An evaluation of thermal Earth observation for characterising urban heatwave event dynamics using the urban heat island intensity metric

Tomas Holderness*†, Stuart Barr†, Richard Dawson†, Jim Hall‡

†School of Civil Engineering and Geosciences, Newcastle University, Newcastle-upon-Tyne, United Kingdom.

‡Environmental Change Institute, University of Oxford, Oxford, United Kingdom

Urban areas have a high sensitivity to extreme temperature events such as heatwaves due to increased absorption and re-radiation of thermal energy from man-made materials as well as anthropogenic heat outputs. Variations in urban form, land use and surface cover result in spatial variability in temperatures across urban areas, meaning that exposure to extreme events is variable at the sub-city scale. Such variability must be quantified in order to better understand urban temperature interactions and identify areas with the greatest potential exposure to extreme heatwave events. Earth observed data offer a spatially complete and homogenous data source to supplement observations from sparse weather station networks in order to quantify the spatial temperature variability across cities.

This paper presents an evaluation of the thermal data acquired by the Advanced Very High Resolution Radiometer (AVHRR) instrument to quantify the spatial temperature dynamics of London. 81 cloud-free AVHRR scenes from summers between 1996 and 2006 are analysed in association with air temperature measurements from four London weather stations in order to characterise the year-on-year temperature dynamics of London. The data were employed to investigate the viability of using AVHRR scenes to distinguish a heatwave year from background years using the commonly employed Urban Heat Island Intensity (UHII) metric. Results show that AVHRR thermal data are highly sensitive to local-meteorological and diurnal effects, requiring temporal averaging to the monthly and seasonal scales to provide robust data for a comparison between different years. Resulting UHII scenes highlight the spatial variability of intensity across London. However, comparison of UHII scenes between summers indicate that the UHII metric is a relatively poor means by which to distinguish between a heatwave summer in London and the 75th, median and 25th percentile summer temperatures of the time series investigated.

1 Introduction

Future climate scenarios predict a rise in average summer temperatures, as well as an increase in the frequency and duration of related heatwave events (IPCC 2007). Major urban conurbations will be particularly sensitive to more extreme and longer heatwaves due to higher population densities and urban induced climatic modifications such as the Urban Heat Island (UHI) effect (Kovats et al. 2006, Di Sabatino et al. 2009), where cities on average
experience increased temperatures compared to surrounding rural hinterlands. A consequence of effects such as the UHI is that cities will experience a greater socio-economic impact from extreme temperature events such as heatwaves, with increased levels of mortality and human discomfort (Kovats et al. 2006, Di Sabatino et al. 2009), such as that seen in the 2003 heatwave in London, where deaths of Londoners over the age of 75 increased by 59% (Johnson et al. 2005). In order to mitigate and adapt to temperature rises and heatwave events that may result from future climate change it is necessary to quantify urban temperature dynamics in order to understand how different components of the urban system, such as energy demand for cooling, respond to thermal stress (Knight et al. 2010). Critically, this needs to be performed in a spatially explicit manner as urban population vulnerability to heat is not just a function of overall population demographics (such as age), but also a function of how they are distributed spatially across the city (Eliasson and Svensson 2003).

Traditionally, urban temperatures have been measured in terms of near surface air temperature (screen-level) using terrestrial weather station networks and quantified using the Urban Heat Island Intensity (UHII) metric (Oke 1987, Kim and Baik 2002, Di Sabatino et al. 2009); the maximum difference between urban and background rural temperatures for a given point in time during one diurnal cycle (Oke 1987, Kim and Baik, 2002, Kolokotroni and Giridharan 2008). In developed nations, weather station networks allow long time-series high temporal frequency data-sets to be generated and analysed. However, in many cities, weather station networks lack a sufficient spatial distribution to characterise intra-urban temperature dynamics (Eliasson and Svensson 2003, Knight et al. 2010). This is a major limiting factor to improving our understanding of intra-urban heat hazard, and ultimately therefore the spatial vulnerability and exposure to heat risks of the population living and working within cities (Eliasson and Svensson 2003, Hung et al. 2006, Knight et al. 2010).

A number of studies in the 1980’s introduced the use of thermal remote sensing to study the UHI effect (e.g., Balling and Brazel 1988, Roth et al. 1989), the success of which have led to thermal images acquired by Earth observation satellites sensors being routinely employed to investigate the UHI effect for cities around the world (e.g., Alabama, Beijing, Birmingham, Indianapolis, Kano, London, Los Angeles, Nagoua, Pyongyang, Seattle, Seoul, Shanghai, Singapore, Tokyo, and Vancouver (Roth et al. 1989, Lo et al. 1997, Kim and Baik 2002, Nichol 2003, Kato and Yamaguchi 2005, Hung et al. 2006, Lu and Weng 2006, Kolokotroni and Giridharan 2008, Tomlinson et al. 2010). Thermal Earth observation techniques are often chosen to replace or supplement in-situ data with estimated measurements of surface temperature due to their spatially complete nature (Tomlinson et al. 2010). They have been employed where it is difficult to monitor certain urban thermal characteristics such as anthropogenic heat distribution using in-situ observations (Dousset and Gourmelon 2003), and when terrestrial observations are unavailable, as is often the case in developing nations (Nichol 2003).

(Voogt and Oke 2003, Lu and Weng 2006). For example, ASTER images have been employed to correlate spatial temperature variations with biophysical descriptors and impervious surface cover (Lu and Weng 2006), to understand the spatial pattern of night-time temperatures (Kato and Yamaguchi 2005), quantify the seasonal variability in UHII over the City of Beijing (Cai et al. 2011) and estimate the storage heat flux in urban areas (Kato and Yamaguchi 2007). Landsat TM data have been used to understand the spatial variation of daytime surface temperatures of African cities (Nichol 2003), as well as undertake analysis of the relationships between land cover change and increased urban surface temperatures (Lo and Quattrochi 2003), and in particular the correlation between vegetation cover and the urban heat island (Weng 2003, Weng et al. 2004).

