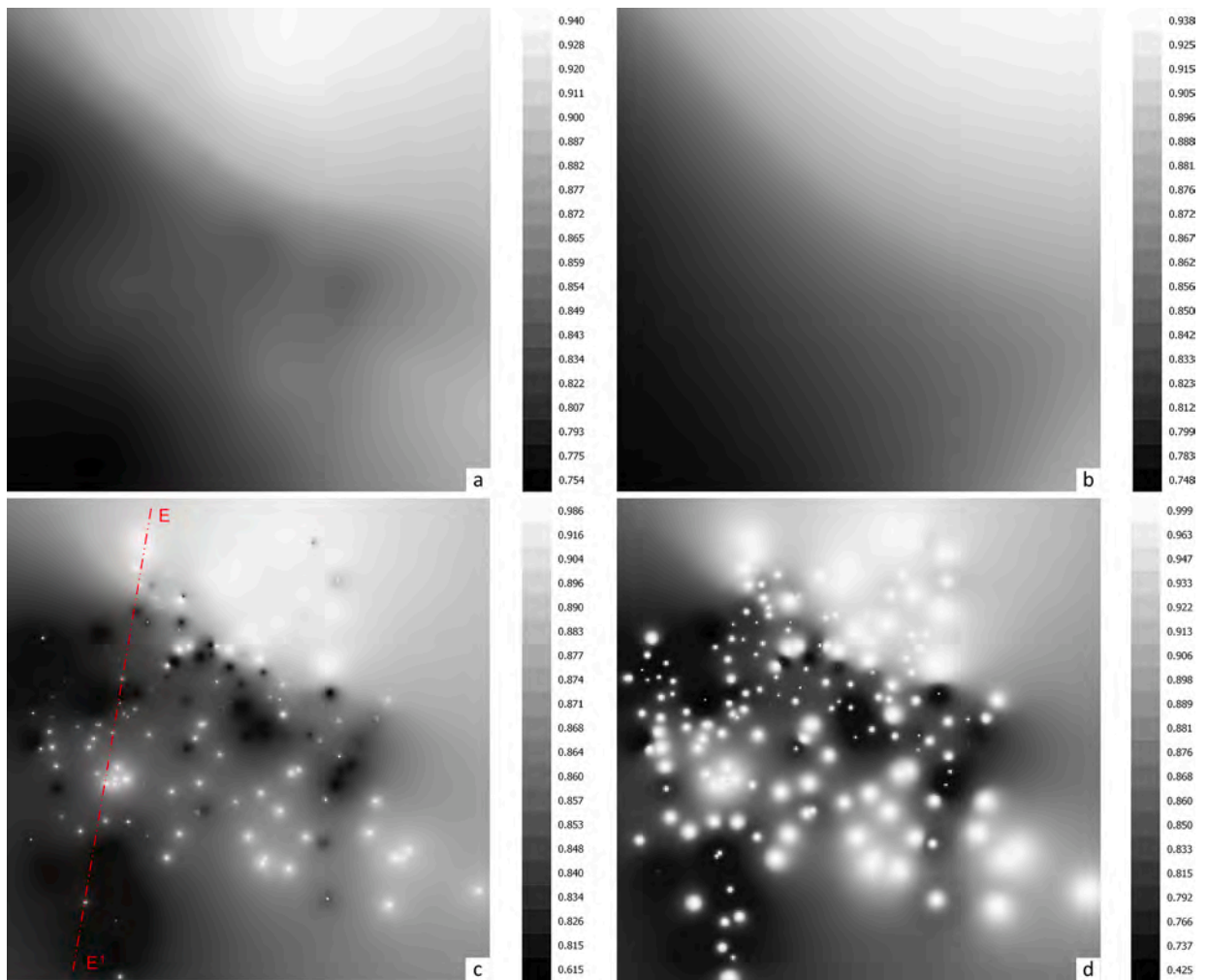
Fig. 11. Spatial distribution of R^2 for GWR models at time 10.40: a). Exponential, b). Gaussian, c). ID weighting, with power 1.0, and d). ID with power 2.0.

a significant instability.

4.5. Model validation with ARPA weather data interpolation

Since October 2017, and based on the regional network of weather stations, ARPA is processing hourly interpolations for each observed weather variable, including T$_a$ at 2 m from the ground. The interpolations are provided as raster data with a spatial resolution of 1.5 km, according to an Optimal Interpolation (OI) approach and three-dimensional correlation functions (Lussana et al., 2009; Uboldi et al., 2008). Such information is the official reference in the Lombardy region regarding weather variables. Hence, an accuracy assessment, of all models, is also provided with respect to the ARPA T$_a$ interpolations as reference data.

Because of the different spatial resolution, NSAT was up-scaled from 900 m to 1.5 km, based on mean pixel values. Additionally, ARPA interpolations were converted to points. A total of 11,107 points were obtained assessment. Finally, because interpolations are hourly-based (integer hour), we selected times closest to the observations under investigation. Table 19 shows R^2, RMSE, MAE, and MBE for the GWR models, namely Exponential, Gaussian, and ID with power 1.0, for MLR, and RK-OK both for linear and logarithmic functions of the variogram.

5. Discussion

Near-surface air temperature is key for UHI studies. Although several investigations address the UHI phenomenon based on LST, it is critical to emphasize that, while at night-time the thermal profiles of NSAT and LST get closer to each other, daytime NSAT is considerably different. Consequently, daytime LST is not a suitable proxy to assess the UHI (Pepin et al., 2016; Sun et al., 2020). In other words, diurnal UHI cannot be associated with the LST. However, despite the increasing attention on the study of UHI, easy-to-
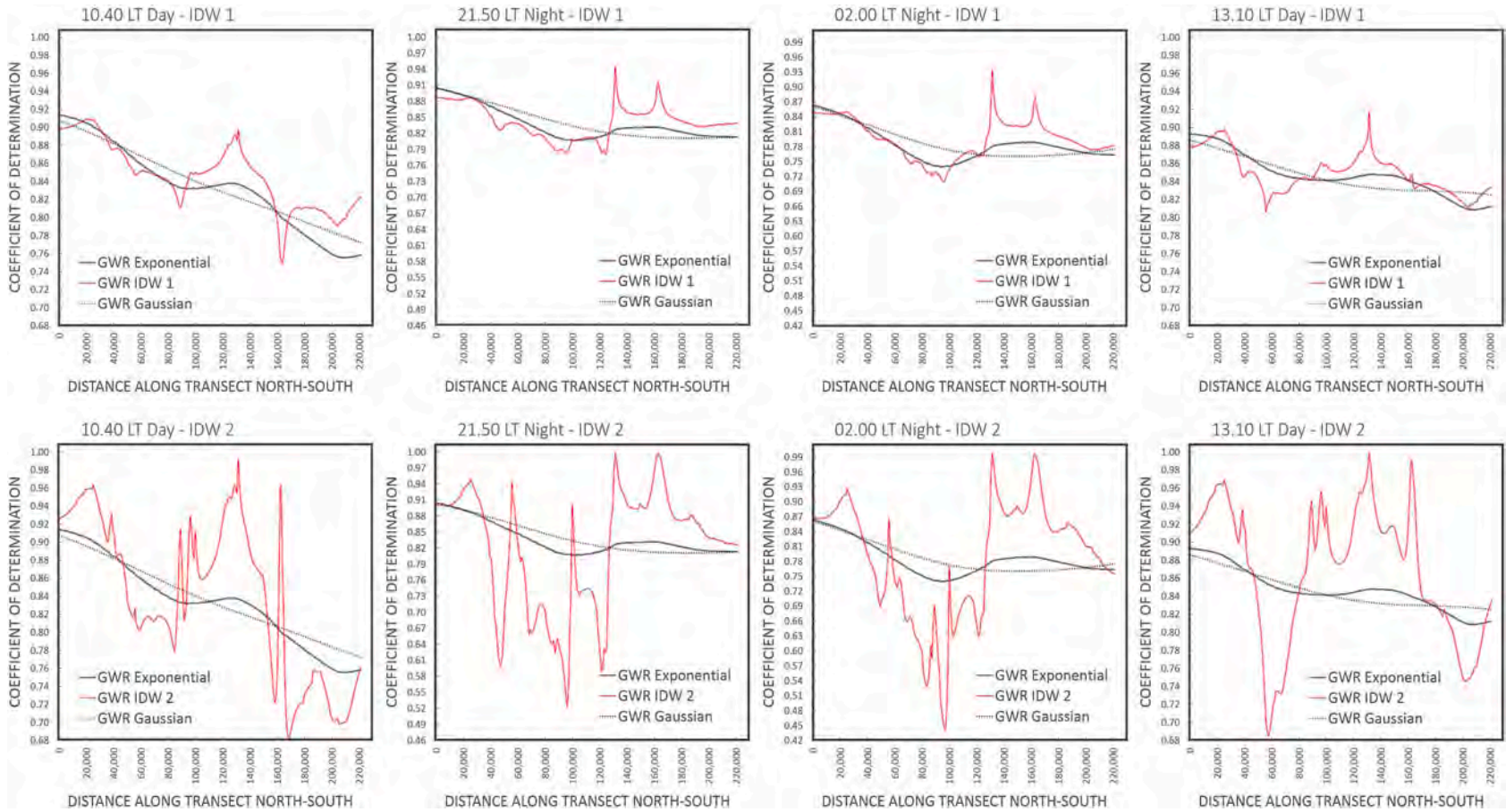
Fig. 12. Sample transect (E-E[1]) through the spatialized R[2] for GWR models with Exponential, Gaussian, and ID weighting with power 1.0 and 2.0. Each time observation is assessed, i.e. 10:40, 21:50, 02:00, and 13:10.

