Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data

Zander S. Venter[a,*], Oscar Brousse[b], Igor Esau[c], Fred Meier[d]

[a] *Terrestrial Ecology Section, Norwegian Institute for Nature Research - NINA, 0349 Oslo, Norway*
[b] *Department of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200E, 3001 Leuven, Belgium*
[c] *Nansen Environmental and Remote Sensing Centre, Thormohlensgt. 47, 5006 Bergen, Norway*
[d] *Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Rothenburgstraße 12, D-12165 Berlin, Germany*

ARTICLE INFO

ABSTRACT

The impacts of climate change such as extreme heat waves are exacerbated in cities where most of the world's population live. Quantifying urbanization impacts on ambient air temperatures (Tair) has relevance for human health risk, building energy use efficiency, vector-borne disease control and urban biodiversity. Remote sensing of urban climate has been focused on land surface temperature (LST) due to a scarcity of data on Tair which is usually interpolated at 1 km resolution. We assessed the efficacy of mapping hyperlocal Tair (spatial resolutions of 10–30 m) over Oslo, Norway, by integrating Sentinel, Landsat and LiDAR data with crowd-sourced Tair measurements from 1310 private weather stations during 2018. Using Random Forest regression modelling, we found that annual mean, daily maximum and minimum Tair can be mapped with an average $RMSE$ of 0.52 °C ($R^2 = 0.5$), 1.85 °C ($R^2 = 0.05$) and 1.46 °C ($R^2 = 0.33$), respectively. Mapping accuracy decreased sharply with < 250 weather stations (approx. 1 station km^{-2}) and remote sensing data averaged within a 100-500 m buffer zone around each station maximized accuracy. Further, models performed best outside of summer months when the spatial variation in temperatures were low and wind velocities were high. Finally, accuracies were not evenly distributed over space and we found the lowest mapping errors in the local climate zone characterized by compact lowrise buildings which are most relevant to city residents. We conclude that this method is transferable to other cities given there was little difference (0.02 °C $RMSE$) between models trained on open- (satellite and terrain) vs closed-source (LiDAR) remote sensing data. These maps can provide a complement to and validation of traditional urban canopy models and may assist in identifying hyperlocal hotspots and coldspots of relevance to urban planners.

1. Introduction

Climate change is introducing unprecedented heat extremes in cities globally with consequences for the socio-economic and ecological sustainability of urban environments (Oke et al., 2017). Heatwaves are one of the most lethal climate-related disasters (Mora et al., 2017). Heat-related mortalities have been documented in Europe (e.g. Schifano et al., 2009), United States (e.g. Schifano et al., 2009), China (e.g. Huang et al., 2010) and Russia (e.g. Trenberth and Fasullo, 2012). A recent global meta-analysis of lethal heatwave events predicts that ca. 30% of the world's population are currently exposed to lethal air temperatures for at least 20 days a year (Mora et al., 2017). Elevated temperatures can also have indirect effects on human respiratory-related illness through enhancing the severity of air pollution in cities (Katsouyanni et al., 1993; Koken et al., 2003). Urban heat affects

building thermal loads and industrial cooling infrastructure which ultimately impacts city-level energy use efficiency and resulting carbon emissions (Bizjak et al., 2018; Tooke et al., 2014). A perhaps lesser-studied aspect of urban heat is the effect on urban biodiversity. This could lead to modified ecosystem services in urban ecosystems – such as pollination (Turrini and Knop, 2015). It has also been demonstrated that urban climates could influence the development of vector borne diseases in tropical cities (Brousse et al., 2019). Modelling the health and biodiversity impacts of urban heat and accounting for urban ecosystem services provided by green infrastructure increasingly needs to be conducted with spatial resolution that captures differences in neighbourhood living conditions and socio-demographics (Barton et al., 2019; Nyelele et al., 2019), Therefore, there is a need for timely and high-resolution temperature data for the sustainable planning and management of climate-resilient cities.

---

* Corresponding author.
E-mail address: zander.venter@nina.no (Z.S. Venter).