However, while such medium spatial resolution thermal images allow a spatially detailed characterisation of surface temperatures (Nichol 1994, Streutker 2002, Streutker 2003, Weng 2003, Weng et al. 2004), it is limited in terms of its utility to monitor temperature dynamics either diurnally, seasonally and at an intra-annual level (Hung et al. 2006, Voogt and Oke 2003, Cheval and Dumitrescu 2009, Cheval et al. 2009). In order to capture, and hence study, the temporal dynamics of urban surface temperatures thermal images acquired by spatially coarser sensors such as AVHRR and MODIS are preferred due to their greater spatial coverage, multiple daily acquisition frequency and long baseline record of acquired scenes (Hung et al. 2006, Stathopoulou and Cartalis 2009, Neteler 2010, Dousset et al. 2011). Such data have been employed to study the UHI effect at a range of temporal sampling intervals including detailed studies of the UHI diurnal cycle within a single year (Roth et al. 1989, Gallo et al. 1993, Dousset et al. 2011) or over multiple years (Lee 1993, Tomlinson et al. 2010). Other studies have investigated the UHI effect at a more aggregated level, performing monthly comparisons between multiple years of data (Gallo et al. 1993, Gallo et al. 1995, Cheval and Dumitrescu 2009, Cheval et al. 2009). Equally, a number of studies have employed a single point-in-time characterisation of the UHI for a number of cities within a specific geographical/climatic region (Hung et al. 2006, Pongracz et al. 2006, Stathopoulou and Cartalis 2009).

Such studies highlight the utility of Earth observation to study urban temperature dynamics and the UHI effect. However, significant issues need to be addressed in order to ensure that Earth observation data are employed in a consistent and objective manner within studies of urban temperature dynamics (Voogt and Oke 2003). For example, a major unresolved issue is how one should combine and utilise both air temperature measurements acquired by weather stations and surface temperature estimates derived from thermal Earth observed images (Voogt and Oke 2003, Tomlinson et al. 2010). At present there is no accepted de-facto means by which such data can be combined outside of empirical relationships (Voogt and Oke 2003). Moreover, while a number of studies have shown some success in characterising and quantifying the spatial relationship between Earth observed surface temperatures of cities and their underlying land cover fabric, land use activities and urban form (Nichol 2003, Kato and Yamaguchi 2005, Tomlinson et al. 2010), considerable work remains to be undertaken before strong and reliable empirical relationships between surface temperatures and urban form and function can be used in heat hazard impact and adaption studies (Weng et al. 2004, Hung et al. 2006, Cai et al., 2011.).

Indeed, even if the interest lies in just employing Earth observation data to characterise the spatial pattern of surface temperatures, careful consideration needs to be taken with regards to choosing a suitable time series of imagery acquired over an appropriate time-period in order to be able to realistically characterise the urban temperature property of interest (Nichol 2003, Kato and Yamaguchi 2005). For example, studies employing MODIS data and Landsat TM data to analyse the temporal variations of UHII between eight ‘mega’ cities in Asia (Hung et al. 2006) highlighted the importance of having a sufficient time-series
of data to be able to not only derive metrics that are insensitive to local diurnal fluctuations, such as metrological conditions, but also able capture the long term spatial and temporal temperature dynamics required for monitoring (Hung et al. 2006). In this regard, it is worth noting that spatial/temporal sampling and time series length has also been recognised as a critical consideration in UHI studies using in-situ weather station screen-level temperature data (Jones and Lister 2009).

With respect to the last issue above, this study investigates the utility of AVHRR thermal images to capture, characterise and quantify the magnitude of a summer heatwave event compared to non-heatwave years. We evaluate the commonly employed UHII metric in this respect using daytime AVHRR scenes. A comparison between AVHRR daytime UHII and UHII derived for in-situ weather stations is performed, along with an analysis of how well both UHII metrics (satellite and weather station) capture a heatwave event. The study is conducted for the city of London UK focusing on a 10-year period between 1996 and 2006. In order to evaluate the ability of AVHRR data to characterise the temperature dynamics of London during the 10-year time period under consideration a detailed analysis is performed of how temporal averaging of the available scenes influences the data’s ability to characterise the underlying temperature regime of London, and how this impacts upon the generation of the UHII metric for different summer periods.

2 Data acquisition and pre-processing

2.1 London weather station data

This project utilized two separate data-sets for analysis. The first was the hourly and daily (diurnal cycle) screen-level (1.25m above surface height) air and surface level temperature observations collected at United Kingdom Meteorological Office weather stations in the British Isles. The data are freely available to researchers within the United Kingdom from the British Atmospheric Data Centre (BADC) and has been widely utilised in a number of urban temperature studies (Jones and Lister 2009, Tomlinson et al. 2010). The data comprise of a series of comma separated value (CSV) files detailing the selected observations and their station of observation. For this study the key measurements of interest were screen-level air temperature and where available grass and concrete surface temperature observations, measured in degrees Celsius (°C).

The raw weather station data were uploaded to a purpose built PostGIS spatial database, which linked temperature observations with geo-referenced weather station locations using UK Ordnance Survey British National Grid coordinates. Analysis of the BADC data showed that only 4 weather stations in London had continuous high quality data for the time period under investigation (1996-2006), namely; London Weather Centre (LWC, elevation: 43 metres), St James’s Park (SJP, elevation: 5 metres), Heathrow airport (LHR, elevation: 25 metres) and Northolt (NTH, elevation: 40 metres). Additionally, for the purposes of generating the UHII metric the rural weather station at High Wycombe (22 kilometres from the Greater London Authority (GLA) boundary) was also selected (Figure 1).
For the 10-year time period under consideration average monthly summer temperatures (1st June to 31st August) were generated for each London weather station from their corresponding daily averages in the PostGIS database. Each summer was then ranked according to its average summer daily temperature and AVHRR scenes from the hottest, 75th, median and 25th percentile summers were selected for further analysis. Selected summers and their corresponding average screen-level air temperatures for all London weather stations were: hottest 2003 (19.36°C), 75th-percentile 1997 (18.24°C), median 2001 (17.93°C) and 25th-percentile 2002 (17.45°C). The four summers were chosen in this manner to keep the number of AVHRR scenes to process viable, while ensuring that the general temperature record of London for the period under consideration was captured.

