A High-Resolution Monitoring Approach of Canopy Urban Heat Island using Random Forest Model and Multi-platform Observations

Shihan Chen[1,2] -, Yuanjian Yang [2*] , Fei Deng[1], Yanhao Zhang [2], Duanyang Liu[3,4], Chao Liu[2] and Zhiqiu Gao[2]

[1]School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China
[2]Collaborative Innovation Centre on Forecast and Evaluation of Meteorological Disasters, School of Atmospheric Physics, Nanjing University of Information Science & Technology, Nanjing 210044, China
[3]Key Laboratory of Transportation Meteorology, China Meteorological Administration, Nanjing 210008, China
[4]Nanjing Joint Institute For Atmospheric Sciences, Nanjing 210008, China

Correspondence to: Yuanjian Yang (yyj1985@nuist.edu.cn)

Abstract Due to rapid urbanization and intense human activities, the urban heat island (UHI) effect has become a more concerning climatic and environmental issue. A high spatial resolution canopy UHI monitoring method would help better understand the urban thermal environment. Taking the city of Nanjing in China as an example, we propose a method for evaluating canopy UHI intensity (CUHII) at high resolution by using remote sensing data and machine learning with a Random Forest (RF) model. Firstly, the observed environmental parameters [e.g., surface albedo, land use/land cover, impervious surface, and anthropogenic heat flux (AHF)] around densely distributed meteorological stations were extracted from satellite images. These parameters were used as independent variables to construct an RF model for predicting air temperature. The correlation coefficient between the predicted and observed air temperature in the test set was 0.73, and the average root-mean-square error was 0.72°C. Then, the spatial distribution of CUHII was evaluated at 30-m resolution based on the output of the RF model. We found that wind speed was negatively correlated with CUHII, and wind direction was strongly correlated with the CUHII offset direction. The CUHII reduced with the distance to the city ~~center~~centre, due to the de-creasing proportion of built-up areas and reduced AHF in the same direction. ~~In some cases, LST is not suitable as a predictive factor for AT.~~ The RF model framework developed for real-time monitoring and assessment of high spatial and temporal -resolution (30m and 1h) CUHII provides scientific support for studying the changes and causes of CUHII, as well as the spatial pattern of urban thermal environments.

1 Introduction

Throughout the world, cities have formed rapidly owing to population growth and people gathering in certain areas to settle and build their lives. Such urbanization brings not only economic development but also the Urban Heat Island (UHI) phenomenon (Oke, 1982; Grimmond and S.U.E., 2007; Mirzaei, 2015; Cao et al., 2016; Zhao et al., 2020). Two major types

of UHIs can be distinguished: (a) the canopy urban heat island (CUHI), and (b) the surface urban heat island (SUHI). The particular type of UHI is defined based on the height above the ground at which the phenomenon is observed and measured (Oke, 1982). The UHI effect has become an indisputable fact and brings adverse impacts on urban ecology and energy

35 consumption (Roth, 2007; Yang et al., 2019; Yang et al., 2020c; Zheng et al., 2020). UHIs amplify thermal stress, so people residing in urban areas are ~~relatively~~ more impacted during heatwave episodes (Koken et al., 2003; Estrada et al., 2017). A recent study on the global UHI predicted that about 30% of the world's population is exposed to lethal high temperatures for at least 20 days per year, and by 2100, this proportion was projected to reach 48% (Mora et al., 2017). UHIs also have the potential to impact vegetation phenology (Kabano et al., 2021), diurnal temperature range (Argüeso et al., 2014), water

40 consumption and general thermal comfort (Salata et al., 2017). Due to its negative impacts, the UHI effect has become a key challenge in achieving urban sustainability and assessing this phenomenon has attracted increasing interest over the last decade or so (Corburn, 2009; Pandey et al., 2014; Malings et al., 2017). In general, both background weather conditions (e.g., the wind vector and heatwaves) and city-specific characteristics (including the presence of urban green space, properties of built-up materials, and the intensity of human activity) influence the UHI's mean intensity and variation (Zhao

45 et al., 2014; Manoli et al., 2019). Concerning these factors, the UHI also shows significant intracity variability since urban areas are highly heterogeneous. Therefore, exploring the formation and causes of UHIs is crucial for decision-makers involved in the planning of urban developments and allocating public resources.

There are two main approaches to studying UHIs: numerical simulation and observation. Numerical simulation can reduce the need for a large number of observations and reveal mechanistic insights by investigating the impacts of cities on

50 meteorological variables (Chun and Guldmann, 2014; Zou et al., 2014; Zhang et al., 2015; Taleghani, 2016; Li et al., 2020). For instance, (Zhang et al., 2015) investigated the influence of land use/land cover (LULC) and anthropogenic heat flux (AHF) on the structure of the urban boundary layer in the Pearl River Delta region, China, through a series of numerical experiments. However, it is important to acknowledge that numerical simulation is a simplification of the real world and cannot replace actual observations. Observational studies of UHIs are arguably more robust in their findings (Hu et al., 2016;

55 Chakraborty and Lee, 2019; Dewan et al., 2021), and can mainly be categorized into the following three methods: (1) in-situ (field) measurement, (2) mobile measurements, and (3) remote sensing technology.

In-situ (field) measurements include conventional measurements from national meteorological stations which are usually located in rural areas, and high-density microclimate observations from experiments or high-density automatic sites over various underlying surfaces~~In-situ (field) measurements include meteorological measurements, which are observed by~~

60 ~~meteorological stations, and high-density observations, which are usually measured by high-spatial-resolution observation stations~~. It is easy to compare long-term series of air temperature (AT) between urban and rural stations based on meteorological observation data (Liu et al., 2006; Liu et al., 2008; Xinfa et al., 2008; Yang et al., 2012; Scott et al., 2018; Nganyiyimana et al., 2020). With the analysis of meteorological data in a long time series, the contribution and trend changes of UHI intensity (UHII) can be clearly discovered. Meanwhile, however, owing to the limitations of meteorological

65 sites in terms of their spatial representation, it is difficult to build a comprehensive understanding on the spatial distribution

of urban thermal environment parameters [such as urban canopy temperature, land surface temperature (LST) and vegetation] (Liu et al., 2008; Nganyiyimana et al., 2020). To overcome these limitations, high-density observation stations are used to explore the spatial distribution of the urban thermal environment and its relationship with the surrounding environment (Hu et al., 2016; Bassett et al., 2016; Ching et al., 2018; An et al., 2020). Deploying denser observation stations

70 or urban microclimate surveys can to some extent compensate for the limitation of a coarse spatial resolution. However, such approaches are usually unsuitable for large-scale studies due to restrictions imposed by certain natural conditions, social activities, as well as the high cost of construction and maintenance (An et al., 2020). For example, mobile transect surveys have been used in many studies (Merbitz et al., 2012; Akdemir and Tagarakis, 2014; Hankey and Marshall, 2015; László and Szegedi, 2015; Al-Ameri et al., 2016; Liu et al., 2017; Popovici et al., 2018), as they can easily obtain the distribution of

75 parameters along a designed route using only a set of equipment attached to a mobile vehicle. However, it is rather costly to obtain observations at a fine resolution, broad coverage and high synchronicity with such an approach.

To overcome these possible issues, LST data from aerial sensors and earth observing satellites are commonly employed in UHI studies, and so remote sensing data such as those from AVHRR (Roth et al., 1989; Caselles et al., 1991; Gallo et al., 1993a), Landsat (Y et al., 2007; Zhou et al., 2015; Zhao et al., 2016), MODIS (Peng et al., 2012; Zhou et al., 2015; Li et al.,

80 2017; Yang et al., 2018; Chakraborty and Lee, 2019), aerial images (Buyadi et al., 2013; Heusinkveld et al., 2014; Yu et al., 2020) and so on (Zhao et al., 2020; Gallo et al., 1993b; Qin et al., 2001; Chakraborty et al., 2020) are widely used to explain the spatial distribution of the surface UHI and its relationship with the local environment (e.g., LULC).Remote sensing data have good application prospects as they can provide fine resolution and wide data coverage at times when other ground-based observations cannot. However, due to the influence of precipitation and clouds, the retrieval of LST sometimes can be

85 challenging. In addition, each satellite remote sensing dataset has its own characteristics (Zhao et al., 2016; Chakraborty and Lee, 2019). For example, Landsat images have a high spatial resolution (30 m) that can show urban block sizes, but the temporal resolution is rather low (16 days). The MODIS LST dataset has the advantage of high temporal resolution (four times per day), but the spatial resolution is only 1 km (Yang et al., 2018Yang et al., 2019).

LST derived by satellites has become an important indicator for exploring variation characteristics of the SUHI, because

90 LST is closely related to the land cover type/structure, population density, anthropogenic heat release, etc., and it also can significantly influence surface air temperature, wind field, humidity and surface fluxes in the urban region (Ho et al., 2016;Yang et al., 2019; Li et al.,2020, 2021). However,Meanwhile, the LST can only quantify the SUHI effect, which is seriously affected by meteorological factors, e.g., cloud and evaporation. In contrast, as an important indicator reflecting the energy exchange between the atmosphere and land in the urban canopy, AT is more representative than LST. In particular,

95 AT is more related with human health and ecological changes in cities (Ho et al., 2016). While

UHI studies based on AT observed by meteorological sites suffer from limited spatial coverage, which impedes a comprehensive understanding on the influencing factors and causes of canopy UHI (CUHI). Thus, there is an urgent need to develop rapid, high spatiotemporal resolution AT and refined CUHI intensity (CUHII) estimation methods to explore the mechanisms under which anthropogenic factors (e.g., urban land-use changes, anthropogenic heat emissions, urban

100   morphology and size) and natural factors (e.g., meteorological conditions and geographical differences) influence the CUHIs of complex and diverse cities.

Therefore, in this study we: (1) based on remote sensing data, AT and wind speed data as well as other environmental information from meteorological observations, retrieved the AT data at a 30-m- spatial and one-hour-temporal resolution in the study area by using machine learning; (2) calculated the CUHII distribution based on the retrieved AT data, and further
105   explored the shape, intensity and influencing factors of the CUHI by combining local LULC, wind vector and urban morphology data.

2 Materials and Methods

2.1 Study areas

Nanjing, the capital city of Jiangsu Province in China, is located along the lower reaches of the Yangtze River and, as part of
110   the Yangtze River Delta Urban Agglomeration, has a high level of urbanization. In fact, Nanjing has been experiencing rapid urbanization since China's economic reform in 1978. According to the National Bureau of Statistics, the population in Nanjing increased from 6.13 million in 2000 to 8.34 million in 2018. In 2016, the built-up area of Nanjing expanded to 773.79 km^2, pushing the city to rank as the ninth-largest among all Chinese cities (Wang et al., 2020b). The total GDP in 2020 was about 1.48 trillion CNY, ranking ninth among all Chinese cities. Along with economic development, Nanjing's
115   UHII was observed to be 0.5°C in 2005 and is increasing at 0.109°C·(10 yr)$^{-1}$ (Xinfa et al., 2008).

2.2 Data

All of the satellite remote sensing data employed in this study are from the geospatial data cloud (https://www.gscloud.cn/), including those gathered by the Landsat 8 Operational Land Imager (OLI). OLI has nine bands, including a coastal band, blue band, green band, red band, near-infrared band, two shortwave infrared bands, a panchromatic band, and a cirrus band.
120   Due to the low temporal resolution (16 days) of the Landsat 8 OLI dataset and the vulnerability to cloud cover, data from three instances of cloudless conditions over Nanjing were selected for use in this paper—namely, 10:43 local time (LT) on 11 August 2013, 2 September 2015 and 21 July 2017. The specific band ranges and uses of Landsat 8 OLI are shown in Table S1 of the supplementary information.

High-density automatic Mmeteorological observation data, including AT (with resolution of the minimum observed scales
125   are0.5°C intervals on 11 August 2013 and 0.1°C intervals on 2 September 2015 and 21 July 2017), wind speed and wind direction, at 11:00 LT on the day closest to the satellite transit time, were selected. All weather stations in operation on those three days were included, numbering 218 totally and 63, 79 and 76 respectively (Figure 1). Figure 1 shows the 2-m AT and LULC on these three days. Compared with the LULC, the spatial patterns of AT on these three days are quite different (Figure 1).

In addition to global climate change, the influence of human activities on the CUHI cannot be ignored. Previous studies have pointed out that AHF is closely related to the change in built-up areas and population density around the stations, which reflects the fact that the effects from both anthropogenic emissions and land-use change are related to latent heat flux and sensible heat flux (Zhou et al., 2012; Yang et al., 2020b; Wang et al., 2020a; Zhang et al., 2021). Therefore, AHF was retrieved via a physical method (Bing Chen and Shi, 2012; Chen et al., 2012; Chen et al., 2014) based on 1000-m spatial resolution NOAA nighttime lighting data and with local economic development and energy consumption data, and the AHF data at the same time in Nanjing were provided by Chen et al (Bing Chen and Shi, 2012; Chen et al., 2012; Chen et al., 2014). Note that the AHF here varied annually. We expect that AHF distribution can shape the main morphology of urban thermal environment. We can not get AHF data at diurnal and seasonal scales. In future, if we got high-temporal-resolution AHF data, we will update them in the model. And lastly, the digital elevation model (DEM) data (30-m spatial resolution) used in this study are based on the second version of ASTER-GDEM, which is provided by the Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences. (http://www.gscloud.cn)

3. Random Forest model framework for air temperature retrieval

3.1. Construction of Random Forest model

The Random Forest (RF) model is a highly flexible machine learning algorithm that can analyze data with missing values or noise and has good anti-interference ability. To date, the RF model has been widely used as a feature selection tool for high-dimensional data to, for example, identify the importance of variables and predict or classify related variables. In this study, an RF model was constructed for each time's dataset to evaluate the AT using the RF package in R language.

3.1.1 Data Preparation

The process of urbanization will have a significant impact on CUHIs (Zhou et al., 2015). To comprehensively take into account the local urban environment, 18 factors were selected as independent variables (Table 1), including anthropogenic parameters (i.e., AHF), geometric parameters (distance from the city center, proportion of LULC area, altitude, longitude, latitude, slope, aspect), and physical parameters [proportion of impervious surface (IS) area, albedo, normalized difference vegetation index (NDVI), normalized difference built-up index (NDBI), green normalized difference vegetation index (gNDVI), soil-adjusted vegetation index (SAVI), normalized difference moisture index (NDMI)]. Their sources and spatial resolution are summarized in table 1. The inversion methods for these environmental variables were as follows: Based on Landsat 8 OLI satellite data, the LULC in Nanjing was divided into four broad categories (built-up, cropland, vegetation, and water body) by combining a support vector machine method and visual interpretation. The remote sensing indices were calculated using corresponding bands (Yang et al., 2012; Shi et al., 2015). The IS and surface albedo data were extracted via multi-band information (Son et al., 2017; Liang, 2001). Then, the geometric center of the built-up area was calculated as the city center, and the distances between the meteorological stations and the city center were calculated. Slope and aspect were

calculated based on the DEM data using ArcMap 10.2. The methods used for extracting the IS data and calculating the remote sensing indices and surface albedo are given in Text S1,together with the accuracy of IS and albedo. All the above data (except for DEM, aspect and slope) were extracted for each of the three years corresponding to the three selected Landsat images. Taking the data on 21 July 2017 as an example, Figure 2 shows the spatial distribution of some of the

165   environmental parameters, i.e., IS, distance from city center, AHF, and NDVI, where high spatial consistency between these parameters and the urban structure can be seen. For example, high-density built-up areas correspond closely to high AHF and low vegetation cover.

Due to advection and turbulent transport, neighborhood surroundings can affect the local temperature (Yang et al., 2012; Shi et al., 2015). Therefore, a fixed buffer zone was built surrounding the meteorological stations. Within the buffer zone of each

170   station the proportion of IS area and that of each LULC type, and the average values of surface albedo, AHF, NDVI, NDBI, SAVI, gNDVI and NDMI were calculated. Together with longitude, latitude, altitude and distance to the city center, these parameters were fed into the RF model as independent variables, with AT as the target variable. In addition, to find out the optimal size of the buffer zones for the model, we compared the model performances for different buffer zone sizes, i.e., buffer zones with a radius of 500 m, 1000 m, 2000 m and 5000 m, respectively. Figure 3 summarizes the research

175   framework of this paper.

3.1.2 Five-Fold Cross-Validation

This paper uses the coefficient of determination (R^2) and root-mean-square error (RMSE) as verification indicators. R^2 indicates the degree of fit between the predicted AT and the observed AT, and the RMSE can reflect the credibility of the prediction result.

180   The cross-validation (CV) method can be used to evaluate the performance of the RF model (Zheng et al., 2020). In this paper, we employ the five-fold CV method, in which the entire dataset is randomly divided into five subsets—each time four subsets are used to train the RF model, and the remaining one is used for validating. After constructing the model, the validation data are used to calculate the current R^2 and RMSE, and the process is repeated until each of the five folds has been used as validation data. The randomness in the process of selecting samples for modeling gives the model the

185   advantage of being robust and highly accurate. With enough decision trees, it can ensure that each sample is used as a training sample and a test sample, effectively avoiding overfitting.

3.1.3 Variable selection and model parameter setting

Since not every variable in the model makes a prominent contribution to the performance, deleting those variables that can reduce the prediction accuracy can improve the performance and simplify the model. Therefore, the number of variables

190   should be minimized on the premise of improving or not affecting the performance of the model. The contribution of each variable is judged by two indicators: the percentage increase in mean-square error (%IncMSE), and the percentage increase in node purity (IncNodePurity). Using the backward selection method, the variable with the smallest contribution is

identified and removed, and the model is re-run. These steps are then repeated until only one variable remains. The R^2 and RMSE under different combinations of variables were evaluated (Figure S1). ~~For each model, the combination of variables with the largest R^2 and smallest RMSE was selected. Using this approach, eight, seven and six variables were selected for the model on 11 August 2013, 2 September 2015 and 21 July 2017, respectively (Table 32).~~

To build an RF model, two important parameters need to be set: the number of decision trees (Ntree) and the number of variables sampled at each node (Mtry). The RF models were established with Ntree from 50 to 1200, with 50 as the step length, and Mtry from 1 to 16 respectively, with 1 as the step length to traverse all the parameters. Figure 4 presents the R^2 and RMSE values in each five-fold CV test.

