Geoscience Letters

Open Access

High-resolution calculation of the urban vegetation fraction in the Pearl River Delta from the Sentinel-2 NDVI for urban climate model parameterization

Michael Mau Fung Wong, Jimmy Chi Hung Fung* and Peter Pak Shing Yeung

Abstract

The European Space Agency recently launched the Sentinel mission to perform terrestrial observations in support of tasks such as monitoring forests, detecting land-cover changes, and managing natural disasters. The resolution of these satellite images can be as high as 10 m depending on the bands. In this study, we used the red and near-infrared bands in 10-m resolution from Sentinel-2 images to calculate the Normalized Difference Vegetation Index (NDVI) and estimate of the green vegetation fraction in urban areas within the Pearl River Delta region (PRD). We used vegetation coverage obtained from high-resolution Google satellite images as a reference to validate the vegetation estimates derived from the Sentinel-2 images, and found the correlation between the two to be as high as 0.97. As such, information from the Sentinel-2 imagery can supplement the urban canopy parameters (UCPs) derived from the World Urban Database and Access Portal Tools (WUDAPT) level-0 dataset, which is used in urban meteorological models. The rapid retrieval and open-source nature of the methodology supports high-resolution urban climate modeling studies.

Keywords: Urban vegetation fraction, Urban climate, Urban fraction, Sentinel-2 satellite image

Introduction

The fraction of green vegetation in an urban environment is a key parameter in the study of urban climate, as it influences an area's microclimates, including moisture levels and temperature (Weng et al. 2004). In recent years, the development of global land-cover datasets (Chen et al. 2015) from MODIS and LANDSAT images and the World Urban Database and Access Portal Tools (WUDAPT) level-0 dataset (Ching et al. 2014) for local climate zones/LCZs has improved the accuracy of urban meteorological modeling. The Landsat/MODIS satellite images have been used to derive global land-cover datasets (Chen et al. 2015; Gong et al. 2013), and are also very useful for monitoring land-use changes, such as the change of agricultural land to urban areas in the Pearl

River Delta region (Seto et al. 2002). In such land-use datasets, urban areas are usually classified as a land-use category; detailed morphologies are not distinguished from one location to another in climate modeling. For example, mesoscale models tend to parameterize an urban area's momentum drag as a representative roughness length for the entire urban class in mesoscale modeling.

In addition to these datasets, many urban meteorological models, such as the Weather Research and Forecasting model with different urban canopy parametrization schemes (e.g., Kusaka et al. 2001; Martilli et al. 2002; Salamanca et al. 2009), need to quantify the urban fraction, usually defined as the fraction of impervious surface in an urban area. Therefore, a good-quality dataset of green/urban fractions is desirable, such as the National Land Cover Database (NLCD) (Homer et al. 2007) adopted in the U.S. However, such data are not available in the public domain for the Pearl River Delta

*Correspondence: majfung@ust.hk
Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

(PRD) region. In recent years, the WUDAPT project has proved valuable for estimating landscape characteristics based on satellite images and machine learning. This project aims to develop a straightforward LCZ classification scheme using free and open data, such as Landsat images and training samples from Google Earth. Once the LCZ classification is in place, building morphology parameters can be estimated for various urban classes using machine learning (Ching et al. 2018). The WUDAPT dataset with estimated building morphology parameters has been applied in urban heat island (UHI) studies in Madrid, with promising results (e.g., Brousse et al. 2016). Hammerberg et al. (2018) also carried out a study comparing the improvements in WRF BEP/BEM performance in Vienna using GIS-extracted building morphology data and WUDAPT level-0 data.

Calculating green/urban fraction for urban climate modeling purposes with such LCZ information would typically require the use of look-up tables, which are highly dependent on the area's geolocation and climatic situation. Obtaining locally available tree data is usually difficult or costly. The European Space Agency's recently launched Sentinel-2 mission (Drusch et al. 2012) provides open-source 10-m resolution data, including the red (visible) and near-infrared (NIR) regions. This imagery has significant potential for estimating the urban green fraction based on the Normalized Difference Vegetation index (NDVI) (Carlson and Ripley 1997), which has been used to create the global land-cover datasets. It has also been widely used in vegetation fraction detection (e.g., Elmore et al. 2000) and in tracking their changes over time (e.g., Eckert et al. 2015). These estimates have been found to impact evapotranspiration modeling, which influences the accuracy of mesoscale weather simulations (Vahmani and Ban-Weiss 2016). Higher-resolution data on urban vegetation fraction would, thus, be beneficial for urban climate modeling.

This study estimates the vegetation fraction in the urban areas of Pearl River Delta region by calculating the NDVI from the 10-m resolution Sentinel-2 images. We obtained the visible red and NIR bands from four tiles of the Sentinel-2 images and merged them at a resolution of 10 m. Additional file 1: Figure S1 shows the corresponding Sentinel-2 RGB image tiles and coverage for this study, which were captured on a clear-sky day (2017-12-31). We then used Google satellite images to validate the urban vegetation fraction estimated from the Sentinel-2 images, as there are no field data available for the target area. Google satellite images have a high resolution (pixel size of 0.59716 m) and are in RGB format, making it possible for the human eye to detect whether a certain region in the imagery has vegetation or not. However, manually identifying vegetation for the whole Pearl

River Delta region is highly labor intensive. Therefore, to reduce validation costs, 200 Google satellite images over the study area were randomly sampled and a color detection algorithm was used to assist in estimating vegetation fractions. These calculations were combined with subjective adjustments to identify and quantify green coverage, and then used the estimates as a reference to validate the Sentinel-2-derived vegetation fraction.