2.2 AVHRR image selection and pre-processing

AVHRR scenes corresponding to the four summers under investigation were selected from an existing archive of 2400 scenes covering May-September, 1985-2008. This archive was provided by the UK Natural Environment Research Council (NERC) Dundee Satellite Receiving Station. Scenes within the archive are cloud free for the centre of the Greater London Area, georectified in GeoTiff format to the British National Grid and calibrated to top of atmosphere albedo (%) or top of atmosphere (TOA) brightness temperature (°K) for the optical and thermal bands respectively.

In this study the interest lay in accessing the utility of AVHRR data to characterise daytime temperature dynamics of London. The choice of daytime scenes was taken for a
number of reasons. Firstly, recent work on the UHII of London had revealed that maximum air temperature UHII in London occurs in the daytime (Kolokotroni and Giridharan 2008). Thus, in this study we evaluate the utility of AVHRR data to express this feature. Secondly, significantly more scenes were available for daytime hours compared to night-time. In relation to this point, the check for cloud contamination employed is more reliable when employing daytime images than night-time data. Finally, the use of daytime images allowed us to employ a surface emissivity correction procedure based on the use of the Normalised Difference Vegetation Index (NDVI) using AVHRR bands 1 and 2 (Van de Griend and Owe, 1993). Clearly, night-time image acquisitions would not allow such a correction procedure to be employed.

A search of the full image archive of daytime (09:00 to 17:00) acquisitions for the four summers under consideration resulted in a total of 162 scenes being extracted. Although all 2400 scenes of the full archive were tagged as cloud free at the centre of London, the 162 selected were further tested for possible cloud contamination at the edge of London. This was achieved by comparing the minimum TOA brightness temperature of each scene to its corresponding daily minimum grass surface temperature recorded by the London Heathrow weather station. If the AVHRR scene temperature was more than 3°C lower than the daily minimum grass surface temperature recorded then it was tagged as cloud contaminated. The 3°C threshold was selected on the basis that a ±3°C uncertainty may be expected to exist between uncorrected AVHRR TOA temperature and corresponding surface temperature (Cooper and Asrar 1989). This procedure resulted in a number of further scenes being rejected primarily due to cloud contamination over the rural station. In total 81 scenes were accepted; 43, 5, 14 and 19 scenes for 2003, 1997, 2001 and 2002 respectively. This procedure was then repeated over the 81 selected scenes to check for cloud contamination at the rural station, resulting in a further 6 scenes being identified as cloudy. In order to maintain as many scenes in each summer as possible the full suite of 81 scenes (cloud free over the GLA area) were used to characterise the surface temperatures of the Greater London Area (GLA). However, when conducting analysis using the rural station location to derive the UHII metric, the 75 scenes that were cloud free over the GLA extent and the rural weather station location were used.

In order to derive estimated surface temperatures from the AVHRR scenes it was necessary to correct for atmospheric attenuation and surface emissivity of the thermal bands. Atmospheric correction can be achieved using either single channel or split-window techniques (Erbersteder et al. 1999, Yu et al. 2008). Single channel techniques using temperature and humidity profiles from observations or atmospheric models can be time consuming to implement operationally (Erbersteder et al. 1999), and there is not a clear consensus within the literature of the approach which gives the best results (Cooper and Asrar 1989). However, split-window techniques which make use of sensors that have two or more thermal bands provide an effective atmospheric correction to brightness temperature measurement and are widely used due to their robustness and simplicity (Yu et al. 2008). Thus, this study limited itself to using multi-channel AVHRR scenes in order to employ a split-window approach.

To correct for surface emissivity a range of methods exist, as reviewed by Dash et al. (2002), many of which require a-priori knowledge of the surface emissivity or provide only relative emissivity values (Sobrino et al. 2008). However the German Aerospace Centre provide an operational approach which derives estimated surface temperatures using a combined split-window technique for correcting atmospheric attenuation (Becker and Li, 1990) and a surface emissivity correction based on the use of the Normalised Difference Vegetation Index (NDVI) using AVHRR bands 1 and 2 (Van de Griend and Owe 1993); an approach which has been found to be computationally efficient when processing large
volumes of scenes in an automated manner (Tungalagsaikhan and Guenther 2007). In this approach, estimated surface temperature (EST) is derived on the basis of bands 4 and 5 of the AVHRR instrument:

\[
T = 1.274 + \frac{(T4+T5)}{2} \times \left(1 + 0.15616 \left(\frac{1-e}{e} - \frac{0.482de}{e^2}\right) + \left(\frac{(T4-T5)}{2 \times (6.26+3.989(1-e)^{1/2}+3.33de)}\right)\right)
\]

(1)

where \(T\) is estimated surface temperature (°C), \(T4\) and \(T5\) are the black-body temperature of bands 4 and 5 of the AVHRR instrument (top of atmosphere (TOA) brightness temperature) and \(e\) and \(de\) represent the emissivity correction factors given by:

\[
e = \frac{(e4+e5)}{2}
\]

(2)

\[
de = e4 - e5
\]

(3)

where \(e4\) and \(e5\) represent the emissivity in bands 4 and 5 of the AVHRR instrument and are estimated on the basis of:

\[
e4 = 1.0094 + 0.047 \ln(NDVI)
\]

(4)

\[
e5 = e4 + 0.01
\]

(5)

In Equation 4 the NDVI (Normalised Difference Vegetation Index) is scaled linearly between 0–255 (Van de Griend and Owe 1993):

\[
NDVI = \left(\frac{b2-b1}{b2+b1} + 1\right) \times 127
\]

(6)

A potential limitation of this approach is the requirement for two thermal bands which were only available on AVHRR versions 2 and 3 from 1994 onwards with the launch of NOAA-14. However, as the period of interest of this study was between 1996 and 2006 dual channel thermal data were available for all years, although as NOAA-15 was not launched until May 1998 the number of dual channel scenes available was limited in the period up to this date compared to post 1998. In the emissivity correction procedure (Equation 1), no atmospheric correction for bands 1 and 2 was undertaken in relation to the calculation of NDVI (Equation 6). Given the number of scenes under investigation (81) and the fact that there is a lack of consensus in the literature as to a standard operational approach to atmospheric correction of historical data in the absence of in-situ atmospheric profile data (Teillet, 1992, Erberstedter et al. 1999) the decision not to perform atmospheric correction could be justified.

