Int J Appl Earth Obs Geoinformation 88 (2020) 102066

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data

Rochelle Schneider dos Santos

Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London, United Kingdom

ARTICLE INFO

Keywords:
Air temperature
Land surface temperature
Urban
MODIS
Gradient boosting
Random forest

ABSTRACT

Urbanisation generates greater population densities and an increase in anthropogenic heat generation. These factors elevate the urban–rural air temperature (T_a) difference, thus generating the Urban Heat Island (UHI) phenomenon. T_a is used in the fields of public health and epidemiology to quantify deaths attributable to heat in cities around the world: the presence of UHI can exacerbate exposure to high temperatures during summer periods, thereby increasing the risk of heat-related mortality. Measuring and monitoring the spatial patterns of T_a in urban contexts is challenging due to the lack of a good network of weather stations. This study aims to produce a parsimonious model to retrieve maximum T_a (T_{max}) at high spatio-temporal resolution using Earth Observation (EO) satellite data. The novelty of this work is twofold: (i) it will produce daily estimations of T_{max} for London at 1 km^2 during the summertime between 2006 and 2017 using advanced statistical techniques and satellite-derived predictors, and (ii) it will investigate for the first time the predictive power of the gradient boosting algorithm to estimate T_{max} for an urban area. In this work, 6 regression models were calibrated with 6 satellite products, 3 geospatial features, and 29 meteorological stations. Stepwise linear regression was applied to create 9 groups of predictors, which were trained and tested on each regression method. This study demonstrates the potential of machine learning algorithms to predict T_{max}: the gradient boosting model with a group of five predictors (land surface temperature, Julian day, normalised difference vegetation index, digital elevation model, solar zenith angle) was the regression model with the best performance ($R^2 = 0.68$, MAE = 1.60 °C, and RMSE = 2.03 °C). This methodological approach is capable of being replicated in other UK cities, benefiting national heat-related mortality assessments since the data (provided by NASA and the UK Met Office) and programming languages (Python) sources are free and open. This study provides a framework to produce a high spatio-temporal resolution of T_{max}, assisting public health researchers to improve the estimation of mortality attributable to high temperatures. In addition, the research contributes to practice and policy-making by enhancing the understanding of the locations where mortality rates may increase due to heat. Therefore, it enables a more informed decision-making process towards the prioritisation of actions to mitigate heat-related mortality amongst the vulnerable population.

1. Introduction

In 2018, 83.4 % of the UK population resided in urban areas; this proportion is projected to reach 90.2 % by 2050 (UN DESA, 2018). Urbanisation leads to greater population densities, reductions in urban greenspace, and an increase in anthropogenic heat sources in cities worldwide. The Urban Heat Island (UHI), a phenomenon where the temperature in urban areas is elevated compared to surrounding rural areas, is one of the consequences of urbanisation that directly impact the urban population (Grimmond et al., 2016). Air temperature (T_a) is a key variable in a wide range of research applications, such as climate change and global warming (Intergovernmental Panel on Climate Change (IPCC, 2018), energy management (Zakšek and Schroedter-Homscheidt, 2009), indoor comfort (Mavrogianni et al., 2012; Bechtel et al., 2017), and human health (Armstrong et al., 2011; Hondula et al., 2012; Macintyre et al., 2018; Nichol and Hang, 2012). T_a is used in the fields of public health and epidemiology to analyse temperature–-mortality associations, since it has been demonstrated to be a significant driver of mortality (Gasparrini et al., 2015). UHI exacerbates exposure to high temperatures during summer periods, thereby increasing the risk of heat-related mortality. Typically, T_a is either measured from networks of meteorological stations or simulated by climate models (Fu and Weng, 2018).

Meteorological stations can provide long-term observational

E-mail address: rochelle.schneider@lshtm.ac.uk.

https://doi.org/10.1016/j.jag.2020.102066
Received 30 October 2018; Received in revised form 31 December 2019; Accepted 3 February 2020
Available online 10 February 2020

weather data; however, their ability to describe the spatial variation of T_a in heterogeneous areas (such as cities) is limited due to their lack of appropriate spatial coverage (Ding et al., 2018; Fu and Weng, 2018; Nichol and Hang, 2012). Several publications have also reported problems with missing T_a values caused by disrupted recordings, poor spatial coverage or a total lack of weather stations, such as in parts of West Africa (Stisen et al., 2007). These problems might be caused by the elevated cost of installation and maintenance of the measurement equipment (Zakšek and Schroedter-Homscheidt, 2009), and sometimes by vandalism. Weather station data can be supplemented by other field measurement approaches; for instance, sensors mounted on lamp posts (Levermore et al., 2012), vehicle-traverses (Hart and Sailor, 2009; Nichol et al., 2009), remotely sensed airborne transects (Lee and Sharples, 2008), and fixed and mobile amateur weather stations (Chapman et al., 2017; Ho et al., 2014; Kloog et al., 2014).

Climate models can simulate long-term spatiotemporal T_a changes over large areas of the Earth. These models can generate grids of meteorological data at varying resolutions. Global Climate Models are used for climate predictions at very low spatial resolution (100–250 km^2) with high temporal resolution (e.g. hourly). Due to the low spatial resolution, they are not able to supply temperature information at city scale (Huth et al., 2015; Wilby and Wigley, 1997). Regional Climate Models are capable of estimating T_a in spatial resolutions of 25–50 km^2; however, they provide only a coarse estimate of the urban climate, since they have a very simplified representation of the urban influence on the model (Levermore et al., 2012). Mesoscale models, with the addition of urban surface representations inside the model, are able to simulate climates at more local levels (1–5 km^2) (Bohnenstengel et al., 2011). Unfortunately, only a few simulated T_a data sets are freely available and open access. The data request is usually made to order for a specific location and time, and the modelling service is very expensive for a small-grant project. Simulation approaches are limited by the need of historical observational data for some locations, the long processing time, the model-code complexity, and computational demand.

For decades, earth observation (EO) satellites have monitored the Earth's changes, building an unprecedented spatiotemporal data set. This data is collected by a variety of sensors on board satellites, which capture the interaction of solar energy with the Earth's surface in the form of reflected, absorbed, and emitted radiation. The need for appropriate spatiotemporal T_a data has driven many researchers to explore new methods to retrieve T_a using EO satellite data for large and small areas (countries and cities, respectively). Land Surface Temperature (LST) is widely reported to be the most relevant satellite predictor for T_a estimation (Benali et al., 2012; Chen et al., 2016; Ho et al., 2014; Vancutsem et al., 2010; Weng, 2009; Xu et al., 2014; Yang et al., 2017; Yoo et al., 2018; Zakšek and Schroedter-Homscheidt, 2009; Zhu et al., 2013). There are three main approaches based on LST:

(1) The Temperature-Vegetation Index (TVX) assumes that vegetated areas have lower temperatures than those without vegetation; this is expressed by the Normalized Difference Vegetation Index (NDVI). Therefore, the surface temperature of the vegetation canopy will be very similar to the surrounding T_a, since areas covered just by leaves actually consist mostly of air (Liu et al., 2016; Nieto et al., 2011; Prihodko and Goward, 1997; Weng, 2009; Zhu et al., 2013). However, in urban areas, the focus on retrieving T_a from remote sensing is actually the opposite; that is, data from unvegetated surfaces. Therefore, this method is unsuitable for cities (Agam et al., 2007; Stisen et al., 2007). (2) Energy balance approaches are grounded in thermodynamics, where the T_a is retrieved by analysing the energy exchanges from the urban land surfaces using LST (Hou et al., 2013; Sun et al., 2005), emissivity, sensible heat flux, latent heat flux and solar radiation. The implementation of this method requires variables that are not measured by satellites (Benali et al., 2012).