After validating the Sentinel-2-derived green fraction, we calculate the region's urban fraction in the 100-m urban grids used in the WUDAPT level-0 dataset. For simplification, we consider the urban fraction to be defined as impermeable surfaces and assume that any area without vegetation is impermeable. We then compare our urban fraction estimates to those derived using the WUDAPT level-0 dataset, with assigned look-up table values for different LCZs, and quantify the new method's benefits. The new urban green fraction dataset should help urban climate researchers/modelers in terms of high-resolution microclimate models and urban air quality/health assessments. The framework developed in this study could be repeated in other regions for similar purposes.

Methods
Generation of a Sentinel-2-resolution green cover dataset
Model

The PRD region's green cover was estimated with the commonly used NDVI. The calculation is as follows:

$$NDVI = (NIR - Red)/(NIR + Red),$$

where Red and NIR are the spectral reflectance measurements acquired in the red (visible) and near-infrared regions, respectively. These spectral reflectances are themselves ratios of reflected over incoming radiation in each individual spectral band; hence, they have values between 0.0 and 1.0. Accordingly, the NDVI varies between -1.0 and $+1.0$. Numerous studies have established the threshold for vegetation as 0.2 (e.g., Sobrino et al. 2004).
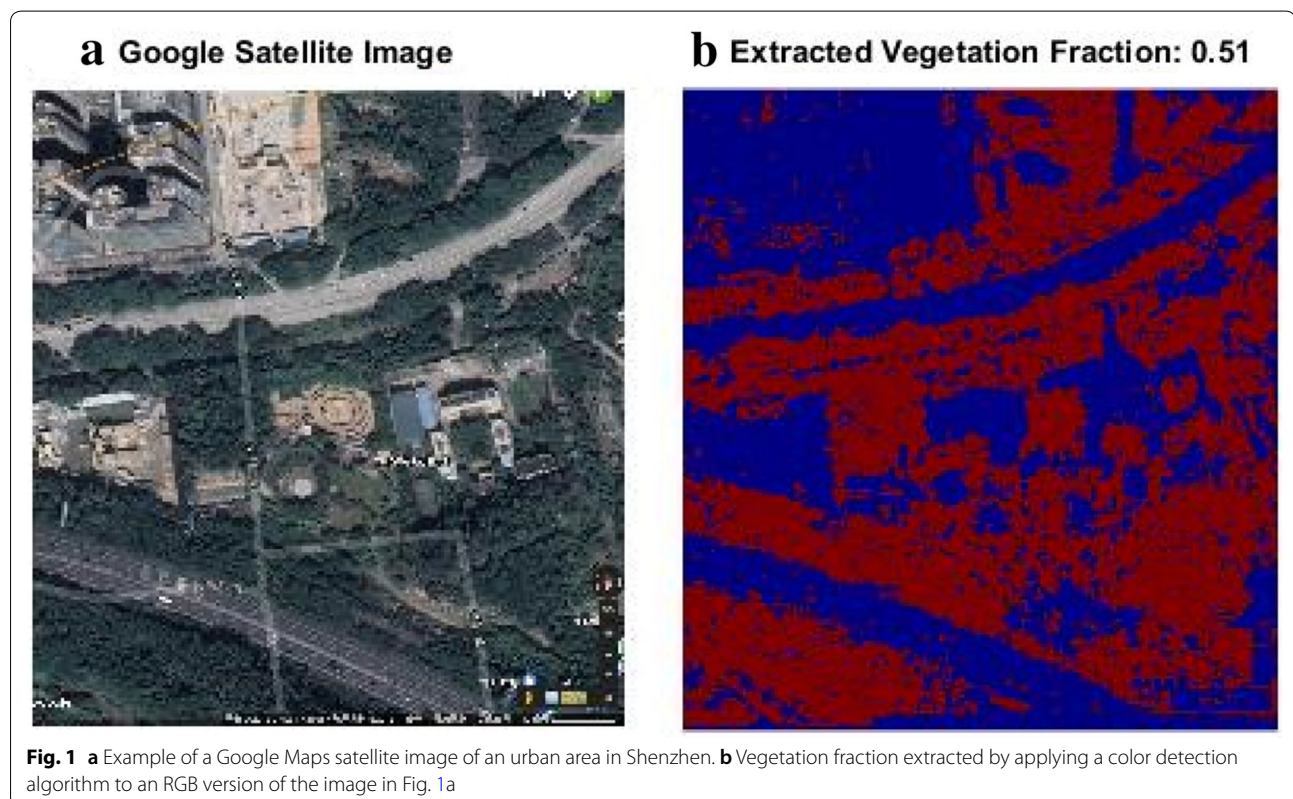
Data

For computation, we merged the Sentinel-2 images' NIR and red bands from four tiles at a resolution of 10 m. Additional file 1: Figure S1 shows the corresponding Sentinel-2 RGB image tiles and their coverage. These data were captured on a clear-sky day (2017-12-31) to ensure the accuracy of the NDVI estimation.

Validation

Due to the absence of field data in the target area, we used high-resolution (pixel size of 0.59716 m) Google satellite images from 2016 as field data to validate the

Wong *et al. Geosci. Lett.*   (2019) 6:2

Page 3 of 10

results. These images were obtained from the Google Maps Static API with a zoom level of 19. We assume that the influence of several months' worth of differences between the field data (Google satellite images) and the Sentinel-2 data would not be significant for the PRD, which is located in a sub-tropical region. Figure 1a shows an example of a high-resolution Google Maps satellite image. Trees can be recognized, and the area that they occupy can be manually measured; these measurements serve as field data or ground truth for validation. However, examining hundreds of sampled images is a highly labor-intensive process. Therefore, we apply a very simple algorithm for color detection (Cheng et al. 2001) to the RGB Google satellite images from the MATLAB image processing toolbox to detect the color green, which usually signals the presence of vegetation. This process automatically generates images with reasonably well-recognized vegetation as an intermediate dataset requiring manual adjustment. However, green areas on the images are not always vegetation; for example, some buildings are green. Therefore, subjective adjustments (step 4) were made to the reference images using brushing tools from ARCGIS to reduce the estimation error. The steps for image processing are summarized as follows.