Nonetheless, the lack of atmospheric correction potentially introduces an uncertainty in the NDVI calculations ranging from ±0.09 during clear conditions (Aerosol Optical Thickness (AOT) < 0.05) to ±0.13 during average conditions (AOT = 0.05 ≥ 0.25) (Nagol et al. 2009). This may result in error propagating into the emissivity correction. In order to assess this, one can invert Equation 4 to obtain the expected NDVI for a typical emissivity of an urban surface (\(\varepsilon \leq 0.92\); Nichol 1994) and vegetation (\(\varepsilon \geq 0.98\); Dash et al. 2005). By adding the expected NDVI error of ±0.13 for average sky conditions to these and applying Equation 4 one obtains an emissivity error of ±0.022 or greater for urban surfaces, and ±0.007 for vegetation. In turn, this emissivity error will propagate into the subsequent estimation of surface temperature. On the basis of work by Sobrino et al. (1991) that found that emissivity accuracy needs to be within ±0.005 to get a surface temperature error below ±0.4K, and the
work of Schadlich et al. (2001) and Dash et al. (2002) who found for mid-latitude areas an emissivity error of ±0.025 gave an error of ±2K in estimated surface temperature, then an initial NDVI error of ±0.13 may be expected to result in ±2K error over urban pixels and ±0.4K over vegetated areas. These errors may be compounded when calculating metrics such as UHII that are based on the difference between urban and rural temperatures, giving a total maximum error in the order of ±2.4K.

3 Generation and analysis of AVHRR derived UHII

3.1 Temporal sensitivity of AVHRR estimated surface temperatures

One of the key strengths of AVHRR data is the high revisit time between scenes, due to the large number of satellites carrying the AVHRR sensor, resulting in up to a maximum of 4 scenes per day being acquired (Cracknell 1997). This offers the potential to characterise temperature dynamics at a daily, weekly, monthly and seasonal time scale. While a positive aspect of the sensor, it is important to recognise that understanding temperature dynamics, particularly in relation to heatwave years compared to non-heatwave years, means that data from different years needs to be compared. This raises the question of what is the ideal temporal frequency of AVHRR data to employ in order to capture consistently and objectively changes in temperature that are a result of a heatwave event rather than other erroneous factors such as local meteorological weather conditions. Temporal averaging has been intensively investigated in marine sea surface temperature studies using AVHRR data (e.g., Gentemann et al. 2003). In such studies, scenes of sea surface temperature are averaged over one week and a weighting applied to night-time scenes in order to minimize diurnal signals (Gentemann et al. 2003).

In order to investigate the sensitivity of urban temperature analysis and measurement of the UHII metric to temporal averaging, statistical comparisons were performed for several levels of aggregation between 2003, the heatwave year, and the three other years under investigation. At the level of individual scenes comparisons were made between scenes acquired on the same Julian day of the year within 65 minutes of each other. At the next level, individual scenes were aggregated to daily averages, and then the 4 summers were aggregated to the monthly averages (June, July and August). The final level of aggregation involved aggregating scenes to summer averages resulting in just a single scene of estimated summer surface temperature per-year. For each level of temporal aggregation/averaging the non-parametric pair-wise Mann-Witney U statistical test of distributions was employed to test if statistically significant differences existed between 2003 scenes and corresponding dates or aggregations (monthly and yearly) of the other 3-years under investigation. This was applied such that spatially coincident pixels from a scene in 2003 was paired with its corresponding location (pixel) in either a monthly or summer scene from one of the other 3-years under investigation. In the statistical tests, a one-tail test was employed such that the expectation was that the temperatures from 2003 were expected to be greater than the other years analysed. In all tests a 95% significance level was employed in order to reject the null hypothesis in favour of an alternative hypothesis that the distributions were drawn from different populations.

3.2 AVHRR estimated surface temperature UHII calculation

For the purposes of this study the Urban Heat Island Intensity metric is formally defined as the maximum difference between urban and background rural temperatures for a given point in time during one diurnal cycle (Oke 1987, Kim and Baik, 2002, Kolokotroni and Giridharan 2008). The weather station air temperature observations from the four urban weather stations
and single rural station were used to generate the UHII metric in this manner via an SQL script that automated this process and populated a new table in the PostGIS BADC database containing one UHII value per day for the Greater London Authority. Given the noted importance of night-time temperatures on the UHII (Oke 1987, Kim and Baik, 2002, Kolokotroni and Giridharan 2008) and the aim of this study to evaluate how well daytime AVHRR scenes can capture and distinguish heatwave years from non-heatwave years compared to established methods employing the full diurnal cycle, air temperature based UHII values were generated from hourly values from the full diurnal cycle at each weather station.

