

# Estimating local-scale urban heat island intensity using nighttime light satellite imageries

Yeran Sun <sup>a,b</sup>, Shaohua Wang <sup>c,\*</sup>, Yu Wang <sup>d</sup>

<sup>a</sup> School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

<sup>b</sup> Department of Geography, College of Science, Swansea University, Swansea SA28PP, United Kingdom

<sup>c</sup> CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

<sup>d</sup> Department of Urban and Rural Planning, School of Architecture, Tianjin University, Tianjin 300072, China

**Abstract:** Urban heat island (UHI) effect tends to harm health, increase anthropogenic energy consumption, and water consumption. Some policies targeting UHI mitigation have been implemented for a few years and thus needs to be evaluated for changes or modifications in the future. A low-cost approach to rapidly monitoring UHI intensity variations can assist in evaluating policy implementations. In this study, we proposed a new approach to local-scale UHI intensity estimates by using nighttime light satellite imageries. We explored to what extent UHI intensity could be estimated according to nighttime light intensity at two local scales. We attempted to estimate district-level and neighbourhood-level UHI intensity across London and Paris. As the geography level rises from district to neighbourhood, the capacity of the models explaining the variations of the UHI intensity decreases. Although the possible presence of residual spatial autocorrelation in the conventional regression models applied to geospatial data, most of the studies are likely to neglect this issue when fitting data to models. To remove negative effects of the residual spatial autocorrelation, this study used spatial regression models instead of non-spatial regression models (e.g., OLS models) to estimate UHI intensity. As a result, district-level UHI intensity was successfully estimated according to nighttime light intensity (approximately  $R^2 = 0.7$ , MAE = 1.16 °C, and RMSE = 1.74 °C).

**Keywords:** urban heat island effect; nighttime light intensity; nighttime light imagery; SNPP-VIIRS; spatial regression model

## 1. Introduction

The urban heat island (UHI) effect refers to the difference in temperature between an urban area and the rural surroundings of a conurbation (Azevedo et al., 2016). The higher temperatures experienced in urban areas compared to the surrounding countryside has enormous consequences for the health and wellbeing of people living in cities (Mohajerani et al., 2017). UHI tends to harm health (Patz et al., 2005; McMichael et al., 2006), increase anthropogenic energy consumption (Rosenfeld et al., 1998), and water consumption (Guhathakurta et al., 2010). Infrared satellite imageries or airborne photos have been used to estimate UHI intensity since they can directly derive the temperature of the surface or the temperature of the atmosphere (Couatts et al., 2016; Chui et al., 2018; Wang et al., 2019; Fabbri and Costanzo, 2020). Apart from some direct measurements like infrared pictures, indirect measurements like visible light pictures are of high potential as well. UHI is contributed by built environment characteristics and heat emissions of economic activity in urban areas. Built environment characteristics, such as impervious surface area density, building area density, building height, and road density, have impacts on both surface temperature and air temperature since impervious surface materials of buildings and roads can influence surface albedo, emissivity, and evapotranspiration (Oke 1987; Pigeon et al., 2007). Constructed impervious surfaces alter sensible and latent heat fluxes, causing urban heat islands (Changnon, 1992). Constructed surfaces are more likely to increase temperature than natural surfaces (Changnon, 1992); whilst urban green spaces can reduce UHI effects (Oke 1987; Debbage & Shepherd, 2015). In the urban areas crossed by the major roads, the traffic is the major source (Pigeon et al., 2007). Aside from vehicle heat emissions, asphalt concrete on road surface is a contributor to the UHI (Mohajerani et al., 2017). Heat emissions are generated by economic activities including domestic, commercial & industrial, and transportation activities. Urban areas have more heat emissions than rural areas due to more economic activities. Urban areas have more domestic activities as urban areas are more densely populated. Urban areas also have more commercial & industrial and transportation activities since urban areas are commercial & industrial centres and have larger traffic volumes. A better understanding of how UHI is contributed by built environment characteristics and economic activity at the local levels can help to inform policy and planning decisions for UHI reduction. Particularly, UHI is usually strongest at night (Azevedo et al., 2016); for example, a study revealed that in Paris the magnitude of the night-time UHI was up to 7 °C more than the daytime UHI (Lac et al., 2013). Compared to nighttime UHI, daytime UHI is less noticeable and far more complicated to characterise (Azevedo et al., 2016). Recent studies demonstrate that nighttime light satellite imagery data may be a good proxy for night-time population (Anderson et al., 2010; Sutton, 1997; Lo, 2001; Pozzi et al., 2003; Zhuo et al., 2009; Liu et al., 2011; Bagan & Yamagata, 2015) or

economic activity (Mellander et al., 2015; Ghosh et al., 2010; Zhang & Seto, 2011). Nighttime light intensity (NTLI) is positively related to constructed impervious surface area (ISA) density and energy consumption due to economic activities at night (Amaral et al., 2005; Shi et al., 2014; Townsend & Bruce, 2010). As both constructed impervious surface area (ISA) density and nighttime consumption make contributions to UHI, NTLI is likely to be positively related to UHI intensity at night.

Therefore, this study attempts to estimate the UHI intensity according to nighttime light intensity (NTLI) offered by open nighttime light satellite imageries. A less costly and time-consuming approach to estimating UHI intensity over space and time is urgently needed for not only researchers but also policy-makers. How to mitigate UHI is vital for improving health and wellbeing of people living in cities. On the one hand, policy-makers are investing more and more on building material research to lower impervious surface area density; and on the other hand, they encourage urban planners to build more green spaces to reduce UHI. Some policies targeting UHI mitigation have been implemented for a few years and thus needs to be evaluated for changes or modifications in the future. A low-cost approach to rapidly monitoring UHI variations can assist in evaluating policy implementations. In this study, we propose a new approach to local-scale UHI intensity estimates by using nighttime light satellite imageries. We choose London as the case study since London is a representative city in the existing UHI studies (e.g., Giridharan and Kolokotroni, 2009; Doick et al., 2014; dos Santos, R.S., 2020).

To the best of our knowledge, this paper is the first study devoted to local-scale UHI intensity estimates using nighttime light intensity directly as the explanatory variable. As satellite overpass time of nighttime light imageries is at night and nighttime UHI is more noticeable than daytime UHI, we selected the annual average nighttime surface UHI intensity to represent the annual average UHI level across London. The year 2015 is selected due to the data availability. Besides, we further choose Paris as a comparable city with London in this study. Owing to similar climatic and meteorological characteristics, economic development levels, energy-related policies, the associations of nighttime light intensity and UHI intensity might be fairly similar between London and Paris. Besides, since Paris is likely more polluted than London (Font et al., 2019), the potential influence of air pollution or sky turbidity conditions on UHI intensity estimates can be discussed in this study.