The principle of parameter selection is to choose a simpler model (smaller Ntree and Mtry) under the premise of good performance. In the end, the optimal Mtry and Ntree based on the datasets on 11 August 2013, 2 September 2015 and 21 July 2017 were 7 and 200, 10 and 150, and 7 and 50, respectively.

3.2. Model testing

Table 2~~1~~ compares the performance of the RF model with different buffer sizes (500 m, 1000 m, 2000 m, and 5000 m) in the five-fold CV. The RF model based on the dataset on 11 August 2013 and 2 September 2015 within 1-km buffer zones performed best, with an R^2 and RMSE of 0.57 and 0.65°C, and 0.59 and 0.69°C, respectively. On 21 July 2017, geospatial data cloud~~meanwhile~~, the R^2 and RMSE with a 2-km buffer zone were 0.47 and 0.80°C respectively, outperforming other buffer sizes. In this study, we chose the 2-km buffer radius. As can be seen from Table 2~~1~~, on 11 August 2013 and 2 September 2015, the R^2 and RMSE with the 1-km buffer zone was very close to that from the optimal buffer size, i.e., the 2-km buffer zone, whereas on 21 July 2017 the R^2 and RMSE with the 1-km buffer zone deteriorated considerably compared to that with the 2-km buffer zone. In addition, according to recent studies, the effective range that can influence local temperature is within 2 km (Yuyu Ren, 2011; Yang et al., 2012; Shi et al., 2015). Therefore, a 2000-m buffer was finally chosen in this study.

In addition, three methods of AT modelling were also compared—two linear regressions (stepwise linear regression (Alonso and Renard, 2019; Mira et al., 2017) and geographically weighted regression (GWR) (Wang et al., 2020a; Li et al., 2021)) and one nonlinear regression (the RF model (Alonso and Renard, 2020)). A detailed description of the linear regression methods is provided in Text S2. For each model, the combination of variables with the largest R^2 and smallest RMSE was selected. Using this approach, eight, seven and six variables were selected for the model on 11 August 2013, 2 September 2015 and 21 July 2017, respectively (Table 3). Table 3~~2~~ also shows the performance of each model based on the dataset within a 2000-m buffer zone. Compared to the other methods, the RF model achieves better R^2 and RMSE, indicating its higher capability in fitting nonlinear and complex data and suitability for predicting AT (Zhu et al., 2019; Yoo et al., 2018).

3.3 Prediction accuracy of RF Models

Figure 5 compares the measured AT of the high-density automatic stations in the training set or testing set and the predicted AT of the RF model in the five-fold CV. In general, a large number of scattered points of predicted and observed AT are clustered around the 1:1 line, indicating good performance of the model. In the training set, the average R^2 and RMSE of the three models are 0.955 and 0.325°C respectively. The R^2 and RMSE using data on 11 August 2013, 2 September 2015 and 21 July 2017 are 0.948 and 0.295°C, 0.954 and 0.310°C, and 0.963 and 0.369°C, respectively, indicating high model accuracy. The result of the testing set shows that the average R^2 and RMSE are 0.535 and 0.719°C, respectively. Among them, the prediction results achieved on 21 July 2017 are slightly less accurate than those obtained on the two other days. A smaller R^2 and larger RMSE were observed on 21 July 2017 (0.802, 0.468°C) compared to 11 August 2013 (0.655, 0.563°C) and 2 September 2015 (0.700, 0.574°C). Based on existing research (Oh et al., 2020; Venter et al., 2020) and follow-up discussion (section 4.2.1), it can be concluded that the model performs best outside of the summer months, when the spatial variation in AT is low and wind velocities are high, corresponding to the model from 2 September 2015. In contrast, during the summer months, the performance of the model constructed with a high spatial variation of AT or low wind speed conditions decreases slightly, corresponding to the datasets on 21 July 2017 and 11 August 2013.

Furthermore, we used %IncMSE and IncNodePurity to determine the contribution of each variable (Table 4~~3~~) and to compare their importance. The NDVI, and the proportion of IS, vegetation and water body area all appeared in the three models, indicating that vegetation, water bodies and human activities have important and universal impacts on the AT distribution. The distance to the city center appeared in the model based on the data on 2 September 2015 and 21 July 2017, and ranked high, implying the impact of urbanization on the heat island.

The absolute error for RF prediction is defined as difference in predicted AT and observed AT at each weather station(See Figure S2).The relative error is defined as that absolute error divided by observed AT, which is shown in Figure 6. In general, the mean relative (absolute) errors by all stations are 0.07% (0.014°C), 0.04% (-0.025°C) and 0.05% (0.003°C) on 11 August 2013, 2 September 2015 and 21 July 2017, respectively. In detail, most of errors are concentrated between -0.49°C and 0.5°C over more than half of all stations for these three days (Figure S2), and more than 39.1 %/71.7%/86.3% of the total stations exhibit predictions with relative errors <1%/2%/3% (Figure 6), indicating good performance of RF models for most areas.

3.4 Model Robustness

~~Based on experiment before, we find reliable variables and parameters that can map high-spatial-solution AT in Nanjing. While the part before mainly focuses on accuracy of the RF model on a specific day, ignoring~~ To validate robustness of this RF framework and its ~~the~~ practicality ~~on~~at a long period ~~of time,~~. ~~Hence, the generality of the model in time needs to be investigated. With~~ hourly meteorological ~~observation~~ AT observations ~~data for~~ during August 2013, September 2015 and

July 2017, ~~and~~ corresponding environment variables were chosen to establish ~~be target and data and time of observed AT~~ added as independent variables, New RF model ~~were established~~.

The temperature differences in a month ~~isare~~ ~~greatlarger, which meansshowing~~ more complicated situations. For five-CV validation, scatterplot of predicted and observed air temperature are given in Figure 7, showing that ~~However, with more samples, the model achieves better performance. The mean RMSE and R² are 0.62°C and 0.99 respectively, indicating higher accuracy compared with using an hour data (mean RMSE: 0.53 and R²: 0.72°C).~~ ~~t~~The mean RMSE are 0.75°C, 0.52°C and 0.59°C, and R^2 are 0.98, 0.99 and 0.99 respectively in August 2013, September 2015 and July 2017. ~~Scatterplot of predicted and observed air temperature are shown in Figure 7. We find~~ In general, for one-month samples, the mean R^2 reached 0.986 and RMSE was 0.620°C. Note that most of the points are clustered around the 1:1 line and the performance is better than the model using ~~only~~ one-day ~~data in a day~~samples. ~~T~~Why ~~t~~he accuracy ~~using data~~ in August 2013 is the lowest is because that resolution ~~the minimum division value~~ of observed AT is 0.5°C in this month, while 0.1°C in other two ~~times~~ months~~is 0.1°C. Larger minimum division value may result in loss of information~~, so the performance is the worst among three ~~times~~ months.

4. Refined CUHII assessment in Nanjing

4.1 Refined AT and CUHII and comparison with LST distribution

After establishing the model, a 2-km buffer area was created for each 30-m-resolution pixel and the same 18 independent variables were calculated. The constructed RF model took these pixel-wise variables as input and output AT for each pixel, and hence we obtained the RF model–predicted AT map at 30-m resolution (Figure 8~~7~~). LST is also a physical manifestation of surface energy and moisture flux exchange between the atmosphere and the biosphere. ~~Some researchers believe~~Previous studies point out that there is a relationship between LST and AT (Mutiibwa et al., 2015; Benali et al., 2012), ~~so~~therefore, figure 9 shows the LST ~~(Figure 9)~~ of Nanjing in the~~se~~ ~~corresponding time~~ days, which w~~as~~ere ~~inverted~~retrieved by using Google Earth Engine~~Radiative Transfer Equation~~ to ~~compare with predicted AT~~. ~~The methods used for calculating LST are given in Text S3.~~ CUHII is an important indicator to quantify the UHI effect, which is usually defined as the difference in AT at the same level between urban and rural areas (Yang et al., 2020c; Nganyiyimana et al., 2020)~~[7,32]~~, as follows:

$$\text{CUHII} = T - T_{\text{rural}}, \quad (3)$$

where T is the predicted AT in each pixel and T_{rural} is the average AT in the reference rural area. A square area of size 10 km × 10 km was selected as the reference rural area in the northern part of Nanjing (Valmassoi and Keller, 2021). It was far from the city center and barely impacted by the UHI effect (Figure 8~~7~~). The average AT in each reference rural area was 36.0 ℃ , 27.8 ℃ and 34.7 ℃ in these three days, respectively. Then, the CUHII distribution in Nanjing was calculated according to Eq. (3) (Figure 10~~8~~).

285 Figure 8~~7~~ shows that the AT on 11 August 2013 and 21 July 2017 was higher and that the AT ranges were 35.4–37.8°C and

33.6–36.4°C, respectively. The corresponding CUHII was strong, with more than 1.5°C in the downtown area (Figure 10~~8~~).

On 2 September 2015, the AT range was 26.8–29.1°C (Figure 8~~7~~) and the CUHI was slightly weaker, with the maximum

value at only 1.3°C (Figure 10~~8~~). ~~While~~In contrast, the LST are higher, ~~does not has the same corresponding relationship,~~

~~with~~ rang~~ing~~ing from 26.1~~9~~2-44.0~~8~~1°C, 21.2~~8~~3-44.0~~4~~1°C and 23.8~~9~~9-42.0~~8~~1°C ~~respectively~~ on 11 August 2013, 2

290 September 2015 and 21 July 2017, respectively (Figure 9). The three images from different seasons and different weather

backgrounds led to significant differences in CUHII, while LST differences are marginal. On 2 September 2015, the overall

CUHI was the weakest among the three days. Consistent with a previous study (Wang et al., 2020b), the summer CUHI in

Nanjing was found to be generally stronger than that in autumn and winter. The difference between the maximal heat island

and cold island intensity on 21 July 2017 was 2.8°C, the largest among the three cases. Generally, the densely populated

295 central city area has a large proportion of IS area, large anthropogenic heat emissions, and higher AT and LST, showing an

obvious UHI phenomenon (Figures 8, 9~~7~~ and 10~~8~~). However, in urban areas with high vegetation coverage or large water

bodies, the AT and LST decreases with weakened CUHII ~~(Figures 7 and 8)~~. The AT and LST gradually decreases from the

city ~~center~~centre to the suburbs. Suburban areas, which are covered by more vegetation and water bodies, have significantly

lower AT and LST than central urban areas. At the boundary of the central city, high-AT areas and heat islands extend

300 outward along built-up areas and roads (Figures 8, 9~~7~~ and 10~~8~~).

Against different weather backgrounds, the spatial distributions of AT and CUHII exhibit heterogeneity in urban Nanjing on

different days. The high-AT area on 11 August 2013 extended from the city center to a wide range, and the extreme value of

AT was the highest (Figure 8~~7~~a), corresponding to the strongest CUHI (Figure 10~~8~~a). Combined with Figure 1, we can see

only a small range of vegetation coverage and water bodies in the central urban area, so the CUHII decreased slightly. Only

305 in the suburban water body and farmland areas were there large cold island areas, and only in this day, the distribution of

LST is corresponding to that of AT~~. Nanjing is traversed by the Yangtze River~~. On 2 September 2015, the high-AT area was

relatively small to the north of the Yangtze River. The AT on the Yangtze River was the lowest (Figure 8~~7~~b), with the

strongest cold island here (Figure 10~~8~~b). The high-AT area extended from the central city to the south, and the cold islands

in the southern water body and vegetation covered areas were not significant. On 21 July 2017, the distribution of the heat

310 island was the opposite. There was a large area of high-AT to the north of the Yangtze River, and the cooling effect of the

Yangtze River was weak (Figure 8~~7~~c). Meanwhile, the AT in the southern suburbs dropped significantly, and cold islands

widely spread in water body and cropland areas (Figure 10~~8~~c). Compared with the distribution of CUHII on 11 August 2013,

the AT over the water bodies and hills in the northeast of the central city was lower, forming a large and strong cold island

area.

315 However, note that the distribution of LST ~~from~~at thi~~s~~ese three times are similar, and they are all strongly related to urban

form and LULC (Li et al., 2021)~~, which are written in the last paragraph~~. This is B~~b~~ecause ~~–~~different ~~influencing~~ factors

cause~~d~~ different spatial distribution between LST and AT. Ground transfers heat to the air through radiation, conduction and

convection after absorbing solar energy, which is the main source of heat in the air (Hong et al., 2018; Khan et al., 2020).

While LST is directly heated by solar energy, which is more sensitive to emissivity, surface material and humidity, which related to LULC, and tendings to have a greater temperature differences for different LULC type (Janatian et al., 2016; Long et al., 2020).The LULC in these periods are similar, so the LST difference are marginal.

To further explore the intensity and coverage of the CUHI on different days, the area (km^2) occupied by different levels of CUHII on the three different days was calculated (Table 54). The CUHI area on 11 August 2013 accounted for 84.1%, and the area of the CUHII in the range of 1–1.5°C and 1.5–2°C was 1486.89 km^2 and 82.96 km^2, respectively. whereas that Oon 2 September 2015, the CUHI area and 21 July 2017 only accounted for 80.2% and the CUHII area at 0–0.5℃ accounted for 57.0%, concentrating in this range, while that at 1–1.5℃ was only 56.97 km^2 . The strongest cold island was lower than −1 ℃, and the overall CUHI effect was relatively weak.81.2 %. Comparing the proportions by area of different CUHII levels, on 2 September 2015 the CUHII area at 0–0.5℃ accounted for 57.0%, concentrating in this range, while that at 1–1.5℃ was only 56.97 km^2, which was much lower than that of the other two days. The strongest cold island was lower than −1℃, and the overall CUHI effect was relatively weak. In contrast, on 11 August 2013, the area of the CUHII in the range of 1–1.5°C and 1.5–2°C was 1486.89 km^2 and 82.96 km^2, respectively, which far exceeded that on the other two days. On 21 July 2017, the CUHI area accounted for 81.2%, and the area where the CUHII was greater than 1.5°C was only 0.14 km^2. Therefore, the distributions of UHII on the three days were quite different and need to be discussed separately.

4.2 Potential drivers of CUHII

According to previous studies, three factors—the wind vector field (He, 2018), LULC (Cao et al., 2018; Wang et al., 2020b) and the urban structure (Shahmohamadi et al., 2011; Li et al., 2020)—are the most important influencing factors of CUHIs. In this section, we explore these three drivers of CUHI in Nanjing city.

4.2.1 Relationship between CUHII and the wind vector field

The horizontal air flow has a significant impact on the intensity and shape of the CUHI (He et al., 2021). Figure 119 shows the wind vector field observed by weather stations on the three days analyzedanalysed in our study.

On 11 August 2013, the average wind speed at the stations was 0.70 m/s, most of which recorded calm wind (0–0.2 m/s) or soft wind (0.3–1.5 m/s) (Figure 119a). The main reason for this was that Nanjing was continuously controlled by the western Pacific subtropical high at this time and was therefore experiencing a continuous heat wave (Jing et al.)—conditions that are usually associated with low wind speeds, descending motion and stable weather, leading to increased CUHI strength (Figure 108a) (Wang et al., 2021). On 2 September 2015, the average wind speed was 1.53 m/s, which was a significant increase (Figure 119b). The overall northwesterly wind direction led to the CUHII being lower than that on 11 August 2013. Indeed, it has been noted in previous work that the wind direction will significantly affect the position and shape of a heat island (Bassett et al., 2016), and in the present study the northwesterly winds resulted in the CUHI extending from the built-up area to the southeast (Figure 108b) whilst weakening significantly in the northwest. On 21 July 2017, the average wind speed

350 reached 3.07 m/s, with a southwesterly wind direction (Figure 119c). The CUHI effect weakened accordingly, extending to the northeast in the downward wind, and the CUHI was significantly weakened in the southwest (Figure 108c). ~~However, whatever the wind speed is, the distribution of LST hardly changed.~~

On all three days the wind speed in the suburban areas was higher than that in the central city, and this is because there is no shelter provided by tall and dense buildings in the suburban areas, which is conducive to cooling from air convection and
355 therefore a weakening of the CUHII (Yang et al., 2020a). That said, records show that, surprisingly, the boundary-layer mean wind speed in a city can be higher than its rural counterpart. On the one hand, Nanjing is traversed by the Yangtze River, and the central city surrounds a large area of water, wherein the low surface roughness of the water is conducive to air convection. On the other hand, channeling/the Venturi effect might be an important factor. When the prevailing wind is parallel to the axis between buildings, it will be forced to enter between the buildings, resulting in higher wind pressure,
360 which increases the wind speed (Droste et al., 2018).

In order to quantify the relationship, the average CUHII and standard deviation under different wind speeds at various meteorological stations were calculated (Figure 120). On 11 August 2013, the maximum wind speed was 2 m/s, which bore no significant relationship with the CUHII (Figure 120a). On 2 September 2015 and 21 July 2017, the maximum wind speed reached 5 m/s and 6 m/s, respectively, which showed a significant negative correlation with the CUHI (Figures 120b and
365 120c). The greater the wind speed, the more significant the negative correlation.