To calculate daytime UHII using AVHRR estimated surface temperatures a slightly different approach had to be taken due to the reduced observation frequency of the AVHRR data compared to the terrestrial weather station data. At best, a maximum of two scenes were available per diurnal cycle. Therefore, the UHII metric was calculated for each individual AVHRR estimated surface temperature scene by subtracting the estimated surface temperature of the pixel containing the rural weather station location (High Wycombe) from the estimated surface temperature values of the pixels that fell within the Greater London Authority boundary in a manner similar to that used by Tomlinson et al. (2010) to derive UHII measurements from MODIS estimated surface temperature scenes. As in the case of the estimated surface temperatures, individual scenes from 2003, 1997, 2001 and 2002 were aggregated to the daily, monthly and yearly averages. Again, the non-parametric pair-wise Mann-Whitney test was applied to test whether the UHII scenes for 2003 were statistically greater than those of other years investigated. As in the statistical testing of the estimated surface temperatures a one-tail 95% significance level was employed for this analysis.
4 Results and Discussion

Figure 2 shows estimated surface temperature for four AVHRR scenes from the 14th August 2002 and 2003 captured within a 65 minute window of each other. For both years there is a notable increase in estimated surface temperature during the time period of observation (24.46°C at 10:56am to 26.90°C at 12:56pm for 2003; 21.22°C at 10:59am to 21.77°C at 12:08pm), although the magnitude of the rise is greater for 2003 (1.2°C/hr) as compared to 2002 (0.48°C/hr). This more dramatic temperature increase for the heatwave year is supported to some extent by the corresponding BADC weather station data which revealed an average increase in air temperature for London weather stations of 1.68°C for 2003 compared to 0.35°C for 2002 over the same time period. It may be tempting, therefore, to consider that the AVHRR data captures the diurnal progression of surface temperature dynamics and differentiates between summers with significantly varying temperatures.

![Figure 2. AVHRR estimated surface temperatures for London scenes acquired on 14th August 2003 and 2002.](image)

However, Table 1 reveals a more complex relationship exists between individual scenes (within a 65-minutes of a 2003 scene) and daily averages of estimated surface
temperatures for the 75th, median and 25th-percentile years compared to the 2003 heatwave year. Both individual scenes and daily mean differences show little correlation with the percentile ordering of the summers in terms of their overall BADC average summer temperatures. For example, mean scene differences between 2003 and the 75th-percentile year of 1997 range from 1.04°C to 9.64°C which is greater than that of the 25th-percentile year of 2002 (-0.04°C to 5.11°C). At a daily level both 1997 (75th-percentile year) and 2001 (50th-percentile year) have noticeably larger absolute mean differences (5.66°C and 5.19°C respectively) than 2002 (25th-percentile year: 2.04°C). Equally, in the case of daily maximum differences, the 25th-percentile year is closer to the heatwave year (9.12°C daily mean maximum difference) relative to the 75th-percentile year (1997: 13.79°C daily mean maximum difference) and 50th-percentile year (2001: 10.55°C daily mean maximum difference). One would expect the greatest difference for individual scenes and daily averages to be for the 25th-percentile year (2002) and smallest for the 75th-percentile year (1997). Moreover, in terms of maximum individual difference, 3 of the 4 largest overall differences (11.80°C, 15.18°C and 14.39°C) are for the 75th-percentile year (1997). This suggests that comparison of AVHRR estimated surface temperatures at a single-scene and daily diurnal level is an unreliable means by which to compare summer temperature dynamics of different years. A probable explanation of these results is that they are strongly influenced by meteorological signals (Hung et al. 2006), and in the case of daily averages the limited number of scenes available (Nichol 2003). In combination this implies that daily averaging is too fine a temporal quantisation for the analysis of estimated summer surface temperatures on a year-by-year basis.

<table>
<thead>
<tr>
<th>2003 Scene</th>
<th>2003 scene time</th>
<th>Comparison Scene</th>
<th>Time difference (minutes)</th>
<th>Mean difference (°C)</th>
<th>Maximum difference (°C)</th>
<th>Daily absolute mean difference (°C)</th>
<th>Daily absolute mean maximum difference (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003/08/09</td>
<td>12:12</td>
<td>1997/08/09</td>
<td>64</td>
<td>1.04</td>
<td>11.80</td>
<td>5.66</td>
<td>13.79</td>
</tr>
<tr>
<td>2003/08/09</td>
<td>13:52</td>
<td>1997/08/09</td>
<td>-36</td>
<td>6.31</td>
<td>15.18</td>
<td>6.31</td>
<td>15.18</td>
</tr>
<tr>
<td>2003/06/24</td>
<td>11:48</td>
<td>2001/06/24</td>
<td>65</td>
<td>-6.06</td>
<td>8.32</td>
<td>-6.06</td>
<td>8.32</td>
</tr>
<tr>
<td>2003/08/14</td>
<td>12:56</td>
<td>2001/08/14</td>
<td>-34</td>
<td>4.32</td>
<td>12.78</td>
<td>4.32</td>
<td>12.78</td>
</tr>
<tr>
<td>2003/08/14</td>
<td>10:56</td>
<td>2002/08/14</td>
<td>4</td>
<td>3.40</td>
<td>15.06</td>
<td>3.40</td>
<td>15.06</td>
</tr>
<tr>
<td>2003/08/14</td>
<td>12:56</td>
<td>2002/08/14</td>
<td>-48</td>
<td>5.11</td>
<td>10.22</td>
<td>5.11</td>
<td>10.22</td>
</tr>
<tr>
<td>2003/08/14</td>
<td>12:56</td>
<td>2002/08/14</td>
<td>53</td>
<td>0.97</td>
<td>6.85</td>
<td>0.97</td>
<td>6.85</td>
</tr>
<tr>
<td>2003/08/14</td>
<td>14:37</td>
<td>2002/08/14</td>
<td>48</td>
<td>-0.04</td>
<td>4.35</td>
<td>-0.04</td>
<td>4.35</td>
</tr>
</tbody>
</table>

Table 1. Individual and daily scene differences between estimated surface temperature for 2003 (heatwave) and 1997, 2001 and 2002.