Table 14

The predictive capacity of MLR model, for each time observation.

Multiple Linear Regression		Observed (Known) values				2-Fold Cross-Validation					
Time	n. Stations	R2	RMSE	MAE	MBE	R2	RMSE	n-RMSE	MAE	n-MAE	MBE	t-Test
10:40 - Day	195	0.71	2.53	**2.01**	−0.16	0.70	2.57	0.12	**2.04**	0.09	−0.15	0.06
21:50 - Night	195	0.86	1.70	**1.36**	−0.14	0.86	1.71	0.08	**1.37**	0.06	−0.14	0.08
02:00 - Night	205	0.83	1.77	**1.37**	−0.16	0.82	1.80	0.09	**1.38**	0.07	−0.17	0.09
13:10 - Day	197	0.62	3.25	**2.67**	−0.23	0.60	3.33	0.14	**2.74**	0.12	−0.19	0.06

Table 15

Moran's I, and Local Geary's C for an MLR model. T_a is the dependent variable, and LST and DEM are the predictors.

Multiple Linear Regression		Global Moran's I						Local Geary's C				
		Observed (Known) values			2-Fold Cross-Validation			2-Fold Cross-Validation				
Time	n. St.	I	z-score	p-value	I	z-score	p-value	No-Sig.	0.05	0.01	0.001	% Sig. pts
10:40 - Day	195	0.18	5.78	0.00	0.17	5.38	0.00	159	22	6	8	18.5
21:50 - Night	195	0.13	4.03	0.00	0.13	3.89	0.00	164	22	6	3	15.9
02:00 - Night	205	0.08	2.63	0.00	0.08	2.45	0.01	164	37	4	0	20.0
13:10 - Day	197	0.20	6.26	0.00	0.18	5.50	0.00	147	28	15	7	25.4

Table 16

Accuracy assessment of Regression Kriging-Ordinary Kriging model, for each time observation, and two variogram functions, namely linear and logarithmic.

Regression Kriging – Ordinary Kriging		Observed (Known) values				2-Fold Cross-Validation						
Variogram Function - Linear												
Time	n. Stations	R2	RMSE	MAE	MBE	R2	RMSE	n-RMSE	MAE	n-MAE	MBE	t-Test
10:40 - Day	195	0.74	2.26	**1.81**	0.05	0.66	2.65	0.12	**2.11**	0.10	0.02	−0.01
21:50 - Night	195	0.88	1.42	**1.14**	0.02	0.86	1.58	0.07	**1.28**	0.06	0.02	−0.01
02:00 - Night	205	0.83	1.64	**1.27**	0.02	0.78	1.84	0.09	**1.43**	0.07	−0.01	0.01
13:10 - Day	197	0.67	2.83	**2.31**	0.09	0.61	3.09	0.13	**2.53**	0.11	0.11	−0.03
Variogram Function - Logarithmic												
Time	n. Stations	R2	RMSE	MAE	MBE	R2	RMSE	n-RMSE	MAE	n-MAE	MBE	t-Test
10:40 - Day	195	0.92	1.25	**0.99**	0.05	0.68	2.55	0.12	**2.05**	0.09	−0.08	0.03
21:50 - Night	195	0.98	0.64	**0.50**	0.02	0.86	1.60	0.08	**1.29**	0.06	0.04	−0.02
02:00 - Night	205	0.87	1.39	**1.08**	0.02	0.01	10.4	0.50	**6.67**	0.32	0.52	−0.05
13:10 - Day	197	0.95	1.04	**0.80**	0.08	0.64	2.94	0.13	**2.36**	0.10	0.10	−0.03

Table 17

Moran's I, and Local Geary's C, for RK-OK models with linear and logarithmic variogram functions.

RK – Ordinary Kriging		Global Moran's I						Local Geary's C				
Variogram Function - Linear		Observed (Known) values			2-Fold Cross-Validation			2-Fold Cross-Validation				
Time	n. Stations	I	z-score	p-value	I	z-score	p-value	No-Sig.	0.05	0.01	0.001	% Sig. pts
10:40 - Day	195	0.04	1.30	0.11	0.07	2.25	0.02	171	15	5	4	12.3
21:50 - Night	195	−0.05	−1.42	0.07	−0.02	−0.47	0.35	182	8	5	0	6.7
02:00 - Night	205	−0.02	−0.33	0.38	0.09	2.97	0.00	154	36	12	3	24.9
13:10 - Day	197	0.04	1.30	0.11	0.05	1.70	0.05	157	29	7	4	20.3
Variogram Function - Logarithmic												
Time	n. Stations	I	z-score	p-value	I	z-score	p-value	No-Sig.	0.05	0.01	0.001	% Sig. pts
10:40 - Day	195	−0.05	−1.30	0.08	−0.01	−0.04	0.49	173	18	4	0	11.3
21:50 - Night	195	−0.01	−2.65	0.00	−0.03	−0.77	0.23	186	5	4	0	4.6
02:00 - Night	205	−0.05	−1.33	0.09	0.41	12.8	0.00	111	47	35	12	45.9
13:10 - Day	197	−0.04	−0.92	0.19	−0.04	−0.92	0.18	167	23	7	0	15.2

replicate standardized methods for automatic modelling of instantaneous NSAT, at a high spatio-temporal resolution such as sub-daily or hourly, and both day- and night-time, are still very few explored. The majority of investigations focus on maximum and minimum, and average monthly, 8-days, and/or daily average air temperature.

Recent studies provided relevant insight toward establishing operative and replicable methods, or tools, for estimating high spatio-temporal resolution T_a. A cost-effective method, based on open data, for spatially continuous mapping of high-resolution T_a is proposed for data on a daily basis (Alvi et al., 2022). Also, a GIS plugin was introduced to allow modelling NSAT from urban meteorological networks, based on a kriging regression model, and suitable at the scale of the neighborhood rather than pixel-based (Touati et al., 2020). A three-stage geographically weighted generalized additive model (GAM) was explored to estimate hourly NSAT, at a

Table 18

Sensitivity analysis for the RK-OK model with logarithmic function for different random combinations of half of the input T_a, i.e. 50% of the points.

Regression Kriging - Ordinary Kriging | Logarithmic											
Time	n. Stations	50% - 1	R^2	RMSE	MAE	MBE	50% - 2	R^2	RMSE	MAE	MBE
10:40 - Day	195	98	0.70	2.66	2.14	−0.29	97	0.69	2.44	1.95	0.13
21:50 - Night	195	98	0.85	1.58	1.24	−0.19	97	0.88	1.63	1.35	0.26
02:00 - Night	205	103	0.79	1.84	1.41	0.38	102	0.02	14.6	11.9	0.67
13:10 - Day	197	99	0.67	2.70	2.19	−0.33	98	0.63	3.16	2.53	0.53
		50% - 3	R^2	RMSE	MAE	MBE	50% - 4	R^2	RMSE	MAE	MBE
10:40 - Day	195	98	0.89	1.54	1.25	0.00	97	0.94	0.95	0.77	0.00
21:50 - Night	195	98	0.99	0.23	0.18	0.00	97	0.15	15.5	13.8	0.00
02:00 - Night	205	103	0.86	1.43	1.12	0.00	102	0.83	1.89	1.44	0.00
13:10 - Day	197	99	0.99	0.56	0.46	0.00	98	0.73	2.63	2.18	0.00

Table 19

Accuracy assessment of GWR (Exponential, Gaussian, and ID with power 1.0), MLR, and RK-OK with respect to ARPA interpolations of T_a.