We attempted to establish models to estimate annual average UHI intensity (urban-rural surface temperature difference) at both the district and neighbourhood levels. We explored to what extent UHI intensity could be estimated according to nighttime light intensity at two geography levels. Moreover, we attempt to replace conventional regression models (non-

spatial regression models) with spatial regression models in this study. Although the possible presence of residual spatial autocorrelation in the conventional regression models applied to geospatial data, most of the studies are likely to neglect this issue when fitting data to models. Therefore, in this study, we attempted to use spatial regression models to estimate UHI intensity according to NTLI if spatial autocorrelation existed in the residuals of non-spatial models estimated. The matrix exponential spatial specification (MESS) models which has analytical, computational, and interpretive advantages over conventional spatial autoregressive models were used in this study. Besides, this study can also pave a new way for estimating UHI intensity in some cities where accurate UHI data is missing.

## **2. Materials and methods**

### *2.1. Research data*

**UHI data:** We used the gridded surface UHI data simulated based on MODIS images (<https://yceo.users.earthengine.app/view/uhimap>). A simplified urban-extent (SUE) algorithm is implemented on the Google Earth Engine platform using MODIS images to calculate the UHI intensity for over 9500 urban clusters using over 15 years of data, making this one of the most comprehensive characterizations of the surface UHI to date (Chakraborty and Lee, 2019). The results from this algorithm have been validated against previous multi-city studies to demonstrate the suitability of the method (Chakraborty and Lee, 2019; Chakraborty et al., 2019). UHI intensity in the dataset refers to the difference in land surface temperature (LST) between the urban area and its surrounding non-urban area. We choose the UHI data across London and Paris in 2015 in this study. Figure 1 shows annual average urban heat island UHI intensity (urban–rural surface temperature difference) across London and Paris in 2015. In addition, city-level annual nighttime UHI intensity is close to annual daytime UHI intensity in London and Paris in 2015 (see <https://yceo.users.earthengine.app/view/uhimap>).

Observed temperature data at sites were downloaded from Met Office (2006; 2018) to validate the model for London in this study. We choose the 6 sites within Greater London as urban sites and 2 sites nearby as rural sites (see Figure 2). The realistic UHI intensity at a site is equal to the difference of temperature at the site and average rural temperature. The average temperature at the two rural sites is used to represent average rural temperature in this study. As the satellite overpass time of SNPP-VIIRS data is at midnight, we choose annual average daily minimum temperature in 2015 to calculate the urban-rural annual average temperature representing observed UHI intensity at the urban sites.

**Nighttime light imagery data:** There are two popular nighttime light satellite data sources: DMSP/OLS and SNPP-VIIRS. As a new generation of nighttime light satellite data, Suomi National Polar-Orbiting Partnership - Visible and Infrared Imager/Radiometer Suite (SNPP-

VIIRS) nighttime light data has been released by National Oceanic and Atmospheric Administration's National Geophysical Data Center (NOAA/NGDC) since 2013. Although SNPP-VIIRS data is superior to DMSP/OLS data, monthly SNPP-VIIRS data is only available from 2012 and annual SNPP-VIIRS data is only available from 2015 onwards ([https://ngdc.noaa.gov/eog/viirs/download\\_dnb\\_composites.html](https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html)). In this study, as we aim to make use of nighttime light satellite data to measure annual nighttime light intensity in 2015, we chose annual SNPP-VIIRS data in 2015. Besides, satellite overpass time of DMSP/OLS data is between 19:30 and 21:30 whilst that of SNPP-VIIRS data is around 01:30. We can download satellite nightlight images from the U.S. National Oceanic and Atmospheric Administration (NOAA) through an *R* package called 'Rnightlights' (Njuguna, 2018). Both Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS) and SNPP-VIIRS datasets are supported (NOAA, 2018). We download the 2015 SNPP-VIIRS annual NTL intensity data across Europe. The raster data is of high spatial resolution (15 arc-second  $\times$  15 arc-second). The average area of grid varies from 0.1 to 0.2 km<sup>2</sup>. Figure 3 shows the annual average nighttime light intensity across London and Paris in 2015. Annual average light intensity is represented by annual average radiance (unit: nW cm<sup>-2</sup> sr<sup>-1</sup>).

Administrative boundaries: In this study, we choose local authority district (LAD) and middle layer super output area (MSOA) to represent two local levels (district and neighbourhood) in London; and we chose arrondissement (ADM) and commune (CM) to represent two local levels (district and neighbourhood) in Paris. A LAD or ADM is simply a district in UK or France respectively. Greater London consists of 33 LADs, including City of London and 32 London boroughs. Grand Paris consists of 30 ADMs, including 20 ones within and 10 ones surrounding City of Paris. London is composed of 983 MSOAs; whilst Paris is composed of 150 CMs. Figure 4 shows the LAD and MSOA boundaries across London in 2011. And Figure 5 shows the ADM and CM boundaries across Paris in 2011.

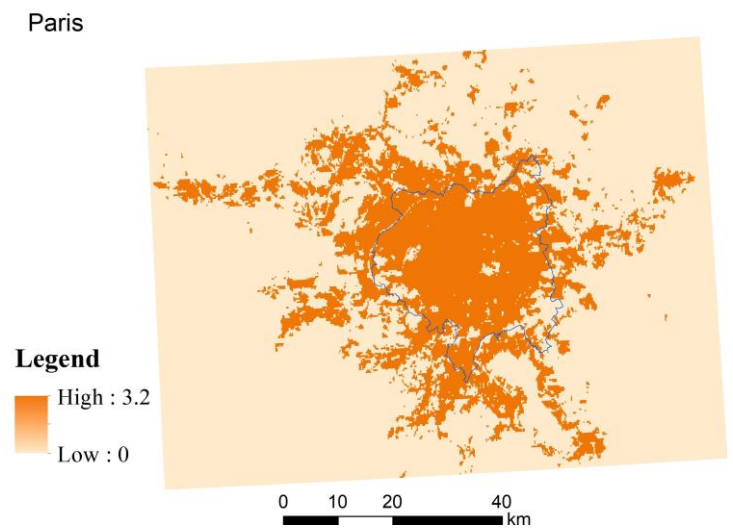
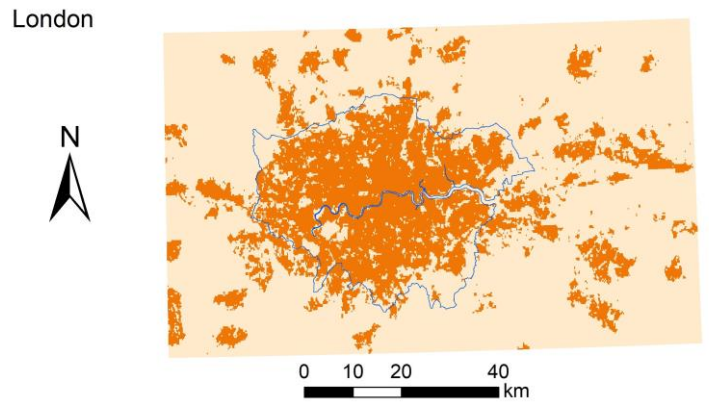


Figure 1: Annual average nighttime urban heat island (UHI) intensity (urban–rural surface temperature difference) across London and Paris (Unit: °C), 2015