(3) Statistical techniques can be divided into two groups: (i) spatial interpolation methods, by which T_a can be predicted at any neighbourhood location within a fixed time, and (ii) regression methods, by which T_a can be predicted at any location and time. The first group is composed of deterministic methods (e.g. Inverse Distance Weighting (IDW) and polynomial functions) and stochastic methods (e.g. kriging-based methods and Geographic Weighted Regression (GWR)) (Chen et al., 2015; Li et al., 2018; Ozelkan et al., 2015; Parmentier et al., 2014, 2015). These methods can be successfully used when the study area has a good distribution of weather stations and T_a is predicted at the same point in time as the model equation is generated. However, long-term daily observational periods would generate several model equations, and study areas with irregular spatial coverage of weather stations would weaken the accuracy of the T_a prediction (Florio et al., 2004; Li et al., 2018; Vancutsem et al., 2010; Vogt et al., 1997). The study area for this research is characterised by the latter scenario; therefore, it is not feasible to predict London's T_{max} through the application of spatial interpolation methods. The second group is composed of simple (one feature) or multiple (two or more features) regression approaches, which are recognised for retrieving T_a when interpolation methods are not adequate. Most of the studies presented in the literature explore different types of regression. These range from parametric models such as Linear Regression (LR) (simple or multiple predictors) (Bechtel et al., 2017; Good, 2015; Ho et al., 2014; Kloog et al., 2014; Lin et al., 2012; Pichierri et al., 2012; Yan et al., 2009; Yang et al., 2017; Xu et al., 2014; Zhang et al., 2016), and Stepwise Regression (Lin et al., 2012; Pichierri et al., 2012); to more complex statistical methods like Machine Learning (ML) methods, such as Random Forest (RF) (Ho et al., 2014; Yoo et al., 2018; Xu et al., 2014; Zhang et al., 2016), Support Vector Machine (SVM) (Ho et al., 2014; Moser et al., 2015; Yoo et al., 2018; Zhang et al., 2016) and Neural Networks (NN) (Jang et al., 2004; Zhang et al., 2016). Relying on a satisfactory relationship between T_a and its predictors, empirical regression algorithms use training and testing processes to learn how to best estimate T_a, even in areas with the most heterogeneous landscape characteristics.

Most of the previous studies focused on estimating T_a in heterogeneous areas have used general and advanced statistical approaches. Yang et al. (2017) used multiple LR to estimate T_a during 2002–2016 in north-east China. The study explored five products from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on board the Aqua satellite; six auxiliary data sets; and eight-day T_{max}, T_{min}, and T_{mean} data from 123 meteorological ground stations. The best LR model for eight-day T_{max} run with LST_{day}, NDVI, mean day length, clear sky, Julian day, and latitude achieved an Adjusted-R^2 of 0.90, RMSE of 4.63 °C, and MAE of 3.69 °C. Xu et al. (2014) applied LR and RF to estimate daily T_{max} for the summer period (June, July, and August) from 2003 to 2012 in British Columbia, Canada. As predictors, they selected five products from Aqua MODIS, four auxiliary data sets, and data from 288 weather stations. RF presented the best performance model (MAE = 2.02 °C, R^2 = 0.74) compared to the LR model (MAE = 2.41 °C, R^2 = 0.64). Benali et al. (2012) aggregated data from 106 meteorological stations for eight-day periods between 2000 and 2009 to estimate Portugal's T_{max}, using as predictors MODIS LST Day, MODIS LST Night, and Day length. They obtained an RMSE of 1.83 °C.

The literature has so far discussed temperature–mortality associations at country (Guo et al., 2018), region (Armstrong et al., 2011), and city levels (Gasparrini et al., 2015). The impasse in assessing summer deaths attributable to high temperatures on a local scale is due to the lack of T_{max} measurements at the same spatio-temporal resolution as mortality data; namely, daily records at census level. This study aims to address the literature gap by investigating the most suitable method and group of predictors to retrieve T_{max} at an appropriate spatio-temporal resolution for census-specific levels. This research brings two novel contributions to knowledge: (i) the daily estimations of T_{max} for London at 1 km^2 within 11 years (2006–2017) using advanced statistical techniques and satellite-derived predictors, and (ii) the application for the first time of the gradient boosting method to estimate T_{max} for an urban area.

2. Study area

The study area is London, which is located in South East England and has a population of nearly nine million (Office for National Statistics (ONS, 2018 within an area of 1572 km^2. According to the Köppen climate classification, the UK is defined as having a warm temperate climate, fully humid with a warm summer Kottek et al., 2006). The Meteorological Office (Met Office) divides the UK into 11 regional climates; each climate region has different maritime and latitudinal influences. London is placed in Southern England (Met Office, 2018). It is located between latitudes 51°40′ and 51°1′ N and longitudes 0°30′ W and 0°20′ E; around 75 km from the English Channel (the UK's south coast), and 60 km from the North Sea (the UK's east coast). London's topography is predominantly flat; excepting the North Downs on the southern border, reaching up to 200 m above sea level, and the Chiltern Hills on the north-western border, reaching up to 160 m. The lowest areas are located in the London Basin, from the Thames estuary (city centre) towards the eastern border (Grawe et al., 2013).

3. Data

3.1. Meteorological data

Daily T_{max} records for June, July, and August between 2006 and 2017 from 29 weather stations were provided by the Met Office Integrated Data Archive System (MIDAS) network (Met Office, 2006). The T_a data was spatiotemporally adjusted with the predictors for the learning process in the regression models. Due to the poor meteorological network coverage, the original boundary of London had to be expanded from 60 × 50 km to 100 × 100 km. (Fig. 1).

3.2. Earth observation satellite data

Terra sun-synchronous polar orbit satellite carries two sensors explored in this study: MODIS and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). Table 1 shows the four MODIS products version 6 level 3 used: Land Surface Temperature (LST), Sun Zenith Angle (SZA), Normalised Difference Vegetation Index (NDVI) (Fig. 2) and Black Sky Albedo (BSA)), and one ASTER product, Digital Elevation Model (DEM) (Fig. 3) (Didan, 2015a, b; Schaaf and Wang, 2015). BSA was selected as the albedo product since it represents the total solar energy reflected by the Earth's surface, also known as directional-hemispherical reflectance (Abraha and Savage, 2008).

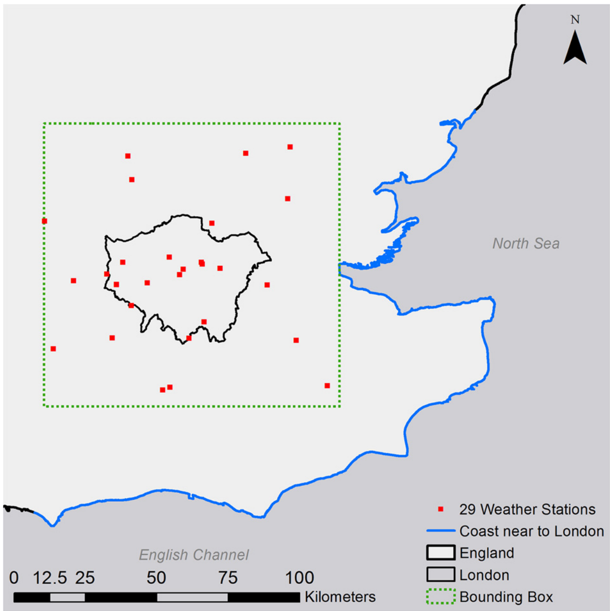

Fig. 1. Location of the weather stations (red square markers) and bounding box (green dashed square). The blue line is the nearest English coast to London (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Some pre-processing steps were undertaken before the data synchronisation. Daily LST pixels with an average error higher than 2 K or cloud-contaminated were removed (Wan et al., 2015). BSA images provided at 500 × 500 m and DEM images at 30 × 30 m were upscaled using bilinear interpolation to match the spatial resolution of the other predictors. Bilinear interpolation is often used to resample data that does not have distinct boundaries (such as temperature, precipitation and DEM) since it computes the average of the four nearest pixels, avoiding assumption based on a single pixel (as the nearest neighbour method).