There are two aspects concerning the influence of air convection on CUHIs. On the one hand, air convection will facilitate horizontal advection cooling between urban and rural areas, thereby weakening the CUHI (T. Brandsma, 2003). The greater the wind speed, the more significant the cooling effect (Figure 120). On the other hand, horizontal convection transfers heat from the upwind to the downwind area, weakening the upwind CUHII and strengthening the downwind CUHII (Bassett et al.,
370 2016) (Figures 108 and 119). Under different wind speeds, the synergy of these two aspects differs significantly. On 11 August 2013, the average wind speed was the smallest among the three days at only 0.7 m/s, and there was no uniform wind direction, corresponding to the strongest CUHI. The distribution of the CUHI was highly correlated with that of built-up areas (Figures 108a and 119a). On 2 September 2015 the average wind speed was 1.53 m/s. Due to the combined effect of horizontal advection cooling and heat transfer, an upwind cold island appeared and, meanwhile, the downwind area received
375 heat from the upwind area and the CUHII increased significantly (Figures 108b and 119b). On 21 July 2017 the average wind speed was 3.07 m/s, and the upwind CUHII also weakened (Figures 108c and 119c). Downwind, however, the urban heat convection was the dominant factor, which reduced the CUHII in some areas.

In contrast, CUHII distribution is in good agreement with LST distribution on 11 August 2013, while large pattern difference during other two days (Figures 9 and 11). This is because clam wind on 11 August 2013 can not induce horizontal ~~However, there is no sufficient evidence show that air con~~advection of urban heat, ~~can affect LST, and from Figure 9 and Figure 11, air convection can hardly influence LST. t~~Therefore, spatial distributions of LST and AT~~-~~ are well matched in this day~~have a good corresponding relationship when there is no wind~~. While under large wind conditions (e.g., larger wind speeds on both

2 September 2015 and 21 July 2017), there are obvious urban heat island advection (Bassett, et al., 2016), resulting in different patterns between CUHII and SUHI during these two days (Figures 9 and 11).

4.2.2 Relationship between CUHII, ~~LST~~ and LULC

LULC also has a significant impact on CUHII (Wang et al., 2020) and LST (Li et al., 2021). The average values and standard deviation of CUHII were calculated for each LULC type on the three days (Figure 13~~1~~). On 11 August 2013, the CUHII in the built-up area was the strongest, exceeding 1.1°C, and in the water body areas it was the weakest at only 0.22°C (Figure 13~~1~~a). On 21 July 2017, the CUHII in the built-up area was the strongest at 0.62°C, and in the vegetation areas it was the weakest at 0.24°C (Figure 13~~1~~b). The CUHII on these two days was highest in the built-up area, followed by cropland, and then water bodies and vegetation. On 2 September 2015 the CUHII in the built-up area was the strongest at 0.32°C, while it was the weakest at −0.06°C in the water body areas (Figure 13~~1~~c).

Different LULC types have different effects on AT owing to their own intrinsic physical properties, mainly reflected in three aspects:

(1) Due to the good thermal conductivity and small specific heat capacity of the surface material in the built-up area, the ability to absorb shortwave radiation during the day is stronger than that of other land uses. The LST is significantly higher than that of the suburbs, and therefore the atmosphere is easily heated (Hong et al., 2018).

(2) Due to sufficient water availability in cropland and vegetation-covered areas, evaporation will increase the latent heat flux and cooling effect (Zhao et al., 2020; Zheng et al., 2018). In contrast, the surface humidity of the built-up area is low, with low corresponding latent heat flux. The difference in latent heat flux will increase the difference in LST and AT between urban and rural areas. The latent heat flux of the water bodies is the largest, and the cooling effect is the most obvious.

(3) There is a significant correlation between LULC and wind speed (Chen et al., 2020). Areas with tall buildings in built-up areas have high surface roughness and low wind speed, whereas water bodies have low surface roughness and high wind speed. The surface roughness of vegetation-covered areas and cropland is somewhere between. The air convection will increase the sensible heat flux and reduce the AT (section 4.2.1). Therefore, LULC and air convection will jointly enhance or weaken the CUHII.

On 11 August 2013, the average wind speed and the difference in wind speed between different LULC types were small, and so was the difference in sensible heat flux. The difference in radiation and sensible heat flux was the main factor. On 21 July 2017, the average wind speed was the highest, and the synergy in the three aspects led to the CUHII over different LULC types being highest in the built-up area, followed by cropland, vegetation, and then water bodies. On 2 September 2015, the CUHII was highest in the built-up areas, followed by vegetation, cropland, and then water bodies. This was due to the influence of low wind speeds, which would have produced heat transfer and made the CUHII shift from the built-up area to other LULC types (section 4.2.1).

4.2.3 Relationship between CUHI and urban structure

Human activities and urbanization have a significant impact on the spatial distribution of UHI (Shahmohamadi et al., 2011; Li et al., 2020). To explore this influence, concentric rings with various radii (5 km, 10 km, 15 km, ..., 40 km) were created surrounding the city center. Within each ring, the average values and error ranges of AHF and CUHII, along with the average proportion of built-up area, were calculated. Figure 142 shows that the CUHII, AHF and proportion of built-up area all significantly decrease with increasing distance to the city center.

From a longitudinal perspective, the AHF and the proportion of built-up areas both increased year by year. The built-up areas of Nanjing on the three days were 982.78 km², 1076.19 km² and 1220.36 km², respectively. The proportion of built-up areas beyond 20 km to the city center increased, especially within the range of 20–25 km. The AHF also showed the same trend, which within the range of 20–25 km even exceeded that in the range of 15–20 km on 2 September 2015 and 21 July 2017. This shows that built-up areas and human influence were spreading from the city center to the surrounding areas during this period. However, the intensity and range of the CUHI did not increase with this trend, because the wind field and weather background have a stronger influence on CUHI than urbanization (Hong et al., 2018; Zong et al., 2021).

5. Discussion

Based on the RF model and combined with local environment and background weather data, the pattern and causes of CUHIs can be analyzed in detail. On 11 August 2013, Nanjing experienced a heat wave, with almost no horizontal convection of air (Figure 119a). In dry areas, such as built-up areas, the latent heat flux remained unchanged, but the high reflectivity of the surface raised the AT. In the heat wave period, the higher AT increased the latent heat flux in rural areas (Khan et al., 2020). For example, vegetation and water bodies alleviated the increase in AT in rural areas. This combined effect exacerbated the difference in AT between the urban and rural areas, making the overall CUHI the strongest (Nganyiyimana et al., 2020; Meili et al., 2021). In Figure 108a, it can be seen that the cooling efficiency of vegetation in the urban area was not high and the coverage of the cooling area was small. This is because the stomata of leaves would have been closed under high AT and dry weather, resulting in reduced evapotranspiration and increased AT (Manoli et al., 2019). On 2 September 2015, northwesterly winds prevailed (Figure 119b), and there was abundant water vapor over the hills of northeast Nanjing and over the Yangtze River. The increase in latent heat flux and horizontal convection cooling lowered the CUHII. Cold islands even appeared to the north of the Yangtze River. The CUHII in the southeast direction was strong (Figure 108b), which was mainly affected by the heat transport of the prevailing winds (Chuanyan et al., 2005), causing the CUHI to shift toward the downwind area. On 21 July 2017, southwesterly winds prevailed in Nanjing, with high wind speed, decreasing the CUHII in the upwind region (Figures 108c and 119c). However, there were large areas of vegetation coverage in the range of 10–20 km in the downwind region, which was affected by the combined effects of land use and horizontal advection cooling leading to lower CUHII than that of 20–30 km. This also confirms the conclusion (Bassett et al., 2016)

that the upwind horizontal advection cooling has the strongest correlation with the weakening of the CUHI effect, and that the downwind region is affected by the wind speed.

There are four main methods for retrieving AT for CUHII assessment:

(1) Statistical methods (Prihodko and Goward, 1997; Alonso and Renard, 2020; Li et al., 2021). Statistical models of environmental factors and temperature are established to evaluate the AT, such as multiple linear regression models, partial least-squares regression and GWR. In previous study (Alonso and Renard, 2020), two methods of AT prediction (namely, stepwise linear regression and GWR) were compared with the RF model. The RF model has the highest accuracy and effectively avoids the problem of autocorrelation by filtering variables, which is consistent with previous work (Yoo et al., 2018; Zhu et al., 2019) and our present work. While conventional statistical methods, in addition, cannot effectively solve nonlinear problems (Oh et al., 2020).

(2) Temperature–vegetation index method (VTX) (Goward, 1997; Stisen et al., 2007; Vancutsem et al., 2010). This refers to inversion using the relationship between AT, LST and vegetation index under the premise that the temperature of a dense vegetation canopy is similar to the AT. While VTX only indicates the relationships between underlying surface, LST and AT. In fact, there are many factors that can affect AT, e.g., anthropogenic heat, altitude, and distance to city. Ignoring these factors, therefore, the accuracy of VTX method was low (Stisen et al., 2007)). In contrast, our RF model input multi-variables, including more affecting AT factors.

(3) Machine learning methods (Venter et al., 2020). Predictions are made by establishing models of various variables and AT, such as RF models or neural networks. Compared with other machine learning methods such as neural networks (Astsatryan et al., 2021), the RF model has better noise immunity and is suitable for small sample sizes in this study. Other machine learning methods usually require a lot of data with little noise, so the data cleaning before modeling will take more time. In future, we would like to compare different machine learning methods to come up with a consistently well-perform model, e.g., SVM and ANN. We will also use stacking ensemble strategy to combine the advantages of different models and get the best prediction results

(4) Physical model methods. This category mainly constitutes the energy balance method (Yang et al., 2018), which refers to the study of AT inversion using the principle of energy balance. Physical model approach is relatively complex, and the performance is highly dependent on the understanding of the mechanism affecting AT, which can only address specific problems. While the RF framework in this paper is relatively simple, comprehensive, and suitable for different weather backgrounds.

Moreover, Tthe RF prediction framework proposed in this work can not only can dynamically predict CUHII in detail and high frequency within highly heterogeneous cities, but can also can be built against different weather backgrounds, mainly because the environmental parameters entered into the model are relatively stable within a certain period (such as the same month or season). As long as the environmental parameters are acquired once, they can be combined with the AT data in real time to establish the RF model, and the spatial distribution characteristics of CUHII with high temporal and spatial resolution

480 can be obtained. For instance, we randomly predicted the 30-m-resolution AT and spatial distribution of CUHII (Figure 153)
with the wind vector field (Figure S3) during the heat wave period of 12–14 August 2012, thereby supporting those involved
in making decisions with respect to urban climate, urban planning and urban energy consumption. Particularly, Tthe
potential that our proposed model can be used cross a short period as most of the environmental parameters fed to the model
probably can remain stable for some time, e.g., one month or even longer (See Figure S4).

485 In addition, Due to changes in local weather conditions (e.g., precipitation and cloud cover) , however, satellite-based LST
samples are various and LST is usually dynamical in one month, leading to uncertainties in predicting AT, therefore, LST is
not suitable to be as an input variable for our present model of CUHIIunder some circumstances, LST cannot be derieved
from used as a predictive factor for AT, such as when there is wind or cloud cover. It is worth mentioning that there are
many factors that affect the CUHII. In addition toExcept for human activities and LULC, the background weather conditions
490 (such as heat waves, air pollution, atmospheric circulation and cloud cover) are also extremely important (Bassett et al., 2016;
Yang et al., 2020a; Khan et al., 2020), which should be introduced to improve the RF model of CUHII.

6. Conclusions

Taking Nanjing as an example and using remote sensing data with data from local weather stations, parameters to
495 characterize the urban environment were constructed—namely, anthropogenic parameters (i.e., AHF), geometric parameters
(distance from city center, proportions of LULC types by area, altitude, and latitude and longitude, slope and aspect), and
physical parameters (proportion of IS, surface albedo, NDVI, NDBI, SAVI, gNDVI and NDMI). A 2-km buffer zone was
created around the meteorological stations, and the observed environmental parameters were extracted. A refined assessment
framework of CUHII was then established by using Random Forest Model with observed AT and environmental variables.
500 Results showed that the correlation coefficient between the predicted and observed AT was 0.731, and the average RMSE
was 0.719°C, indicating accurate results. Based on one--month datasamples, the R^2 reached 0.986 and RMSE was 0.620°C.
Finally, the high-spatial-resolution (30m) CUHII distribution was analyzed. It was found that the shape of the CUHII was
highly correlated with the spatial distribution of AHF and built-up area under calm wind conditions. Under the prevailing
wind conditions, the CUHII should be discussed separately in upwind and downwind areas divided by the central city. In the
505 upwind area, there was a significant negative correlation between the wind speed and CUHII. The higher the wind speed, the
more obvious the negative correlation. In the downwind area, horizontal convection cooling was found to be the leading
factor under high wind speed weather, and heat transfer was the leading factor under low wind speed weather. The combined
effects of built-up areas, heatwaves, and human factors can strengthen the CUHII, while the vegetation canopy and water
bodies will weaken it. Vegetation and water bodies in the central urban area were found to have a significant cooling effect,
510 providing a reference for urban development. With increasing distance from the city center, the CUHII decreased sharply. In
some weather condition with wind or cloud, LST is not suitable for predict factor of AT.

In general, overlapping the refined CUHII with local environmental variables and weather conditions helps to explore the causes of CUHIs in more detail, instead of being limited to the location of meteorological sites and frequent changes in various types of weather. The new 30-m resolution CUHII evaluation framework developed in this study has strong portability, and important practical value. Our findings are helpful towards improving our understanding of the relationship between human activities and regional climate change, which can provide important guidance for urban development planning and allocation of public resources in the context of global warming and rapid urbanization.

Code availability

The model in this paper is based on the random Forest data package in the R language, and our implementation and analysis code are available upon request to the corresponding author (yyj1985@nuist.edu.cn).

Data Availability

Landsat 8 OLI datasets (http://www.gscloud.cn/sources/index?pid=263&rootid=1&label=Landsat8&sort=priority&page=1 (access on 10 April 2021)) were used to retrieve IS area, LULC and NDVI. nighttime light satellite datasets (http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html (access on 10 April 2021)) were used to retrieve anthropogenic heat flux (AHF), surface meteorological observations were collected from the China Meteorological Data Service Center (http://data.cma.cn/en(access on 10 April 2021)), DEM were achieved from geospatial data

Supplement

The following are available online, Text S1: Specific inversion steps of related environmental variables. Table S1: band ranges and the main use of Landsat 8 OLI. Figures S1: The performance of the RF models under different variable combinations. Figure S2. the predicted error of the air temperature by random forest: (a) 11 August 2013; (b) 2 September 2015; (c) 21 July 2017. Figure S3: Spatial distribution of air temperature and wind vector field in Nanjing and the reference rural area during 12-14 August, 2013.

Author contributions

Conceptualization, Y.Y.; methodology, S.C. and Y.Y.; Data Curation: D.L. and S.C.; software, S.C.; validation, S.C. and Y.Y.; formal analysis, Y.Z., S.C., D.L., C.L. and Y.Y.; writing—original draft preparation, S.C.; writing—review and editing, F.D, S.C., Y.Z., C.L., D.L. and Y.Y.; visualization, S.C. and Y.Z.; supervision, Y.Y. and F.D. ; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

540 **Financial support**

This research was supported by the National Key Research and Development Program of China (grant no. 2018YFC1506502), the National Natural Science Foundation of China (grant no. 42175098 and 42061134009), and the University Student Innovation Training Project of Nanjing University of Information Science and Technology (grant no. 201910300283).

545 **Acknowledgements**

References

Akdemir, S. and Tagarakis, A.: Investigation of Spatial Variability of Air Temperature, Humidity and Velocity in Cold Stores by Using Management Zone Analysis, Tarim Bilimleri Dergisi, 20, 175-186, 10.1501/Tarimbil_0000001277, 2014.

550 Al-ameri, A. Q., Lamit, H., Ossen, D., and Raja Shahminan, R. N.: Urban Heat Island and Thermal Comfort Conditions at Micro-climate Scale in a Tropical Planned City, Energy and Buildings, 133, 10.1016/j.enbuild.2016.10.006, 2016.

Alonso, L. and Renard, F.: Integrating Satellite-Derived Data as Spatial Predictors in Multiple Regression Models to Enhance the Knowledge of Air Temperature Patterns, Urban Science, 3, 10.3390/urbansci3040101, 2019.

Alonso, L. and Renard, F.: A New Approach for Understanding Urban Microclimate by Integrating Complementary Predictors at Different

555 Scales in Regression and Machine Learning Models, Remote Sensing, 12, 10.3390/rs12152434, 2020.

An, N., Dou, J., González-Cruz, J. E., Bornstein, R. D., Miao, S., and Li, L.: An Observational Case Study of Synergies between an Intense Heat Wave and the Urban Heat Island in Beijing, Journal of Applied Meteorology and Climatology, 59, 605-620, 10.1175/jamc-d-19-0125.1, 2020.

Argüeso, D., Evans, J. P., Fita, L., and Bormann, K. J.: Temperature response to future urbanization and climate change, Climate

560 Dynamics, 42, 2183-2199, 10.1007/s00382-013-1789-6, 2014.

Astsatryan, H., Grigoryan, H., Poghosyan, A., Abrahamyan, R., Asmaryan, S., Muradyan, V., Tepanosyan, G., Guigoz, Y., and Giuliani, G.: Air temperature forecasting using artificial neural network for Ararat valley, Earth Science Informatics, 14, 1-12, 10.1007/s12145-021-00583-9, 2021.

Bassett, R., Cai, X., Chapman, L., Heaviside, C., Thornes, J. E., Muller, C. L., Young, D. T., and Warren, E. L.: Observations of urban

565 heat island advection from a high-density monitoring network, Quarterly Journal of the Royal Meteorological Society, 142, 2434-2441, 10.1002/qj.2836, 2016.

Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., and Santos, A.: Estimating air surface temperature in Portugal using MODIS LST data, Remote Sensing of Environment, 124, 108-121, 10.1016/j.rse.2012.04.024, 2012.