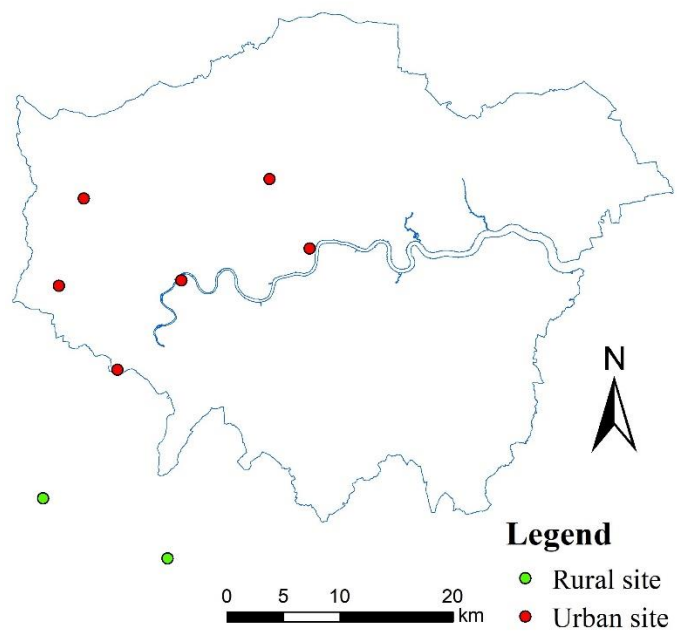


Figure 2: Locations of weather stations across Greater London

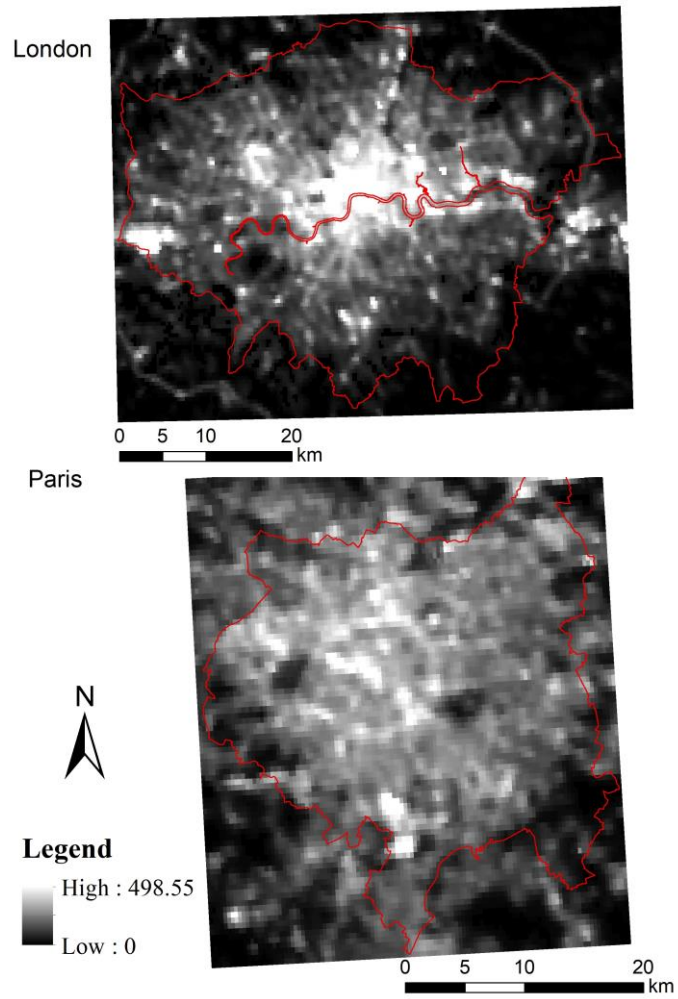


Figure 3: Annual average nighttime light intensity across London and Paris, 2015 (unit:  $\text{nW cm}^{-2} \text{sr}^{-1}$ )

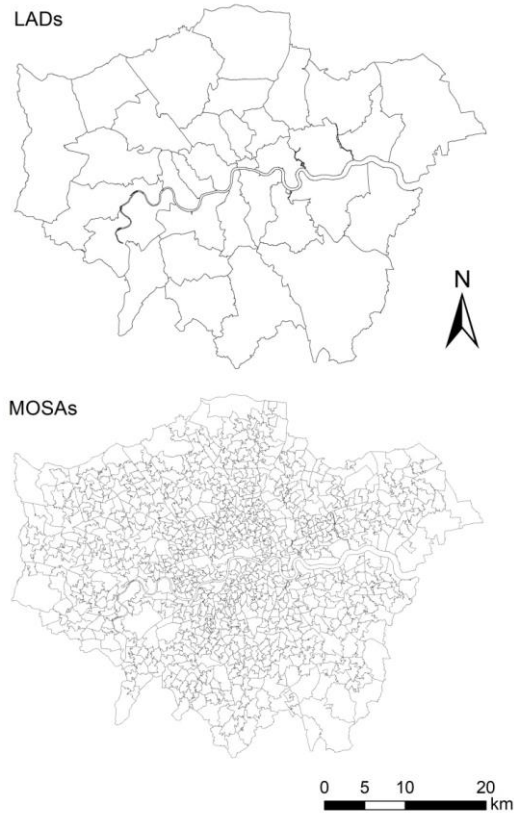


Figure 4: LAD and MSOA boundaries across Greater London

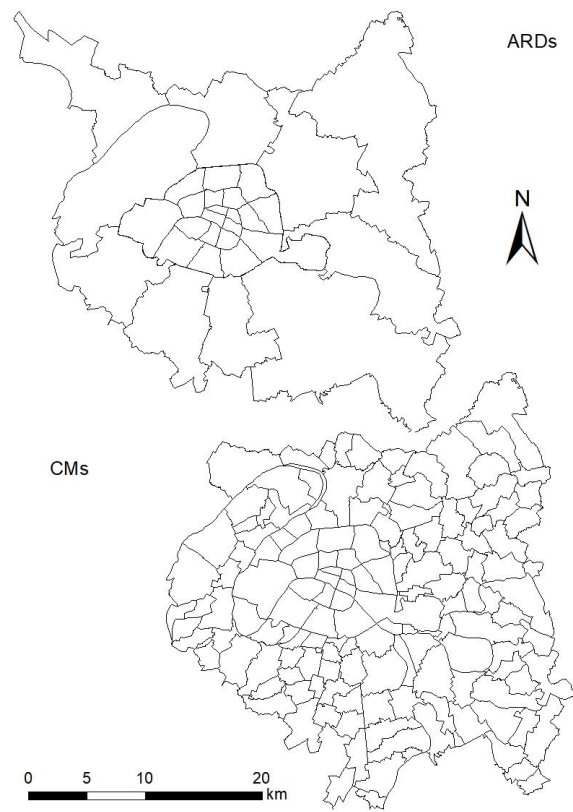


Figure 5: ADM and CM boundaries across Grand Paris



## 2.2. Explanatory and response variables

The explanatory variable is annual average nighttime light intensity (AANTLI); whilst the response variable is annual average urban heat island intensity (AAUHII). AAUHII and AANTLI are both measured at the LAD, MSOA, ADM, and CM levels.