There has been substantial work on mapping urban heat with moderate-resolution satellites worldwide (e.g. Chakraborty and Lee, 2019), and with higher-resolution satellites for several cities (e.g. Esau et al., 2019; Miles and Esau, 2017; Zhou et al., 2015). However, the majority of this work has focussed on satellite measures of land surface temperature (LST) with less focus on near-surface air temperature (Tair) (Zhou et al., 2018). This is largely because satellites provide spatially explicit LST data whereas Tair is measured at single locations where meteorological stations are available. While LSTs are an important boundary condition for urban energy balance and for quantifying urban heat island (UHI) effects, their relevance to human health is not well known and the use of LST to quantify health risk appears to be due to the relative ease with which it can be mapped and not due to its appropriateness as an indicator of human heat exposure (Ho et al., 2016). Urban epidemiology studies on heat-related morbidity and mortality often employ Tair time series or spatially interpolated Tair using a handful of meteorological stations (Hajat et al., 2010; Kosatsky et al., 2012). These methods do not account for the thermal complexity of urban environments, where micro- to local-scale variability of Tair is caused by factors such as land cover, building properties, vegetation density, sun exposure, and wind dynamics (Oke and Maxwell, 1975; Saaroni and Ziv, 2010; Voogt and Oke, 2003). The spatial variability of these factors exists at very small scales (1–100 m), a variability that is not captured by weather stations that are typically spread far apart (1–100 km, Muller et al., 2013, Oke, 2007). Some attempts to overcome this limitation include deploying high-density temperature monitoring networks (Chapman et al., 2015) or to use mobile sampling (motorized vehicle or bicycle) of air temperatures (Tsin et al., 2016; Yang and Bou-Zeid, 2019), however, these methods are costly to install and maintain. The recent proliferation of private weather stations in urban centres provide crowd-sourced temperature data as an efficient and inexpensive alternative to previous sampling efforts (Fenner et al., 2019; Hammerberg et al., 2018; Meier et al., 2017; Muller et al., 2015).

The integration of satellite remote sensing and high density crowd-sourced temperature holds great potential for mapping urban air temperatures. Previous methods to map Tair have had limited access to dense weather station networks and have included (1) temperature vegetation index methods (TVX), (2) thermodynamic balance models, and (3) statistical and machine learning regression models. TVX models rely on semi-empirical relationships between vegetation cover, LST and Tair, combined with spatial interpolation methods (Stisen et al., 2007; Zhu et al., 2013). Thermodynamic models estimate Tair based on theoretical understanding of energy balance, sensible and latent heat fluxes and require substantial parametrization with data that are rarely available in a spatially explicit manner (Oke, 1988; Sun et al., 2005). Statistical models that relate satellite brightness temperatures and LST data to ground-reference Tair data have also been used to map Tair and UHI (Hafner and Kidder, 1999; Weng, 2009). Most recently, attempts to map Tair have adopted machine learning regression models using relationships between satellite measures of surface reflectance (visible and infrared wavelengths) and Tair to interpolate Tair over space (Janatian et al., 2017; Tsin et al., 2016; Yoo et al., 2018; Zhu et al., 2019). The majority of these studies use the Moderate Resolution Imaging Spectroradiometer (MODIS) satellites to produce maps at 1 km resolution. Very few utilize higher resolution satellites like Landsat 7 and 8 and Sentinel 2 and none to date have employed crowd-sourced Tair to train and validate hyperlocal Tair models.