In general the monthly data seems to exhibit a slightly better expected pattern of temperature distribution (Table 2). Firstly, apart from June, the monthly mean estimated surface temperatures follow the expected pattern of the percentile years derived from an analysis of the BADC weather station air temperature data of London. Only June 2001 falls outside this ordering with a mean estimated surface temperature of 26.79°C. The stability of the percentile year ordering is not, however, as evident in the mean minimum and maximum monthly temperatures (Table 2), showing a greater variability with months for cooler years such as 2001 and 2002 exhibiting quite high mean minimum and maximum monthly estimated surface temperatures (e.g., June 2001 mean minimum of 17.98°C (3.00°C>2003).
and mean maximum of 32.15°C (4.87°C>2003). When further averaging is performed to the full summer season level (1st June to 31st August; Figure 3) the pattern of mean temperatures being consistent with the weather station percentile ordering persists (Table 2). Again, the mean minimum and maximum yearly estimated surface temperatures show a less consistent relationship, and in the case of the mean maximum estimated temperatures the similarity of the values perhaps suggests that at the summer level distinguishing extreme temperature differences may be problematic. However, Mann-Whitney U tests revealed that all 2003 months were statistically hotter than their corresponding month in 1997, 2001 and 2002 at a 95% significance level (Table 3). This implies that it is possible spatially on the basis of the overall distribution of estimated surface temperatures at a monthly level to distinguish the heatwave year as the statistical test applied is pair-wise based on corresponding pixel locations. In a similar manner, at the level of averaging over an entire summer year, all but 2001 (the median year) were found to be statistically cooler than 2003 at a 95% significance level (Table 3).

<table>
<thead>
<tr>
<th>Month/Year</th>
<th>Mean EST (°C)</th>
<th>Mean minimum EST (°C)</th>
<th>Mean maximum EST (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 2003</td>
<td>22.17</td>
<td>14.98</td>
<td>27.28</td>
</tr>
<tr>
<td>June 1997</td>
<td>20.52</td>
<td>13.92</td>
<td>24.61</td>
</tr>
<tr>
<td>June 2001</td>
<td>26.79</td>
<td>17.98</td>
<td>32.15</td>
</tr>
<tr>
<td>July 2003</td>
<td>27.55</td>
<td>21.54</td>
<td>31.94</td>
</tr>
<tr>
<td>July 1997</td>
<td>Too few scenes to generate monthly statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>July 2001</td>
<td>22.90</td>
<td>17.14</td>
<td>27.05</td>
</tr>
<tr>
<td>July 2002</td>
<td>17.17</td>
<td>10.98</td>
<td>24.18</td>
</tr>
<tr>
<td>August 2003</td>
<td>26.35</td>
<td>21.52</td>
<td>29.62</td>
</tr>
<tr>
<td>August 1997</td>
<td>Too few scenes to generate monthly statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>August 2001</td>
<td>24.50</td>
<td>18.90</td>
<td>28.25</td>
</tr>
<tr>
<td>August 2002</td>
<td>21.56</td>
<td>15.12</td>
<td>26.96</td>
</tr>
<tr>
<td>Summer 2003</td>
<td>25.94</td>
<td>20.75</td>
<td>29.46</td>
</tr>
<tr>
<td>Summer 1997</td>
<td>25.31</td>
<td>19.32</td>
<td>29.62</td>
</tr>
<tr>
<td>Summer 2001</td>
<td>25.83</td>
<td>19.84</td>
<td>30.32</td>
</tr>
<tr>
<td>Summer 2002</td>
<td>21.07</td>
<td>16.16</td>
<td>24.93</td>
</tr>
</tbody>
</table>

Table 2. Monthly and summer spatial averages of estimated surface temperatures (EST) for the Greater London Authority for 2003, 1997, 2001 and 2002 summers derived from AVHRR data.

<table>
<thead>
<tr>
<th>Year</th>
<th>Umin Calculated 1-tail X&gt;Y</th>
<th>Reject Null @ 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 2003-1997</td>
<td>713661</td>
<td>Yes</td>
</tr>
<tr>
<td>June 2003-2001</td>
<td>239243</td>
<td>Yes</td>
</tr>
<tr>
<td>June 2003-2002</td>
<td>1039891</td>
<td>Yes</td>
</tr>
<tr>
<td>July 2003-2001</td>
<td>119851</td>
<td>Yes</td>
</tr>
<tr>
<td>July 2003-2002</td>
<td>1835</td>
<td>Yes</td>
</tr>
<tr>
<td>August 2003-2001</td>
<td>498324</td>
<td>Yes</td>
</tr>
<tr>
<td>August 2003-2002</td>
<td>108144</td>
<td>Yes</td>
</tr>
<tr>
<td>Summer 2003-1997</td>
<td>988104</td>
<td>Yes</td>
</tr>
<tr>
<td>Summer 2003-2001</td>
<td>1113169</td>
<td>No</td>
</tr>
<tr>
<td>Summer 2003-2002</td>
<td>98090</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Reject Null if Umin Calculated < Umin Critical (1110722 for 1-tail test X>Y @ 95%; X=2003 summer heatwave year).
Table 3. Mann-Whitney statistical results comparing monthly and yearly AVHRR estimated surface temperatures.

Figure 3 AVHRR Summer estimated surface temperatures for the Greater London Authority with temporal averaging at the seasonal level.

The importance of choosing the correct temporal and spatial aggregation level is further highlighted in Table 4 where the summer season average estimated surface temperatures of individual pixels containing the urban and rural weather stations are presented. This shows that at the summer season level there is relatively little agreement between the ordered magnitude of overall mean, mean minimum and mean maximum estimated surface temperatures compared to the order of the screen-level temperatures (hottest (2003), 75th-percentile 1997, 50th-percentile 2001 and 25th-percentile 2002). For most weather stations the 50th-percentile (2001) has the highest overall mean and mean minimum temperatures (i.e., St James Park, Northolt and London Weather Centre). More encouraging in this case, is the fact that the hottest mean maximum temperature is recorded for the heatwave year of 2003 for all weather station locations, although again for the other years there is an inconsistent ordering. Thus, while Table 3 shows that all years apart from 2001 were statistically cooler than 2003 at the summer season level of aggregation, the results of Table 4 raise doubts as to whether averaging over an entire summer allows intra-annual temperature dynamics to be captured reliably.