1. High-resolution satellite RGB images are downloaded from Google Static Map API.
2. The RGB images are converted into HSV space for color detection.
3. Green areas in the RGB images are detected by setting an optimum threshold (manually determined after a few iterations) for hue value, ranging 61−210. This green range image is then filtered with image processing tools in MATLAB to (i) remove noise using the "bwareaopen" command, (ii) enhance the green signal in the presence of shadows between trees using the dilution command "imdilate," and (iii) connect the trees in images close to on another using the command "imclose."
4. Procedures 1–3 are automatically repeated 200 times randomly throughout the PRD dataset to provide a georeferenced images containing vegetation information.
5. Finally, these images are overlaid with Google Maps images on ARCGIS to fine-tuned the results by removing objects (e.g., green buildings) and adding trees that were not identified by the color detection scheme.
6. The vegetation fraction of the resulting image (an example is shown in Fig. 1b) was then calculated as follows:

Fig. 1 **a** Example of a Google Maps satellite image of an urban area in Shenzhen. **b** Vegetation fraction extracted by applying a color detection algorithm to an RGB version of the image in Fig. 1a

Vegetation fraction

$$= \frac{\text{Number of green pixels}}{\text{Total number of pixels in the sampled image}}.$$

Note that the vegetation data can be sampled manually. However, steps 1–3 automate the tedious job of sampling the vegetation component in RGB satellite images by generating a set of initial images that capture a majority of green cover. The subjective adjustment in Step 4 then maximizes the accuracy of the estimated vegetation fraction. In this study, the 200 images obtained through this procedure were used as a reference to validate the vegetation data obtained from the Sentinel-2 10-m imagery.

We subsampled to a 360×360 m grid size (suitable for fine-scale urban climate modeling). Two hundred points are selected and the corresponding 200 Google images with the same size are obtained through steps 1–5 to obtain the vegetation fraction. These values are compared to determine the accuracy of the Sentinel-2-derived green fraction for urban climate modeling purposes.

Comparison with WUDAPT level-0
After validation, the 10-m vegetation product for the PRD region was compared to the WUDAPT level-0 (100 m) data (Cai et al. 2016) to demonstrate the benefits of the new method. Specifically, we compared the urban fraction (defined as the percentage of ground covered by impervious surfaces) estimated by these two datasets is compared. For the WUDAPT level-0 data, the urban fraction was assigned according to the local climate zones 1–10 (urban categories from Brousse et al. 2016). For the Sentinel images, the urban fraction was estimated as follows:

Urban fraction $= 1 -$ vegetation fraction.

To facilitate comparison, we subsampled the Sentinel-2-derived vegetation fraction to the WUDAPT grid resolution (100 m).

Figure 2 shows a flow chart depicting the process used in this study. Table 1 shows the default look-up table values used to assign the urban fraction in each LCZ in the WUDAPT level-0 estimation.

Results
Accuracy of generated vegetation fraction in urban areas
Spatial comparison
Figure 3 compares selected examples of the vegetation fraction calculated using the NDVI and Google static map images, respectively, for areas around Hong Kong with urban fraction ranging from about 0.23–0.97. As can be seen in Fig. 3, the color detection algorithm

with subjective adjustment based on the Google Satellite images (reference data) produces finer details than the Sentinel-2 NDVI method, likely due to the relatively higher image resolution and the manual image adjustments. Nevertheless, the NDVI calculated from the Sentinel-2 images still identifies comparable sizes of green fractions in urban areas with a large range of urban densities.

Random sample correlation
This sampling process was repeated for a smaller sample of images with vegetation fractions between 0 and 1 with a grid size of 360×360 m^2. The aim is to validate the Sentinel-2's ability to retrieve urban vegetation fractions by comparing its results with those of a WUDAPT dataset sampled 200 times. Figure 4 shows a scatter plot comparing vegetation fractions calculated using Google images and the Sentinel-2 images. The correlation coefficient R for the two samples is as high as 0.97, thereby quantitatively demonstrating the quality of the Sentinel-2 data. The 95% confidence interval around R ranges from 0.97 to 0.98.

Cross-comparison with WUDAPT level-0 After validating the methodology for retrieving green fractions from Sentinel images, we used the estimated vegetation fraction product at a resolution of 100 m to estimate the urban fraction and compare it with that derived from the WUDAPT level-0 data using a default look-up table for each LCZ (Brousse et al. 2016). As shown in Fig. 5 (white represents non-urban regions, according to the WUDAPT dataset), our urban fraction estimate differs significantly from the urban fraction derived from the WUDAPT level-0 default look-up table (more than 0.5 absolute difference). As an example, the blue and purple circles in Fig. 5a show relatively high (Kowloon) and relatively low (Tsing Yi) density urban areas in Hong Kong, respectively. A comparison of the corresponding areas in Fig. 5c (white denoting a difference less than 0.02) shows that in high-density urban areas (Kowloon), the WUDAPT level-0 data tend to overestimate the urban fraction, because the resolution (100 m) cannot resolve the tree clusters, courts, or parks that are partially resolved in Figs. 1 and 3. In contrast, for less dense urban areas (in this example, Tsing Yi), the WUDAPT level-0 data tend to underestimate the urban fraction, as there are container terminals, oil depots, and large parking lots where vegetation coverage is rare.