Accuracy Assessment with ARPA interpolations												
	11,107 points	GWR - Exponential				GWR - Gaussian				GWR - IDW 1		
Time	Interpolation Time	R2	RMSE	MAE	MBE	R2	RMSE	MAE	MBE	R2	RMSE	MAE	MBE
10:40 - Day	11:00	0.96	1.17	**0.72**	0.02	0.96	1.19	**0.74**	0.04	0.96	1.18	**0.72**	0.03
21:50 - Night	22:00	0.94	1.22	**0.87**	−0.33	0.94	1.25	**0.88**	−0.32	0.95	1.18	**0.86**	−0.32
02:00 - Night	02:00	0.94	1.10	**0.78**	−0.26	0.93	1.14	**0.80**	−0.27	0.94	1.10	**0.77**	−0.26
13:10 - Day	13:00	0.95	1.30	**0.92**	0.38	0.95	1.33	**0.95**	0.41	0.95	1.31	**0.91**	0.40
	11,107 points	MLR				Kriging - Linear				Kriging - Logarithmic		
Time	Interpolation Time	R2	RMSE	MAE	MBE	R2	RMSE	MAE	MBE	R2	RMSE	MAE	MBE
10:40 - Day	11:00	0.79	2.39	**1.83**	−0.53	0.79	2.31	**1.72**	−0.26	0.83	2.14	**1.56**	−0.19
21:50 - Night	22:00	0.90	1.63	**1.20**	−0.58	0.92	1.41	**1.06**	−0.39	0.92	1.35	**1.02**	−0.39
02:00 - Night	02:00	0.92	1.52	**1.05**	−0.51	0.92	1.27	**0.91**	−0.31	0.92	1.22	**0.88**	−0.30
13:10 - Day	13:00	0.74	2.82	**2.18**	−0.40	0.74	2.69	**2.06**	−0.09	0.80	2.41	**1.77**	−0.02

downscaled spatial resolution of 1 km per pixel, combining machine learning algorithms and a non-parametric approach (Zhou et al., 2020).

In line with these studies, we establish an easy-to-replicate, yet effective, original operational model, based on open data, for systematic estimation of high-resolution instantaneous NSAT, during heat wave events, and for large geographical areas. The model employs a GWR algorithm, with exponential weighting, which combines on-site measured T_a, from local fixed weather station networks, as the dependent variable, and satellite-derived LST and elevation (DEM) as predictors. The achieved NSAT is time instantaneous data with daily and hourly frequency consistent with MODIS revisiting time. Here, we tested the model for four hours, and for 2 consecutive days, namely on the 31st of July and 1st of August 2020, during a heat wave event in the Lombardy region.

To calibrate the model, we first experimented with different predictors, namely Wind Speed, LST, and DEM. After proving the irrelevant correlation among Ws and T_a, for the area and periods under investigation, we also proved the enhanced predictive capacity of combining LST and DEM as predictors. Indeed, based on our findings, we point out that the use of DEM, in combination with LST, significantly increased the model performance during a heat wave event. Mostly daytime, when T_a and LST are less close to each other, the use of DEM is relevant. The use of both variables improves all performance indicators either for (known) observed values, as well as for estimated (unknown) values, both day- and night-time. Nevertheless, LST reconstruction, if needed, should be carefully addressed because thermal data is extremely sensitive to short-term time changes.

To prove the effectiveness of the model, we provided a deep unambiguous validation, including the assessment of spatial autocorrelation of the residuals, seldom explored to validate spatial modelling. Because of the spatial resolution of the data, ad-hoc in-situ validation, by traverse measurement with a car, for instance, is neither feasible nor useful. Hence, we have designed a 2-fold Cross-Validation protocol. Deep testing is provided, including the comparison with different weighting schemes of the GWR, MLS, and Regression Kriging.

Among the tested models, the GWR with an exponential weighting scheme showed higher performance. Very high R^2, which reports the percentage of explained variance, was observed both for known points, with values ranging from 0.85 to 0.90, as well as unknown (estimated) points with values ranging from 0.82 to 0.88 for 2-Fold CV. Bias was also found to be very low, with MBE very close to 0. RMSE ranges from around 1.36 to 1.55 and MAE from 1.08 to 1.20 for observed points, while RMSE is 1.45 to 1.77 and MAE 1.15 to 1.36 for 2-Fold CV. Generally, 1–2 °C is considered an accurate level of precision (Benali et al., 2012). Normalized RMSE and MAE are roughly 0.05 to 0.08, which means 5% to 8% of the range of the observed temperatures. Further, no spatial autocorrelation of CV residuals was found for the models globally, in three out of four models. Only little significant correlation arises for the model at 13:10. To reduce such a concern, which mainly regards hottest hours, either 2-Fold CV Q-Q plots, RMSE or Local Geary's C could be used for calibrating the models. According to Geary's C, very few points were found with little significant spatial autocorrelation. Such points should be further observed.

Finally, we provided a sensitivity analysis to prove the model's stability. Although GWR shows good stability either for Exponential or Gaussian weighting, with the Exponential scheme performing higher, ID weighting reveals a critical instability. Besides, we report a

reduced predictive capacity of global models, such as MLR and Regression Kriging and, despite we underline an improved capacity for a RK-OK with a logarithmic function of the variogram, we also report the latter to be critically unstable in terms of predictivity.

Among the kriging functions, the Universal Kriging (UK) should be explored to better respond to non-stationarity, but calibration is not reliable, thus increasing complexity (Bostan et al., 2012; Hengl et al., 2007). Yet, it is reported that the GWR model is more effective than the Kriging model in estimating T_a during warmer months (Wang et al., 2017). Several pieces of research suggested a hybrid GWR combined with interpolated residual kriging (GWR-K) for spatial prediction of T_a, reporting to be more effective for modelling non-stationary phenomena (Harris et al., 2010; Szymanowski et al., 2013; Szymanowski and Kryza, 2011).

6. Conclusions

Due to the increased frequency and intensity of heat waves, the availability of updated, reliable, and detailed data on urban air temperature is urgent. NSAT is key for studying urban heat islands and related human health, well-being, and comfort. However, a standardized and widely recognized method for systematically generating instantaneous NSAT, globally, is still lacking.

The main contribution of this research to the state-of-the-art is to outline an operational cost- and time-effective and easy-to-replicate model for estimating high spatio-temporal resolution instantaneous NSAT, suitable for regional studies, worldwide. A GWR model, with an exponential weighting scheme, is employed, and rigorous validation is provided. Although the assessment is given for an extreme heat event, it covers different thermal conditions, both day- and night-time. Nonetheless, the model should be evaluated for different seasons. Also, to improve the assessment and report on the actual predictive capacity, further Cross-Validation schemes should be tested, including 10-Fold CV or Leave-One-Out Cross-Validation (LOOCV). Additionally, while we only addressed regression methods to define the most suitable model, further investigations will focus on testing GWR algorithms with respect to machine learning methods, including Random Forest.

This research is the first step to designing an effective GIS tool, either stand-alone open software or plugin, based on Python or R scripting, for systematic estimation of instantaneous NSAT. Through this work, we demonstrate that it is possible to build a model that can fully rely on open data and open source software, which is crucial for broad replication. Besides, further improvements such as a 24-h NSAT based on the Sine-Index model for hourly T_a (Hou et al., 2013), and downscaling at a very high resolution, will be investigated and introduced. Finally, future experiments will also consider the use of Sentinel-3 thermal infrareds, at a spatial resolution of 1 km, as a complementary LST data source to increase the number of time observations, and to reduce the possible lack of information due to cloud coverage.