Operationalizing Tair mapping in monitoring and forecasting contexts requires effective quantification and communication of mapping error and accuracy which is something that has not been a focus of previous efforts. Studies often report overall model accuracies but do not explore the temporal and spatial variation in mapping accuracies. For example, Ho et al. (2014) advanced the knowledge on Tair mapping substantially by using machine learning to map Tair over Vancouver, Canada, however the analysis was limited to 6 days scattered over 5 years and thus determinants of model accuracies were difficult to explain. Aggregating explanatory variables over space is important to account for the neighbourhood effects of land cover on Tair measured at any one point in space (Guo and Moore, 1998; Zakšek and Oštir, 2012). Therefore, apart from the temporal changes in model accuracy, the spatial scale of data aggregation can also affect model performance (Ho et al., 2016). Finally, the type of predictor variables used (e.g. satellite, terrain, urban canopy measures) may alter Tair mapping and studies that identify how data source influence model accuracies are lacking. The broader applicability and transferability of Tair modelling methods requires that predictor variables are open-source and globally available (e.g. Sentinel satellite data) and not city-specific and closed-source or proprietary (e.g. building height and other GIS data).

In this study, we integrate satellite and LiDAR remote sensing data with crowd-sourced Tair data from Netatmo private weather stations to create spatially contiguous maps of hyperlocal Tair over Oslo, Norway. We adopt the term "hyperlocal" from literature on smart cities (Marble, 2018; Mora et al., 2019) and use it to define the geographical scale of temperature mapping at micro- to local-scales of 10–30 m spatial resolution. The study was conducted 2018, during which Oslo experienced extreme summer heat waves in which July temperatures exceeded the long-term day (22 °C) and night (12 °C) averages by 14 and 8 °C, respectively. We used a machine learning framework to train and validate Random Forest regression models to interpolate point temperatures over the city for annual mean, daily minimum and maximum Tair using all available imagery during 2018. We aimed to assess how Tair mapping accuracies are influenced by (1) data source and predictor variable type, (2) number of available weather stations, (3) scale of spatial data aggregation, and (4) season and weather conditions. We did this by collecting a range of predictor variables from Landsat, Sentinel and LiDAR remote sensing datasets. Data were collected within buffer zones of varying size around each weather station and at different times of year. Model accuracies were assessed by validating against a withheld sample of weather stations and these accuracies were related to weather conditions on the day of satellite fly-over.

2. Data and methodology

2.1. Study site

Oslo, located on the Oslo Fjord in Eastern Norway (59′55 N, 10′45E), with a mean annual temperature of 6 °C, experiences mild summers (average for July of 18 °C) with 177 frost-free days per year, and relatively cold winters (January average of −3 °C). Oslo municipality houses ca. 673,000 residents (Statistikkbanken, 2018) and the population has increased by 20% between 2007 and 2018 (Oslo kommune, 2017). The built-up zone tracks the fjord and contains most of the city's industry, commerce, and residents and consequently privately-owned weather stations (Fig. 1). This is surrounded by coniferous forest zoned as non-residential and thus protected from further development (Miljødepartementet, 2009). Oslo was selected for this study because of its relevance to ecosystem accounting with ongoing projects quantifying the extent and condition of urban ecosystems relevant to air temperatures that determine highly localized ecosystem services (Barton et al., 2015). In addition, very few urban climate studies have been performed on high latitude cities (Konstantinov et al., 2018).