Figure 6 further emphasizes the discrepancy between the urban fractions calculated using each dataset for the entire PRD region. The *Y*-axis represents the proportion of urban grids (100 m resolution) in the whole study

Wong *et al. Geosci. Lett.* (2019) 6:2

Page 5 of 10
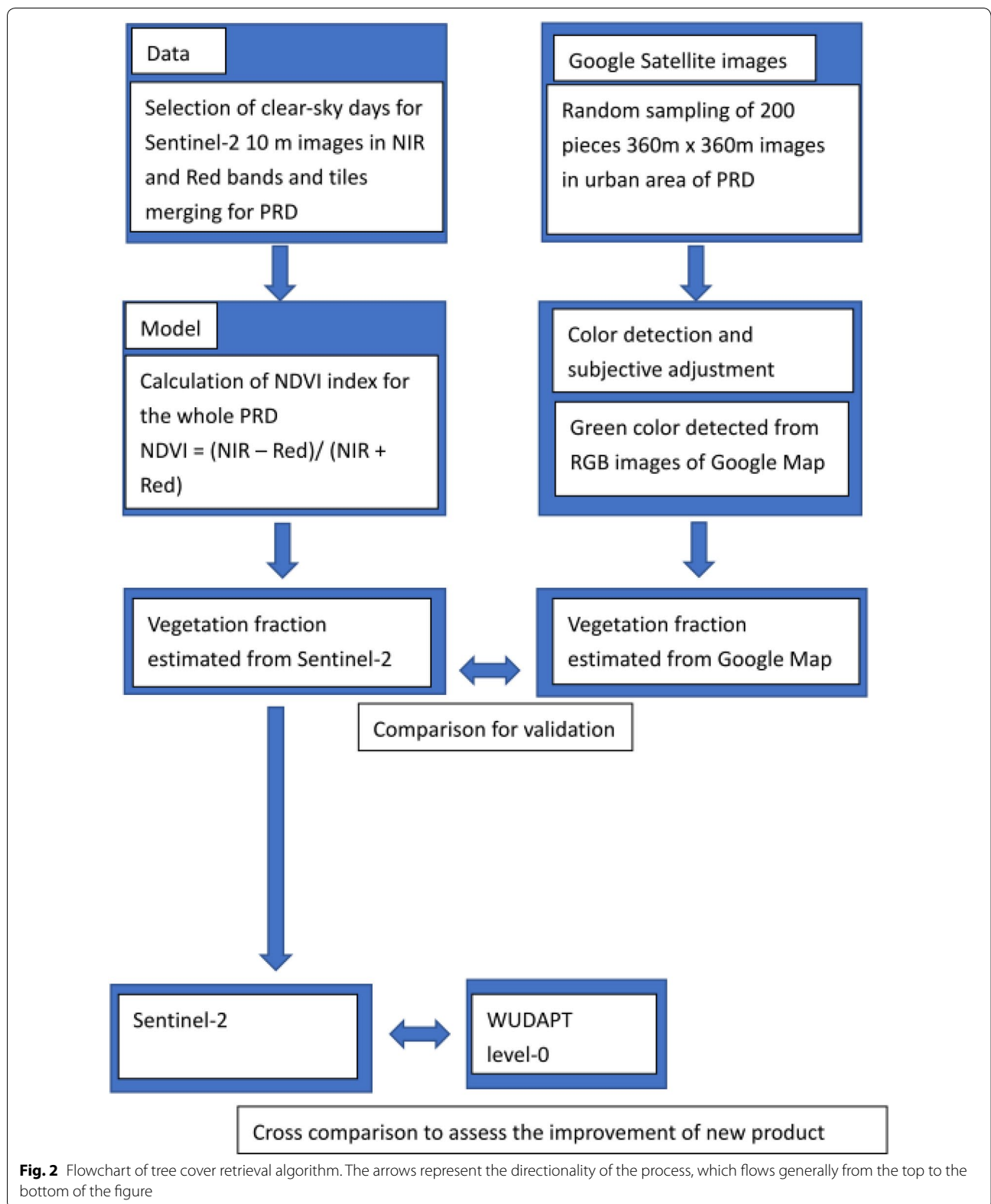
Fig. 2 Flowchart of tree cover retrieval algorithm. The arrows represent the directionality of the process, which flows generally from the top to the bottom of the figure

Wong *et al. Geosci. Lett.* (2019) 6:2

Page 6 of 10

Table 1 Assigned urban fraction values for different LCZs. adopted from Brousse et al. (2016)

LCZ code	LCZ categories	Urban fraction
1	Compact high-rise	1.00
2	Compact mid-rise	0.95
3	Compact low-rise	0.90
4	Open high-rise	0.65
5	Open mid-rise	0.70
6	Open low-rise	0.65
7	Lightweight low-rise	0.85
8	Large low-rise	0.85
9	Sparsely built	0.30
10	Heavy industry	0.55

region (PRD) with a given difference (indicated by the x axis) between the estimates derived from the WUDAPT look-up table and the Sentinel-2 images, respectively. For example, a Y-axis value of 0.1 denotes 10% of the urban area throughout the entire region. The differences between the two methods were found to be as high as -0.9 to 1 depending on the location. This suggest that the Sentinel-based method estimates high impervious coverage in some regions where WUDAPT level-0 data indicate coverage of pervious surfaces and vice versa. Nevertheless, the approximate normal distribution of the histogram shows that for almost half of the domain, the differences (46% of the urban area in PRD) center at -0.2 to 0.2; this suggests that the datasets actually have a reasonable level of agreement, despite the coarser resolution of the WUDAPT level-0 data.