CRediT authorship contribution statement

Nicola Colaninno: Conceptualization, Investigation, Methodology, Data curation, Validation, Writing – original draft, Writing – review & editing. **Eugenio Morello:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Nicola Colaninno reports financial support, administrative support, article publishing charges, and equipment were provided by Polytechnic of Milan Department of Architecture and Urban Studies. Nicola Colaninno reports a relationship with Politecnico di Milan Department of Architecture and Urban Studies that includes: employment.

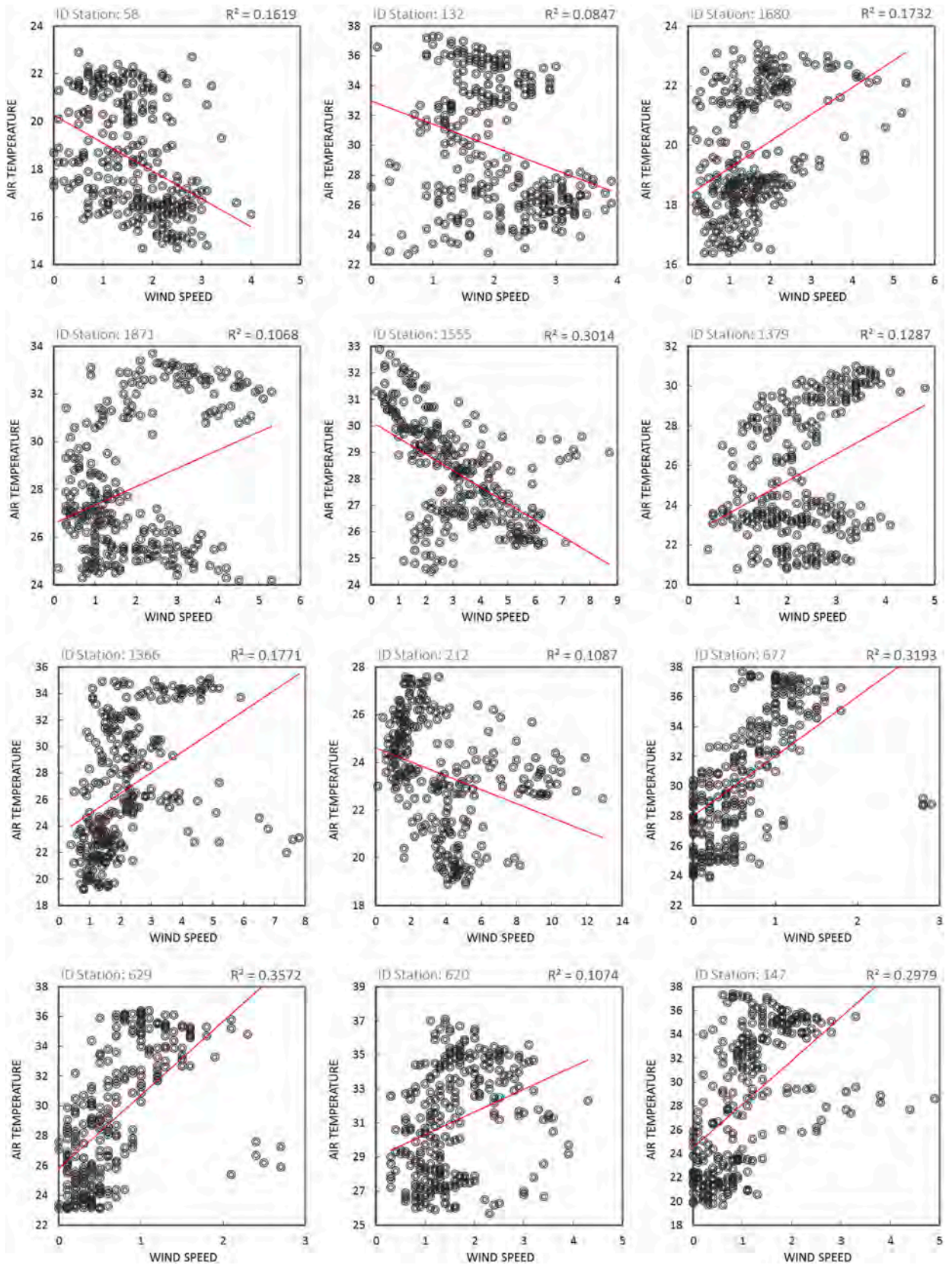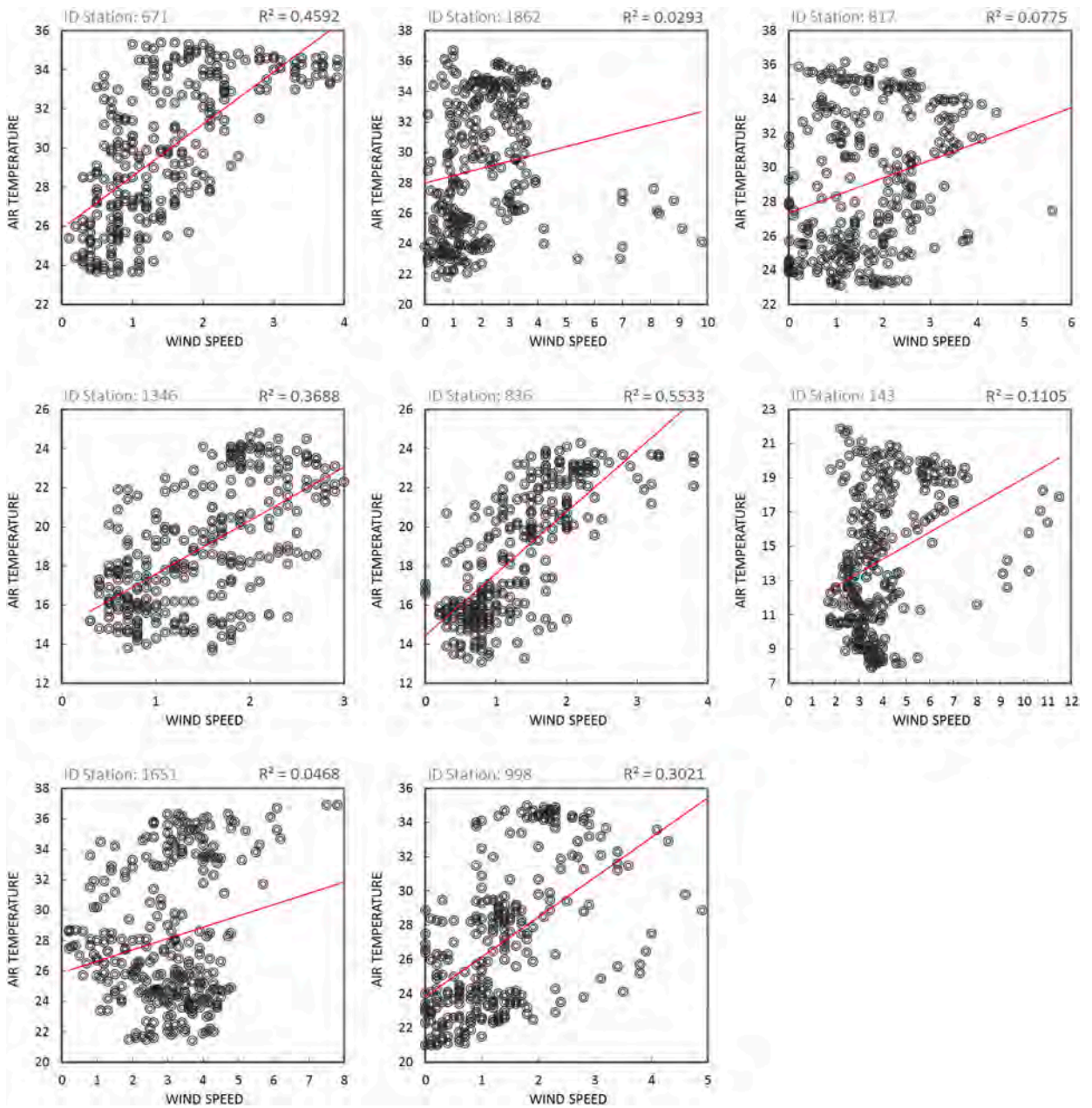## Data availability

Data will be made available on request.