3.3. Auxiliary data

Four auxiliary variables were also included based on their performance reported in the literature. Distance to the coast (metres) was calculated for each 1 km^2 pixel inside the bounding box area (Fig. 1). Julian day has been used in the literature to reflect temporal and seasonal variations in T_{max} (Jang et al., 2004; Noi et al., 2016; Yang et al., 2017; Zeng et al., 2015). Latitude and longitude coordinates from each 1 km^2 pixel were used as spatial components in the regression methods (Benali et al., 2012; Ding et al., 2018; Good, 2015; Yang et al., 2017; Zeng et al., 2015).

4. Methodology

4.1. Regression methods

Traditional statistical parametric regressions (i.e. LR) describes the relationship between the dependent variable (T_{max}) and the predictors by using a known function, a fixed set of parameters, and predefined assumptions about the data (for example, the mean and standard deviation of a normal distribution). Conversely, non-parametric regressions (such as ML) make minimal assumptions about the data and try to build the mathematical model (an optimal regression function for a specific data) based on what the model has learnt from the training samples.

A variable importance ranking was constructed using a stepwise linear regression to explore the potential contribution of each predictor to T_{max} retrieval in each model. This regression was only used in this study for this purpose; therefore, it was defined as Method 0. One parametric regression and five ML regression methods are investigated here to predict T_{max}: (i) Method 1 – LR, (ii) Method 2 – Decision Trees (DT), (iii) Method 3 – RF, (iv) Method 4 – Gradient Boosting (GB), (v) Method 5 – SVM and (vi) Method 6 – NN.

Method 1 is an intrinsically linear model, where the relationship between the dependent variable and the predictors can be explained by a straight line. LR is the most common regression approach explored in the literature to predict T_a since it produces a performance diagnostic with many outputs. Although ML algorithms are best known for dealing with complex nonlinear relationships, they can also be implemented to enhance the understanding of linear relationships. Method 2 – DT is a supervised, tree-building ML algorithm, designed to determine the most logical data splits into smaller and smaller subsets. The first split considers all predictors (but it might change within different optimisation strategies) and the decision of which predictor should be selected in the next split is based on the candidate with the lowest burden to the model accuracy. This burden is measured by the cost function that tries to find the predictor with less heterogeneity, which means with small variability. After the selection of the first predictor, the following splits will perform under the same strategy. Method 3 – RF is a supervised, tree-building ML algorithm which operates using many decision trees

Table 1
Summary of the satellite products used and the list of previous studies which have used them to predict T_a. London is spread over two MODIS tiles.

Product	Name	Pixel Size	Temporal Granularity	Total images	Previous studies which have used the following satellite products
MODIS					
LST	MOD11A1	1 km^2	Daily	2208	Benali et al., 2012; Chen et al., 2016; Vancutsem et al., 2010; Yang et al., 2017; Zhu et al., 2013
SZA	MOD13A2	1 km^2	16-Day	144	Jang et al., 2004; Vancutsem et al., 2010; Zeng et al., 2015
NDVI	MOD13A3	1 km^2	Monthly	72	Chen et al., 2016
BSA	MCD43A3	500 m^2	Daily	2208	Chen et al., 2016; Ding et al., 2018; Noi et al., 2016; Yang et al., 2017
ASTER					
DEM	ASTGTM	30 m^2	Single (2011)	2	Chen et al., 2016; Jang et al., 2004; Noi et al., 2016; Yang et al., 2017; Zeng et al., 2015

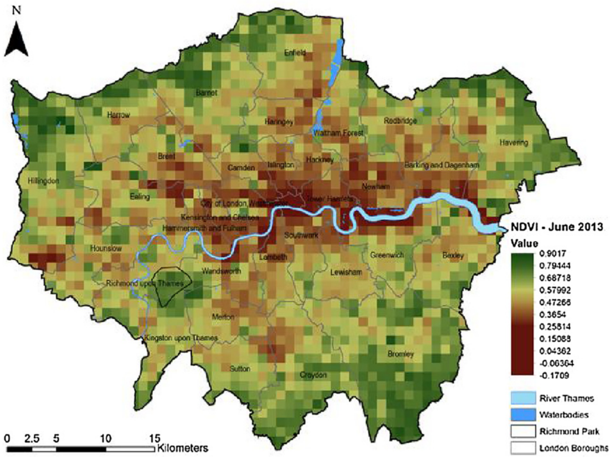

Fig. 2. NDVI from the month of June 2013, obtained from MODIS sensor.

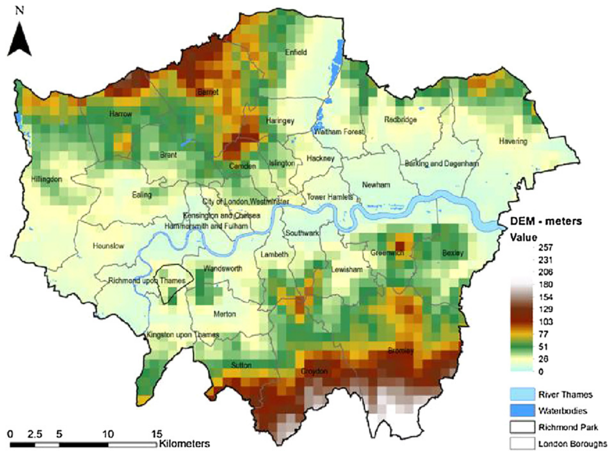

Fig. 3. DEM obtained from ASTER sensor.

trained in parallel. Therefore, multiple independent decision trees (known as bagging ensemble method) use training data provided from a random sampling with replacement from the original training set. The final accuracy results are the average of the performance of all decision trees. This ML model is the second-most-common regression approach in T_a estimation from satellite-derived products, after LR. This multiple tree-based structure helps the model to become more robust, compared to a single decision tree, and also reduces the chances of overfit on the training data. Method 4 – GB is a supervised, tree-building ML algorithm which operates using many decision trees trained in sequence. Therefore, multiple dependent decision trees (known as boosting ensemble method) use training data provided from a random sampling with replacement from the original training set based on a weighted condition. Differently from RF, each decision tree receives a performance weight which is used to build the next decision tree model. Models with low performance are assigned a lower weight, then the subsequent decision trees concentrate on these weak learners during their training. The final accuracy results are taken from the weighted average of all decision trees' performance.

Linear–SVM (Method 5) and LR (Method 1) are similar regression methods since both methods try to explain the relationship between the dependent variable and the predictors using a linear kernel. The difference is Linear–SVM algorithm adds two auxiliary straight lines to define which are the training set points located on or close to the decision boundaries. Therefore, the points that fall inside this margin are those training samples already explained by the model and do not generate any cost for the model performance. Therefore, the algorithm focuses on those points outside the decision boundary to build the model. Method 6 – NN is a ML method inspired by the biological learning process of a complex set of interconnected neurons. A multilayer perceptron (MLP) is a class of feedforward NN composed of multiple layers of computational units and utilises a supervised learning algorithm technique called backpropagation. Each input (predictor) is multiplied by a specific weight and the sum of these weighted inputs on each hidden neuron is then multiplied by a specific bias. If this final result from each hidden neuron passes the threshold defined by the activation function, the information will then move towards the output layer. All information that arrives on this last layer is summed up, returning a prediction in the range of the dependent variable.

4.1.1. Model fitting

The 29 weather stations did not provide a complete time series of T_{max} measurements from 1 June 2006 to 31 August 2017 since some weather stations were activated and others deactivated during the observed period. Therefore, data collected from the weather stations and the predictors were combined in a unique and large database to produce a robust sample size, providing enough training and testing examples for the regression models. All the T_{max} measurements were treated as independent observations where the latitude, longitude and Julian day, were included as spatial and temporal components of the measured T_{max}.