Figure 7 shows a spatial comparison of selected locations in Hong Kong (Tsing Yi and Kowloon), where the urban fraction retrieved from the WUDAPT level-0 is either underestimated or overestimated by more than 0.2 relative to our estimates.

The Sentinel images successfully resolve individual tree clusters and our estimates suggest a much lower urban fraction when aggregated to a resolution of 100 m (down to 0.2, compared to 0.8 for WUDAPT). WUDAPT identified urban areas that have relatively larger vegetation coverage (the open low-/high-rise classes) but the data
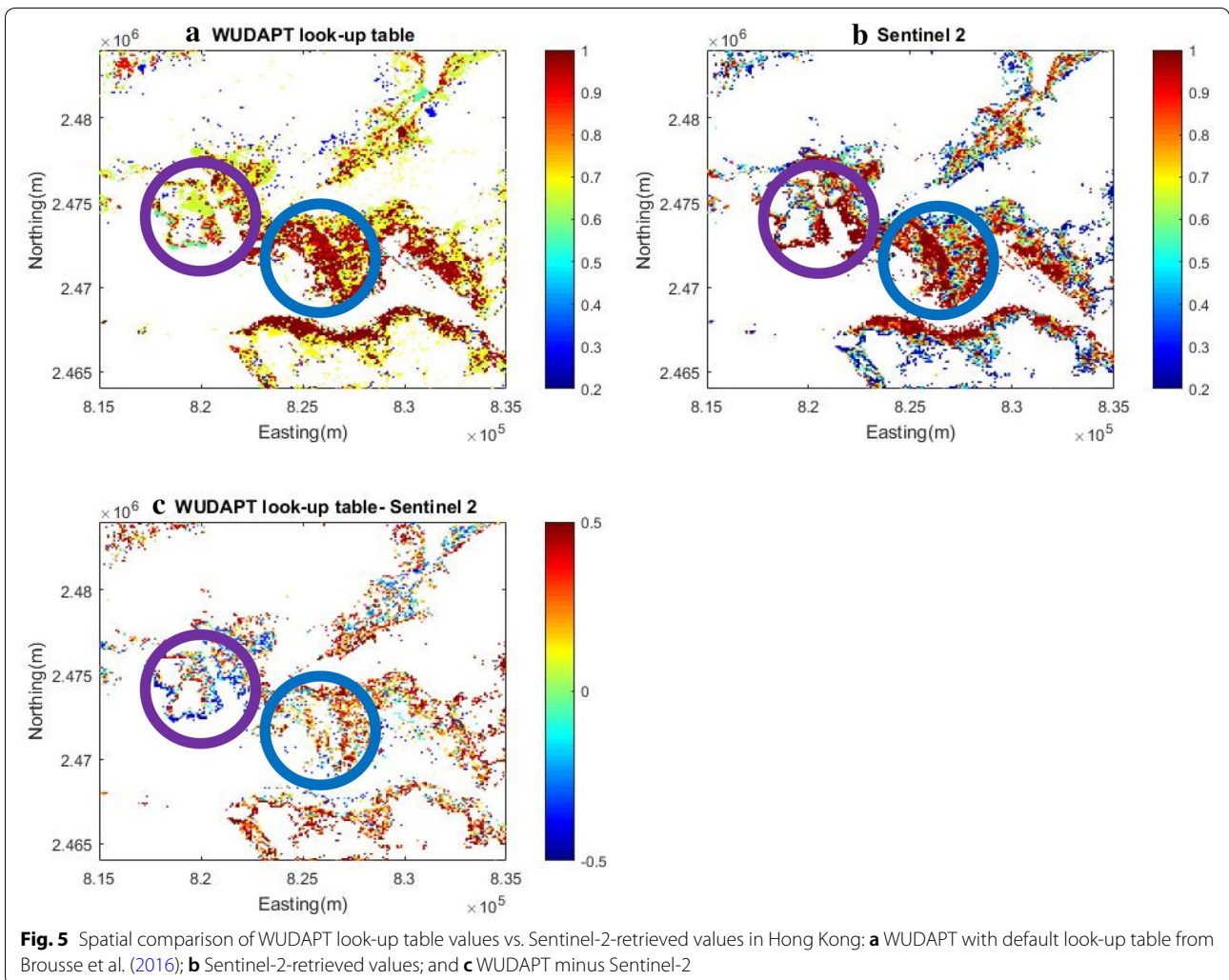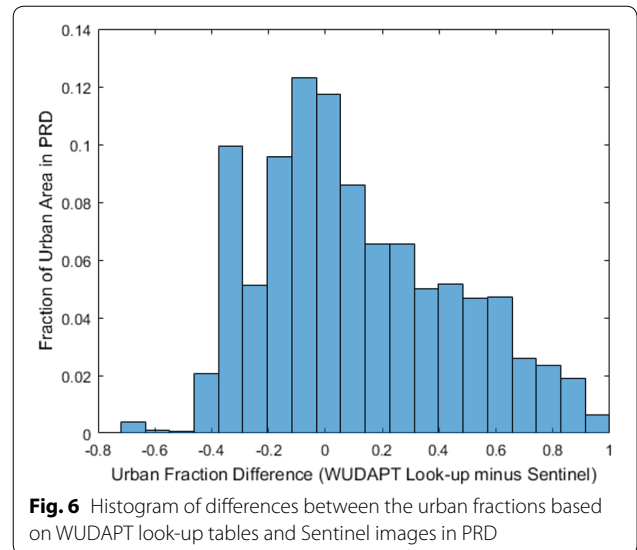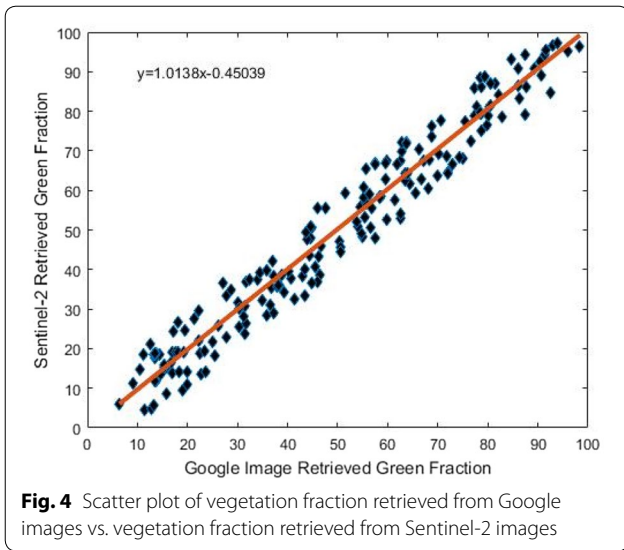
were less accurate or detailed than the coverage retrieved from the Google Satellite images (Fig. 7a). For Tsing Yi, WUDAPT underestimated the urban fraction despite the correct identification of the LCZs (large low-rise or