LAD-level AANTLI is represented by the area-weighted mean level of AANTLI in a LAD. Supposing that  $i$  is a LAD, we calculated its AANTLI as:

$$LAD\_AANTLI(i) = \sum_{j \in S(i)} AANTLI(j) * \frac{area(i,j)}{LAD\_area(i)} \quad (1)$$

where  $AANTLI(j)$  represents the AANTLI of the grid  $j$ .  $Area(i,j)$  represents the area of the overlapping part of grid  $j$  and LAD  $i$ ;  $S(i)$  is the set of grids overlapping with LAD  $i$ .

LAD-level AAUHII is represented by the area-weighted mean level of AAUHII in a LAD. Supposing that  $i$  is a LAD, we calculated its AAUHII as:

$$LAD\_AAUHII(i) = \sum_{j \in S(i)} AAUHII(j) * \frac{area(i,j)}{LAD\_area(i)} \quad (2)$$

where  $AAUHII(j)$  represents the AAUHII of the grid  $j$ .  $Area(i,j)$  represents the area of the overlapping part of grid  $j$  and LAD  $i$ ;  $S(i)$  is the set of grids overlapping with LAD  $i$ .

Likewise, MSOA-level AANTLI is represented by the area-weighted mean level of AANTLI in a MSOA. Supposing that  $i$  is a MSOA, we calculated its AANTLI as:

$$MSOA\_AANTLI(i) = \sum_{j \in S(i)} AANTLI(j) * \frac{area(i,j)}{MSOA\_area(i)} \quad (3)$$

where  $AANTLI(j)$  represents the AANTLI of the grid  $j$ .  $Area(i,j)$  represents the area of the overlapping part of grid  $j$  and MSOA  $i$ ;  $S(i)$  is the set of grids overlapping with MSOA  $i$ .

MSOA-level AAUHII is represented by the area-weighted mean level of AAUHII in a MSOA. Supposing that  $i$  is a MSOA, we calculated its AAUHII as:

$$MSOA\_AAUHII(i) = \sum_{j \in S(i)} AAUHII(j) * \frac{area(i,j)}{MSOA\_area(i)} \quad (4)$$

where  $AAUHII(j)$  represents the AAUHII of the grid  $j$ .  $Area(i,j)$  represents the area of the overlapping part of grid  $j$  and MSOA  $i$ ;  $S(i)$  is the set of grids overlapping with MSOA  $i$ .

Likewise, ADM-level AANTLI is represented by the area-weighted mean level of AANTLI in a ADM. Supposing that  $i$  is a ADM, we calculated its AANTLI as:

$$ADM\_AANTLI(i) = \sum_{j \in S(i)} AANTLI(j) * \frac{area(i, j)}{ADM\_area(i)} \quad (5)$$

where  $AANTLI(j)$  represents the AANTLI of the grid  $j$ .  $Area(i, j)$  represents the area of the overlapping part of grid  $j$  and ADM  $i$ ;  $S(i)$  is the set of grids overlapping with ADM  $i$ .

ADM-level AAUHII is represented by the area-weighted mean level of AAUHII in a ADM. Supposing that  $i$  is a ADM, we calculated its AAUHII as:

$$ADM\_AAUHII(i) = \sum_{j \in S(i)} AAUHII(j) * \frac{area(i, j)}{ADM\_area(i)} \quad (6)$$

where  $AAUHII(j)$  represents the AAUHII of the grid  $j$ .  $Area(i, j)$  represents the area of the overlapping part of grid  $j$  and ADM  $i$ ;  $S(i)$  is the set of grids overlapping with ADM  $i$ .

Likewise, CM-level AANTLI is represented by the area-weighted mean level of AANTLI in a CM. Supposing that  $i$  is a CM, we calculated its NTLI as:

$$CM\_AANTLI(i) = \sum_{j \in S(i)} AANTLI(j) * \frac{area(i, j)}{CM\_area(i)} \quad (7)$$

where  $AANTLI(j)$  represents the AANTLI of the grid  $j$ .  $Area(i, j)$  represents the area of the overlapping part of grid  $j$  and CM  $i$ ;  $S(i)$  is the set of grids overlapping with CM  $i$ .

CM-level AAUHII is represented by the area-weighted mean level of AAUHII in a CM. Supposing that  $i$  is a CM, we calculated its AAUHII as:

$$CM\_AAUHII(i) = \sum_{j \in S(i)} AAUHII(j) * \frac{area(i, j)}{CM\_area(i)} \quad (8)$$

where  $AAUHII(j)$  represents the AAUHII of the grid  $j$ .  $Area(i, j)$  represents the area of the overlapping part of grid  $j$  and CM  $i$ ;  $S(i)$  is the set of grids overlapping with CM  $i$ .

### 2.3. Moran's I: Testing for spatial dependence

Moran (1950) first proposed a test statistic to quantify the level of spatial autocorrelation between adjacent locations (Moran, 1950). Specifically, Moran's (1950) test statistic is defined as (Getis and Ord, 1992)

$$I = \frac{n}{W} \frac{\sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum w_{ij} (x_i - \bar{x})^2} \quad (9)$$

where  $n$  is the number of spatial units (points or areas),  $x_i$  is the value of the variable  $x$  at the location  $i$ ;  $w_{ij}$  is the spatial weight of the location  $i$  and  $j$  (simply,  $w_{ij} = 1$  if  $i$  and  $j$  are near or adjacent;  $w_{ij} = 0$  otherwise);  $W$  is the sum of all  $w_{ij}$ .

#### *2.4. Spatial regression model*

Conventional regression models such as OLS models sometimes cannot work well for geospatial data due to the possible presence of residual spatial autocorrelation in the models established. Presence of residual spatial autocorrelation undermines the assumption of regression model that all the observations should be independent of each other. Compared to conventionally nonspatial regression models that might have auto-correlated residuals, spatial regression models can remove the impact of residual autocorrelation. We should replace conventional regression models (non-spatial regression models) with spatial regression models if the residuals of the conventional regression models are spatially autocorrelated. Conventionally non-spatial models like OLS models assume that regression residuals should be random and independent with each other. However, sometimes spatial autocorrelation exists in the regression residuals when conventionally non-spatial models are applied to geospatial data. If one residual can be predicted by its neighbouring residuals, some predictive information is not captured by the model but it is included in the residuals. In other words, the models are not well established to explain the contributions of explanatory variables to the response variable. A well-established model should have residuals which are random distributed rather than are correlated with each other. As the applications of conventionally non-spatial models to geospatial data are likely to cause non-randomly distributed residuals, spatial regression models are proposed to address this issue. As the most popular form of spatial regression models, spatial autoregressive models are widely used to fit linear models with autoregressive errors and spatial lags of the dependent and independent variables. There are some specifications of spatial autoregressive models with different ways of generating weighting matrices, such as inverse distance or nearest neighbour, or creating custom matrices.