Bing Chen and Shi, G.-Y.: Estimation of the Distribution of Global Anthropogenic Heat Flux, Atmospheric and Oceanic Science Letters,

570 5, 108-112, 10.1080/16742834.2012.11446974, 2012.

Buyadi, S. N. A., Mohd, W. M. N. W., and Misni, A.: Green Spaces Growth Impact on the Urban Microclimate, Procedia - Social and Behavioral Sciences, 105, 547-557, https://doi.org/10.1016/j.sbspro.2013.11.058, 2013.

Cao, C., Lee, X., Liu, S., Schultz, N., Xiao, W., Zhang, M., and Zhao, L.: Urban heat islands in China enhanced by haze pollution, Nat Commun, 7, 12509, 10.1038/ncomms12509, 2016.

575 Cao, Q., Yu, D., Georgescu, M., Wu, J., and Wang, W.: Impacts of future urban expansion on summer climate and heat-related human health in eastern China, Environ Int, 112, 134-146, 10.1016/j.envint.2017.12.027, 2018.

Caselles, V., López García, M. J., Meliá, J., and Pérez Cueva, A. J.: Analysis of the heat-island effect of the city of Valencia, Spain, through air temperature transects and NOAA satellite data, Theoretical and Applied Climatology, 43, 195-203, 10.1007/BF00867455, 1991.

Chakraborty, T. and Lee, X.: 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, 10.1016/j.jag.2018.09.015, 2019.

Chakraborty, T., Hsu, A., Manya, D., and Sheriff, G.: A spatially explicit surface urban heat island database for the United States: Characterization, uncertainties, and possible applications, ISPRS Journal of Photogrammetry and Remote Sensing, 168, 74-88, 10.1016/j.isprsjprs.2020.07.021, 2020.

Chen, B., Dong, L., Shi, G., Li, L.-J., and Chen, L.-F.: Anthropogenic Heat Release: Estimation of Global Distribution and Possible Climate Effect, Journal of the Meteorological Society of Japan. Ser. II, 92A, 157-165, 10.2151/jmsj.2014-A10, 2014.

Chen, B., Shi, G., Wang, B., Zhao, J., and Tan, S.: Estimation of the anthropogenic heat release distribution in China from 1992 to 2009, Acta Meteorologica Sinica, 26, 507-515, 10.1007/s13351-012-0409-y, 2012.

Chen, X., Jeong, S., Park, H., Kim, J., and Park, C.-R.: Urbanization has stronger impacts than regional climate change on wind stilling: a lesson from South Korea, Environmental Research Letters, 15, 10.1088/1748-9326/ab7e51, 2020.

Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X., Ren, C., Brousse, O., Martilli, A., Neophytou, M., Mouzourides, P., Stewart, I., Hanna, A., Ng, E., Foley, M., Alexander, P., Aliaga, D., Niyogi, D., Shreevastava, A., Bhalachandran, P., Masson, V., Hidalgo, J., Fung, J., Andrade, M., Baklanov, A., Dai, W., Milcinski, G., Demuzere, M., Brunsell, N., Pesaresi, M., Miao, S., Mu, Q., Chen, F., and Theeuwes, N.: WUDAPT: An Urban Weather, Climate, and Environmental Modeling Infrastructure for the Anthropocene, Bulletin of the American Meteorological Society, 99, 1907-1924, 10.1175/bams-d-16-0236.1, 2018.

Chuanyan, Z., Zhongren, N., and Guodong, C.: Methods for modelling of temporal and spatial distribution of air temperature at landscape scale in the southern Qilian mountains, China, Ecological Modelling, 189, 209-220, 10.1016/j.ecolmodel.2005.03.016, 2005.

Chun, B. and Guldmann, J. M.: Spatial statistical analysis and simulation of the urban heat island in high-density central cities, Landscape and Urban Planning, 125, 76-88, 10.1016/j.landurbplan.2014.01.016, 2014.

Corburn, J.: Cities, Climate Change and Urban Heat Island Mitigation: Localising Global Environmental Science, Urban Studies, 46, 413-427, 10.1177/0042098008099361, 2009.

Dewan, A., Kiselev, G., Botje, D., Mahmud, G. I., Bhuian, M. H., and Hassan, Q. K.: Surface urban heat island intensity in five major cities of Bangladesh: Patterns, drivers and trends, Sustainable Cities and Society, 71, 10.1016/j.scs.2021.102926, 2021.

Droste, A. M., Steeneveld, G. J., and Holtslag, A. A. M.: Introducing the urban wind island effect, Environmental Research Letters, 13, 10.1088/1748-9326/aad8ef, 2018.

Estrada, F., Botzen, W. J. W., and Tol, R. S. J.: A global economic assessment of city policies to reduce climate change impacts, Nature Climate Change, 7, 403-406, 10.1038/nclimate3301, 2017.

Gallo, K. P., McNab, A. L., Karl, T. R., Brown, J. F., Hood, J. J., and Tarpley, J. D.: The use of NOAA AVHRR data for assessment of the urban heat sland effect, Journal of Applied Meteorology and Climatology, 32, 899-908, 10.1175/1520-0450(1993)032<0899:TUONAD>2.0.CO;2, 1993a.

Gallo, K. P., McNab, A. L., Karl, T. R., Brown, J. F., Hood, J. J., and Tarpley, J. D.: The use of a vegetation index for assessment of the urban heat island effect, International Journal of Remote Sensing, 14, 2223-2230, 10.1080/01431169308954031, 1993b.

Goward, L. P. S. N.: Estimation of Air Temperature from Remotely Sensed Surface Observations, 1997.

Grimmond and S.U.E.: Urbanization and global environmental change: local effects of urban warming, Geographical Journal, 173, 83-88, 2007.

Hankey, S. and Marshall, J. D.: Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM2.5, Particle Size) Using Mobile Monitoring, Environmental Science & Technology, 49, 9194-9202, 10.1021/acs.est.5b01209, 2015.

He, B.-J.: Potentials of meteorological characteristics and synoptic conditions to mitigate urban heat island effects, Urban Climate, 24, 26-33, 10.1016/j.uclim.2018.01.004, 2018.

He, B. J., Wang, J., Liu, H., and Ulpiani, G.: Localized synergies between heat waves and urban heat islands: Implications on human thermal comfort and urban heat management, Environ Res, 193, 110584, 10.1016/j.envres.2020.110584, 2021.

Heusinkveld, B. G., Steeneveld, G. J., van Hove, L. W. A., Jacobs, C. M. J., and Holtslag, A. A. M.: Spatial variability of the Rotterdam urban heat island as influenced by urban land use, Journal of Geophysical Research: Atmospheres, 119, 677-692, 10.1002/2012jd019399, 2014.

Ho, H. C., Knudby, A., Xu, Y., Hodul, M., and Aminipouri, M.: A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area, Science of The Total Environment, 544, 929-938, 10.1016/j.scitotenv.2015.12.021, 2016.

Hong, J.-S., Yeh, S.-W., and Seo, K.-H.: Diagnosing Physical Mechanisms Leading to Pure Heat Waves Versus Pure Tropical Nights Over the Korean Peninsula, Journal of Geophysical Research: Atmospheres, 10.1029/2018jd028360, 2018.

Hu, X.-M., Xue, M., Klein, P. M., Illston, B. G., and Chen, S.: Analysis of Urban Effects in Oklahoma City using a Dense Surface Observing Network, Journal of Applied Meteorology and Climatology, 55, 723-741, 10.1175/jamc-d-15-0206.1, 2016.

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., and Ghazanfari, S.: A statistical framework for estimating air temperature using

MODIS

land surface temperature data, International Journal of Climatology, 37, 1181-1194, 10.1002/joc.4766, 2016.

Jing, S. U. N., Yueqin, S. H. I., Miao, C. A. I., Yuquan, Z., and Lin, T.: Analysis on Cloud Structure Forecast and Seeding Conditions of 3 Types of Cloud System in South China, 气象英文版, 41, 1356-1366,

Kabano, P., Lindley, S., and Harris, A.: Evidence of urban heat island impacts on the vegetation growing season length in a tropical city, Landscape and Urban Planning, 206, 103989, https://doi.org/10.1016/j.landurbplan.2020.103989, 2021.

Khan, H. S., Paolini, R., Santamouris, M., and Caccetta, P.: Exploring the Synergies between Urban Overheating and Heatwaves (HWs) in Western Sydney, Energies, 13, 10.3390/en13020470, 2020.

Koken, P. J., Piver, W. T., Ye, F., Elixhauser, A., Olsen, L. M., and Portier, C. J.: Temperature, air pollution, and hospitalization for cardiovascular diseases among elderly people in Denver, Environ Health Perspect, 111, 1312-1317, 10.1289/ehp.5957, 2003.

László, E. and Szegedi, S.: A multivariate linear regression model of mean maximum urban heat island: A case study of Beregszász (Berehove), Ukraine, Idojaras (Budapest, 1905), 199, 409-423, 2015.

Li, X., Zhou, Y., Asrar, G. R., Imhoff, M., and Li, X.: The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States, Science of The Total Environment, 605-606, 426-435, https://doi.org/10.1016/j.scitotenv.2017.06.229, 2017.

Li, Y., Schubert, S., Kropp, J. P., and Rybski, D.: On the influence of density and morphology on the Urban Heat Island intensity, Nat Commun, 11, 2647, 10.1038/s41467-020-16461-9, 2020.

Li, Y., Zhou, B., Glockmann, M., Kropp, J. P., and Rybski, D.: Context sensitivity of surface urban heat island at the local and regional scales, Sustainable Cities and Society, 74, 103146, https://doi.org/10.1016/j.scs.2021.103146, 2021.

Liang, S.: Narrowband to broadband conversions of land surface albedo I: Algorithms, Remote Sensing of Environment, 76, 213-238, https://doi.org/10.1016/S0034-4257(00)00205-4, 2001.

Liu, L., Lin, Y., Liu, J., Wang, L., Wang, D., Shui, T., Chen, X., and Wu, Q.: Analysis of local-scale urban heat island characteristics using an integrated method of mobile measurement and GIS-based spatial interpolation, Building and Environment, 117, 191-207, 10.1016/j.buildenv.2017.03.013, 2017.

Liu, W., Ji, C., Zhong, J., Jiang, X., and Zheng, Z.: Temporal characteristics of the Beijing urban heat island, Theoretical and Applied Climatology, 87, 213-221, 10.1007/s00704-005-0192-6, 2006.

Liu, X., Guo, J., Zhang, A., Zhou, J., Chu, Z., Zhou, Y., and Ren, G.: Urbanization Effects on Observed Surface Air Temperature Trends in North China, Journal of Climate, 21, 1333-1348, 10.1175/2007jcli1348.1, 2008.

Long, D., Yan, L., Bai, L., Zhang, C., Li, X., Lei, H., Yang, H., Tian, F., Zeng, C., Meng, X., and Shi, C.: Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach, Remote Sensing of Environment, 246, 10.1016/j.rse.2020.111863, 2020.

Malings, C., Pozzi, M., Klima, K., Bergés, M., Bou-Zeid, E., and Ramamurthy, P.: Surface heat assessment for developed environments: Probabilistic urban temperature modeling, Computers, Environment and Urban Systems, 66, 53-64, https://doi.org/10.1016/j.compenvurbsys.2017.07.006, 2017.

Manoli, G., Fatichi, S., Schlapfer, M., Yu, K., Crowther, T. W., Meili, N., Burlando, P., Katul, G. G., and Bou-Zeid, E.: Magnitude of urban heat islands largely explained by climate and population, Nature, 573, 55-60, 10.1038/s41586-019-1512-9, 2019.

Meili, N., Manoli, G., Burlando, P., Carmeliet, J., Chow, W. T. L., Coutts, A. M., Roth, M., Velasco, E., Vivoni, E. R., and Fatichi, S.: Tree effects on urban microclimate: Diurnal, seasonal, and climatic temperature differences explained by separating radiation, evapotranspiration, and roughness effects, Urban Forestry & Urban Greening, 58, 10.1016/j.ufug.2020.126970, 2021.

Merbitz, H., Fritz, S., and Schneider, C.: Mobile measurements and regression modeling of the spatial particulate matter variability in an urban area, Science of The Total Environment, 438, 389-403, https://doi.org/10.1016/j.scitotenv.2012.08.049, 2012.

Mira, M., Ninyerola, M., Batalla, M., Pesquer, L., and Pons, X.: Improving Mean Minimum and Maximum Month-to-Month Air Temperature Surfaces Using Satellite-Derived Land Surface Temperature, Remote Sensing, 9, 10.3390/rs9121313, 2017.

Mirzaei, P. A.: Recent challenges in modeling of urban heat island, Sustainable Cities and Society, 19, 200-206, 10.1016/j.scs.2015.04.001, 2015.

Mora, C., Dousset, B., Caldwell, I. R., Powell, F. E., Geronimo, R. C., Bielecki, Coral R., Counsell, C. W. W., Dietrich, B. S., Johnston, E. T., Louis, L. V., Lucas, M. P., McKenzie, M. M., Shea, A. G., Tseng, H., Giambelluca, T. W., Leon, L. R., Hawkins, E., and Trauernicht, C.: Global risk of deadly heat, Nature Climate Change, 7, 501-506, 10.1038/nclimate3322, 2017.

Mutiibwa, D., Strachan, S., and Albright, T.: Land Surface Temperature and Surface Air Temperature in Complex Terrain, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8, 4762-4774, 10.1109/jstars.2015.2468594, 2015.

Nganyiyimana, J. N. J., Kim, I., Geun, M. S., and YunID, Y.: Synergies between urban heat island and heat waves in Seoul: The role of wind speed and land use characteristics, 10.1371/journal.pone.0243571, 2020.

Oh, J. W., Ngarambe, J., Duhirwe, P. N., Yun, G. Y., and Santamouris, M.: Using deep-learning to forecast the magnitude and characteristics of urban heat island in Seoul Korea, Sci Rep, 10, 3559, 10.1038/s41598-020-60632-z, 2020.

Oke, T. R.: The energetic basis of the urban heat island, Royal Meteorological Society, 108, 1-24, https://doi.org/10.1002/qj.49710845502, 1982.

Pandey, A. K., Singh, S., Berwal, S., Kumar, D., Pandey, P., Prakash, A., Lodhi, N., Maithani, S., Jain, V. K., and Kumar, K.: Spatio – temporal variations of urban heat island over Delhi, Urban Climate, 10, 119-133, https://doi.org/10.1016/j.uclim.2014.10.005, 2014.

Peng, S., Piao, S., Ciais, P., Friedlingstein, P., Ottle, C., Bréon, F.-M., Nan, H., Zhou, L., and Myneni, R. B.: Surface Urban Heat Island Across 419 Global Big Cities, Environmental Science & Technology, 46, 696-703, 10.1021/es2030438, 2012.

Popovici, I., Goloub, P., Podvin, T., Blarel, L., Loisil, R., Unga, F., Mortier, A., Deroo, C., Victori, S., Ducos, F., Torres, B., Delegove, C., Choël, M., Pujol-Söhne, N., and Pietras, C.: Description and applications of a mobile system performing on-road aerosol remote sensing and in situ measurements, Atmospheric Measurement Techniques, 11, 4671-4691, 10.5194/amt-11-4671-2018, 2018.

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

Qin, Z., Dall'Olmo, G., Karnieli, A., and Berliner, P.: Derivation of split window algorithm and its sensitivity analysis for retrieving land surface temperature from NOAA-advanced very high resolution radiometer data, Journal of Geophysical Research: Atmospheres, 106, 22655-22670, https://doi.org/10.1029/2000JD900452, 2001.

Roth, M.: Review of urban climate research in (sub)tropical regions, International Journal of Climatology, 27, 1859-1873, 10.1002/joc.1591, 2007.

Roth, M., Oke, T. R., and Emery, W. J.: Satellite-derived urban heat islands from three coastal cities and the utilization of such data in urban climatology, International Journal of Remote Sensing, 10, 1699-1720, 10.1080/01431168908904002, 1989.

Salata, F., Golasi, I., Petitti, D., de Lieto Vollaro, E., Coppi, M., and de Lieto Vollaro, A.: Relating microclimate, human thermal comfort and health during heat waves: An analysis of heat island mitigation strategies through a case study in an urban outdoor environment, Sustainable Cities and Society, 30, 79-96, https://doi.org/10.1016/j.scs.2017.01.006, 2017.

Scott, A. A., Waugh, D. W., and Zaitchik, B. F.: Reduced Urban Heat Island intensity under warmer conditions, Environ Res Lett, 13, 10.1088/1748-9326/aabd6c, 2018.

Shahmohamadi, P., Che-Ani, A. I., Maulud, K. N. A., Tawil, N. M., and Abdullah, N. A. G.: The Impact of Anthropogenic Heat on Formation of Urban Heat Island and Energy Consumption Balance, Urban Studies Research, 2011, 1-9, 10.1155/2011/497524, 2011.

Shi, T., Huang, Y., Wang, H., Shi, C.-E., and Yang, Y.-J.: Influence of urbanization on the thermal environment of meteorological station: Satellite-observed evidence, Advances in Climate Change Research, 6, 7-15, 10.1016/j.accre.2015.07.001, 2015.

Son, N.-T., Chen, C.-F., Chen, C.-R., Thanh, B.-X., and Vuong, T.-H.: Assessment of urbanization and urban heat islands in Ho Chi Minh City, Vietnam using Landsat data, Sustainable Cities and Society, 30, 150-161, 10.1016/j.scs.2017.01.009, 2017.

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

T. BRANDSMA, G. P. K. a. H. R. A. W.: EMPIRICAL ESTIMATION OF THE EFFECT OF URBAN HEAT ADVECTION ON THE TEMPERATURE SERIES OF DE BILT (THE NETHERLANDS), INTERNATIONAL JOURNAL OF CLIMATOLOGY, 23, 829-845, 10.1002/joc.902, 2003.