Annex I

I.A. Sub-hourly spatial correlation for a selection of 20 samples weather stations during the two days, namely 31st of July and 1st of August 2020. 288 time observations are examined for each station, i.e. at each 10 min through 48 h.
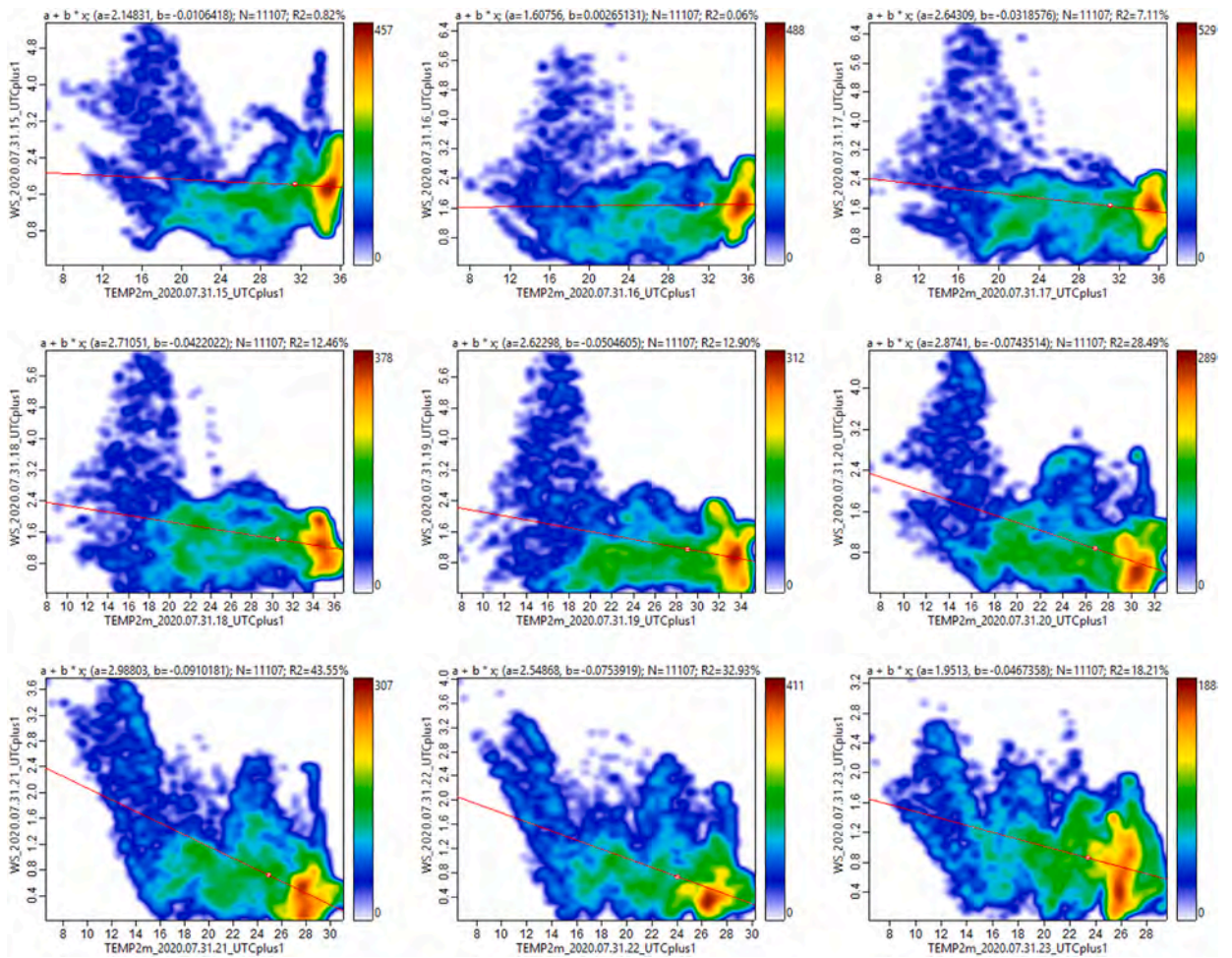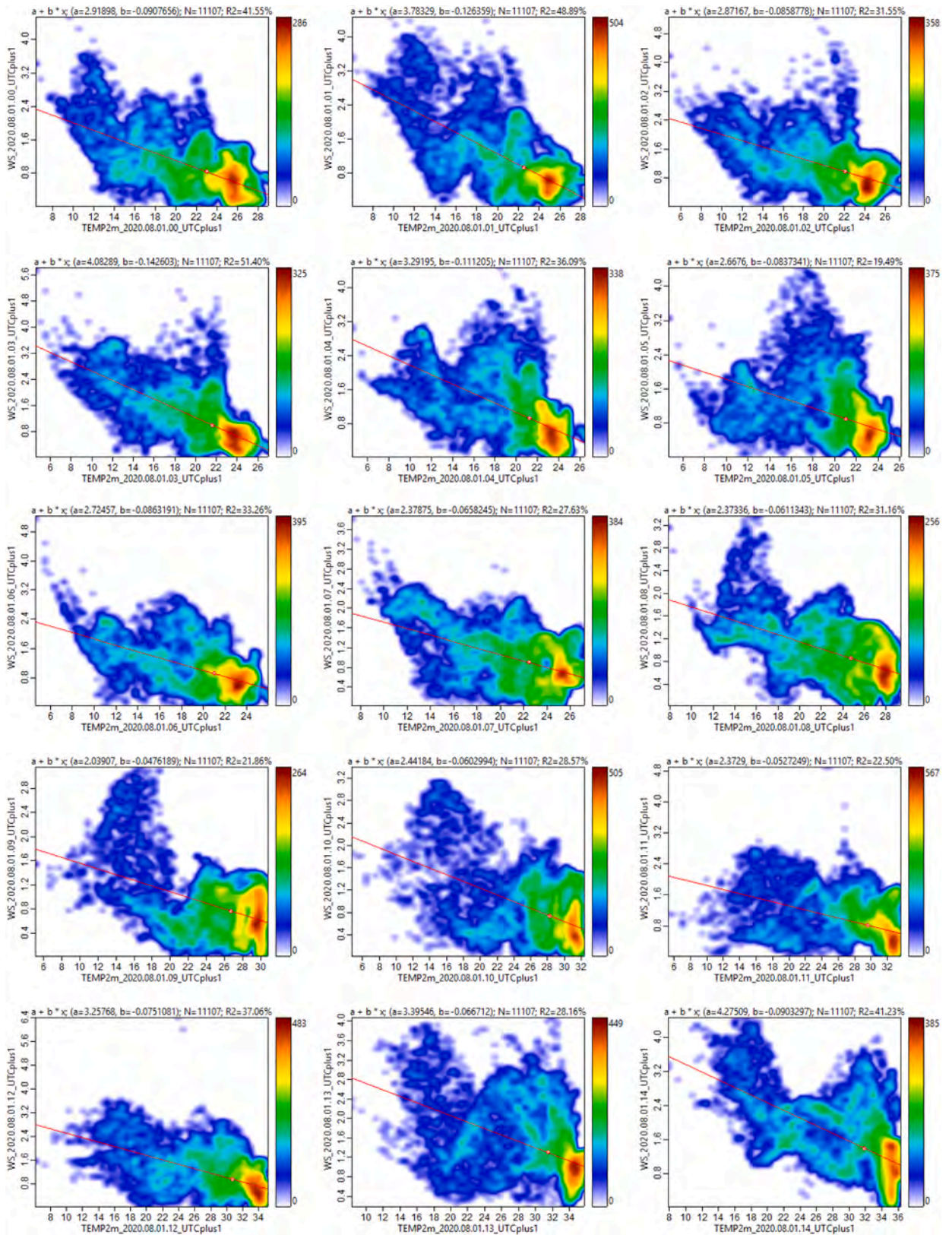
. (*continued*).