In this study, we employed the matrix exponential spatial specification (MESS) models, which are developed to model spatially dependent data. The MESS model can produce estimates and inferences similar to those from conventional spatial autoregressive models, but has analytical, computational, and interpretive advantages (LeSage & Pace, 2007). Compared to conventional spatial autoregressive (SAR) models, the MESS model advocates the use of a matrix exponential spatial specification (MESS) of dependence that replaces the conventional

geometric decay of influence over space with an exponential pattern of decay (LeSage & Pace, 2009). The matrix exponential spatial specification (MESS) simplifies the log-likelihood allowing a closed form solution to the problem of maximum-likelihood estimation, and greatly simplifies the Bayesian estimation of the model (LeSage & Pace, 2009). This specification has theoretical as well as computational advantages over the SAR specification due to the ease of inversion, differentiation, and integration of the matrix exponential (LeSage & Pace, 2009). The MESS model can be described as follow (LeSage & Pace, 2007; LeSage & Pace, 2009):

A spatial regression mode can be expressed as

$$Sy = X\beta + \varepsilon \quad (10)$$

where the vector  $y$  contains  $n$  observations on the dependent variable, each associated with one region or point in space. The matrix  $X$  represents an  $n \times k$  full column rank matrix of constants which correspond to observations on  $k$  independent variables for each region. The  $n$ -element vector  $\varepsilon$  is distributed as  $N(0, \sigma^2 I_n)$ . The  $k$  element vector  $\beta$  is a vector of corresponding parameters, and  $S$  denotes an  $n \times n$  non-singular matrix of constants that may depend on an unknown real, scalar parameter.

The MESS specification replaces the conventional geometric decay of influence from higher-order neighboring relationships implied by the spatial autoregressive process with an exponential pattern of decay in influence from higher-order neighbouring relationships. Specifically, the MESS model transforms  $S$  to model spatial dependence among the elements of the vector  $y$ :

$$S = e^{\alpha W} = \sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!} \quad (11)$$

where  $W$  is an  $n \times n$  non-negative matrix with zeros on the diagonal and  $\alpha$  represents a scalar real parameter.  $W$  represents a spatial weight matrix, and  $W_{ij} > 0$  indicates that observation  $j$  is a neighbour of observation  $i$ . The matrix exponential  $S$ , along with matrix  $W$ , imposes a decay of influence for higher-order neighbouring relationships.

### 3. Results and Discussion

#### 3.1. Relationships of UHII and NTLI

We explored the relationships of UHII and NTLI through the scatterplots. In Figure 6, both the scatterplots and the corresponding Pearson correlation coefficients indicate the positive relationship of UHII and NTLI exists in London at both the LAD and MSOA levels (See Figure 6). Similarly, in Figure 7, both the scatterplots and the corresponding Pearson correlation coefficients indicate the positive relationship of UHII and NTLI exists in Paris at both the ADM and CM levels (See Figure 7).