Taleghani, M., Sailor, D., & Ban-Weiss, G. A: Micrometeorological simulations to predict the impacts of heat mitigation strategies on pedestrian thermal comfort in a Los Angeles neighborhood, Environmental Research Letters, 11, 024003, 10.1088/1748-9326/11/2/024003 2016.

Valmassoi, A. and Keller, J.: How to visualize the Urban Heat Island in Gridded Datasets?, Advances in Science and Research, 18, 41-49, 10.5194/asr-18-41-2021, 2021.

Vancutsem, C., Ceccato, P., Dinku, T., and Connor, S. J.: Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa, Remote Sensing of Environment, 114, 449-465, 10.1016/j.rse.2009.10.002, 2010.

Venter, Z. S., Brousse, O., Esau, I., and Meier, F.: Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data, Remote Sensing of Environment, 242, 10.1016/j.rse.2020.111791, 2020.

Wang, L., Fan, S., Hu, F., Miao, S., Yang, A., Li, Y., Liu, J., Liu, C., Chen, S., Ho, H. C., Duan, Z., Gao, Z., and Yang, Y.: Vertical Gradient Variations in Radiation Budget and Heat Fluxes in the Urban Boundary Layer: A Comparison Study Between Polluted and Clean Air Episodes in Beijing During Winter, Journal of Geophysical Research Atmospheres, 10.1029/2020JD032478, 2020a.

Wang, R., Hou, H., Murayama, Y., and Derdouri, A.: Spatiotemporal Analysis of Land Use/Cover Patterns and Their Relationship with Land Surface Temperature in Nanjing, China, Remote Sensing, 12, 10.3390/rs12030440, 2020b.

Wang, W., Chen, B., Xu, Y., Zhou, W., and Wang, X.: Urban heat islands in Hong Kong: Bonding with atmospheric stability, Atmospheric Science Letters, 10.1002/asl.1032, 2021.

Xinfa, Q., H, G. L., Y, Z., J, J. A., and Jian, H. Y.: Study on Urban Heat Island Effect of Nanjing, Climatic and Environmental Research (in Chinese), 13, 807-814, 10.3878/j.issn.1006-9585.2008.06.12, 2008.

Y, C., Sui, D. Z., T, F., and W, D.: Fractal analysis of the structure and dynamics of a satellite-detected urban heat island, International Journal of Remote Sensing, 28, 2359-2366, 10.1080/01431160500315485, 2007.

Yang, J., Yin, P., Sun, J., Wang, B., Zhou, M., Li, M., Tong, S., Meng, B., Guo, Y., and Liu, Q.: Heatwave and mortality in 31 major Chinese cities: Definition, vulnerability and implications, Sci Total Environ, 649, 695-702, 10.1016/j.scitotenv.2018.08.332, 2019.

Yang, P., Ren, G., Yan, P., and Deng, J.: Tempospatial Pattern of Surface Wind Speed and the "Urban Stilling Island" in Beijing City, Journal of Meteorological Research, 34, 986-996, 10.1007/s13351-020-9135-5, 2020a.

Yang, Y.-J., Gao, Z., Shi, T., Wang, H., Li, Y., Zhang, N., Zhang, H., and Huang, Y.: Assessment of urban surface thermal environment using MODIS with a population-weighted method: a case study, Journal of Spatial Science, 64, 287-300, 10.1080/14498596.2017.1422155, 20182019.

Yang, Y.-J., Wu, B.-W., Shi, C.-e., Zhang, J.-H., Li, Y.-B., Tang, W.-A., Wen, H.-Y., Zhang, H.-Q., and Shi, T.: Impacts of Urbanization and Station-relocation on Surface Air Temperature Series in Anhui Province, China, Pure and Applied Geophysics, 170, 1969-1983, 10.1007/s00024-012-0619-9, 2012.

Yang, Y., Zhang, M., Li, Q., Chen, B., Gao, Z., Ning, G., Liu, C., Li, Y., and Luo, M.: Modulations of surface thermal environment and agricultural activity on intraseasonal variations of summer diurnal temperature range in the Yangtze River Delta of China, Science of The Total Environment, 736, 139445, https://doi.org/10.1016/j.scitotenv.2020.139445, 2020b.

Yang, Y., Zheng, Z., Yim, S. Y. L., Roth, M., Ren, G., Gao, Z., Wang, T., Li, Q., Shi, C., Ning, G., and Li, Y.: PM 2.5 Pollution Modulates Wintertime Urban Heat Island Intensity in the Beijing‐Tianjin‐Hebei Megalopolis, China, Geophysical Research Letters, 47, 10.1029/2019gl084288, 2020c.

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

Yu, R., Lyu, M., Lu, J., Yang, Y., Shen, G., and Li, F.: Spatial Coordinates Correction Based on Multi-Sensor Low-Altitude Remote Sensing Image Registration for Monitoring Forest Dynamics, IEEE Access, 8, 18483-18496, 10.1109/ACCESS.2020.2968335, 2020.

Yuyu Ren, G.: A Remote-Sensing Method of Selecting Reference Stations for Evaluating Urbanization Effect on Surface Air Temperature Trends, Journal of Climate, 24, 3179-3189, 10.1175/2010JCLI3658.1, 2011.

Zhang, N., Wang, X., Chen, Y., Dai, W., and Wang, X.: Numerical simulations on influence of urban land cover expansion and anthropogenic heat release on urban meteorological environment in Pearl River Delta, Theoretical and Applied Climatology, 126, 469-479, 10.1007/s00704-015-1601-0, 2015.

Zhang, Y., Ning, G., Chen, S., and Yang, Y.: Impact of Rapid Urban Sprawl on the Local Meteorological Observational Environment Based on Remote Sensing Images and GIS Technology, Remote Sensing, 13, 10.3390/rs13132624, 2021.

Zhao, J., Zhao, X., Liang, S., Zhou, T., Du, X., Xu, P., and Wu, D.: Assessing the thermal contributions of urban land cover types, Landscape and Urban Planning, 204, 10.1016/j.landurbplan.2020.103927, 2020.

Zhao, L., Lee, X., Smith, R. B., and Oleson, K.: Strong contributions of local background climate to urban heat islands, Nature, 511, 216-219, 10.1038/nature13462, 2014.

Zhao, M., Cai, H., Qiao, Z., and Xu, X.: Influence of urban expansion on the urban heat island effect in Shanghai, International Journal of Geographical Information Science, 30, 2421-2441, 10.1080/13658816.2016.1178389, 2016.

Zheng, Z., Zhao, C., Lolli, S., Wang, X., Wang, Y., Ma, X., Li, Q., and Yang, Y.: Diurnal variation of summer precipitation modulated by air pollution: observational evidences in the beijing metropolitan area, Environmental Research Letters, 15, 10.1088/1748-9326/ab99fc, 2020.

Zheng, Z., Ren, G., Wang, H., Dou, J., Gao, Z., Duan, C., Li, Y., Ngarukiyimana, J. P., Zhao, C., Cao, C., Jiang, M., and Yang, Y.: Relationship Between Fine-Particle Pollution and the Urban Heat Island in Beijing, China: Observational Evidence, Boundary-Layer Meteorology, 169, 93-113, 10.1007/s10546-018-0362-6, 2018.

Zhou, D., Zhao, S., Zhang, L., Sun, G., and Liu, Y.: The footprint of urban heat island effect in China, Sci Rep, 5, 11160, 10.1038/srep11160, 2015.

Zhou, Y., Weng, Q., Gurney, K. R., Shuai, Y., and Hu, X.: Estimation of the relationship between remotely sensed anthropogenic heat discharge and building energy use, ISPRS Journal of Photogrammetry and Remote Sensing, 67, 65-72, https://doi.org/10.1016/j.isprsjprs.2011.10.007, 2012.

Zhu, X., Zhang, Q., Xu, C.-Y., Sun, P., and Hu, P.: Reconstruction of high spatial resolution surface air temperature data across China: A new geo-intelligent multisource data-based machine learning technique, Science of The Total Environment, 665, 300-313, https://doi.org/10.1016/j.scitotenv.2019.02.077, 2019.

Zong, L., Liu, S., Yang, Y., Ren, G., Yu, M., Zhang, Y., and Li, Y.: Synergistic Influence of Local Climate Zones and Wind Speeds on the Urban Heat Island and Heat Waves in the Megacity of Beijing, China, Frontiers in Earth Science, 9, 10.3389/feart.2021.673786, 2021.

Zou, J., Sun, J., Zhang, N., and Zhao, W.: Evaluation and Parameter-Sensitivity Study of a Single-Layer Urban Canopy Model (SLUCM) with Measurements in Nanjing, China, Journal of Hydrometeorology, 15, 1078-1090, 10.1175/jhm-d-13-0129.1, 2014.

795

Tables and captions

Table 1: Independent variables with their sources and spatial resolution

	Parameters	Source	Spatial resolution(m)
geometric parameters	Proportion of LULC area Latitude and longitude distance from the city center Altitude, Slope and aspect	Landsat 8 data LULC data DEM data	30
physical parameters	Proportion of IS area Albedo NDVI, NDBI, gNDVI, SAVI and NDMI	Landsat 8 data	30
anthropogenic parameters	AHF data	NOAA nighttime lighting data	1000

Notes: DEM, digital elevation model; IS, impervious surface; NDVI, normalized difference vegetation index; NDBI , normalized difference built-up index; gNDVI, green normalized difference vegetation index; SAVI, soil-adjusted vegetation index; NDMI, normalized difference moisture index; AHF, anthropogenic heat flux.

800

Table 2⅟: R^2 and RMSE of the RF model with different buffer sizes (500 m, 1000 m, 2000 m, 5000 m).

	500 m		1000 m		2000 m		5000 m	
	R^2	RMSE (°C)	R^2	RMSE (°C)	R^2	RMSE (°C)	R^2	RMSE (°C)
11/08/2013	0.33	0.75	**0.57**	**0.65**	0.56	0.65	0.36	0.74
02/09/2015	0.58	0.70	**0.59**	**0.69**	0.57	0.70	0.49	0.76
21/07/2017	0.19	0.92	0.17	0.91	**0.47**	**0.80**	0.16	0.93

Table 32: R^2 and RMSE of stepwise regression, GWR (geographically weighted regression) and the RF model within a 2-km buffer zone.

	Stepwise regression		GWR		RF model	
	R^2	RMSE (°C)	R^2	RMSE(°C)	R^2	RMSE (°C)
11/08/2013	0.30	0.69	0.33	0.77	**0.56**	**0.65**
02/09/2015	0.47	0.74	0.44	0.82	**0.57**	**0.70**
21/07/2017	0.27	0.90	0.12	0.93	**0.47**	**0.80**

810

Table 43. Importance of input variables for the RF model of AT estimation on the three different days.

11/08/2013	%IncMSE	IncNodePurity
Water-body	9.23	4.71
NDVI	8.38	4.22
NDBI	7.15	7.13
IS	6.93	6.46
Built-up	4.19	2.05
Vegetation	2.35	1.38
AHF	0.89	2.91
Cropland	0.27	1.70

02/09/2015	%IncMSE	IncNodePurity
Cropland	5.10	9.45
Distance to city center	4.57	8.59
Water body	4.00	11.05
NDVI	3.18	5.34
NDBI	2.44	4.05
Built-up	2.41	2.78
SAVI	1.49	2.67
Vegetation	1.44	2.24
IS	0.40	2.59

21/07/2017	%IncMSE	IncNodePurity
Distance to city center	20.01	16.22
IS	18.36	15.75
Vegetation	11.52	8.08
NDVI	9.89	3.85
gNDVI	7.86	3.28
SAVI	6.78	2.38
Water-body	6.45	6.24

Notes: NDVI, normalized difference vegetation index; IS, impervious surface; AHF, anthropogenic heat flux; DEM, digital
815 elevation model; NDBI , normalized difference built-up index; gNDVI, green normalized difference vegetation index; SAVI, soil-
adjusted vegetation index.

Table 54. Area occupied by different levels of urban heat island intensity on different days (km².)

CUHII level(°C)	−1.5 to −1	−1 to −0.5	−0.5 to 0	0 to 0.5	0.5 to 1	1 to 1.5	1.5 to 2
11/08/2013	0.00	0.15	1047.43	1517.19	2446.03	1486.89	82.96
02/09/2015	0.02	192.13	1109.89	3751.88	1472.26	56.97	0.00
21/07/2017	0.23	232.52	1005.04	2670.11	2040.98	634.13	0.14

820
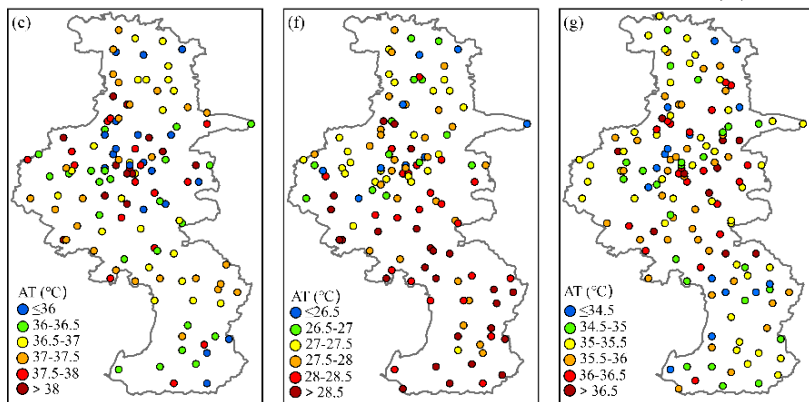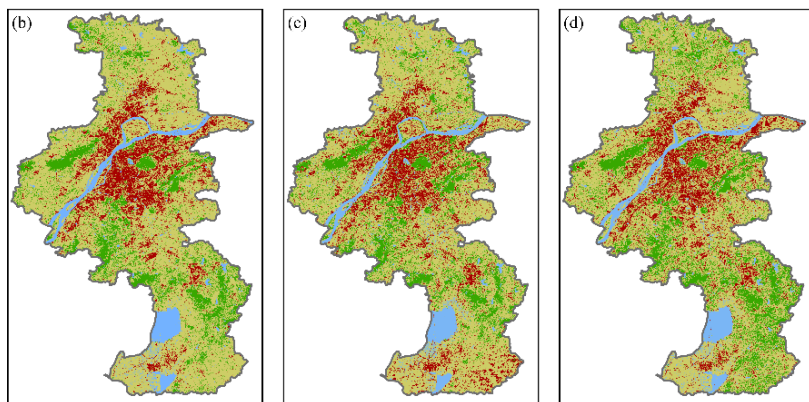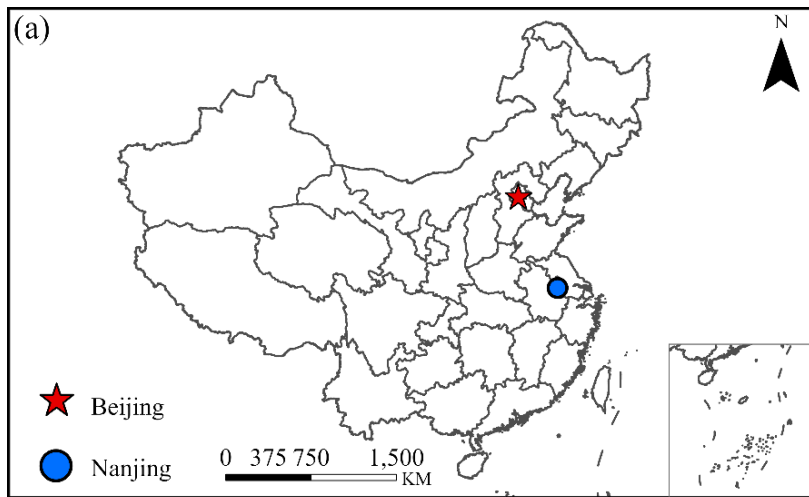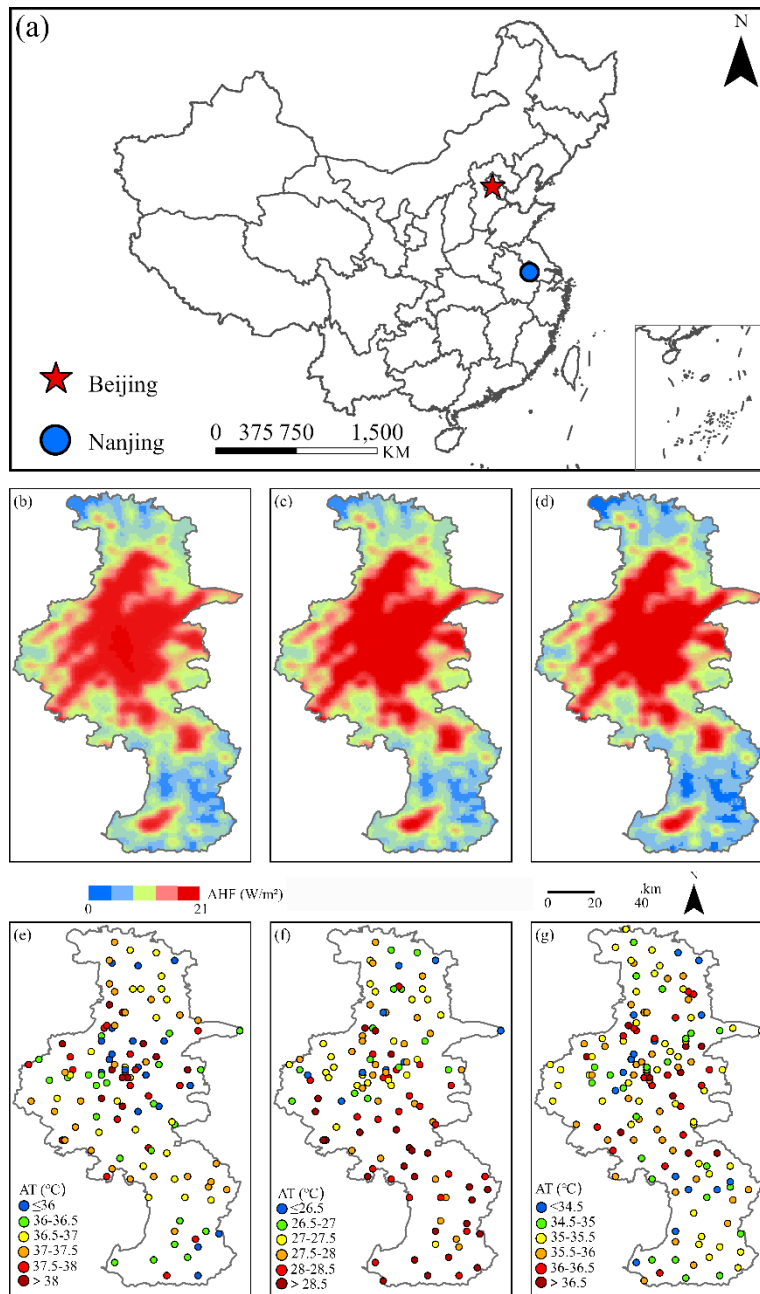
Figure 1: anthropogenic heat flux~~Land use/land cover~~ of Nanjing city and locations of high-density automatic meteorological stations in Nanjing with recorded air temperature: (a) location map of Nanjing in China; (b, e) 11:00 local time (LT) 11 August 2013; (c, f) 11:00 LT 2 September 2015; (d, g) 11:00 LT 21 July 2017.