4.1.2. Model calibration, validation, and accuracy

An exploratory analysis was initially conducted to compute the predictors' Pearson correlation coefficient (r). Sequentially, a stepwise regression approach was performed to rank the predictors and verify their significance for the model (Lin et al., 2012; Yoo et al., 2018; Zhang et al., 2016). In all regression models, the data set was randomly split into 70 % training (calibration) and 30 % testing (validation). In this study, R squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) were used as metrics for validation and comparison of the six statistical models. These performance measurements were adopted due to their universal use in the literature to compare regression models within the same study or with other papers of the same scope. The difference between MAE and RMSE depends on the variance in the individual errors in the sample. Both metrics have the same unit of the variable of interest; they are known as negatively oriented scores (Chai and Draxler, 2014).

Table 2
Pearson correlation (*r*) results.

NDVI	DEM	Lat	BSA	SZA	Long	Coast	Julian	LST
−0.21	−0.158	−0.095	−0.074	−0.027	−0.001	0.069	0.127	0.74

5. Results

5.1. Feature contribution

Table 2 presents the Pearson correlation results, showing the LST was the predictor with the highest degree of correlation with London's T_{max}, demonstrating a positive correlation of 0.74, followed by NDVI (*r* = -0.21) and DEM (*r* = -0.158). Although the other predictors presented less than ± 0.15 correlation, some of them demonstrated a significant contribution to estimating T_{max} in the regression models.

Using a forward selection approach, the stepwise process creates a predictor ranking: a sequence of groups, starting with the first model using only the predictor that had the smallest probability of *F* (ρ-value), which in this study was LST, up to the total number of predictors considered in the analysis. The summary of the stepwise regression model is presented in Table 3. As can be noticed, even obtaining the fourth-strongest correlation with T_{max}, Julian day was considered as the second-most relevant predictor. All predictors proposed in this study were relevant to estimate London's T_{max}, using a confidence level of 95 %. The predictors' list was ranked as follows: LST, Julian day, NDVI, DEM, SZA, BSA, distance from the coast, latitude and longitude. The B column contains the unstandardized coefficients which represent the magnitude and direction (positive or negative) of the predictor's effect on the dependent variable. As demonstrated by Table 3, even latitude and longitude having a significant association with T_{max}, their effect on the T_{max} estimations was negligible.

A histogram of regression standardised residual (Fig. 4) and the normal P-P Plot (Fig. 5) were created based on the regression model performed with Model 9. As can be noticed, they presented a normal distribution behaviour, with the classical bell shape and a linear pattern plot, respectively.

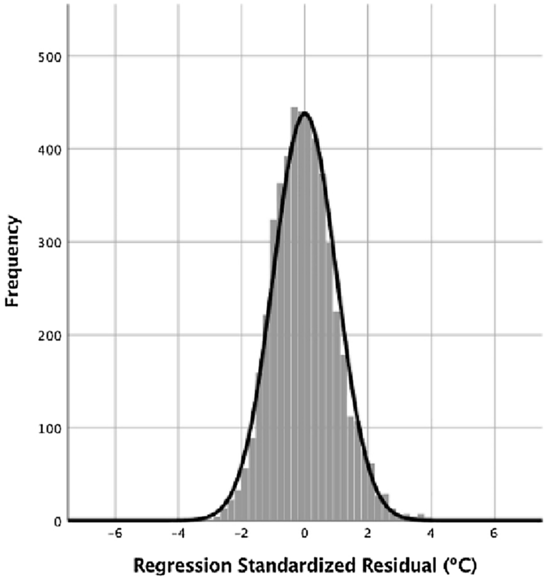

Fig. 4. Histogram of regression standardized residual of a stepwise regression model with all predictors included.

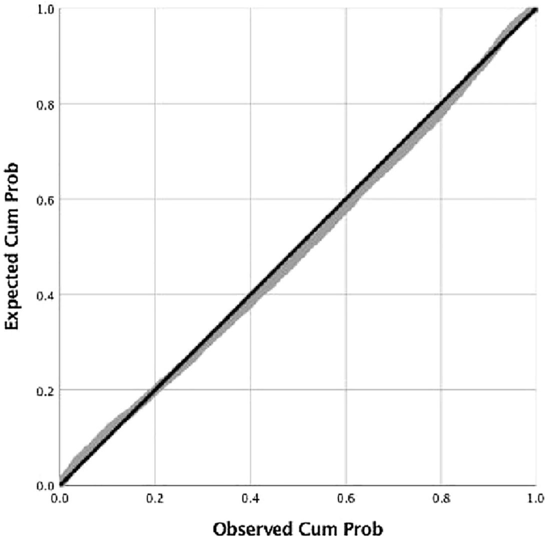

Fig. 5. Normal P-P Plot of regression standardized residual of a stepwise regression model with all predictors included.

5.2. Performance of models

Based on the stepwise regression results, nine groups of predictors were created to check the performance of the six regression methods using a sample size of 6.442 observations. For comparison purposes, Table 4 shows the performance metrics (R^2, MAE, and RMSE) of the 54 regression models. The R^2 results from Method 0 are displayed together with Method 1, since both are based on a linear regression approach.

The R^2 range of all models was between 0.32 (Method 2, Groups 2 and 3) and 0.68 (Method 4, Groups 4 and 5). The RMSE ranged from 2.99 °C (Method 2, Group 2) to 2.03 °C (Method 4, Group 5). The MAE ranged from 2.27 °C (Method 2, Group 3) to 1.59 °C (Method 4, Group 4). The GB method outperformed all methods in all groups except Group 1, where Methods 5 and 6 presented the lowest RMSE (2.46 °C) and MAE (1.92 °C). The models using a single DT (Method 2) obtained the worst performance of the 54 models, except for Method 2/Group 1, which presented a close result but still lower than the other models. As can be noticed, a single decision tree (Method 2) was unable to capture the variability of T_{max}; therefore, in the following comparative results, the models performed with Method 2 were not considered. The method and group combination which presented the best performance was Method 4 (GB) and Group 5 (LST, Julian day, NDVI, DEM, and SZA) with R^2 = 0.68, MAE = 1.60 °C, and RMSE = 2.03 °C. Although Group 9 offered good statistical results in almost all models, the parsimonious model was achieved using fewer predictors, as in Group 5.

Table 3
Stepwise regression results for the model with all predictors.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	15.592	1.624		9.599	0
LST	0.589	0.008	0.796	77.94	0
Julian day	0.037	0.002	0.265	20.344	0
NDVI	4.863	0.289	0.202	16.817	0
DEM	−0.007	0.001	−0.105	−8.789	0
SZA	−0.061	0.008	−0.095	−7.279	0
BSA	−9.914	1.649	−0.063	−6.012	0
Distance from the coast	−0.0390	0.004	−0.210	−9.254	0
Latitude	−2.54E-11	0.000	−0.180	−8.652	0
Longitude	1.21E-11	0.000	0.083	6.655	0

Table 4
Performance results of 9 groups of predictors on each regression method, computing 54 regression models. The best result is highlighted in grey.

		Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9
Method 0/1 LR	R²	0.53	0.55	0.56	0.57	0.57	0.56	0.55	0.57	0.57
	MAE	1.96	1.92	1.90	1.86	1.86	1.89	1.89	1.89	1.88
	RMSE	2.49	2.45	2.42	2.38	2.37	2.40	2.40	2.39	2.39
Method 2 DT	R²	0.40	0.32	0.32	0.40	0.36	0.38	0.37	0.36	0.37
	MAE	2.18	2.26	2.27	2.16	2.16	2.17	2.16	2.20	2.19
	RMSE	2.80	2.99	2.98	2.82	2.89	2.87	2.88	2.90	2.89
Method 3 RF	R²	0.42	0.54	0.60	0.64	0.65	0.66	0.65	0.64	0.64
	MAE	2.15	1.89	1.77	1.67	1.66	1.61	1.65	1.69	1.68
	RMSE	2.76	2.46	2.29	2.17	2.13	2.12	2.16	2.18	2.17
Method 4 GB	R²	0.54	0.62	0.64	0.68	0.68	0.67	0.67	0.67	0.67
	MAE	1.93	1.76	1.69	1.59	1.60	1.62	1.63	1.62	1.63
	RMSE	2.47	2.24	2.17	2.05	2.03	2.08	2.09	2.08	2.08
Method 5 SVM	R²	0.54	0.57	0.61	0.63	0.63	0.65	0.65	0.64	0.64
	MAE	1.92	1.84	1.76	1.70	1.71	1.67	1.68	1.68	1.69
	RMSE	2.46	2.37	2.28	2.20	2.20	2.16	2.16	2.17	2.17
Method 6 NN	R²	0.54	0.58	0.61	0.63	0.64	0.64	0.65	0.64	0.64
	MAE	1.92	1.85	1.77	1.71	1.71	1.70	1.70	1.70	1.71
	RMSE	2.46	2.40	2.27	2.20	2.18	2.19	2.17	2.19	2.19

5.3. Performance of the parsimonious model

Using the parsimonious model, T_{max} predictions for two summer days of 2013 were provided for all London's grid cells. Fig. 6 shows the distribution of predicted T_{max} for 6 June, ranging from 19.90 °C to 24 °C, and Fig. 7 shows the distribution of predicted T_{max} for 6 July, from 24.7 °C to 29.7 °C. In both figures, the light grey polygons are the London boroughs and the green polygon is the Richmond Park (approximately 10 km^2). The interesting finding in plotting daily T_{max} was the variability in the T_{max} intensity inside the London boundary. The T_{max} predictions match with the physical and thermal characteristic of the city.

Most of the grid cells located on the outskirts of the city presented cooler T_{max} and these areas are known as the London's green belt (London First, 2018). Other areas that also presented lower T_{max} are located at higher elevation (Fig. 3), for example those located at the north of the Borough of Camden, the west of the Borough of Haringey, the south of the Borough of Lambeth and Southwark, the west of Lewisham, and the majority of the grid cells inside the boroughs of Sutton, Croydon and Bromley. Conversely, it is possible to recognise the urban influence that has warmed up many grid cells throughout the city. There is a noticeable difference in temperature amplitude between the cooler pixels previously described and the warmer pixels located at densely built-up areas spread out both north and south of the River Thames.

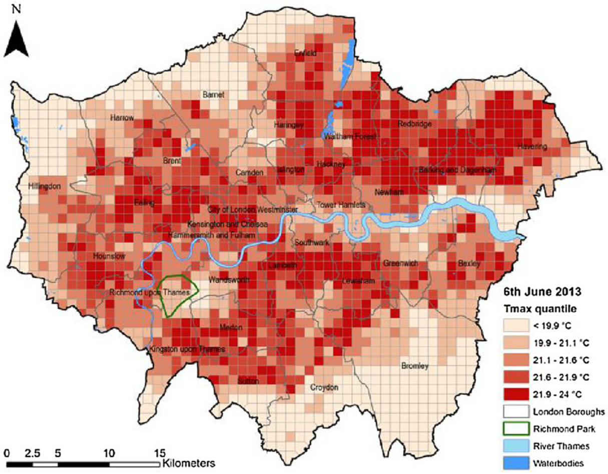

Fig. 6. T_{max} predictions across London on 6 June 2013. The daily T_{max} average was 20.8 °C.

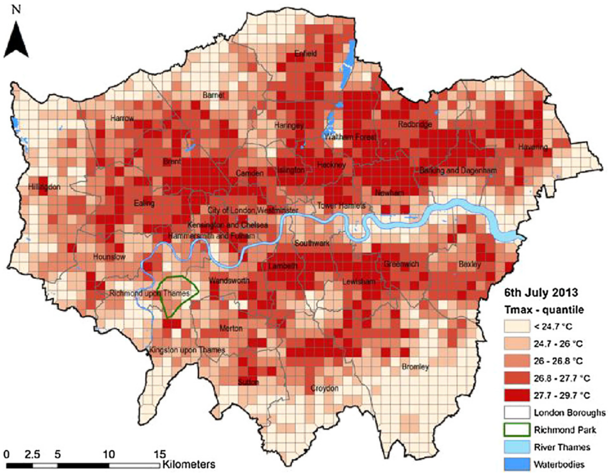

Fig. 7. T_{max} predictions across London on 6 July 2013. The daily T_{max} average was 26.1 °C.

6. Discussion

6.1. Variable importance and significance

According to Zakšek and Schroedter-Homscheidt (2009), T_a is indirectly driven by the LST. The exploratory analysis showed that the LST is the most relevant predictor for estimating London's T_{max}. This strong relationship has also been found in many other studies which have retrieved T_a from satellites (Benali et al., 2012; Chen et al., 2016; Ho et al., 2014; Vancutsem et al., 2010; Weng, 2009; Xu et al., 2014; Yang et al., 2017; Yoo et al., 2018; Zakšek and Schroedter-Homscheidt, 2009; Zhu et al., 2013). Due to the small number of investigated months (only June, July, and August), the initial assumption was that Julian day would not be a relevant predictor; however; it was found to be more important than NDVI, DEM and SZA, in agreement with Jang et al. (2004). The stepwise results also demonstrated that even latitude and longitude being significant to T_{max} estimations, their contribution to predict T_{max} for London was not relevant. As mentioned before, climate is influenced by latitude, but its relevance as a predictor might be associated with a study area of hundreds of thousands of kilometres. The size of London (approx. 1.572 km^2) is relatively small where not much variability in the coordinate values is seen between 1 km^2 pixels. However, previous studies have presented evidence of the latitude and longitude importance on providing T_a prediction for very large areas (e.g. country level) (Benali et al., 2012; Ding et al., 2018; Good, 2015; Yang et al., 2017; Zeng et al., 2015).

6.2. Performance comparison between models

Table 4 displays the performance results of 54 regression models that were investigated to determine the parsimonious model to predict T_{max} for London. The hypothesis was that all the selected predictors would be relevant to retrieving T_{max} for London. This assumption was confirmed through statistical investigations; however, the parsimony principle relies on the selection of the simplest scientific explanation that best illustrates a relationship, with the fewest assumptions and variables but with high informative power (Forster, 1999). Therefore, even though all the proposed predictors were significant, the selected model was the one which required the smallest number of predictors to perform T_{max} estimation with the lowest accuracy error.

Method 1 was the only parametric regression method included, where the description of the performance was already explained by Method 0 since both are linear regression methods. Method 2 was the first ML method implemented, which uses a single decision tree algorithm and presented the lowest performance comparing to all other regression methods. The underperformance and instability are often mentioned in the literature when this method is compared to other tree-building algorithms (Dwyer and Holte, 2007). The meaning of instability is defined as the large difference in accuracy between the training and testing sets. When using many predictors with a wide range of values, the tree also becomes too large and deep which makes the interpretation of each predictor's effect more difficult, weakening the strengths of the decision. The good performance demonstrated by RF (Method 3) and GB (Method 4) models is explained by their ability to use a combination (in parallel or in sequence) of hundreds or thousands of trees to efficiently determine which are the strong and weak training trees' settings, allowing more opportunities to find the best model fit and providing higher stability.