heavy industry) are correctly identified, which owes to the way that the look-up table values for the urban canopy parameters (UCP) are assigned (in this case the urban fraction). When the look-up table is used, the median value is usually assigned to the whole study area, the Sentinel-2 retrieval method, in contrast, explicitly specifies a sub-grid scale for the vegetation/urban fraction, dependent on the geolocation. Therefore, the urban fraction retrieved from the Sentinel-2 images can provide information to represent the heterogeneity of the landscape at a sub-grid scale that is typically unavailable when using WUDAPT level-0 data for urban climate model parameterization, in which a single value (usually the mean or median) represents each LCZ. Our method, thus, offers an improvement for determining urban fractions for climate modeling research.
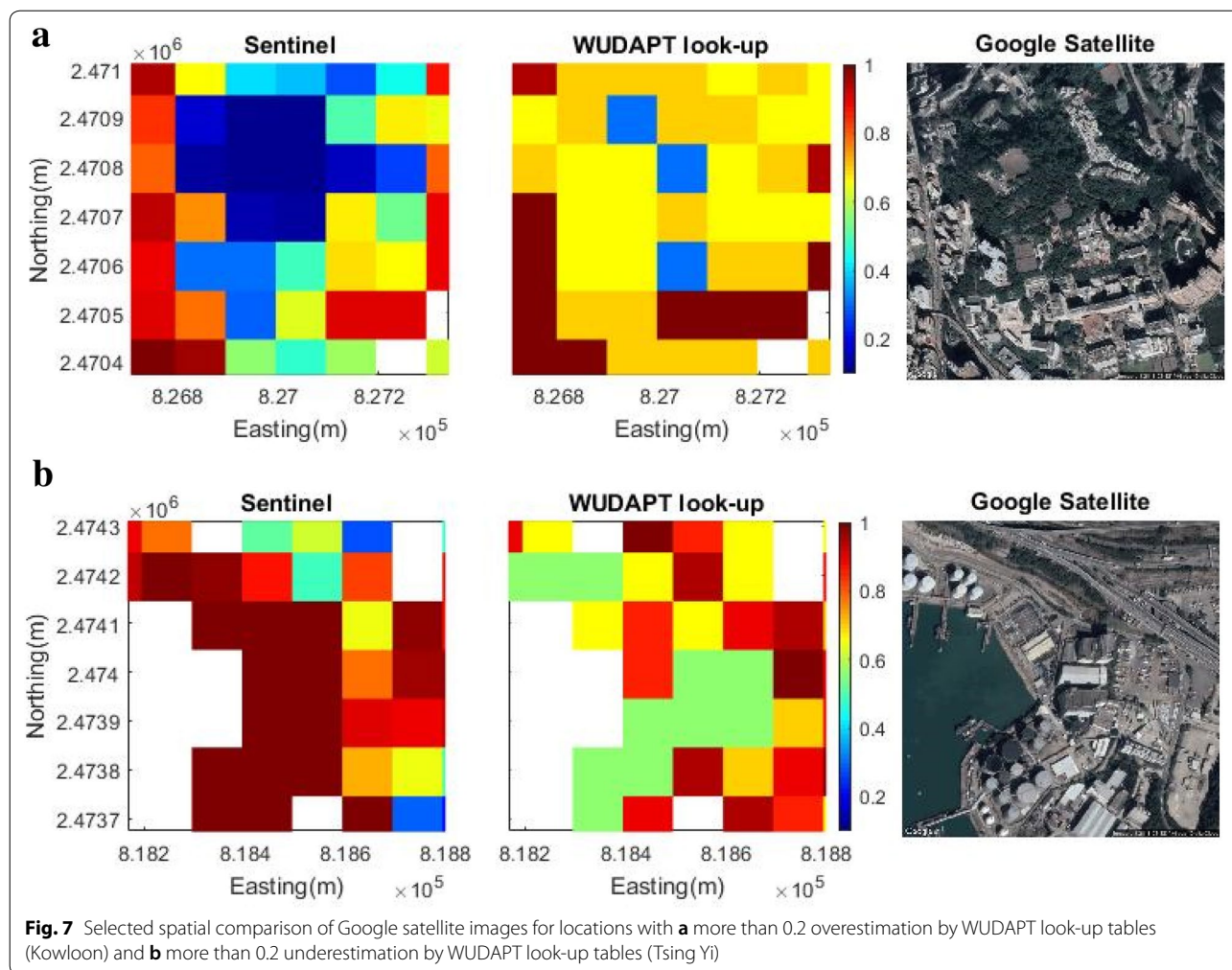
Conclusion