825
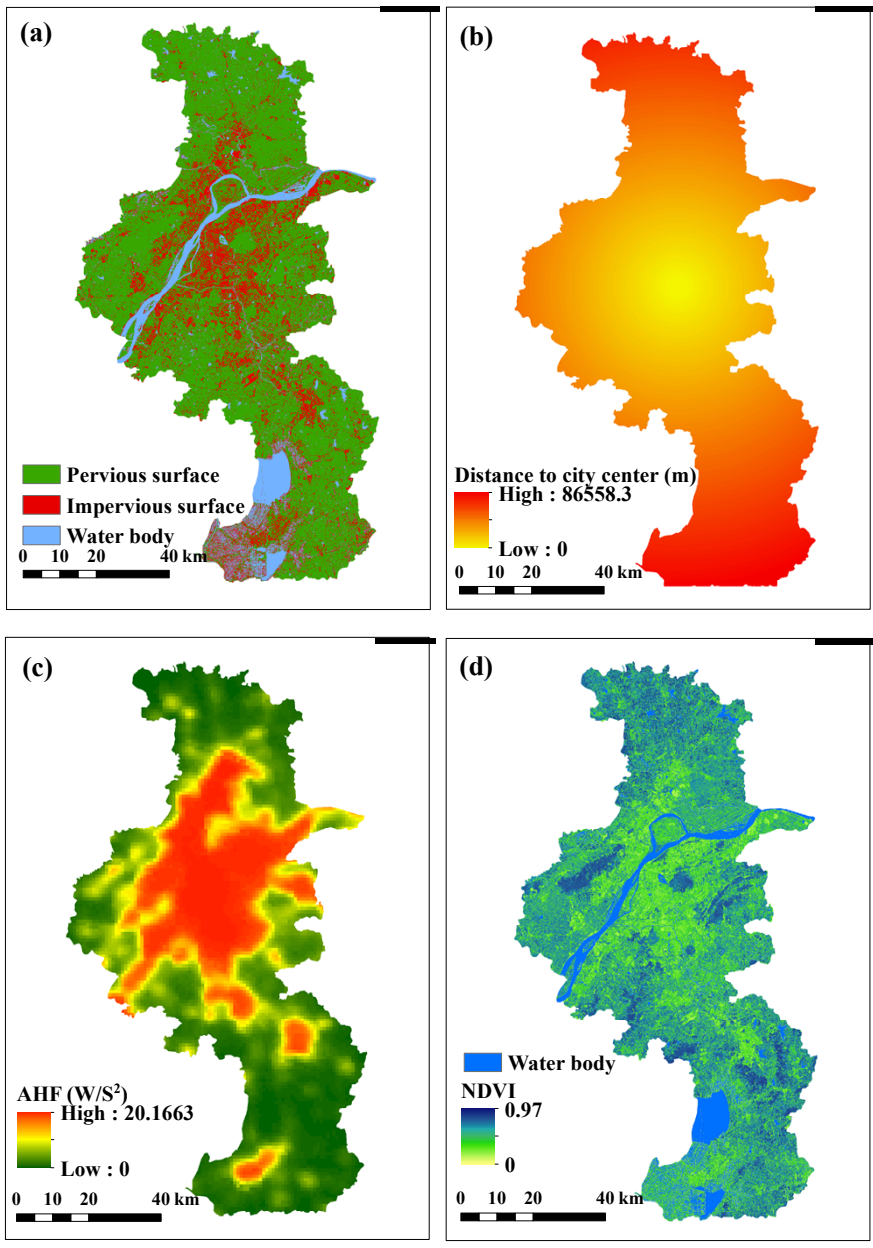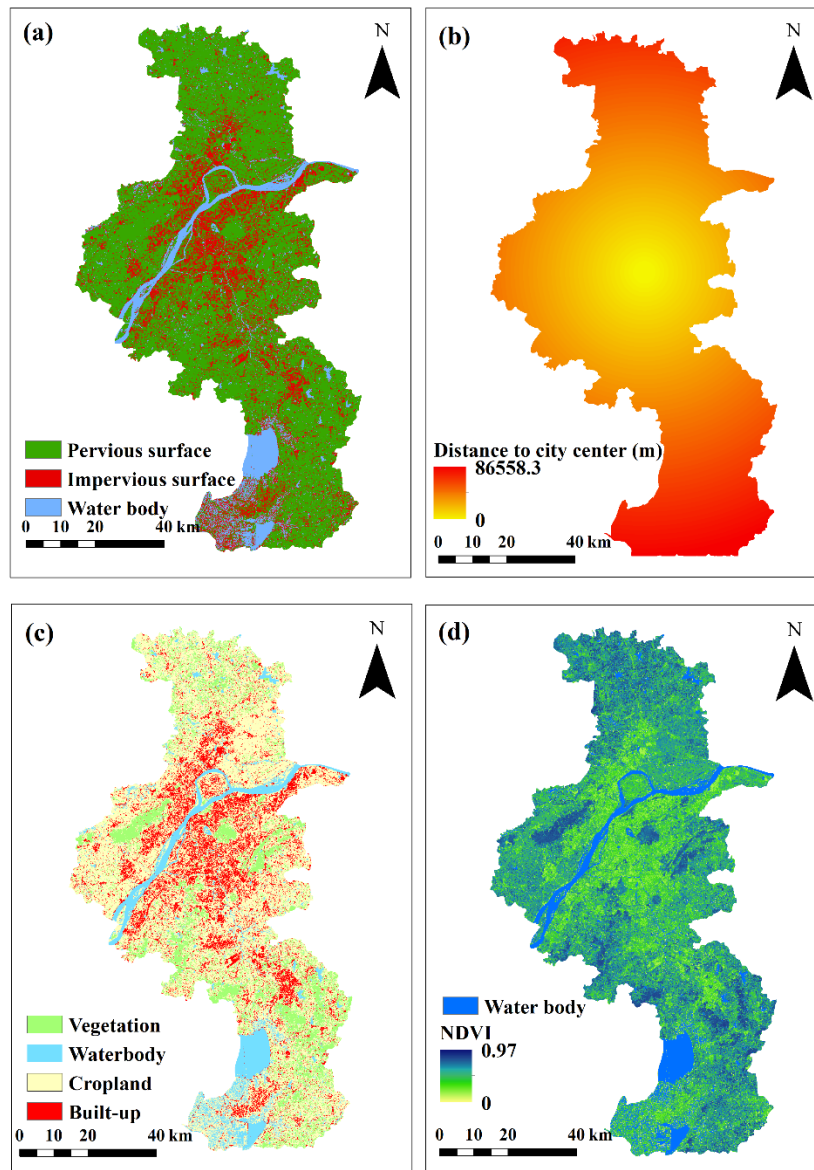
Figure 2: Spatial distribution of ~~some~~ typical **environmental variables on 21 July 2017 in Nanjing city: (a) impervious surface; (b) distance from city** ~~center~~centre**; (c)** ~~AHF (anthropogenic heat flux)~~LULC**; (d) NDVI (normalized difference vegetation index).**

830

```
┌─────────────────────────────────────────────────┐
│                    Data set                     │
│  ┌──────────────────┐    ┌──────────────────┐   │
│  │   Independent    │    │    Dependent     │   │
│  │     Variable     │    │  Variable (AT)   │   │
│  └──────────────────┘    └──────────────────┘   │
└─────────────────────────────────────────────────┘
                         │
                         ▼
┌─────────────────────────────────────────────────┐
│               Random forest model               │
│  ┌───────────────────────────────────────────┐  │
│  │     5-fold cross-validation (CV) method   │  │
│  └───────────────────────────────────────────┘  │
│  ┌───────────────────────────────────────────┐  │
│  │          1.Filtering variables            │  │
│  └───────────────────────────────────────────┘  │
│  ┌───────────────────────────────────────────┐  │
│  │    2.Selecting the parameters (Ntree and  │  │
│  │                  Mtry)                    │  │
│  └───────────────────────────────────────────┘  │
│  ┌───────────────────────────────────────────┐  │
│  │ 3. Repeating the steps above with different│  │
│  │    buffer radius to select buffer size    │  │
│  └───────────────────────────────────────────┘  │
└─────────────────────────────────────────────────┘
                         │
                         ▼
┌─────────────────────────────────────────────────┐
│                   prediction                    │
│  ┌──────────────────┐    ┌──────────────────┐   │
│  │  Predicting AT   │    │ Calculating CUHII │   │
│  └──────────────────┘    └──────────────────┘   │
└─────────────────────────────────────────────────┘
```
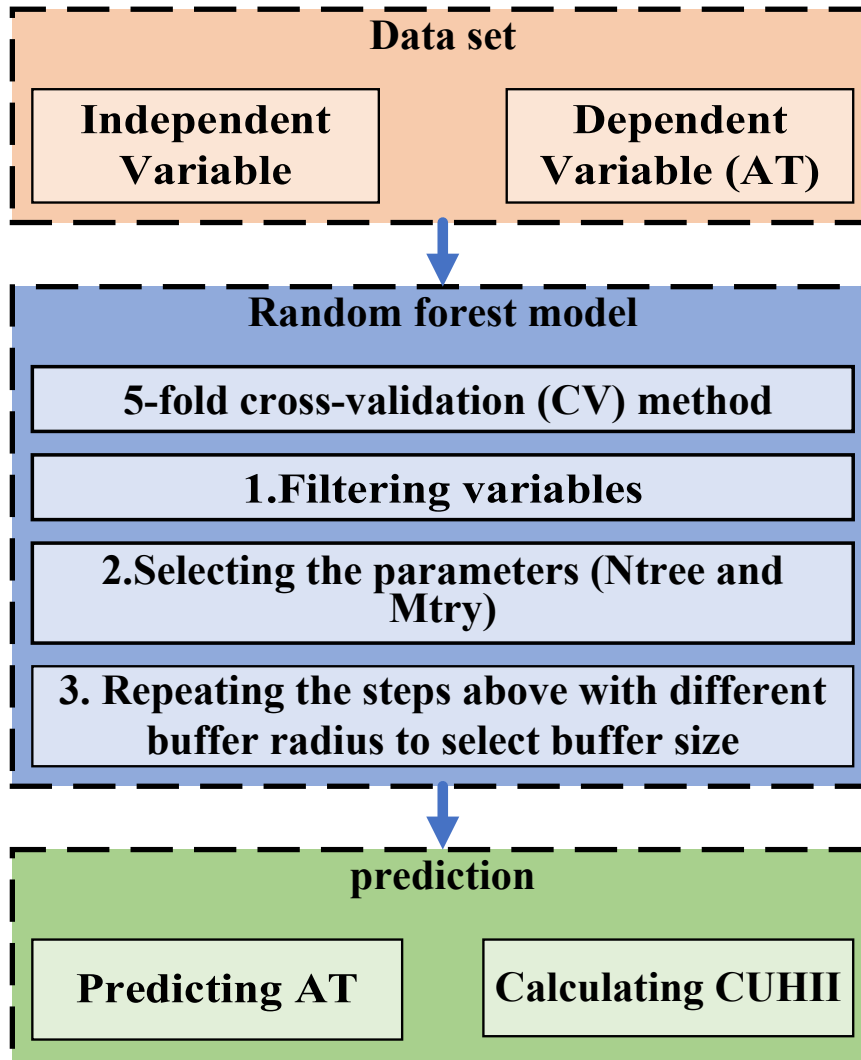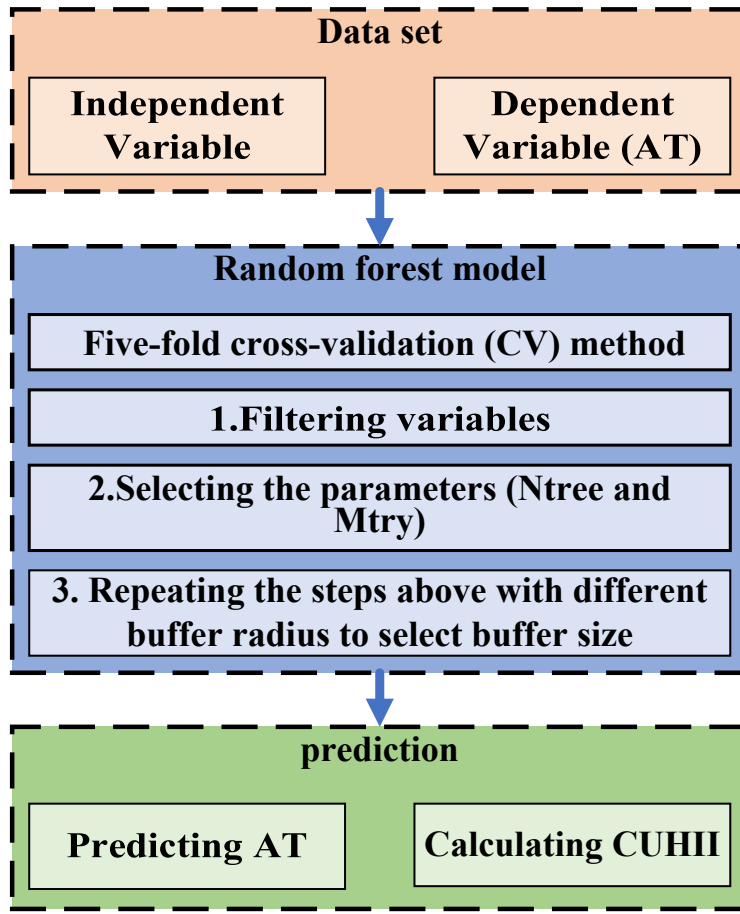
Figure 3: Flowchart for constructing the RF model and evaluating the CUHII (Canopy ~~urban~~ layer urban heat island ~~intensity~~).
Abbreviations: DEM, digital elevation model; AHF, anthropogenic heat flux; LULC, land use/land cover; IS, impervious surface; AT, air temperature; RF, Random Forest.

835
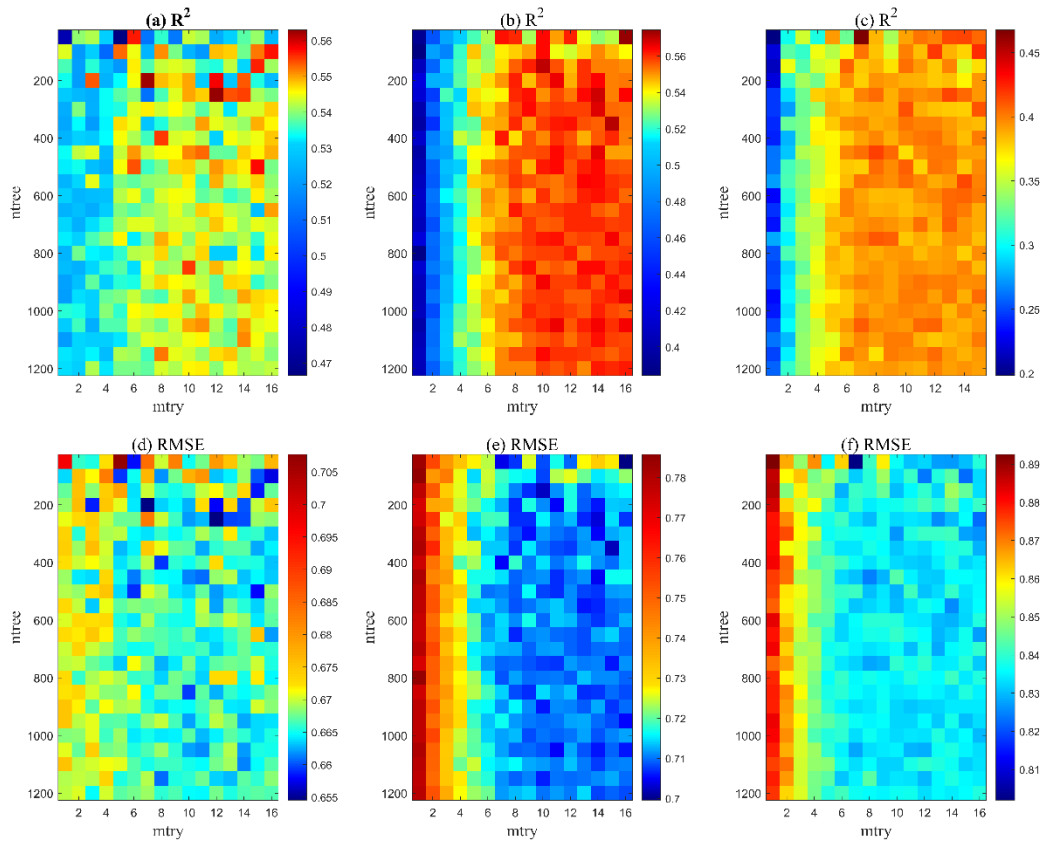
Figure 4. The (a–c) R^2 (coefficient of determination) and (d–f) RMSE (root-mean-square error) changes with the parameters Ntree and Mtry of the model using the dataset on (a, d) 11 August 2013, (b, e) 2 September 2015, and (c, f) 21 July 2017.