An additional examination was performed in four ML methods (Method 2 was excluded from this analysis). Using the results presented in Table 4, the first analysis was performed by column, keeping the group of predictors for all methods constant. As evidenced, Method 4 had a higher performance than Methods 3, 5 and 6 for eight groups (from Group 2 to Group 9). An examination of the groups provides evidence of improvements in the models caused by the use of a ML method (Methods 2–6) rather than a parametric method (Method 1). An RMSE improvement of 10 % from the reference (Method 1) was defined to determine in which models the use of ML methods was relevant, based on the default parameters configuration of each algorithm. Method 4 was the only ML method to present an RMSE improvement above 10 % in all groups (except Group 1) with the highest improvement on Group 5 (14.35 %). Method 3 reached the threshold on Groups 5–7, Method 5 on Groups 6–7 and Method 6 on Group 7. These results explain why the linear regression approach is still the most popular regression method implemented in the literature to predict T_a. The reasons draw upon not only the small RMSE improvement from ML methods, but mostly because it is easier to implement this parametric regression method and to interpret the performance indicators. However, the disadvantage of this parametric method is in its default assumptions about the data, for example, normality of residuals, constant variance and a true linearity of the modelled relationship (Helsel and Hirsch, 1992). In the cases where the relationship between the dependent variable and the predictors is not linear, these default assumptions cannot be accepted; therefore, the linear regression is not a suitable method to perform the predictive analysis.

The second analysis was performed by row, keeping the regression method constant for all groups. It is important to highlight that for Group 1 the performance from Method 1 to Method 6 was very similar. Analysing the results based on this horizontal structure, the outperformance of ML methods in relation to the parametric method is evident. Method 1 did not improve with the addition of more predictors from Group 1 to Group 9, with the RMSE falling by 4 %. However, Method 4 obtained 16 %, the second-highest RMSE improvement (Method 3 was 21 %), followed by Method 5 (12 %) and Method 6 (11 %). Evaluating the RMSE improvements group by group, it was found that the addition of Julian brought the biggest improvements for Methods 3–4 (around 10 %) while it was only around 3 % for Methods 5 and 6. The addition of NDVI and DEM also generated a bigger improvement to all methods (around 7 % and 3 %). After LST, NDVI is the most common variable used in LST-based modelling approaches: (i) TVX models (Agam et al., 2007; Vancutsem et al., 2010; Zhu et al., 2013), (ii) energy balance models (Zakšek and Schroedter-Homscheidt, 2009; Sun et al., 2005) and (iii) advanced analytic models (Chen et al., 2016; Didari et al., 2017; Ding et al., 2018; Lin et al., 2012; Noi et al., 2016; Yan et al., 2009; Yang et al., 2017). After Group 5, the addition of a new predictor produced only small improvement in RMSE. The variation in the method performance between groups demonstrates that each ML method has a different learning process and a different way to understand the data, and each of them provide their own predictive algorithm which best approximates to the real estimations.

6.3. Comparison with previous studies

The examination of the remote sensing literature shows that none of the studies using satellite-based ML models to predict T_a are exactly the same, mostly because of the following differences: (i) T_a type (T_{mean}, T_{max}, or T_{min}), (ii) geographical location, (iii) size of the case study, (iv) statistical modelling approach, (v) temporal resolution (daily and monthly), (vi) predictors type and source, and (vii) observation period (Bechtel et al., 2017). Therefore, it is difficult to compare the results with previous studies. However, the literature uses similar statistical metrics (R^2, RMSE and MAE) to report the results.

The literature on the use of satellite-based ML models is concentrated on providing T_a prediction for very large areas (hundreds of thousands of square kilometres) by reason of the coverage limitations of meteorological stations (Benali et al., 2012; Chen et al., 2016; Jang et al., 2004; Kloog et al., 2014; Vancutsem et al., 2010; Xu et al., 2014; Yang et al., 2017; Zhang et al., 2016; Zhu et al., 2013). However, these studies did not extract a smaller area to discuss their results at local level. Inside the hundreds of thousand km^2, there are several cities (such as London with 1.569 km^2) for which the T_a pattern, intensity and variation could be reported and compared with other cities from the same country. Only a few studies have investigated urban areas using satellite-based ML models to retrieve T_a (Ho et al., 2014; Yoo et al., 2018).

Ho et al. (2014) used LST, DEM, solar view factor, solar radiation and normalised difference water index to predict daily T_{max} for Greater Vancouver (Canada) (2.700 km^2). The data set was built based on T_{max} records from 59 weather stations for the summertime period between 2001 and 2010. The study explored three regression models and RF presented the best results (RMSE = 2.3 °C), which was higher comparing the RMSE obtained for London using the novel approach (gradient boosting (Method 4), RMSE = 2.03 °C) and the RMSE obtained by the Method 3 (RF, RMSE = 2.13 °C). Yoo et al. (2018) used a RF method to estimate daily T_{max} for two cities: (i) Los Angeles (USA), with 1.290 km^2, and (ii) Seoul (South Korea), with 650 km^2. Their temporal coverage (July and August from 2006 to 2016) was similar to the observational period used for London (which is June, July and August from 2006 to 2017). Yoo et al. (2018) also used LST, NDVI, DEM, latitude and longitude as T_{max} predictors; however, eight types of LST images were included per day from different Terra and Aqua passing times. Since LST is the most important variable to predict T_{max}, the addition of multiple-types of LST will increase the model accuracy; therefore, as the result their model presented an RMSE lower than 2 °C. However, there is a trade-off between the RMSE improvement, by adding more LST types, and the sample size reduction since the prediction can only be performed if all LST types are cloud-free for each grid cell. Based on the parsimony principle, the most suitable model to predict London's T_{max} did not only have to present the lowest accuracy error, as the majority of the studies in the literature aimed, but also to provide the maximum number of 1 km^2 grid cells able to have the T_{max} predicted by the model.

6.4. *Spatial distribution of T_{max}*

Figs. 6 and 7 showed the spatial distribution of London's predicted T_{max} on 6 June and 6 July 2013, respectively. The results provide evidence of the cool and warm effects from land cover elements on the T_{max} variability at the local level. Examples of cooler influence include high elevation, large vegetation coverage (such as Richmond Park), and water bodies (such as rivers, canals, and reservoirs). Examples of warmer influence include built-up areas and intense traffic that where located at the west of the reservoirs, in the Boroughs of Enfield and Haringey, as well as to the east in the borough of Waltham Forest. These areas have high population density. Along the west side of the reservoirs there are also a lot of manufacturing sites, which are embedded within the housing stock. This high density of residential addresses continues towards the River Thames, extending across the boroughs of Hackney, Islington, Tower Hamlets, and Newham. The lowest-elevation areas are located in the London Basin, from the Thames estuary (city centre) towards the eastern border. The River Thames is an important pervious surface that crosses London from west to east; however, its daytime contribution on the T_{max} predictions is not evidenced along all its path. Two hypotheses are highlighted as plausible explanations. First, the river's width is only 250 m at Tower Bridge (between the Boroughs of Tower Hamlets and Southwark) and less than 100 m next to Richmond Park; therefore, the heat mitigation caused by the river might have been lost in a 1 km^2 grid cell, not being able to compute its effects on the surrounding areas. Second, the intensity of traffic flows and densely built-up areas along the river (a mix of residential, commercial, and industrial uses) might have inhibited the cooling effects from this water body during the day.

7. Conclusion

This research analysed the performance of 54 regression models, assembled from 6 regression models, 6 satellite products, 3 geospatial features, and 29 meteorological stations, to select the parsimonious model to retrieve daily T_{max} for London. It contributes to the literature through the development of a novel ML approach, by utilising for the first time the GB algorithm calibrated with EO data to retrieve T_{max} at high spatio-temporal resolution in an urban area. The predictive errors of the parsimonious model (MAE = 1.60 °C and RMSE = 2.03 °C) are comparable to those reported in previous studies, despite the small study area and the small number of weather stations. The contribution of satellite products has proven to be crucial for the estimation of essential climate variables; therefore, more research attention must be given to the exploration of these data sources.

The research findings provide benefits for public health policies regarding adaptation and mitigation responses to climate change. They will assist epidemiologists and health professionals to improve their estimations of the mortality burden attributed to high temperatures at intra-city levels. Practitioners and policymakers can also rely on this data set to better understand the spatial distribution of T_{max} and detect UHI anomalies, allowing the prioritisation of actions to mitigate heat-related mortality amongst the vulnerable population.