Global land-cover and local climate zone mapping is very useful for urban meteorological modeling. The urban fraction is a particularly important parameter as it affects the accuracy of urban meteorological simulations (e.g., Cui and De Foy 2012). This study uses relatively new and high-resolution satellite images from Sentinel-2 (in the red and near-infrared bands) to estimate the urban vegetation fraction based on the commonly used NDVI. The results show a high correlation (0.97) with reference estimates derived from a random sample of 200 high-resolution true color Google satellite images of the Pearl River Delta region. We compared the urban fractions estimated from this dataset with those derived from WUDAPT level-0 LCZ and default look-up table values, and found that the discrepancies between the two datasets was less than 0.2 for about half of the sampled urban areas in the PRD; however, the differences were also found to be as high as 0.8 for some sampled locations, either due to the WUDAPT data's inability to resolve individual parks or tree clusters in highly dense urban areas or because of
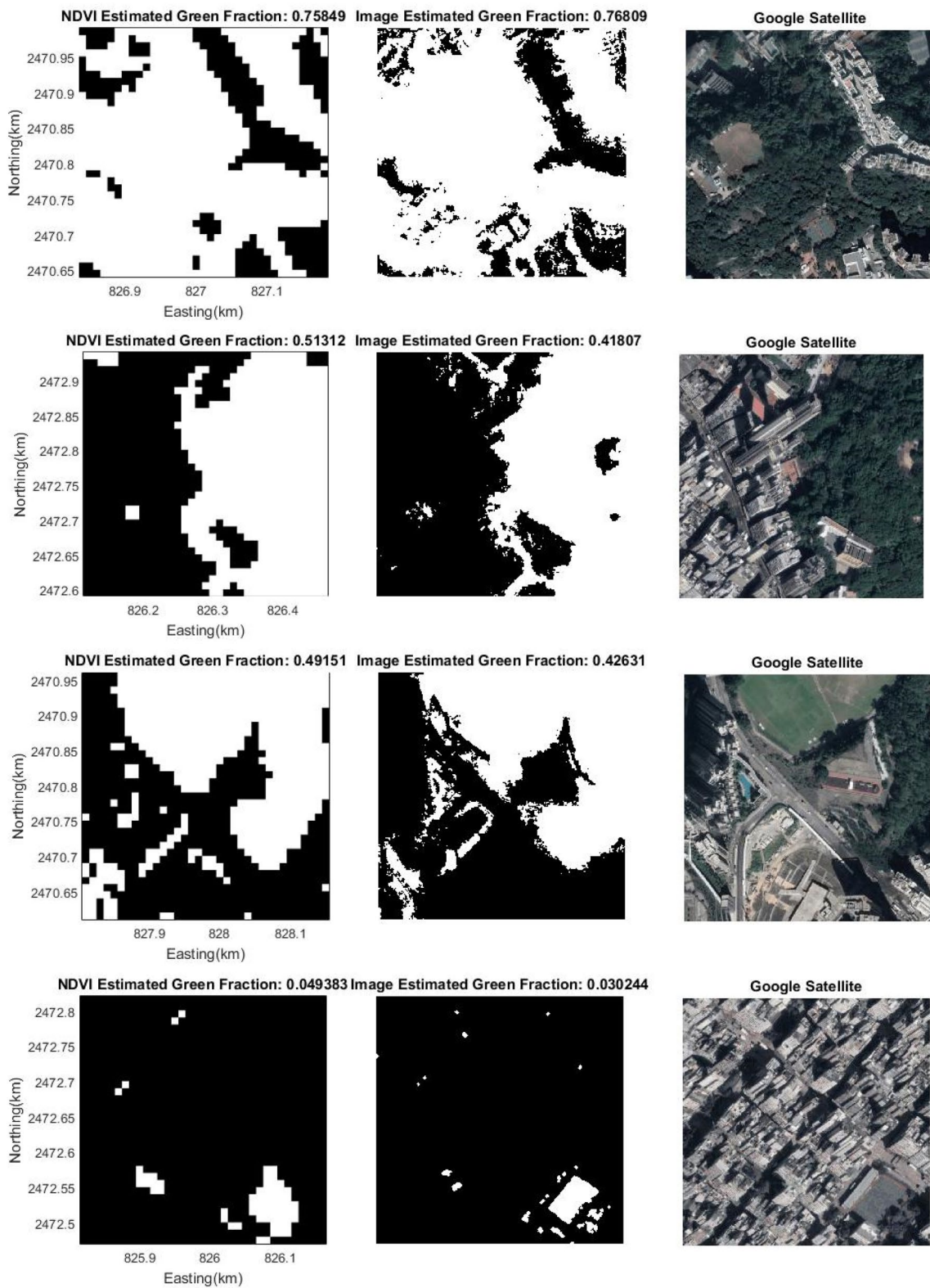
Fig. 3 Spatial comparisons of several locations in Hong Kong. Column 1: vegetation fraction retrieved from Sentinel-2. Column 2: vegetation fraction retrieved from Google Images. Column 3: Google static satellite images