While the results above demonstrate the utility of monthly estimated surface temperatures to capture urban temperature dynamics, it is common to employ the UHII metric in order to express the Urban Heat Island phenomena (Jones and Lister 2009, Tomlinson et al. 2010). Again, at the individual scene level (Figure 4) AVHRR derived UHII seems to capture the spatial variability in temperature seen in Figure 3. In particular analysis of Figure 4 showed that urban green areas across the city have lower UHII values (≈1.5–0.5 °C).
compared to high density built up areas (~6.0–8.7 °C), highlighting the importance of characterising UHII in a spatially explicit manner so as to capture the spatial variability seen in the estimated surface temperature values (Figure 3). However, at the location of the weather stations AVHRR estimated surface temperature UHII values were found to be greatest annually in 2002 (the lower quartile summer) at 8.38 °C (LWC – HWC) compared to 7.41 °C (LHR – HWC) for the heatwave year of 2003. Nonetheless, spatially maximum AVHRR estimated surface temperature UHII values were found to be consistently higher at the London Weather Centre (LWC) in the centre of London for all years except for 2003 where it is found at London Heathrow (LHR) on the urban-rural fringe of London. This suggests that at the intra-annual scale the AVHRR estimated surface temperature UHII is capturing a change in the spatial dependence of maximum UHII. In the case of air temperature UHII similar results were found; with maximum UHII being captured at LWC over all four years but with a marked increase in UHII at LHR during the heatwave summer of 2003, resulting in a marginal difference between LWC and LHR of 0.1 °C. However, Table 5 shows for the location of the London Weather Centre that 2003 (the hottest and heatwave year) does not exhibit either the highest maximum or highest mean UHII for either weather station air temperature (highest maximum and mean occurs for 2002, the 25th-percentile year) or estimated surface temperature UHII (highest maximum occurs for 2001 (median year) and highest mean occurs for 2002). Figure 5, shows the AVHRR derived seasonal average UHII values for each of the four summers, and exemplifies this fact. Moreover, Figure 5 also shows that spatially much higher UHIIIs are pervasive in 2002 compared to 2003.

Figure 4. Urban Heat Island Intensity (UHII) for London using estimated surface temperature from AVHRR data acquired on the 8th August 2003 14:04 (GMT).
<table>
<thead>
<tr>
<th>Year</th>
<th>Number of scenes</th>
<th>SJP</th>
<th>LHR</th>
<th>NTH</th>
<th>LWC</th>
<th>HWC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>Min</td>
<td>Max</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>2003</td>
<td>43</td>
<td>10.60</td>
<td>34.91</td>
<td>25.81</td>
<td>4.57</td>
<td>11.31</td>
</tr>
<tr>
<td>2001</td>
<td>12</td>
<td>19.87</td>
<td>31.56</td>
<td>26.43</td>
<td>3.55</td>
<td>18.81</td>
</tr>
</tbody>
</table>

[SJP – St James’s Park, LHR – London Heathrow, NTH – Northolt, LWC – London Weather Centre, HWC – High Wycombe (rural)].

Table 4. Summer seasonal AVHRR estimated surface temperature values from pixel locations containing urban and rural weather stations.
Table 5. Summer season UHII values derived using BADC air temperatures from London Weather Centre and AVHRR estimated surface temperatures (EST) from the pixel location containing London Weather Centre.

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean Air UHII (°C)</th>
<th>Minimum Air UHII (°C)</th>
<th>Maximum Air UHII (°C)</th>
<th>Mean EST UHII (°C)</th>
<th>Minimum EST UHII (°C)</th>
<th>Maximum EST UHII (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>4.74</td>
<td>3.10</td>
<td>8.30</td>
<td>6.31</td>
<td>1.20</td>
<td>19.77</td>
</tr>
<tr>
<td>1997</td>
<td>3.73</td>
<td>3.60</td>
<td>4.00</td>
<td>6.46</td>
<td>3.54</td>
<td>8.66</td>
</tr>
<tr>
<td>2001</td>
<td>4.35</td>
<td>2.50</td>
<td>5.90</td>
<td>8.80</td>
<td>1.66</td>
<td>22.83</td>
</tr>
<tr>
<td>2002</td>
<td>4.97</td>
<td>3.90</td>
<td>9.00</td>
<td>8.98</td>
<td>3.05</td>
<td>18.21</td>
</tr>
</tbody>
</table>

As in the case of the AVHRR estimated surface temperatures, Mann-Whitney U tests were performed between the 2003 UHII values and the other years investigated for monthly and yearly aggregations. Again a one-tail test based on 2003 UHII being greater than other years was employed at a 95% significance level (Table 6). This revealed that at a monthly level only 4 of 7 pair-wise combinations showed 2003 to have statistically greater UHII values with 3 of these being for the month of June alone. For the other months (July and August) 3 out of the 4 tests showed no statistically significant difference (Table 6). It should be noted that during the exceptionally warm summer of 2003, the greatest temperatures were recorded in August (Johnson et al. 2005); in this case highlighting the potential inability of the UHII metric to capture significant temperature changes at the monthly level. At the yearly level of aggregation the inability to distinguish between the heatwave year and the other years on the basis of the UHII metric is even more striking with all 3 pair-wise combinations being found to have no statistical difference at a 95% significance level (Table 6).

The above analysis of the UHII values using the AVHRR estimated surface temperature values reveal that it is possible to capture at a single point in time the spatial representation of the UHI. However, the results also pose serious questions about the utility of the UHII metric when derived using thermal Earth observed data. The data show there is no significant change in the difference between urban and rural temperatures during a heatwave summer, resulting in the metric’s inability to quantify the changes between heatwave and non-heatwave summers. In the case of the AVHRR UHII results such differences maybe the result of vegetation loss at the rural station during the heatwave summer due to drought, which will change the cover-surface temperature interaction (Lu and Weng 2006) and cause the metric to improperly quantify the ‘true’ urban heat island. However, this does not explain the similar observations from corresponding weather station screen-level air temperature derived UHII results. Taken together, Table 5 suggests that the UHII derived either from AVHRR scenes or from weather station records is questionable in capturing intra-annual temperature dynamics and in particular heatwave events. Interestingly, similar questions have been raised when using solely air-temperature derived UHII (Jones and Lister 2009). While the AVHRR data represent daytime scenes, the comparative air temperature based UHII values in Table 5 are generated from hourly values over the full diurnal cycle at each weather station, suggesting that the above described differences in UHII between summers are manifested in both daytime surface and diurnal air temperature based UHII observations, further supporting the evidence that UHII is not a suitable metric for analysis of urban temperature dynamics within the remit of urban climate and extreme temperature event analysis.
Table 6. Mann-Whitney statistical results comparing monthly and yearly UHII values for London derived from AVHRR estimated surface temperatures.