Future research on advanced statistical techniques might explore the development of an appropriate T_{max} data set for census-specific levels, aiming to provide correct local climate data for epidemiological studies on the associations between temperature and mortality.

Author contributions

Rochelle Schneider dos Santos designed the study, performed the data collection, conducted the analysis, and wrote the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author is grateful for the support received from the Brazilian research agency CNPq (National Council for Scientific and Technological Development). This research had free and open access to satellite data provided by the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), and to weather data provided by the UK Centre for Environmental Data Analysis. Finally, the author thanks the anonymous reviewers for their constructive comments.

References

Abraha, M.G., Savage, M.J., 2008. Comparison of estimates of daily solar radiation from air temperature range for application in crop simulations. Agric. For. Meteorol. 148 (3), 401–416. https://doi.org/10.1016/j.agrformet.2007.10.001.

Agam, N., Kustas, W.P., Anderson, M.C., Li, F., Neale, C.M., 2007. A vegetation index based technique for spatial sharpening of thermal imagery. Remote Sens. Environ. 107 (4), 545–558. https://doi.org/10.1016/j.rse.2006.10.006.

Armstrong, B.G., Chalabi, Z., Fenn, B., Hajat, S., Kovats, S., Milojevic, A., Wilkinson, P., 2011. Association of mortality with high temperatures in a temperate climate: England and Wales. J. Epidemiol. Community Health 65 (4), 340–345. https://doi.org/10.1136/jech.2009.093161.

Bechtel, B., Zaksek, K., Oßenbrugge, J., Kaveckisa, K., Bohnera, J., 2017. Towards a satellite based monitoring of urban air temperatures. Sustain. Cities Soc. 34, 22–31. https://doi.org/10.1016/j.scs.2017.05.018.

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

Bohnenstengel, S.I., Evans, S., Clark, P.A., Belcher, S.E., 2011. Simulations of the London Urban Heat Island. Q. J. R. Meteorol. Soc. 137 (659), 1625–1640. https://doi.org/10.1002/qj.855.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geosci. Model. Dev. 7, 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.

Chapman, L., Bell, C., Bell, S., 2017. Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. Int. J. Climatol. 37, 3597–3605. https://

doi.org/10.1002/joc.4940.

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.

Chen, Y., Quan, J., Zhan, W., Guo, Z., 2016. Enhanced statistical estimation of air temperature incorporating nighttime light data. Remote Sens. 8 (8), 1–23. https://doi.org/10.3390/rs8080656.

Didan, K., 2015a. MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAChttps://doi.org/10.5067/MODIS/MOD13A2.006.

Didan, K., 2015b. MOD13A3 MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAChttps://doi.org/10.5067/MODIS/MOD13A3.006.

Ding, L., Zhou, J., Zhang, X., Liu, S., Cao, R., 2018. Downscaling of surface air temperature over the Tibetan Plateau based on DEM. Int. J. Appl. Earth Obs. Geoinf. 73, 136–147. https://doi.org/10.1016/j.jag.2018.05.017.

Dwyer, K., Holte, R., 2007. Decision tree instability and active learning. In: In: Kok, J.N., Koronacki, J., Mantaras, R.L., Matwin, S., Mladenič, D., Skowron, A. (Eds.), Machine Learning: ECML 2007. ECML 2007. Lecture Notes in Computer Science 4701. pp. 128–139. https://doi.org/10.1007/978-3-540-74958-5_15.

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 (15), 2979–2994. https://doi.org/10.1080/01431160310001624593.

Forster, M.R., 1999. Parsimony and simplicity. In: Wilson, R.A., Keil, F.C. (Eds.), The MIT Encyclopedia of the Cognitive Sciences. The MIT Press, Cambridge, MA, pp. 627–629.

Fu, P., Weng, Q., 2018. Variability in annual temperature cycle in the urban areas of the United States as revealed by MODIS imagery. ISPRS J. Photogramm. Remote. Sens. 146, 65–73. https://doi.org/10.1016/j.isprsjprs.2018.09.003.

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., Forsberg, B., Leone, M., De Sario, M., Bell, M.L., Guo, Y.L.L., Wu, C., Kan, H., Yi, S.M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P.H.N., Honda, Y., Kim, H., Armstrong, B., 2015. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. Lancet 386, 369–375. https://doi.org/10.1016/S0140-6736(14)62114-0.

Good, E., 2015. Daily minimum and maximum surface air temperatures from geostationary satellite data. J. Geophys. Res. Atmos. 120, 2306–2324. https://doi.org/10.1002/2014JD022438.

Grawe, D., Thompson, H.L., Salmond, J.A., Caia, X., Schlunzen, K.H., 2013. Modelling the impact of urbanisation on regional climate in the Greater London Area. Int. J. Climatol. 33, 2388–2400. https://doi.org/10.1002/joc.3589.

Grimmond, C.S.B., Ward, H.C., Kotthaus, S., 2016. How is urbanization altering local and regional climate? In: Seto, K.C., Solecki, W.D., Griffith, C.A. (Eds.), The Routledge Handbook of Urbanization and Global Environmental Change. Routledge.

Guo, Y., Gasparrini, A., Li, S., Sera, F., Vicedo-Cabrera, A.M., Coelho, M., de Sousa, Z.S., Saldiva, P.H.N., Lavigne, E., Tawatsupa, B., Punnasiri, K., et al., 2018. Quantifying excess deaths related to heatwaves under climate change scenarios: a multicountry time series modelling study. PLoS Med. 15 (7), e1002629. https://doi.org/10.1371/journal.pmed.1002629.

Hart, M.A., Sailor, D.J., 2009. Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island. Theor. Appl. Climatol. 95, 397–406. https://doi.org/10.1007/s00704-008-0017-5.

Helsel, D.R., Hirsch, R.M., 1992. Statistical Methods in Water Resources, vol. 49 546 p., ISBN: 9780080875088.

Ho, H.C., Knudby, A., Sirovyak, P., Xub, Y., Hodul, M., Henderson, S.B., 2014. Mapping maximum urban air temperature on hot summer days. Remote Sens. Environ. 154, 38–45. https://doi.org/10.1016/j.rse.2014.08.012.

Hondula, D.M., Davis, R.E., Leisten, M.J., Saha, M.V., Veazey, L.M., Wegner, C.R., 2012. Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983-2008: a case-series analysis. Environ. Health 11, 16. https://doi.org/10.1186/1476-069X-11-16.

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.

Huth, R., Miksovsky, J., Stepanek, P., Belda, M., Farda, A., Chladova, Z., Pisoft, P., 2015. Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature. Theor. Appl. Climatol. 120 (3–4), 533–553. https://doi.org/10.1007/s00704-014-1190-3.

Intergovernmental Panel on Climate Change (IPCC), 2018. Global Warming of 1.5ºC. (accessed 9 February 2019). https://www.ipcc.ch/sr15/.

Jang, J.D., Viau, A.A., Anctil, F., 2004. Neural network estimation of air temperatures from AVHRR data. Int. J. Remote Sens. 25, 4541–4554. https://doi.org/10.1080/01431160310001657533.

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.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130.

Lee, S.E., Sharples, S., 2008. An analysis of The Urban Heat Island of Sheffield - The impact of a changing climate. In: PLEA 2008 – 25th Conference on Passive and Low Energy Architecture. Dublin, 22nd to 24th October 2008.

Levermore, G.J., Courtney, R., Watkins, R., Cheung, H., Parkinson, J.B., Laycock, P., Natarajan, S., Nikolopoulou, M., McGilligan, C., Muneer, T., Tham, Y., Underwood, C.P., Edge, J.S., Du, H., Sharples, S., Kang, J., Barclay, M., Sanderson, M., 2012. Deriving and Using Future Weather Data for Building Design from UK Climate Change Projections – An Overview of the COPSE Project. Manchester University, UK. .