840
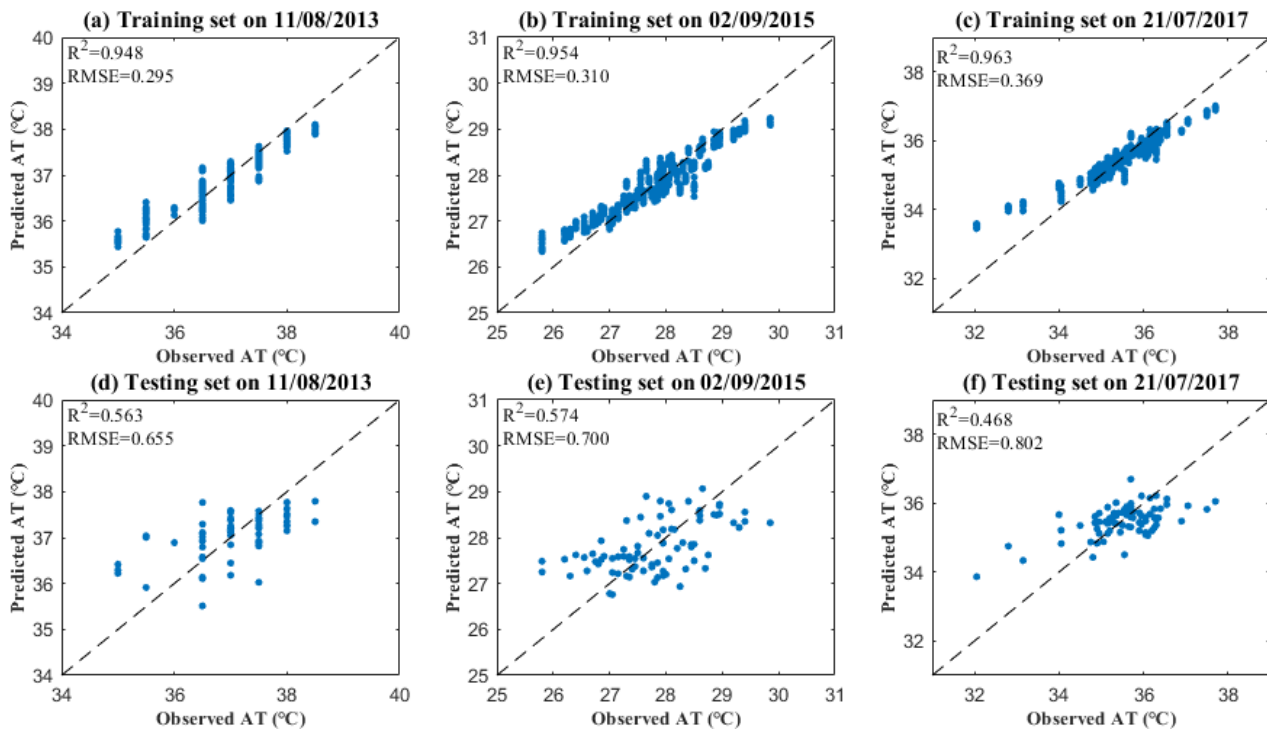
Figure 5: Scatterplot of predicted and observed air temperature: five-fold cross-validation (CV) for the training set on (a) 11 August 2013, (b) 2 September 2015 and (c) 21 July 2017; five-fold CV for the testing set on (d) 11 August 2013, (e) 2 September 2015 and (f) 21 July 2017.

Error(%)   ● -5.00 ~ -2.00   ○ -1.99 ~ 2.00   ● 2.01 ~ 5.00

Land use ▢ Water Body ▢ Vegetation ▢ Cropland ▢ Built-up

0   10   20 km
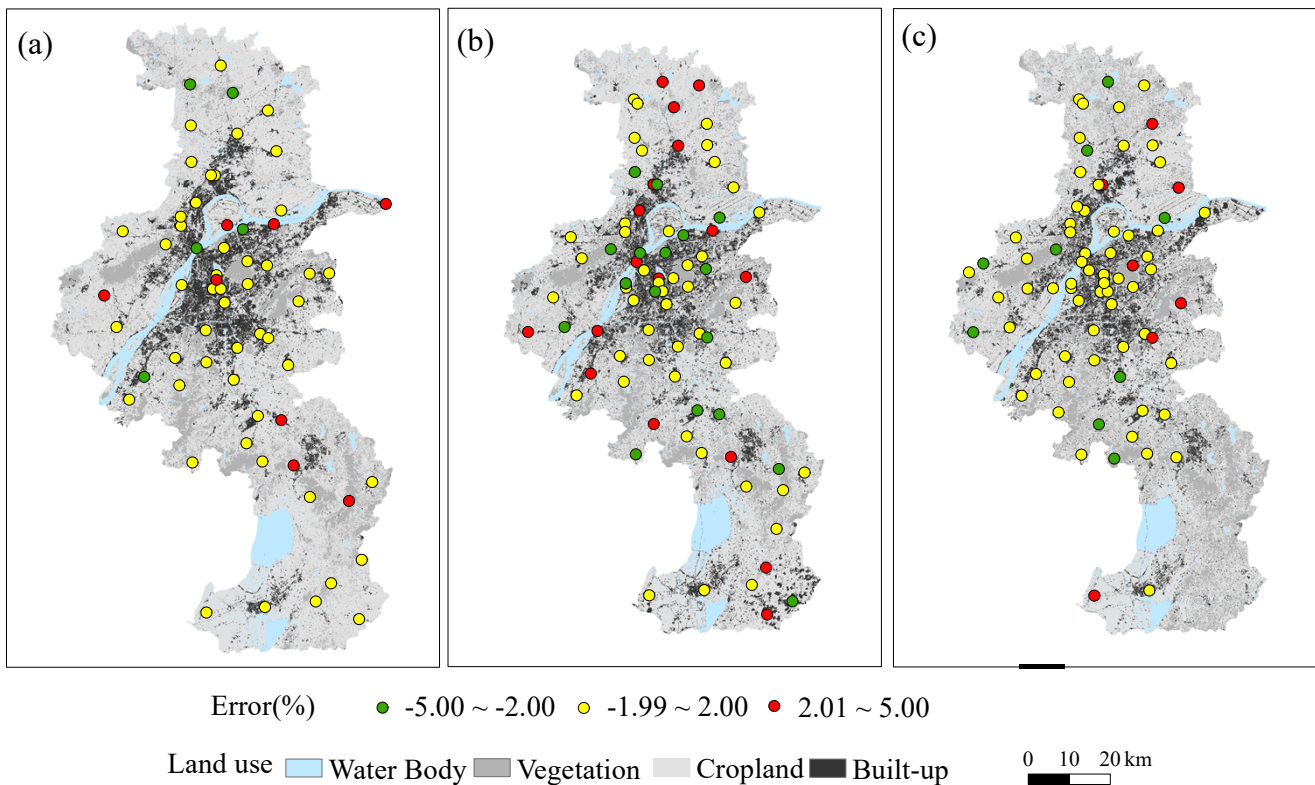
Figure 6: the predicted relative error of the air temperature by random forest: (a) 11 August 2013; (b) 2 September 2015; (c) 21 July 2017.

850
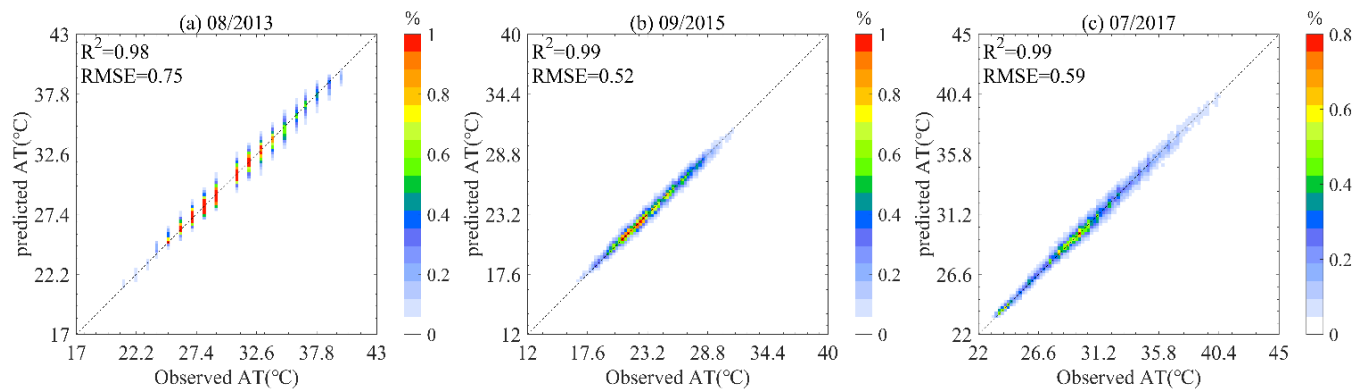
Figure 7: Scatterplot of predicted and observed air temperature using data in a month five-fold CV for the testing set on (a) August 2013, (b) September 2015 and (c) July 2017.
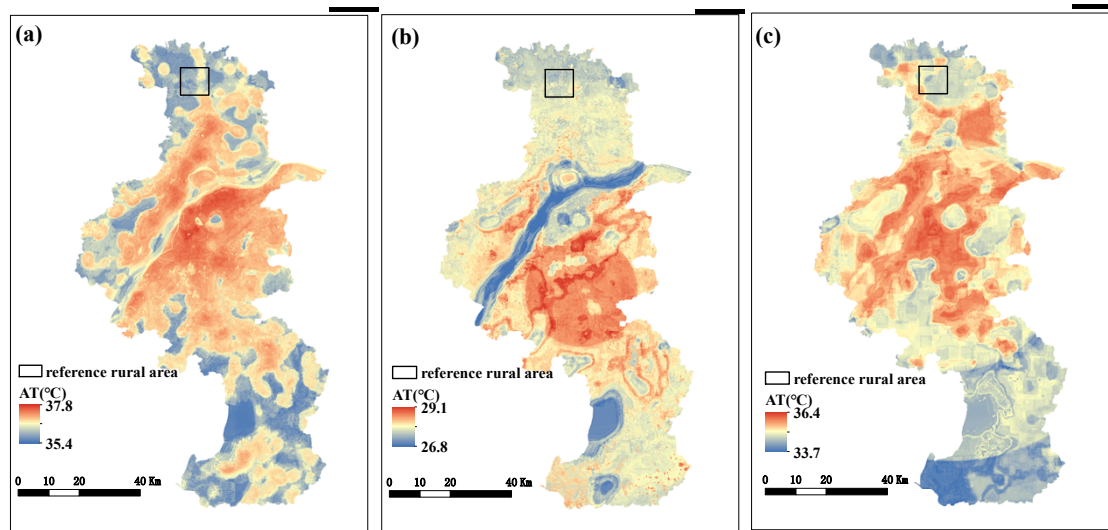
Figure 87: Spatial distribution of air temperature (AT) in Nanjing and the reference rural area: (a) 11 August 2013; (b) 2 September 2015; (c) 21 July 2017.

860
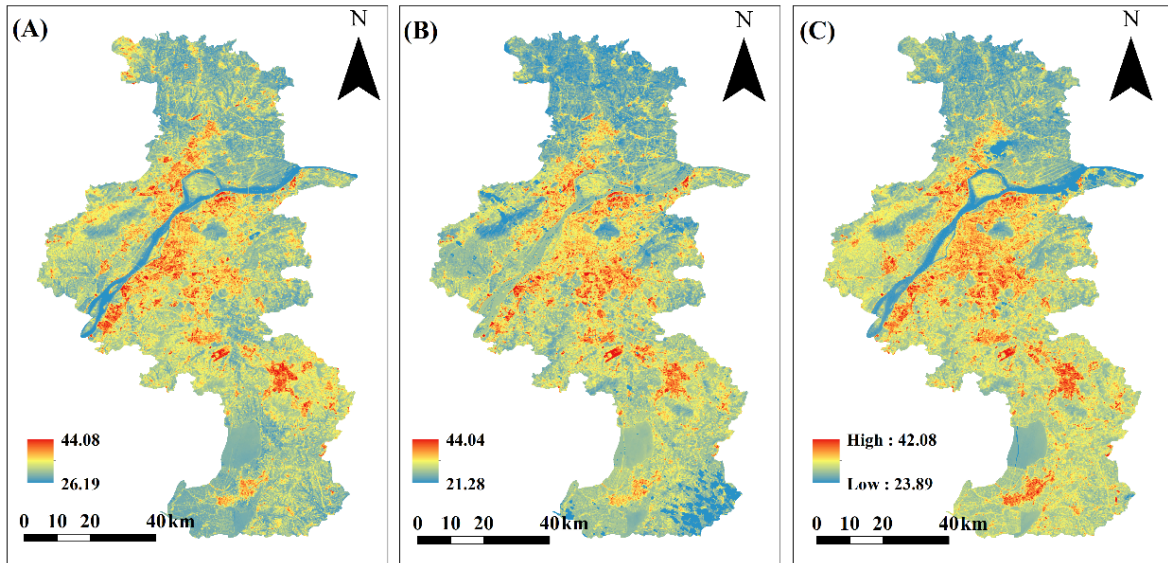
Figure 9: Spatial distribution of the LST in Nanjing: (a) 11 August 2013; (b) 2 September 2015; (c) 21 July 2017.
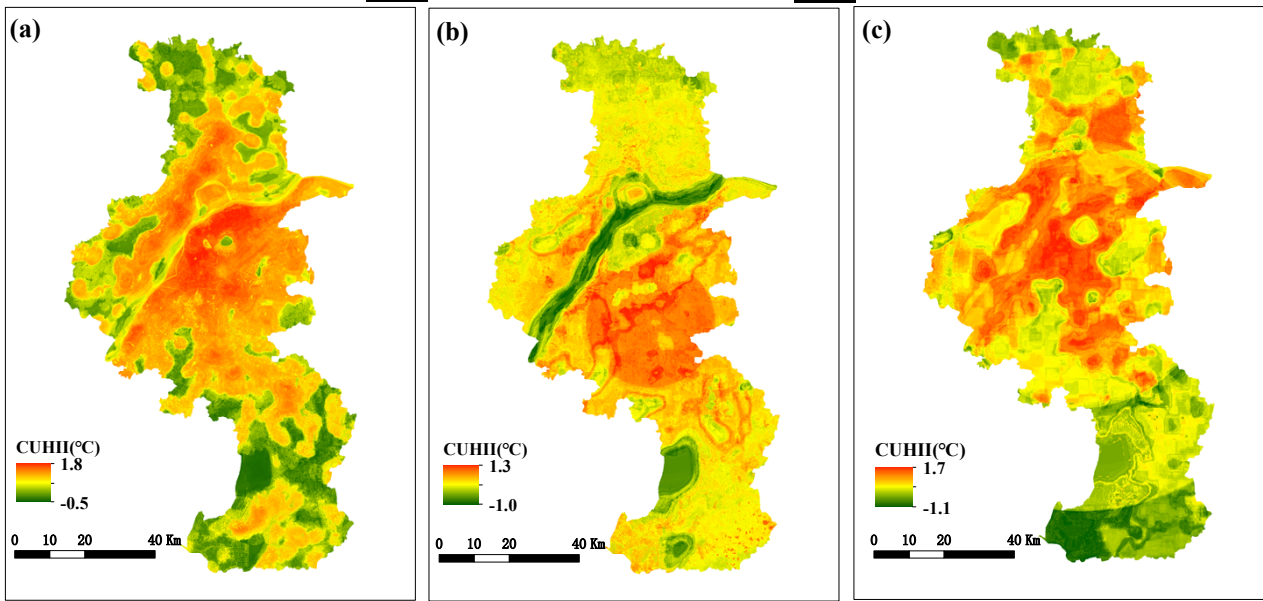
Figure 108: Spatial distribution of the CUHII in Nanjing: (a) 11 August 2013; (b) 2 September 2015; (c) 21 July 2017.

865
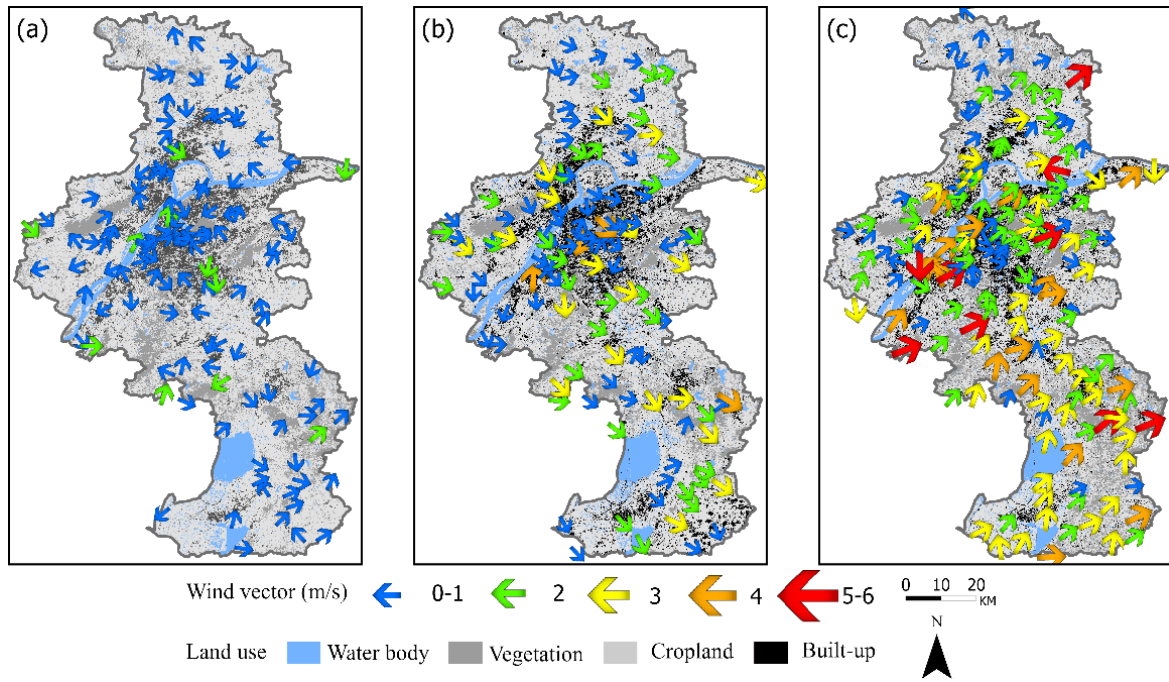
Figure 119: Wind vector field in Nanjing on (a) 11 August 2013, (b) 2 September 2015, and (c) 21 July 2017.