I.B. Spatial correlation among hourly T_a and Ws based on interpolations at a spatial resolution of 1.5 km. The analysis is hourly, for 48 h (two days), on 30,798 pixels.
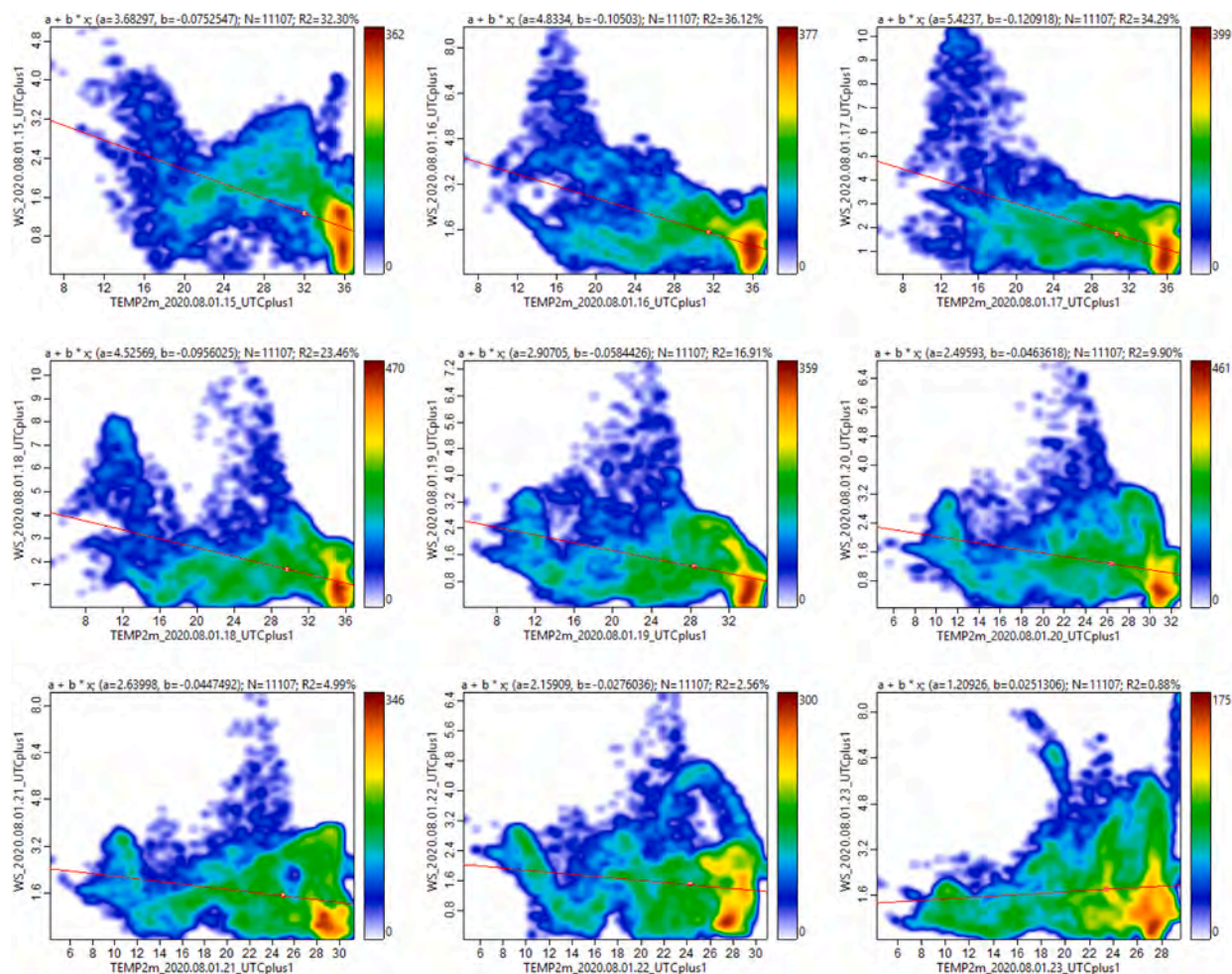
31 July 2020
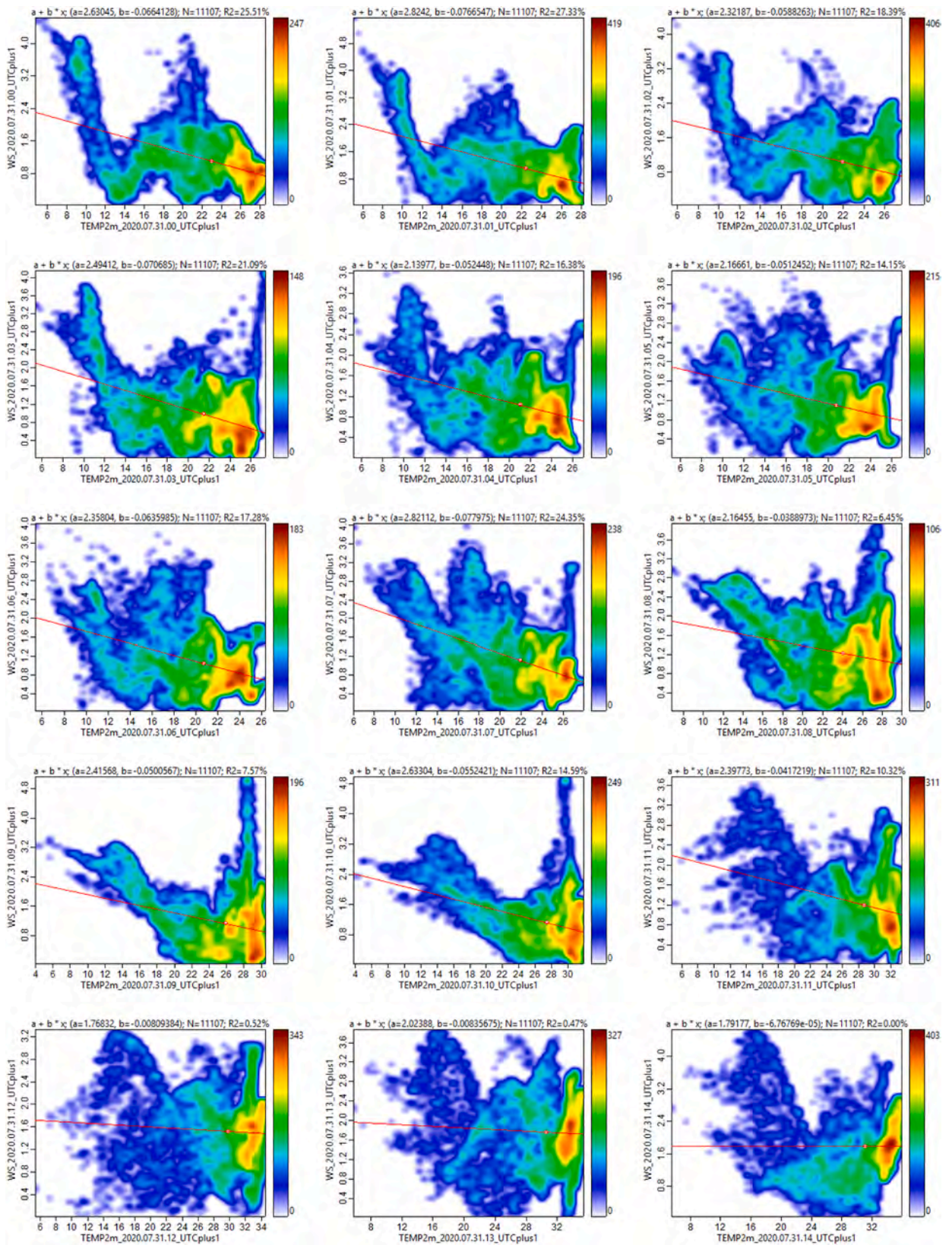
. (*continued*).

01 August 2020

. (*continued*).

References

Abbassi, Y., Ahmadikia, H., Baniasadi, E., 2022. Impact of wind speed on urban heat and pollution islands. Urban Clim. 44, 1–16. https://doi.org/10.1016/j.uclim.2022.101200.

Alvi, U., Suomi, J., Käyhkö, J., 2022. A cost-effective method for producing spatially continuous high-resolution air temperature information in urban environments. Urban Clim. 42 https://doi.org/10.1016/j.uclim.2022.101123.

Anselin, L., 2018. Spatial Weights as Distance Functions [WWW Document]. GeoDa An Introd. to Spat, Data Anal. URL https://geodacenter.github.io/workbook/4c_distance_functions/lab4c.html (accessed 2.26.19).

Anselin, L., 2019. A local Indicator of multivariate spatial association: extending Geary's c. Geogr. Anal. 51, 133–150. https://doi.org/10.1111/gean.12164.