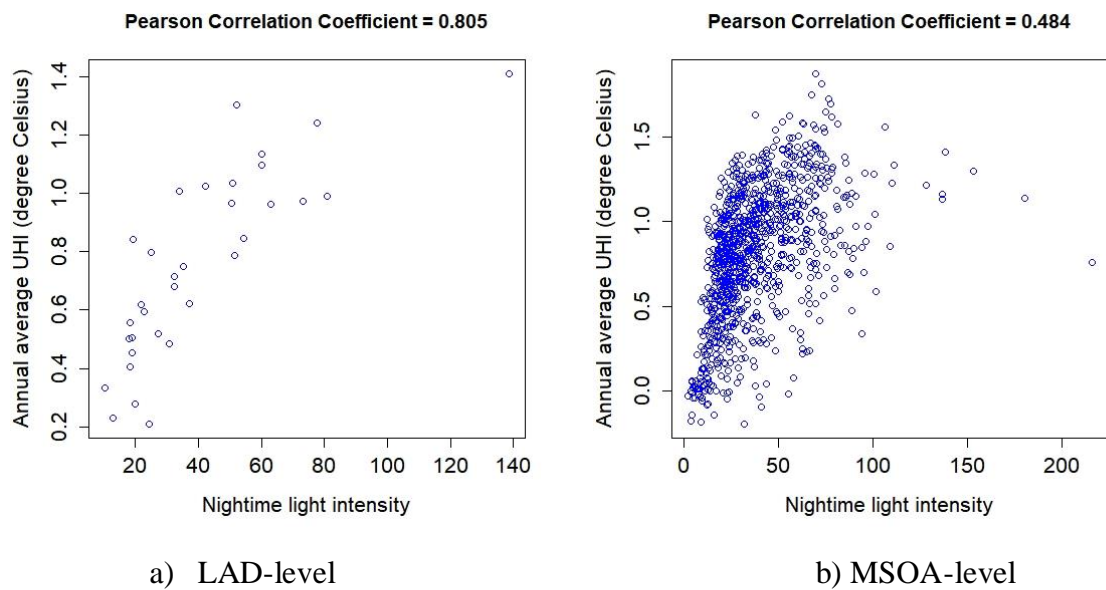


Figure 6: Scatterplots for the relationships of UHII and NTLI in London

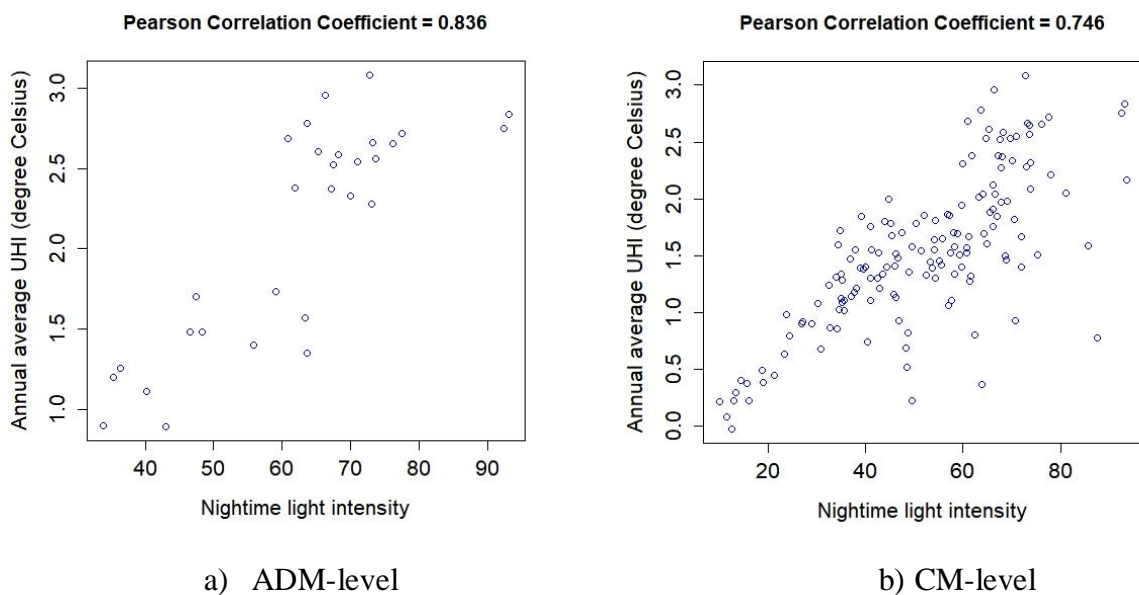


Figure 7: Scatterplots for the relationships of UHII and NTLI in Paris

### 3.2. Estimating UHII using OLS models

We first established OLS models to estimate UHII. In this study, 33 observations (33 LADs) and 983 observations (983 MSOAs) were input into OLS models respectively. We selected the three most popular models: linear, exponential, and quadratic models to explain the relationships of NTLI and UHII. The quadratic and log-linear models outperform the linear models at all the levels due to higher *R*-squared values. Table 1 shows estimation results for the linear, quadratic, and log-linear models at LAD and MSOA levels. In Table 1, the coefficient is the coefficient estimated for each independent variable; and \*, \*\*, \*\*\*, and \*\*\*\* mean corresponding *p*-values are below 0.05, 0.01, 0.001, and 0.0001 respectively. An *R*-squared value of over 0.7 indicates that LAD- or ADM-level UHII was successfully estimated (see Table 1). MSOA- or CM-level UHII was not estimated as well as LAD- or ADM-level UHII since the *R*-squared values of the OLS models for MSOA- or CM-level UHII are less than 0.6 (see Table 1). Furthermore, compared to quadratic form log-linear form is more strongly recommended since some coefficients in the quadratic models are not statistically significant (see Table 1).

**Table 1.** Estimation results for the OLS models

	Coefficient	London		Paris	
		LAD-level	MSOA-level	ADM-level	CM-level
<b>Linear form</b>	Intercept	0.364647 ***	0.5038810****	-0.214947	0.118564
	NTLI	0.009649 ***	0.0081545****	0.037404 ***	0.027102 ***
	Multiple <i>R</i> -squared	0.6476	0.2345	0.6995	0.5559
	Adjusted <i>R</i> -squared	0.6363	0.2337	0.6887	0.5529
<b>Quadratic form</b>	Intercept	1.552e-01	2.828e-01****	-1.691796	-2.654e-01
	NTLI <sup>2</sup>	-7.594e-05 **	-9.083e-05****	-0.000423	-1.794e-04 *
	NTLI	1.918e-02 ****	1.866e-02****	0.089001 **	4.496e-02 ****
	Multiple <i>R</i> -squared	0.7262	0.3158	0.7319	0.5685
	Adjusted <i>R</i> -squared	0.708	0.3144	0.7121	0.5626
<b>Log-linear form</b>	Intercept	-0.84090 ****	-0.49734 ****	-6.7778 ****	-2.78019 ****
	Ln(NTLI)	0.45303 ****	0.37809 ****	2.1694 ****	1.11339 ****
	Multiple <i>R</i> -squared	0.7328	0.3519	0.7215	0.5608
	Adjusted <i>R</i> -squared	0.7241	0.3512	0.7115	0.5579

Note: \*, \*\*, \*\*\*, and \*\*\*\* mean corresponding *p*-values are below 0.05, 0.01, 0.001, and 0.0001 respectively.

### 3.3. Testing for residual spatial autocorrelation in the OLS models

In the previous subsection, log-linear form is strongly recommended. Table 2 lists the testing results for residual spatial autocorrelations in the non-spatial regression models (OLS models in the log-linear form). More specifically, Table 2 lists the testing results for residual spatial autocorrelations in the log-linear models. A positive observed value of Moran's  $I$  and a  $p$ -value of below 0.00001 indicate statistically significant spatial autocorrelations in the residuals of the OLS regression models. This suggests the necessity of usage of spatial regression models.

**Table 2.** Testing results for residual spatial autocorrelation in the OLS models (log-linear form)

a) London-wide LAD-level			
Moran's $I$		Variance	$p$ -value
Observed	Expected		
0.3943	-0.0477	0.0131	< 0.0001 ****

b) London-wide MSOA-level			
Moran's $I$		Variance	$p$ -value
Observed	Expected		
0.6658	-0.0019	0.0004	< 0.0001 ****

c) Paris-wide ADM-level			
Moran's $I$		Variance	$p$ -value
Observed	Expected		
0.2770	-0.0497	0.015	0.0038

d) Paris-wide CM-level			
Moran's $I$		Variance	$p$ -value
Observed	Expected		
0.6063	-0.0116	0.0025	< 0.0001 ****

### 3.4. Estimating UHII using spatial regression models

Likewise, 33 observations (33 LADs), 983 observations (983 MSOAs), 30 observations (30 ADMs), and 150 observations (150 CMs) were input into the spatial regression models (matrix exponential spatial specification, MESS models) respectively. Table 3 lists the estimation results for the MESS models. In London, LAD-level UHII was successfully estimated using the MESS model as the  $R$ -squared value is around 0.7; whilst MSOA-level

UHII was not estimated as well as LAD-level UHII. Similarly, in Paris, ADM-level UHII was successfully estimated using the MESS model as the *R*-squared value is around 0.7; whilst CM-level UHII was not estimated as well as ADM-level UHII.

The London-wide LAD-level MESS model can be written as

$$\text{UHII} = 0.41289\text{Ln}(\text{NTLI}) - 0.85762$$

And Paris-wide ADM-level MESS model can be written as

$$\text{UHII} = 1.63081\text{Ln}(\text{NTLI}) - 5.45717$$

According to some previous studies (Fabbri and Costanzo, 2020; dos Santos, R.S., 2020), root mean square error (RMSE) and mean absolute error (MAE) are used to validate the estimated UHII. As a result, we validated the estimated UHII against the observed UHII at the 6 urban sites in Figure 2 (MAE = 1.16 °C and RMSE = 1.74 °C). The results on London in this study is comparable to a recent study of London based on a small number of stations as well (MAE = 1.60 °C and RMSE = 2.03 °C). In conclusion, district-level UHII is successfully estimated according to nighttime light intensity.

**Table 3.** Estimation results for the MESS models

Coefficient	London		Paris	
	LAD-level	MSOA-level	ADM-level	CM-level
Intercept	-0.85762 ****	-0.27442 ****	-5.45717 ****	-1.37524 ****
Ln(NTLI)	0.41289 ****	0.15331 ****	1.63081 ****	0.47947 ****
Multiple <i>R</i> -squared	0.70319	0.17482	0.71282	0.35957
Adjusted <i>R</i> -squared	0.69362	0.17398	0.70257	0.35524

Note: \*, \*\*, \*\*\*, and \*\*\*\* mean corresponding *p*-values are below 0.05, 0.01, 0.001, and 0.0001 respectively.

### 3.5. Discussion

The modifiable areal unit problem (MAUP) widely exists in statistical analysis of aggregate geospatial data (Nakaya, T., 2000; Zhang and Kukadia, 2005; Dark and Bram, 2007; Wong, 2009). Several studies have confirmed that statistical results vary based on scale and aggregation which is a cause for concern for anyone conducting research with geospatial data (Dark and Bram, 2007). When areal units are aggregated into fewer, larger units for statistical analysis, values associated with the variation of the data decrease which will affect any associated statistical analysis (Dark and Bram, 2007). In this study, as raster-based grids were aggregated to district units (LADs and ADMs) and neighbourhood units (MSOAs and CMs),



modelling results might tend to differ from district to neighbourhood levels. Specifically, compared with aggregation of grids into neighbourhood units, aggregation of grids into district units tends to lose more information associated with the variation of the data, leading to a decrease in statistical analysis results (modelling results) in this study.