870
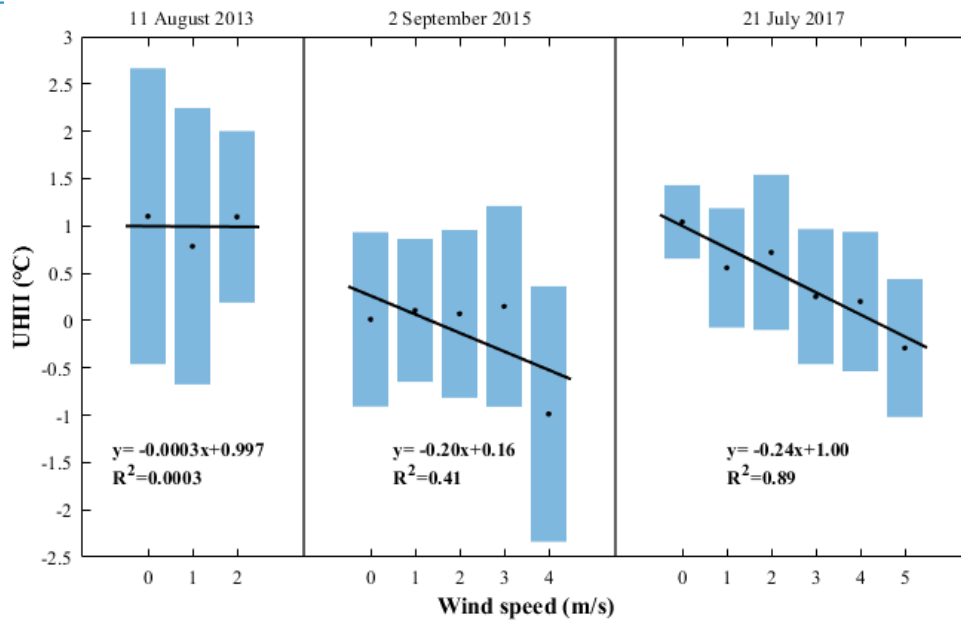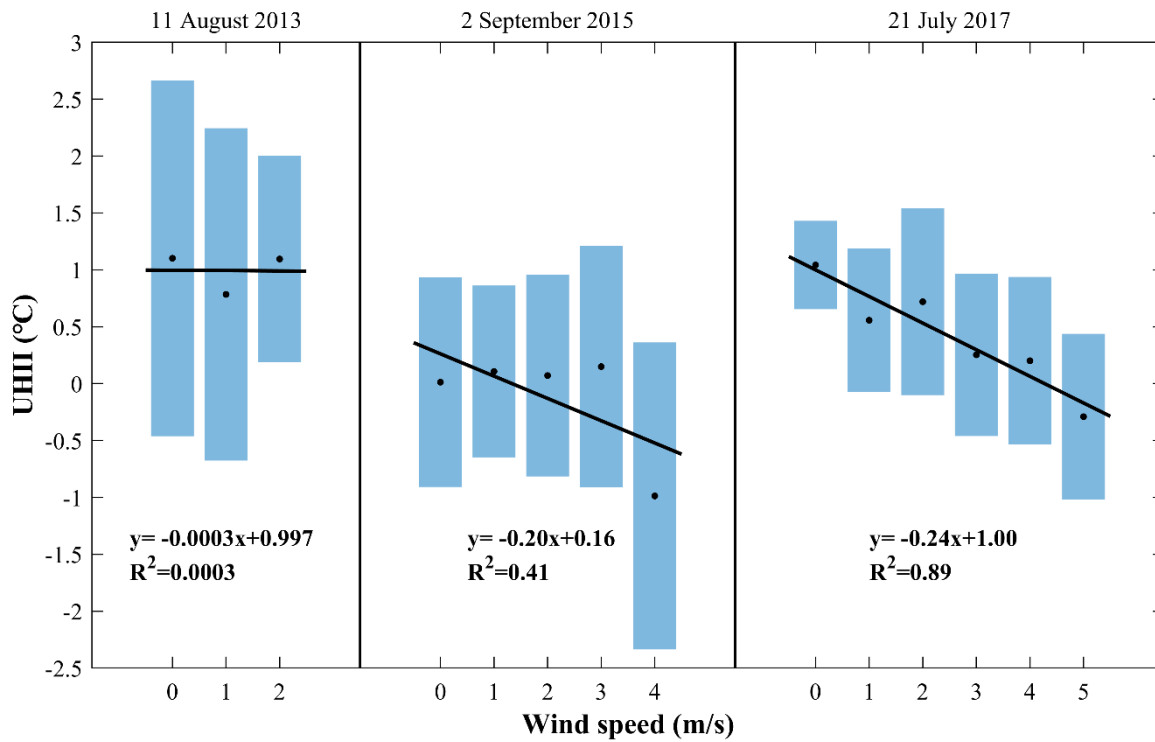
Figure 12~~0~~: Relationship between canopy urban heat island intensity (CUHII) and wind speed around all meteorological stations on (a) 11 August 2013, (b) 2 September 2015, and (c) 21 July 2017. The black dots represent the mean canopy UHII, while the bottom and top edges of the box indicate the uncertainties of one standard deviation~~value~~ from the mean
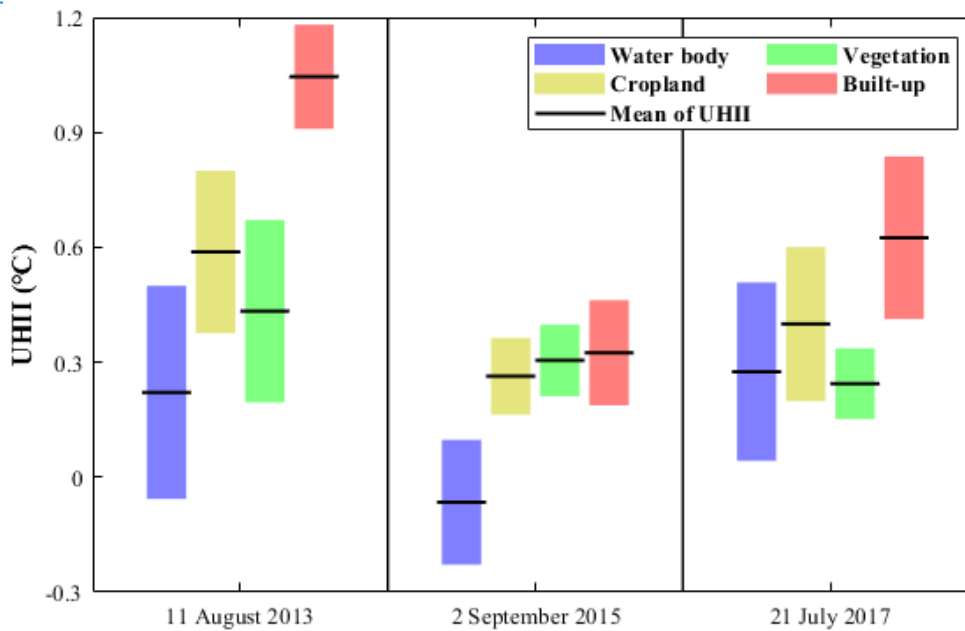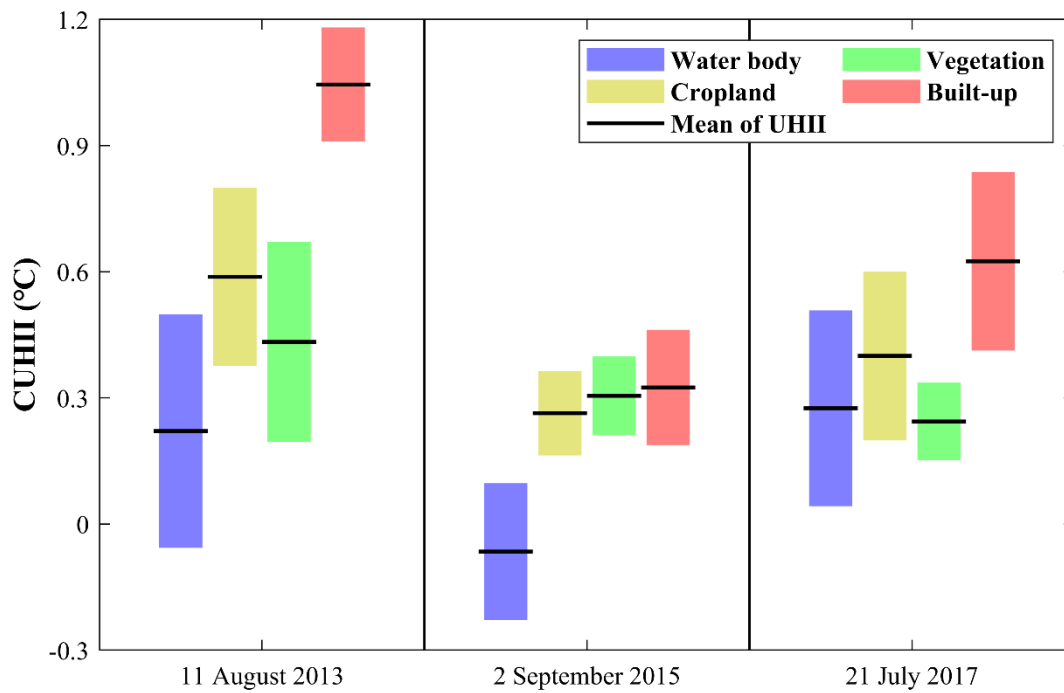
Figure 134: Histogram of canopy urban heat island intensity (CUHII) over different land use/land cover on (a) 11 August 2013, (b) 2 September 2015, and (c) 21 July 2017. Horizontal lines represent the mean intensity, while shaded regions are the uncertainties of one standard value from the mean.
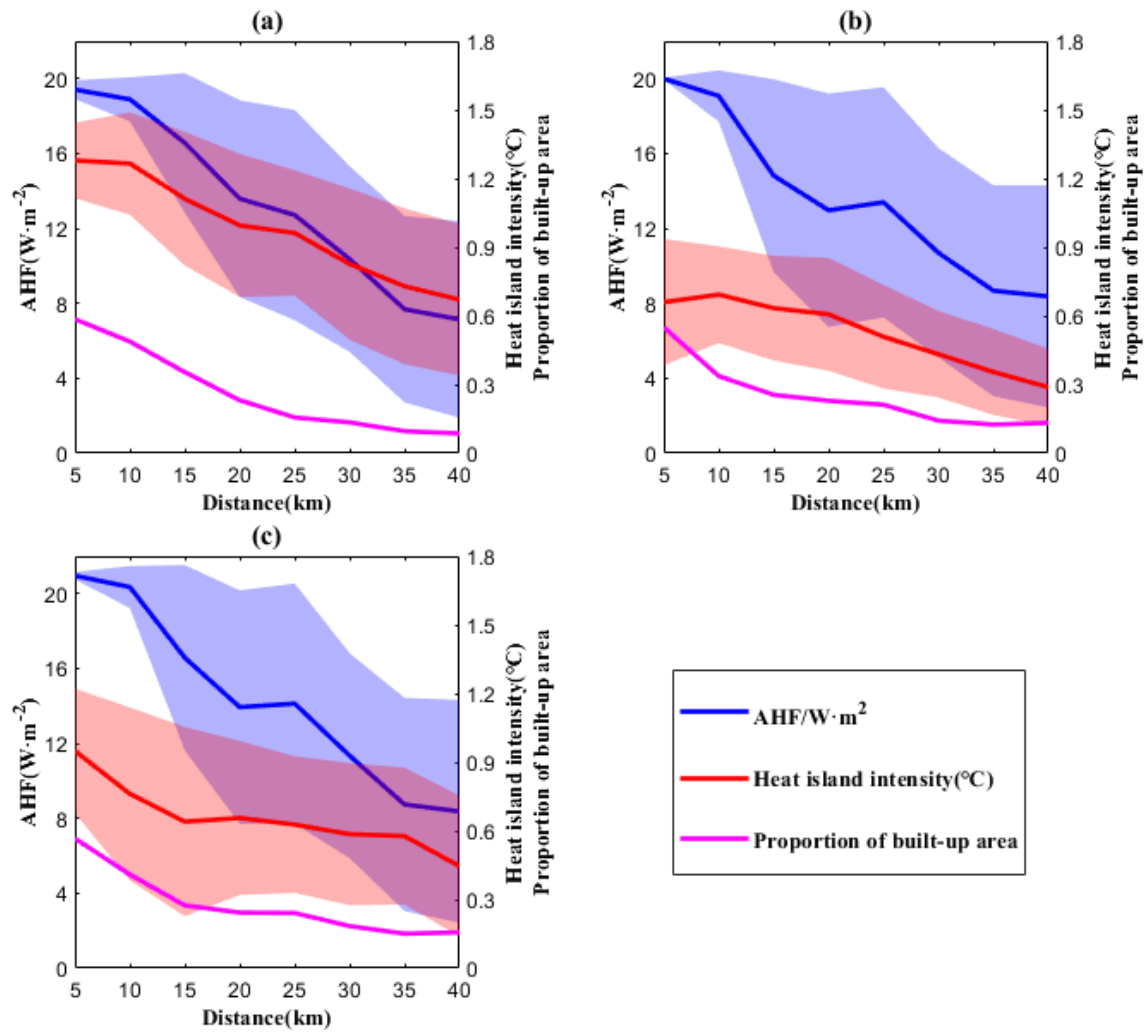
Figure 1~~4~~**2**: Changes in air temperature, anthropogenic heat flux (AHF) and the proportion of built-up areas with distance from the city center on (a) 11 August 2013, (b) 2 September 2015, and (c) 21 July 2017. Thick lines represent mean values, while shaded regions are the uncertainties of one standard value from the mean.

885
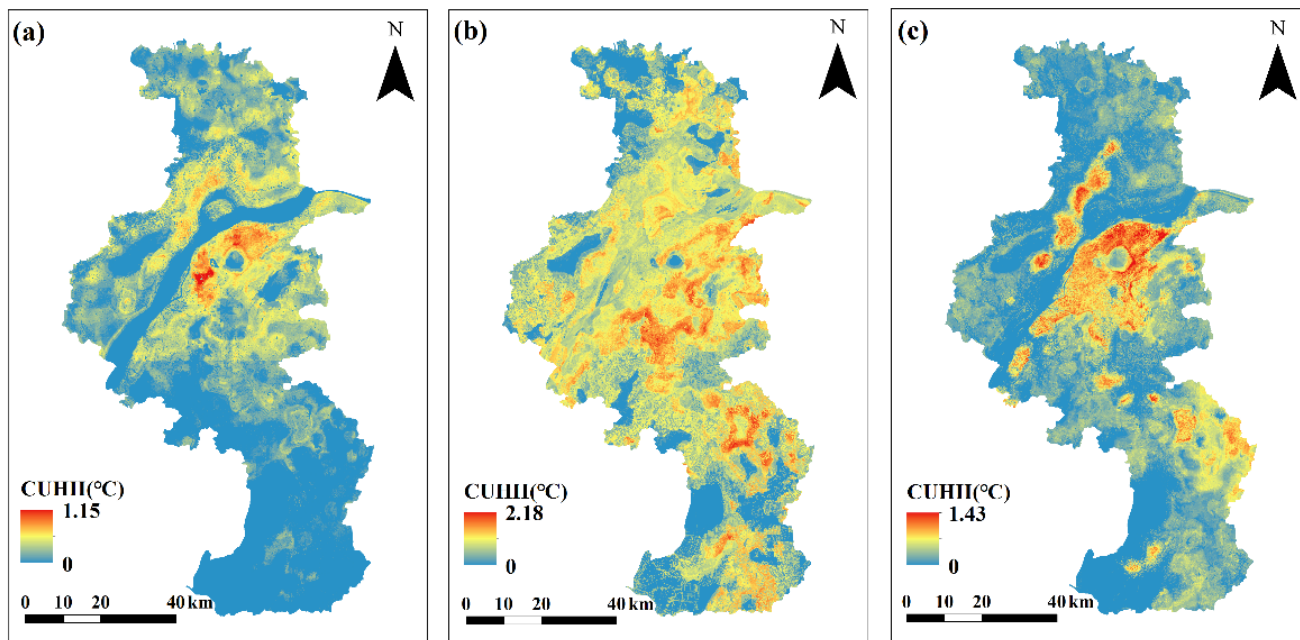
Figure 153: Spatial distribution of canopy urban heat island intensity (CUHII) in Nanjing during a heat wave period: (a) 12 August 2013; (b) 13 August 2013; (c) 14 August 2013.