2.2. Weather station data

Open-access air temperature data from 1452 privately-owned weather stations within Oslo and surrounds were used as reference data for satellite models of Tair. The stations are sold by the company Netatmo (https://www.netatmo.com) and the data for outdoor modules, which upload weather measurements via Wi-Fi to the cloud, are made freely available for download through an application
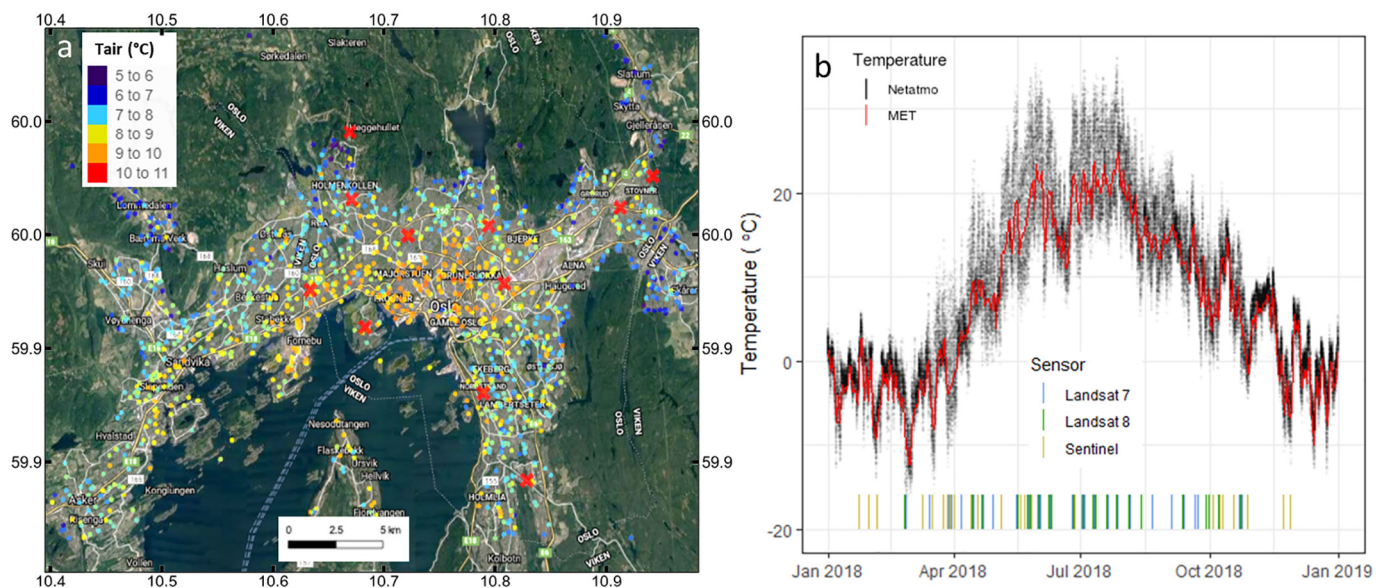
Fig. 1. Mean annual near-surface air temperature (Tair) for private Netatmo weather stations (coloured circles) over Oslo for 2018 (a). The location of reference meteorological stations managed by the Norwegian Meteorological Institute is indicated with red crosses. Hourly temperatures for all stations are plotted in b as grey points with the mean daily Tair measurements by Meteorological Institute stations overlaid in red. Vertical bars on the x-axis indicate the presence of Sentinel and Landsat scenes utilized in the modelling procedure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

programming interface (API) with the Netatmo servers. We retrieved hourly Tair data for all stations for all available dates during 2018, along with latitude and longitude coordinates for their location.

Quality assessment is an indispensable requirement for the use of crowd-sourced air temperature data (Meier et al., 2017). The quality control (QC) procedure developed by Napoly et al. (2018) was used for identification of statistically implausible Tair due to misplacement of sensors, solar exposition (radiative errors), inconsistent meta data and device malfunctions (Fig. S1). This procedure needs no reference meteorological data and is easy to apply via an available software R-package "CrowdQC" (Grassmann et al., 2018). The first step (M1) of the QC flags Netatmo stations with equivalent longitude and latitude coordinates. The second step (M2) applies a modified z-score approach for the detection of statistical outliers from the hourly Tair distribution. As an intermediate step in M2, Tair is converted to account for natural variation of Tair due to different elevations of Netatmo stations. Elevation data were derived from the Shuttle Radar Topography Mission 30 m DEM (Farr et al., 2007). The third step (M3) removes months where > 20% of data points are flagged during QC step M2. The fourth step (M4) excludes any station that, when correlated against spatial median of hourly Tair over a month, produce a Pearson's correlation coefficient of < 0.9. Following quality control procedures, there were 1310 stations available over 2018. Due to time gaps in some station time series, there were on average 593 stations available per month (Fig. S1) and each station was operational for an average of 118 days during the year (Fig. S2).

As a further means of Netatmo Tair validation, we collected daily Tair data from 11 weather stations owned by the Norwegian Meteorological Institute. These meteorological stations are spread across the urban extent of Oslo (Fig. 1) and provide temperature data collected according to standards outlined by the World Meterological Organisation. We performed a linear regression of city-wide Netatmo daily mean Tair on the daily means derived from meteorological stations (Fig. S3). This produced a strong correlation (R^2 = 0.99, RMSE = 0.355 °C, Fig. S3), validating the further use of Netatmo data in our study.