Li, X., Zhou, Y., Asrar, G.R., Zhu, Z., 2018. Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United States. Remote Sens. Environ. 215, 74–84. https://doi.org/10.1016/j.rse.2018.05.034.

Lin, S., Moore, N.J., Messina, J.P., DeVisser, M.H., Wu, J., 2012. Evaluation of estimating daily maximum and minimum air temperature with MODIS data in East Africa. Int. J. Appl. Earth Obs. Geoinf. 118, 128–140. https://doi.org/10.1016/j.jag.2012.01.004.

Liu, S., Su, H., Tian, J., Zhang, R., Wang, W., Wu, Y., 2016. Evaluating four remote sensing methods for estimating surface air temperature on a regional scale. J. Appl. Meteorol. Climatol. 56, 803–814. https://doi.org/10.1175/JAMC-D-16-0188.1.

London First, 2018. The Green Belt: A Place for Londoners? Retrieved December 12, 2019. Acessed from. https://www.londonfirst.co.uk/sites/default/files/documents/2018-05/Green-Belt.pdf.

Macintyre, H., Heaviside, C., Taylor, J., Picetti, R., Symond, P., Cai, X.M., Vardoulakis, S., 2018. Assessing urban population vulnerability and environmental risks across an urban area during heatwaves – implications for health protection. Sci. Total Environ. 610–611, 678–690. https://doi.org/10.1016/j.scitotenv.2017.08.062.

Mavrogianni, A., Wilkinson, P., Davies, M., Biddulph, P., Oikonomou, E., 2012. Building characteristics as determinants of propensity to high indoor summer temperatures in London dwellings. Build. Environ. 55, 117–130. https://doi.org/10.1016/j.buildenv.2011.12.003.

Met Office, 2006. MIDAS: UK Hourly Weather Observation Data. Accessed from. NCAS British Atmospheric Data Centre. http://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c.

Met Office, 2018. UK Regional Climates. https://www.metoffice.gov.uk/climate/uk/regional-climates.

Moser, G., De Martino, M., Serpico, S.B., 2015. Estimation of air surface temperature from remote sensing images and pixelwise modeling of the estimation uncertainty through support vector machines. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 8, 332–349. https://doi.org/10.1109/JSTARS.2014.2361862.

Nichol, J.E., Hang, T.P., 2012. Temporal characteristics of thermal satellite images for urban heat stress and heat island mapping. ISPRS J. Photogramm. Remote. Sens. 74, 153–162. https://doi.org/10.1016/j.isprsjprs.2012.09.007.

Nichol, J.E., Fung, W.Y., Lam, K., Wong, M.S., 2009. Urban heat island diagnosis using ASTER satellite images and 'in situ' air temperature. Atmos. Res. 94 (2), 276–284. https://doi.org/10.1016/j.atmosres.2009.06.011.

Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E., Stisen, S., 2011. Air temperature estimation with MSG-SEVIRI data: calibration and validation of the TVX algorithm for the Iberian Peninsula. Remote Sens. Environ. 115, 107–116. https://doi.org/10.1016/j.rse.2010.08.010.

Noi, P.T., Kappas, M., Degener, J., 2016. Estimating daily maximum and minimum land air surface temperature using MODIS land surface temperature data and ground truth data in Northern Vietnam. Remote Sens. 8, 1002. https://doi.org/10.3390/rs8121002.

Office for National Statistics (ONS), 2018. Population Estimates for the UK, England and Wales, Scotland and Northern Ireland: Mid-2017. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2017#nearly-12-million-uk-residents-aged-65-years-and-over.

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 (4), 1000–1025. https://doi.org/10.1080/01431161.2015.1007248.

Parmentier, B., McGill, B., Wilson, A.M., Regetz, J., Jetz, W., Guralnick, R.P., Tuanmu, M., Robinson, N., Schildhauer, M., 2014. An assessment of methods and remotely sensed derived covariates for regional predictions of 1 km daily maximum air temperature. Remote Sens. 6 (9), 8639–8670. https://doi.org/10.3390/rs6098639.

Parmentier, B., McGill, B.J., Wilson, A.M., Regetz, J., Jetz, W., Guralnick, R., Tuanmud, M.N., Schildhauer, M., 2015. Using multi-timescale methods and satellite-derived land surface temperature for the interpolation of daily maximum air temperature in Oregon. Int. J. Climatol. 35, 3862–3878. https://doi.org/10.1002/joc.4251.

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

Prihodko, L., Goward, S.N., 1997. Estimation of air temperature from remotely sensed surface observations. Remote Sens. Environ. 60 (3), 335–346. https://doi.org/10.1016/S0034-4257(96)00216-7.

Schaaf, C., Wang, Z., 2015. MCD43A3 MODIS/Terra + Aqua BRDF/Albedo Daily L3 Global - 500m V006 [Data set]. NASA EOSDIS Land Processes DAAChttps://doi.org/10.5067/MODIS/MCD43A3.006.

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., Eklundh, L., 2007. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sens. Environ. 110, 262–274. https://doi.org/10.1016/j.rse.2007.02.025.

Sun, Y.J., Wang, J.F., Zhang, R.H., Gillies, R.R., Xue, Y., Bo, Y.C., 2005. Air temperature retrieval from remote sensing data based on thermodynamics. Theor. Appl. Climatol. 80 (1), 37–48. https://doi.org/10.1016/S0034-4257(96)00216-7.

United Nations, Department of Economic and Social Affairs (UN DESA), 2018. 2018 Revision of World Urbanization Prospects. https://population.un.org/wup/. (Accessed 9 February 2019).

Vancutsem, C., Ceccato, P., Dinku, T., Connor, S.J., 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sens. Environ. 114, 449–465. https://doi.org/10.1016/j.rse.2009.10.002.

Vogt, V.J., Viau, A.A., Paquet, F., 1997. Mapping regional air temperature fields using satellite-derived surface skin temperatures. Int. J. Climatol. 14 (14), 1559–1579. https://doi.org/10.1002/(SICI)1097-0088(19971130)17:14 < 1559::AID-JOC211 > 3.0.CO;2-5.

Wan, Z., Hook, S., Hulley, G., 2015. MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAChttps://doi.org/10.5067/MODIS/MOD11A1.006.

Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: methods, applications, and trends. ISPRS J. Photogramm. Remote. Sens. 64, 335–344. https://doi.org/10.1016/j.isprsjprs.2009.03.007.

Wilby, R., Wigley, T., 1997. Downscaling general circulation model output: a review of methods and limitations. Prog. Phys. Geogr. 21, 530–548. https://doi.org/10.1177/030913339702100403.

Xu, W., Knudby, A., Ho, H.C., 2014. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. Int. J. Remote Sens. 35 (24), 8108–8121. https://doi.org/10.1080/01431161.2014.978957.

Yan, H., Zhang, J., Hou, Y., He, Y., 2009. Estimation of air temperature from MODIS data in east China. Int. J. Remote Sens. 30, 23. https://doi.org/10.1080/01431160902842375.

Yang, Y.Z., Cai, W.H., Yang, J., 2017. Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in Northeast China. Remote Sens. 9, 410. https://doi.org/10.3390/rs9050410.

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.

Zakšek, K., Schroedter-Homscheidt, M., 2009. Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. ISPRS J. Photogramm. Remote. Sens. 64, 414–421. https://doi.org/10.1016/j.isprsjprs.2009.02.006.

Zeng, L., Wardlow, B.D., Tadesse, T., Shan, J., Hayes, M.J., Li, D., Xiang, D., 2015. Estimation of daily air temperature based on MODIS land surface temperature products over the corn belt in the US. Remote Sens. 7 (1), 951–970. https://doi.org/10.3390/rs70100951.

Zhang, H., Zhang, F., Ye, M., Che, T., Zhang, G., 2016. Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. J. Geophys. Res. Atmos. 121 (19), 11,425–11,441. https://doi.org/10.1002/2016JD025154.

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.