Belda, M., Holtanová, E., Halenka, T., Kalvová, J., 2014. Climate classification revisited: from Köppen to Trewartha. Clim. Res. 59, 1–13. https://doi.org/10.3354/cr01204.

Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N., Santos, A., 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ. 124, 108–121. https://doi.org/10.1016/j.rse.2012.04.024.

Bostan, P.A., Heuvelink, G.B.M., Akyurek, S.Z., 2012. Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. Int. J. Appl. Earth Obs. Geoinf. 19, 115–126. https://doi.org/10.1016/j.jag.2012.04.010.

Chen, F., Liu, Y., Liu, Q., Qin, F., 2015. A statistical method based on remote sensing for the estimation of air temperature in China. Int. J. Climatol. 35, 2131–2143. https://doi.org/10.1002/joc.4113.

Cho, D., Yoo, C., Im, J., Lee, Y., Lee, J., 2020. Improvement of spatial interpolation accuracy of daily maximum air temperature in urban areas using a stacking ensemble technique. GISci. Remote Sens. 57, 633–649. https://doi.org/10.1080/15481603.2020.1766768.

Corona, P., Fattorini, L., Franceschi, S., Chirici, G., Maselli, F., Secondi, L., 2014. Mapping by spatial predictors exploiting remotely sensed and ground data: a comparative design-based perspective. Remote Sens. Environ. 152, 29–37. https://doi.org/10.1016/j.rse.2014.05.011.

Cristóbal, J., Ninyerola, M., Pons, X., 2008. Modeling air temperature through a combination of remote sensing and GIS data. J. Geophys. Res. 113, D13106. https://doi.org/10.1029/2007JD009318.

Czarnota, J., Wheeler, D.C., Gennings, C., 2015. Evaluating geographically weighted regression models for environmental chemical. Risk Anal. 14, 117–127. https://doi.org/10.4137/CIN.S17296.Received.

EarthData, 2020. LP DAAC - MOD11A2 [WWW Document]. NASA LP DAAC USGS EROS Cent. URL https://lpdaac.usgs.gov/products/mod11a2v006/ (accessed 9.30.21).

Elmes, A., Healy, M., Geron, N., Andrews, M.M., Rogan, J., Martin, D.G., Sangermano, F., Williams, C.A., Weil, B., 2020. Mapping spatiotemporal variability of the urban heat island across an urban gradient in Worcester, Massachusetts using in-situ thermochrons and Landsat-8 Thermal Infrared Sensor (TIRS) data. GISci. Remote Sens. 57, 845–864. https://doi.org/10.1080/15481603.2020.1818950.

Florio, E.N., Lele, S.R., Chang, Y.C., Sterner, R., Glass, G.E., 2004. Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach. Int. J. Remote Sens. 25, 2979–2994. https://doi.org/10.1080/01431160310001624593.

Fotheringham, A.S., Charlton, M.E., Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ. Plan. A 30, 1905–1927. https://doi.org/10.1068/a301905.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically weighted regression : the analysis of spatially varying relationships, p. 269.

Good, E.J., Ghent, D.J., Bulgin, C.E., Remedios, J.J., 2017. A spatiotemporal analysis of the relationship between near-surface air temperature and satellite land surface temperatures using 17 years of data from the ATSR series. J. Geophys. Res. Atmos. 122, 9185–9210. https://doi.org/10.1002/2017JD026880.

Harris, P., Fotheringham, A.S., Crespo, R., Charlton, M., 2010. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. Math. Geosci. 42, 657–680. https://doi.org/10.1007/s11004-010-9284-7.

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Comput. Geosci. 33, 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001.

Hou, P., Chen, Y., Qiao, W., Cao, G., Jiang, W., Li, J., 2013. Near-surface air temperature retrieval from satellite images and influence by wetlands in urban region. Theor. Appl. Climatol. 111, 109–118. https://doi.org/10.1007/s00704-012-0629-7.

Jato-Espino, D., Manchado, C., Roldán-Valcarce, A., Moscardó, V., 2022. ArcUHI: a GIS add-in for automated modelling of the Urban Heat Island effect through machine learning. Urban Clim. 44 https://doi.org/10.1016/j.uclim.2022.101203.

Kambezidis, H.D., 2012. The solar resource. In: Comprehensive Renewable Energy. Elsevier Ltd., pp. 27–84. https://doi.org/10.1016/B978-0-08-087872-0.00302-4

Kawashima, S., Ishida, T., Minomura, M., Miwa, T., 2000. Relations between surface temperature and air temperature on a local scale during winter nights. J. Appl. Meteorol. 39, 1570–1579. https://doi.org/10.1175/1520-0450(2000)039<1570:RBSTAA>2.0.CO;2.

Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2014. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. Remote Sens. Environ. 150, 132–139. https://doi.org/10.1016/j.rse.2014.04.024.

Kloog, I., Nordio, F., Lepeule, J., Padoan, A., Lee, M., Auffray, A., Schwartz, J., 2017. Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data. Int. J. Climatol. 37, 296–304. https://doi.org/10.1002/joc.4705.

Lai, J., Zhan, W., Quan, J., Bechtel, B., Wang, K., Zhou, J., Huang, F., Chakraborty, T., Liu, Z., Lee, X., 2021. Statistical estimation of next-day nighttime surface urban heat islands. ISPRS J. Photogramm. Remote Sens. 176, 182–195. https://doi.org/10.1016/j.isprsjprs.2021.04.009.

Leung, Y., Mei, C.L., Zhang, W.X., 2000a. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environ. Plan. A 32, 9–32. https://doi.org/10.1068/a3162.

Leung, Y., Mei, C.L., Zhang, W.X., 2000b. Testing for spatial autocorrelation among the residuals of the geographically weighted regression. Environ. Plan. A 32, 871–890. https://doi.org/10.1068/a32117.

Li, L., Zha, Y., 2019. Estimating monthly average temperature by remote sensing in China. Adv. Sp. Res. 63, 2345–2357. https://doi.org/10.1016/j.asr.2018.12.039.

Lloyd, C.D., 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. J. Hydrol. 308, 128–150. https://doi.org/10.1016/j.jhydrol.2004.10.026.

Lorenzi, L., Melgani, F., Mercier, G., 2011. Inpainting strategies for reconstruction of missing data in VHR images. IEEE Geosci. Remote Sens. Lett. 8, 914–918. https://doi.org/10.1109/LGRS.2011.2141112.

Lu, B., Charlton, M., Harris, P., Fotheringham, A.S., 2014. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. Int. J. Geogr. Inf. Sci. 28, 660–681. https://doi.org/10.1080/13658816.2013.865739.

Lussana, C., Salvati, M., Pellegrini, U., Uboldi, F., 2009. Efficient high-resolution 3-D interpolation of meteorological variables for operational use. Adv. Sci. Res. 105–112.

Malbéteau, Y., Merlin, O., Gascoin, S., Gastellu, J.P., Mattar, C., Olivera-Guerra, L., Khabba, S., Jarlan, L., 2017. Normalizing land surface temperature data for elevation and illumination effects in mountainous areas: a case study using ASTER data over a steep-sided valley in Morocco. Remote Sens. Environ. 189, 25–39. https://doi.org/10.1016/j.rse.2016.11.010.

Metz, M., Andreo, V., Neteler, M., 2017. A new fully gap-free time series of land surface temperature from MODIS LST data. Remote Sens. 9, 1333. https://doi.org/10.3390/rs9121333.

Mutiibwa, D., Strachan, S., Albright, T., 2015. Land surface temperature and surface air temperature in complex terrain. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 4762–4774. https://doi.org/10.1109/JSTARS.2015.2468594.

Nichol, J., 2005. Remote sensing of urban heat islands by day and night. Photogramm. Eng. Remote. Sens. 71, 613–621. https://doi.org/10.14358/PERS.71.5.613.

Ninyerola, M., Pons, X., Roure, J.M., 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. Int. J. Climatol. 20, 1823–1841. https://doi.org/10.1002/1097-0088(20001130)20:14<1823::AID-JOC566>3.0.CO;2-B.

Oke, T.R., 1981. Canyon geometry and the nocturnal urban heat island: comparison of scale model and field observations. J. Climatol. 1, 237–254. https://doi.org/10.1002/JOC.3370010304.