The coefficient for Ln(NTLI) in Paris is much larger than that in London (1.63081 > 0.41289). There are two possible reasons. One is that, on average, LAD is larger than ADM. The coefficient for Ln(NTLI) tends to increase as the aggregation unit grows. In Table 3, coefficients for Ln(NTLI) increase as the aggregation unit grows from neighbour level (LAD- and ADM-level) to district level (MSOA- and CM-level) in both London and Paris (from 0.15331 to 0.41289 and from 0.47947 to 1.63081 respectively). The other is that the attenuation of light from ground to satellite might be stronger in Paris than in London. Owing to more severe air pollution or atmospheric turbidity, Paris is likely to have stronger attenuation of light than London. As a consequence, a potential increase in the coefficient for Ln(NTLI) is likely to be an offset against a stronger attenuation of light.

The coefficients estimated are likely to differ from one city to another. One challenge to the extension of the models is that the coefficients estimated is likely to differ from one city to another as empirical results show. Therefore, it is important to estimate the coefficients in different cities to explore whether and how the coefficients vary over space and time. The key step of extending the models established in this study is to apply these models to other European cities including both megacities and smaller cities. More specifically, geographically weighted regression (GWR) models may be applied to Europe-wide local-scale UHI intensity estimates in order to investigate the coefficients estimated vary across Europe.

Governments dedicated to UHI mitigation proposed and implemented policies such as increasing green spaces and green buildings (O'Malley et al., 2015; Costanzo et al., 2016; Park et al., 2017; He, 2018; Qi et al., 2019; He, 2019). It is vital to assess the implementations of policies targeting UHI mitigation. This study demonstrates that nighttime light satellite imagery data like NPP-VIIRS imageries offers an approach to rapidly and easily monitoring UHI intensity which would make policy implementation assessment easy and low cost. Using open satellite imageries to estimate UHI intensity would play an increasingly important role in UHI-related policy implementation assessment in the cities whose UHI data is not publicly available.

## 4. Conclusions

In this study, we estimated district-level and neighbourhood-level UHI intensity across London and Paris using nighttime light satellite imagery data. Due to the relationships of UHII and NTLI uncovered, we first established conventionally nonspatial models to estimate UHI intensity. Consequently, due to the presence of residual spatial autocorrelation in the conventional regression models (OLS models) estimated, we used spatial regression models instead of OLS models to estimate the UHI intensity at both LAD and MSOA levels. This study demonstrated a new approach to estimating UHI intensity at the local scales. We answer two methodological questions in the existing studies on the estimation of UHI intensity: 1) how the spatial resolution impacts the capacity of models; 2) how to overcome the residual spatial autocorrelation in the conventionally nonspatial regression models. As the geography level rises from district (LAD or ADM) to neighbourhood (MSOA or CM), the capacity of the models explaining the variations of the UHI intensity decreases. To remove negative effects of the residual spatial autocorrelation, this study uses spatial regression models instead of non-spatial regression models (e.g., OLS models) to estimate UHI intensity. District-level UHI intensity is successfully estimated according to nighttime light intensity (approximately  $R^2 = 0.7$ , MAE = 1.16 °C, and RMSE = 1.74 °C).

There are some limitations in this study. First, there might exist a gap between the time of simulated UHI data, the time of observed UHI data, and the satellite overpass time of SNPP-VIIRS data. Specifically, the time of simulated UHI data and the satellite overpass time of SNPP-VIIRS data are during the night; whilst the time of observed UHI data could be either midnight or early morning (e.g., 4 am). Second, this study focused on annual UHI intensity estimates whilst monthly UHI intensity estimates is of much interest as well. Third, this study applied the approach only to European megacities, such as London and Paris; whilst whether UHI intensity could be estimated by NTLI in smaller cities as well need to be empirically examined.

We will take account of some aspects in the future. First, we would apply the approach to smaller European cities. We may attempt to estimate Europe-wide or UK-wide local-scale UHI intensity by using SNPP-VIIRS imageries. Second, we would estimate monthly UHI intensity according to monthly NTLI observed by SNPP-VIIRS data if monthly UHI data was available in the future. Subsequently, we will further empirically examine whether the model parameters would differ from one month to another.

## References:

- Amaral, S., Câmara, G., Monteiro, A.M.V., Quintanilha, J.A. & Elvidge, C.D., 2005. Estimating population and energy consumption in Brazilian Amazonia using DMSP nighttime satellite data. *Computers, Environment and Urban Systems*, 29, 179-195.
- Anderson, S.J., Tuttle, B.T., Powell, R.L. & Sutton, P.C., 2010. Characterizing relationships between population density and nighttime imagery for Denver, Colorado: issues of scale and representation. *International Journal of Remote Sensing*, 31(21), 5733-5746.
- Azevedo, J.A., Chapman, L. & Muller, C.L., 2016. Quantifying the daytime and night - time urban heat island in Birmingham, UK: A comparison of satellite derived land surface temperature and high resolution air temperature observations. *Remote Sensing*, 8, 153.
- Bagan, H. & Yamagata, Y., 2015. Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. *GIsci Remote Sensing*, 52(6), 765-780.
- Changnon, S.A., 1992. Inadvertent weather modification in urban areas: Lessons for global climate change. *Bulletin of the American Meteorological Society*, 73, 619-627.
- Chakraborty, T. and Lee, X., 2019. A simplified urban-extent algorithm to characterize surface urban heat islands on a global scale and examine vegetation control on their spatiotemporal variability. *International Journal of Applied Earth Observation and Geoinformation*, 74, 269-280.
- Chakraborty, T., Hsu, A., Manya, D. and Sheriff, G., 2019. Disproportionately higher exposure to urban heat in lower-income neighborhoods: a multi-city perspective. *Environmental Research Letters*, 14(10), 105003.
- Chui, A.C., Gittelsohn, A., Sebastian, E., Stamler, N. and Gaffin, S.R., 2018. Urban heat islands and cooler infrastructure—measuring near-surface temperatures with hand-held infrared cameras. *Urban climate*, 24, 51-62.
- Costanzo, V., Evola, G., & Marletta, L. (2016). Energy savings in buildings or UHI mitigation? Comparison between green roofs and cool roofs. *Energy and buildings*, 114, 247-255.
- Coutts, A.M., Harris, R.J., Phan, T., Livesley, S.J., Williams, N.S. and Tapper, N.J., 2016. Thermal infrared remote sensing of urban heat: Hotspots, vegetation, and an assessment of techniques for use in urban planning. *Remote Sensing of Environment*, 186, 637-651.
- Dark, S.J. and Bram, D., 2007. The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, 31(5), 471-479.
- Debbage, N. & Shepherd, J.M., 2015. The urban heat island effect and city contiguity. *Computers, Environment and Urban Systems*, 54, 181-194.
- Doick, K.J., Peace, A. and Hutchings, T.R., 2014. The role of one large greenspace in mitigating London's nocturnal urban heat island. *Science of the Total Environment*, 493, 662-671.