<table>
<thead>
<tr>
<th>Year</th>
<th>Umin Calculated 1-tail X&gt;Y</th>
<th>Reject Null @ 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 2003-1997</td>
<td>313792</td>
<td>Yes</td>
</tr>
<tr>
<td>June 2003-2001</td>
<td>1034638</td>
<td>Yes</td>
</tr>
<tr>
<td>June 2003-2002</td>
<td>777175</td>
<td>Yes</td>
</tr>
<tr>
<td>July 2003-2001</td>
<td>2139493</td>
<td>No</td>
</tr>
<tr>
<td>July 2003-2002</td>
<td>991436</td>
<td>Yes</td>
</tr>
<tr>
<td>August 2003-2001</td>
<td>1169920</td>
<td>No</td>
</tr>
<tr>
<td>August 2003-2002</td>
<td>1654360</td>
<td>No</td>
</tr>
<tr>
<td>Summer 2003-1997</td>
<td>1425958</td>
<td>No</td>
</tr>
<tr>
<td>Summer 2003-2001</td>
<td>1307869</td>
<td>No</td>
</tr>
<tr>
<td>Summer 2003-2001</td>
<td>1641385</td>
<td>No</td>
</tr>
</tbody>
</table>

Reject Null if Umin Calculated < Umin Critical (1110722 for 1-tail test X>Y @ 95%; X=2003 summer heatwave year).

Figure 5. AVHRR Summer Urban Heat Island Intensity values for the Greater London Authority with temporal averaging at the seasonal level

5 Conclusions

This study has demonstrated how a long-temporal baseline of daytime AVHRR data can be employed to capture the summer temperature regime of the city of London, UK, including the response to a known heatwave event. By processing a large number of daytime scenes for a sample of years that characterise the temperature distribution of London over a decade period we have managed to evaluate the intra-annual temperature dynamics of London. The results presented show, as in other studies (e.g., Tomlinson et al. 2010), that both temporally and
spatially significant surface temperature variations can be captured using Earth observation data. In this work, by evaluating temporal averaging at the daily, monthly and summer season level we have shown that there is in the case of London, a high level of sensitivity in the AVHRR data to diurnal and localised meteorological effects (Nichol 2003, Hung et al. 2006). Equally at the other end of the level of aggregation we found that summer season averages fail to capture the ordered magnitude of intra-annual temperature dynamics recorded by established weather station screen-level air temperatures. Our analysis reveals that characterising estimated surface temperatures at the monthly level gives the best statistical discrimination of the intra-annual temperature of London and allows a heatwave summer year to be distinguished. The results presented highlight the importance of generating robust temporal averages from multiple scenes to remove noise and quantify the underlying urban temperature regime.

With regards to the UHII metric the research has demonstrated the utility of AVHRR scenes to generate Urban Heat Island Intensity surface maps, which can be seen to represent spatial variability of intensity over the urban surface. However, the utility of the metric is questioned due to the lack of response seen in UHII data derived from both AVHRR and air temperature measurements between different summer temperature regimes. Testing of these data in a statistically robust manner showed that the 2003 heatwave UHII data-sets for both image surface and ground air temperatures did not exhibit significantly greater intensities than the other three years under consideration. However, several other studies have reported distinctive UHII values during heatwaves for several cities (Cheval and Dumitrescu 2009, Cheval et al. 2009, Tomlinson et al. 2010). As such, further research is required before a definitive evaluation of the utility of the UHII metric can be made. Nonetheless, on the basis of the results presented here for both AVHRR estimated surface temperatures and weather station screen-level air temperatures, along with the corresponding findings of other studies (e.g., Jones and Lister 2009), a more suitable metric for daytime AVHRR scenes to characterise over time urban temperature dynamics and perform intra-annual comparisons may be one based on summer season monthly average or monthly average maximum temperatures.

Finally, while the focus of this work has been to investigate how one may capture and characterise the intra-annual temperature dynamics of cities using Earth observation data, it is important to realise that characterisation is only part of providing an understanding of the drivers of urban surface and air temperatures. In this regard, while we have presented a repeatable approach for other cities where a long time series of AVHRR data exists, the work needs to be extended to address a number of important issues. First, we need to improve our understanding spatially of the relationship between Earth observed estimated surface temperatures and screen-level air temperatures; an issue that has been recognised as being important if thermal Earth observation is to be used in temperature hazard and risk assessment studies of cities (Voogt and Oke 2003, Tomlinson et al. 2010). Second, and related, we need to understand spatially at the city-scale how surface and air temperatures are affected by land cover, anthropogenic drivers and urban form differences. While significant bodies of work have been established in this area (Oke 1987, Eliasson and Svensson 2003, Lu and Weng 2006, Tomlinson et al. 2010), again much remains to be done before a full understanding is obtained. In both cases, the 30-year daytime cloud free AVHRR archive used to provide the selected scenes for this study offers the potential to not only assess these issues spatially but also to assess their temporal consistency not just for London, but also for other UK cities.
7 Acknowledgements

This work is funded by a UK Engineering and Physical Sciences Research Council (EPSRC) PhD studentship, awarded to Tomas Holderness. Additionally, Dundee Satellite receiving station (A UK Natural Environment Research Council facility (NERC)) is acknowledged for generating and providing the original AVHRR data-sets. The Remote Sensing and Photogrammetry Society is acknowledged for the award of their travel bursary for travel to the First International Conference on Sustainable Urbanisation, 2010, Hong Kong China, where this work was initially presented.

8 References


ERBERTSEDER, T., TUNGALAGSAIKHAN, P., BITTNER, M., MEISNER, R., SCHROEDTER, M., DECH, S., 1999, Towards an operational atmospheric correction for