Fig. 4 Scatter plot of vegetation fraction retrieved from Google images vs. vegetation fraction retrieved from Sentinel-2 images

Fig. 6 Histogram of differences between the urban fractions based on WUDAPT look-up tables and Sentinel images in PRD

Fig. 5 Spatial comparison of WUDAPT look-up table values vs. Sentinel-2-retrieved values in Hong Kong: **a** WUDAPT with default look-up table from Brousse et al. (2016); **b** Sentinel-2-retrieved values; and **c** WUDAPT minus Sentinel-2

Wong *et al. Geosci. Lett.* (2019) 6:2

Page 9 of 10

Fig. 7 Selected spatial comparison of Google satellite images for locations with **a** more than 0.2 overestimation by WUDAPT look-up tables (Kowloon) and **b** more than 0.2 underestimation by WUDAPT look-up tables (Tsing Yi)

the limited range in the median values for certain study areas. Overall, the vegetation fraction estimated using the Sentinel-2 images complements the existing WUDAPT level-0 data for urban meteorological modeling.

Additional file

Additional file 1: Figure S1. The study region and the corresponding four tiles of Sentinel-2 images (in RGB band).

Abbreviations
NDVI: Normalized Difference Vegetation Index; WUDAPT: World Urban Database and Access Portal Tools; PRD: Pearl River Delta; UHI: urban heat island; WRF BEP/BEM: Weather Research and Forecasting model with multi-layer building effect parameterization and the building energy model; NLCD: National Land Cover Database; RGB: red green blue; HSV: hue, saturation, value.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
Yes.

Consent for publication
Yes.

Ethics approval and consent to participate
Not applicable.

Wong *et al. Geosci. Lett.*   (2019) 6:2

Page 10 of 10

Publisher's Note

References

Brousse O, Martilli A, Foley M, Mills G, Bechtel B (2016) WUDAPT, an efficient land use producing data tool for mesoscale models? integration of urban LCZ in WRF over madrid. Urban Clim 17:116–134

Cai M, Ren C, Xu Y, Dai W, Wang XM (2016) Local climate zone study for sustainable megacities development by using improved WUDAPT methodology–a case study in Guangzhou. Proc Environ Sci 36:82–89

Carlson TN, Ripley DA (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens Environ 62(3):241–252

Chen J, Chen J, Liao A, Cao X, Chen L, Chen X et al (2015) Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J Photogr Remote Sens 103:7–27

Cheng H, Jiang XH, Sun Y, Wang J (2001) Color image segmentation: advances and prospects. Pattern Recogn 34(12):2259–2281

Ching J, See L, Mills G, Alexander P, Bechtel B, Feddema J et al (2014) WUDAPT: facilitating advanced urban canopy modeling for weather, climate and air quality applications. In: Proc. Amer. Meteorol. Soc. Symp. Urban Environ., pp 1–7

Ching J, Mills G, Bechtel B, See L, Feddema J, Wang X et al (2018) WUDAPT: an urban weather, climate and environmental modeling infrastructure for the anthropocene. Bull Am Meteorol Soc 99(9):1907–1924

Cui YY, De Foy B (2012) Seasonal variations of the urban heat island at the surface and the near-surface and reductions due to urban vegetation in Mexico City. J Appl Meteorol Climatol 51(5):855–868

Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F et al (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens Environ 120:25–36

Eckert S, Hüsler F, Liniger H, Hodel E (2015) Trend analysis of MODIS NDVI time series for detecting land degradation and regeneration in Mongolia. J Arid Environ 113:16–28

Elmore AJ, Mustard JF, Manning SJ, Lobell DB (2000) Quantifying vegetation change in semiarid environments: precision and accuracy of spectral mixture analysis and the normalized difference vegetation index. Remote Sens Environ 73(1):87–102

Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L et al (2013) Finer resolution observation and monitoring of global land cover: first mapping results with landsat TM and ETM data. Int J Remote Sens 34(7):2607–2654

Hammerberg K, Brousse O, Martilli A, Mahdavi A (2018) Implications of employing detailed urban canopy parameters for mesoscale climate modelling: a comparison between WUDAPT and GIS databases over Vienna, Austria. Int J Climatol 38:e1241–e1257

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C et al (2007) Completion of the 2001 national land cover database for the counterminous united states. Photogr Eng Remote Sens 73(4):337

Kusaka H, Kondo H, Kikegawa Y, Kimura F (2001) A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. Bound-Layer Meteorol 101(3):329–358

Martilli A, Clappier A, Rotach MW (2002) An urban surface exchange parameterisation for mesoscale models. Bound-Layer Meteorol 104(2):261–304

Salamanca F, Krpo A, Martilli A, Clappier A (2009) A new building energy model coupled with an urban canopy parameterization for urban climate simulations—part I. Formulation, verification, and sensitivity analysis of the model. Theor Appl Climatol 99(3):331. https://doi.org/10.1007/s00704-009-0142-9

Seto KC, Woodcock C, Song C, Huang X, Lu J, Kaufmann R (2002) Monitoring land-use change in the pearl river delta using landsat TM. Int J Remote Sens 23(10):1985–2004

Sobrino JA, Jimenez-Munoz JC, Paolini L (2004) Land surface temperature retrieval from LANDSAT TM 5. Remote Sens Environ 90(4):434–440

Vahmani P, Ban-Weiss GA (2016) Impact of remotely sensed albedo and vegetation fraction on simulation of urban climate in WRF-urban canopy model: a case study of the urban heat island in Los Angeles. J Geophys Res Atmos 121(4):1511–1531

Weng Q, Lu D, Schubring J (2004) Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. Remote Sens Environ 89(4):467–483