- dos Santos, R.S., 2020. Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102066.
- Fabbri, K. and Costanzo, V., 2020. Drone-assisted infrared thermography for calibration of outdoor microclimate simulation models. *Sustainable Cities and Society*, 52, 101855.
- Font, A., Guiseppin, L., Blangiardo, M., Ghersi, V. and Fuller, G.W., 2019. A tale of two cities: is air pollution improving in Paris and London?. *Environmental pollution*, 249, 1-12.
- Giridharan, R. and Kolokotroni, M., 2009. Urban heat island characteristics in London during winter. *Solar Energy*, 83(9), 1668-1682.
- Getis, A. & Ord, J.K., 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24, (3), 189-206.
- Ghosh, T., L Powell, R., D Elvidge, C., E Baugh, K., C Sutton, P. & Anderson, S., 2010. Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3(1).
- Guhathakurta, S. & Gober, P., 2010. Residential Land Use, the Urban Heat Island, and Water Use in Phoenix: A Path Analysis. *Journal of Planning Education and Research*, 30(1), 40–51.
- He, B. J. (2018). Potentials of meteorological characteristics and synoptic conditions to mitigate urban heat island effects. *Urban climate*, 24, 26-33.
- He, B. J. (2019). Towards the next generation of green building for urban heat island mitigation: Zero UHI impact building. *Sustainable Cities and Society*, 50, 101647.
- Lac, C., Donnelly, R.P., Masson, V., Pal, S., Riette, S., Donier, S., Queguiner, S., Tanguy, G., Ammoura, L., & Xueref-Remy, I., 2013. CO2 dispersion modelling over paris region within the CO2 -megaparis project. *Atmospheric Chemistry and Physics*, 13, 4941–4961.
- Lauwaet, D., Hooyberghs, H., Maiheu, B., Lefebvre, W., Driesen, G., Van Looy, S., & De Ridder, K. (2015). Detailed Urban Heat Island projections for cities worldwide: dynamical downscaling CMIP5 global climate models. *Climate*, 3(2), 391-415.
- LeSage, J.P. & Pace, R.K., 2007. A matrix exponential specification. *Journal of Econometrics*, 140, 190-214.
- LeSage, J.P. & Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, Chapter 9.
- Liu, Q., Sutton, P.C. & Elvidge, C.D., 2011. Relationships between night-time imagery and population density for Hong Kong. *APAN*, 31, 79-90.
- Lo, C.P., 2001. Modeling the population of China using DMSP operational linescan system nighttime data. *Photogrammetric Engineering & Remote Sensing*, 67(9), 1037-1047.
- Mellander, C., Lobo, J., Stolarick, K. & Matheson, Z., 2015. Night-time light data: A good proxy measure for economic activity?. *PloS one*, 10(10), e0139779.
- Met Office, 2006. MIDAS: UK Hourly Weather Observation Data. Accessed from. NCAS British Atmospheric Data Centre. <https://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c>

- Met Office, 2018. HadUK-Grid gridded and regional average climate observations for the UK. Accessed from Centre for Environmental Data Analysis. <http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb>
- McMichael, A.J., Woodruff, R.E. & Hales, S., 2006. Climate change and human health: Present and future risks. *The Lancet*, 367 (9513), 859-869.
- Mohajerani, A., Bakaric, J. & Jeffrey-Bailey, T., 2017. The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete. *Journal of Environmental Management*, 197, 522–538.
- Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1), 17–23.
- Nakaya, T., 2000. An information statistical approach to the modifiable areal unit problem in incidence rate maps. *Environment and Planning A*, 32(1), 91-109.
- Njuguna, C., 2018. Rnightlights: Satellite Nightlight Data Extraction. <https://cran.r-project.org/web/packages/Rnightlights/index.html>
- NOAA, 2018. Version 4 DMSP-OLS Nighttime Lights Time Series. <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>
- Oke, T., 1987. *Boundary Layer Climates*, 2nd ed.; Methuen: London, UK; p. 289.
- O'Malley, C., Piroozfar, P., Farr, E. R., & Pomponi, F. (2015). Urban Heat Island (UHI) mitigating strategies: A case-based comparative analysis. *Sustainable Cities and Society*, 19, 222-235.
- Park, J., Kim, J. H., Lee, D. K., Park, C. Y., & Jeong, S. G. (2017). The influence of small green space type and structure at the street level on urban heat island mitigation. *Urban forestry & urban greening*, 21, 203-212.
- Patz, J.A., Campbell-Lendrum, D., Holloway, T. & Foley, J.A., 2005. Impact of regional climate change on human health. *Nature*, 438 (7066), 310-31.
- Pigeon, G., Legain, D., Durand, P. & Masson, V., 2007. Anthropogenic heat release in an old European agglomeration (Toulouse, France). *International Journal of Climatology*, 27, 1969–1981.
- Pozzi, F., Small, C. & Yetman, G., 2003. Modeling the distribution of human population with nighttime satellite imagery and gridded population of the world. *Earth Observation Magazine*, 12(4), 24-30.
- Qi, J. D., He, B. J., Wang, M., Zhu, J., & Fu, W. C. (2019). Do grey infrastructures always elevate urban temperature? No, utilizing grey infrastructures to mitigate urban heat island effects. *Sustainable Cities and Society*, 46, 101392.
- Rosenfeld, A.H., Akbari, H., Romm, J.J. & Pomerantz, M., 1998. Cool communities: Strategies for heat island mitigation and smog reduction. *Energy Buildings*, 28 (1), 51-62.
- Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., Chen, L. and Wu, J., 2014. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data. *Remote Sens (Basel)* 6, 1705-1724.

- Sutton, P., 1997. Modeling population density with night-time satellite imagery and GIS. *Computers, Environment and Urban Systems*, 21(3-4), 227-244.
- Townsend, A.C. & Bruce, D.A., 2010. The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution. *International Journal of Remote Sensing*, 31(16), 4459-4480.
- Wang, K., Jiang, Q.G., Yu, D.H., Yang, Q.L., Wang, L., Han, T.C. and Xu, X.Y., 2019. Detecting daytime and nighttime land surface temperature anomalies using thermal infrared remote sensing in Dandong geothermal prospect. *International Journal of Applied Earth Observation and Geoinformation*, 80, 196-205.
- Wong, D., 2009. The modifiable areal unit problem (MAUP). *The SAGE handbook of spatial analysis*, 105(23), 2.
- Zhang, Q. & Seto, K.C., 2011. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment*, 115(9), 2320-2329.
- Zhang, M. and Kukadia, N., 2005. Metrics of urban form and the modifiable areal unit problem. *Transportation Research Record*, 1902(1), 71-79.
- Zhuo, L., Ichinose, T., Zheng, J., Chen, J., Shi, P.J. & Li, X., 2009. Modelling the population density of China at the pixel level based on DMSP/OLS non - radiance - calibrated night-time light images. *International Journal of Remote Sensing*, 30(4), 1003-1018.