Ozelkan, E., Bagis, S., Ozelkan, E.C., Ustundag, B.B., Yucel, M., Ormeci, C., 2015. Spatial interpolation of climatic variables using land surface temperature and modified inverse distance weighting. Int. J. Remote Sens. 36, 1000–1025. https://doi.org/10.1080/01431161.2015.1007248.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.

Pepin, N.C., Maeda, E.E., Williams, R., 2016. Use of remotely sensed land surface temperature as a proxy for air temperatures at high elevations: findings from a 5000m elevational transect across Kilimanjaro. J. Geophys. Res. Atmos. 121, 9998–10015. https://doi.org/10.1038/175238c0.

Propastin, P., Kappas, M., Erasmi, S., 2008. Application of geographically weighted regression to investigate the impact of scale on prediction uncertainty by modelling relationship between vegetation and climate *. Int. J. Spat. Data Infrastruct. Res. 3, 73–94.

Rao, Y., Liang, S., Wang, D., Yu, Y., Song, Z., Zhou, Y., Shen, M., Xu, B., 2019. Estimating daily average surface air temperature using satellite land surface temperature and top-of-atmosphere radiation products over the Tibetan plateau. Remote Sens. Environ. 234, 111462 https://doi.org/10.1016/j.rse.2019.111462.

Rasul, A., Balzter, H., Smith, C., Remedios, J., Adamu, B., Sobrino, J., Srivanit, M., Weng, Q., 2017. A review on remote sensing of urban heat and cool islands. Land 6, 38. https://doi.org/10.3390/land6020038.

Roth, M., Oke, T.R., Emery, W.J., 1989. Satellite-derived urban heat islands from 3 coastal cities and the utilization of such data in urban climatology. Int. J. Remote Sens. 10, 1699–1720.

Samanta, Sailesh, Pal, D.K., Lohar, D., Pal, B., 2011a. Preparation of digital data sets on land use/land cover, soil and digital elevation model for temperature modelling using remote sensing and GIS techniques. Indian J. Sci. Technol. 4, 636–642.

Samanta, S., Pal, D.K., Lohar, D., Pal, B., 2011b. Modeling of temperature and rainfall of West Bengal through remote sensing and GIS techniques. Int. J. Geoinform. 7, 31–40.

Schwarz, N., Schlink, U., Franck, U., Großmann, K., 2012. Relationship of land surface and air temperatures and its implications for quantifying urban heat island indicators - an application for the city of Leipzig (Germany). Ecol. Indic. 18, 693–704. https://doi.org/10.1016/j.ecolind.2012.01.001.

Shen, H., Jiang, Y., Li, T., Cheng, Q., Zeng, C., Zhang, L., 2020. Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data. Remote Sens. Environ. 240, 111692 https://doi.org/10.1016/j.rse.2020.111692.

Souch, C., Grimmond, S., 2006. Applied climatology: urban climate. Prog. Phys. Geogr. 30, 270–279. https://doi.org/10.1191/0309133306pp484pr.

Su, Y.F., Foody, G.M., Cheng, K.S., 2012. Spatial non-stationarity in the relationships between land cover and surface temperature in an urban heat island and its impacts on thermally sensitive populations. Landsc. Urban Plan. 107, 172–180. https://doi.org/10.1016/j.landurbplan.2012.05.016.

Sun, T., Sun, R., Chen, L., 2020. The trend inconsistency between land surface temperature and near surface air temperature in assessing urban heat island effects. Remote Sens. 12, 1271. https://doi.org/10.3390/rs12081271.

Szymanowski, M., Kryza, M., 2009. GIS-based techniques for urban heat island spatialization. Clim. Res. 38, 171–187. https://doi.org/10.3354/cr00780.

Szymanowski, M., Kryza, M., 2011. Application of geographically weighted regression for modelling the spatial structure of urban heat island in the city of Wroclaw (SW Poland). Procedia Environ. Sci. 3, 87–92. https://doi.org/10.1016/j.proenv.2011.02.016.

Szymanowski, M., Kryza, M., Spallek, W., 2013. Regression-based air temperature spatial prediction models: an example from Poland. Meteorol. Z. 22, 577–585. https://doi.org/10.1127/0941-2948/2013/0440.

Terrádez Mas, J., Fraschini, F., Rossetto, M., 2016. Rapporto di sintesi della Strategia Regionale di adattamento ai cambiamenti climatici.

Touati, N., Gardes, T., Hidalgo, J., 2020. A GIS plugin to model the near surface air temperature from urban meteorological networks. Urban Clim. 34, 100692 https://doi.org/10.1016/j.uclim.2020.100692.

Uboldi, F., Lussana, C., Salvati, M., 2008. Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks. Meteorol. Appl. 15, 331–345. https://doi.org/10.1002/met.76.

Voogt, J., 2014. How Researchers Measure Urban Heat Islands.

Wang, Xiaofei, Wang, Xiaoyi, 2020. Spatiotemporal fusion of remote sensing image based on deep learning. J. Sensors 2020. https://doi.org/10.1155/2020/8873079.

Wang, M., He, G., Zhang, Zhaoming, Wang, G., Zhang, Zhengjia, Cao, X., Wu, Z., Liu, X., 2017. Comparison of spatial interpolation and regression analysis models for an estimation of monthly near surface air temperature in China. Remote Sens. 9 https://doi.org/10.3390/rs9121278.

Wang, Y., Zhang, Z., Feng, L., Du, Q., Runge, T., 2020. Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous United States. Remote Sens. 12 https://doi.org/10.3390/RS12081232.

Wu, M., Niu, Z., Wang, C., Wu, C., Wang, L., 2012. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. J. Appl. Remote. Sens. 6, 063507 https://doi.org/10.1117/1.JRS.6.063507.

Xing, Z., Li, Z.L., Duan, S.B., Liu, X., Zheng, X., Leng, P., Gao, M., Zhang, X., Shang, G., 2021. Estimation of daily mean land surface temperature at global scale using pairs of daytime and nighttime MODIS instantaneous observations. ISPRS J. Photogramm. Remote Sens. 178, 51–67. https://doi.org/10.1016/j.isprsjprs.2021.05.017.

Xu, Y., Qin, Z., Shen, Y., 2012. Study on the estimation of near-surface air temperature from MODIS data by statistical methods. Int. J. Remote Sens. 33, 7629–7643. https://doi.org/10.1080/01431161.2012.701351.

Xu, Y., Knudby, A., Ho, H.C., 2014. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. Int. J. Remote Sens. 35, 8108–8121. https://doi.org/10.1080/01431161.2014.978957.

Xue, Y., Fung, T., Tsou, J., 2012. An investigation of local effects on surface warming with geographically weighted regression (Gwr). Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XXXIX-B8, 131–136.

Yan, H., Zhang, J., Hou, Y., He, Y., 2009. Estimation of air temperature from MODIS data in east China. Int. J. Remote Sens. 30, 6261–6275. https://doi.org/10.1080/01431160902842375.

Yang, J.S., Wang, Y.Q., August, P.V., 2004. Estimation of land surface temperature using spatial interpolation and satellite-derived surface emissivity. J. Environ. Inf. 4, 40–47. https://doi.org/10.3808/jei.200400035.

Yoo, C., Im, J., Park, S., Quackenbush, L.J., 2018. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. ISPRS J. Photogramm. Remote Sens. 137, 149–162. https://doi.org/10.1016/j.isprsjprs.2018.01.018.

Zhang, Z., Du, Q., 2022. Hourly mapping of surface air temperature by blending geostationary datasets from the two-satellite system of GOES-R series. ISPRS J. Photogramm. Remote Sens. 183, 111–128. https://doi.org/10.1016/j.isprsjprs.2021.10.022.

Zhang, L., Gove, J.H., 2005. Spatial assessment of model errors from four regression techniques. For. Sci. 51, 334–346. https://doi.org/10.1093/forestscience/51.4.334.

Zhou, B., Erell, E., Hough, I., Shtein, A., Just, A.C., Novack, V., Rosenblatt, J., Kloog, I., 2020. Estimation of hourly near surface air temperature across Israel using an ensemble model. Remote Sens. 12 https://doi.org/10.3390/rs12111741.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114, 2610–2623. https://doi.org/10.1016/j.rse.2010.05.032.

Zhu, W., Lu, A., Jia, S., 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sens. Environ. 130, 62–73. https://doi.org/10.1016/j.rse.2012.10.034.