2.3. Satellite remote sensing data

All satellite remote sensing data were processed and analysed within the Google Earth Engine (GEE) cloud-computing platform (Gorelick et al., 2017). The satellite datasets used included surface reflectance Landsat 7 Enhanced Thematic Mapper + and Landsat 8 Operational Land Imager data, and surface reflectance Sentinel-2 MutiSpectral Instrument data. All available scenes intersecting Oslo during 2018 were included and masked for clouds using the 'pixel_qa' band. Images have been orthorectified and atmospherically corrected by GEE. We used the bands listed along with spatial resolutions in Table 1 and calculated the normalized difference vegetation index (NDVI, Tucker, 1979) and the index-based built-up index (IBI, Xu, 2008) for each available scene. Although there are other spectral indices to capture bare land and impervious surfaces, IBI compares favourably to the normalized difference built-up index (NDBI) for mapping built up land cover (Assyakur et al., 2012).

Landsat land surface temperatures (LST) were obtained from Landsat 7 and Landsat 8 (thermal data resampled from 100 m native resolution to 30 m resolution) by implementing the single channel algorithm from Jiménez-Muñoz et al. (2014) following the workflow described in Parastatidis et al. (2017) in GEE. In short, the algorithm derives LSTs from the Landsat infrared thermal bands by calculating land cover emissivities out of NDVI values. Atmospheric corrections are performed by using 6-hourly reanalysis data of atmospheric water vapor contents provided by the National Centres for Environmental Prediction (Kalnay et al., 1996). LSTs are obtained only for overpasses with < 50% cloud coverage. A cloud-masking using the 'pixel_qa' band is applied prior to running the algorithm. Water streams and reservoirs are masked out.

In addition to satellite spectral responses, we also derived terrain variables from the Shuttle Radar Topography Mission 30 m digital elevation model (DEM, Farr et al., 2007). These included height above sea level and the standard deviation in the 60 × 60 m pixel neighbourhood for elevation values as a measure of terrain ruggedness/roughness. We also used the Landsat-derived 30 m global water occurrence dataset (Pekel et al., 2016) along with a fast distance transform to calculate distances from the coast given that proximity to water

Table 1
List of predictor variables used in the modelling of air temperatures in Oslo. Abbreviations: NDVI: normalized difference vegetation index; IBI: index-based built-up index; STRM: shuttle radar topography mission; CHM: canopy height model; LiDAR: light detection and ranging; sd: standard deviation.

Variable	Data source	Data type	Spatial resolution
Red	Landsat 7, 8 and	Open	
Green	Sentinel 2	source	L: 30 m, S: 10 m
Blue			
Near infrared			
Short-wave infrared 1			L: 30 m, S: 20 m
Short-wave infrared 2			
NDVI			L: 30 m, S: 10 m
IBI			L: 30 m, S: 20 m
Land surface temperature	Landsat 7, 8		30 m
Elevation above sea	STRM		30 m
Terrain aspect			
Terrain slope			
Terrain ruggedness			
CHM	LiDAR	Closed	1 m
CHM slope		source	
CHM aspect			
CHM shadow/SVI			
Building height	LiDAR + building		
Building height sd 1–4 m	footprint		
Building height sd 4–20 m			
Building height sd 20–100 m			
Fractional tree cover	LiDAR +		
Tree height	orthophoto		
Distance to coast	Global water	Open	30 m
Distance to fresh water	occurrence	source	

is an important determinant of Tair (Stewart, 2000) and that it might increase the explanatory power of the modelling procedure.

2.4. Urban heat island calculation

Although the urban heat island (UHI) effect is not part of our study scope, we decided to calculate the mean annual UHI intensity to contextualize Oslo's Tair gradient within the broader literature on urban climates. We defined UHI in line with recent developments in the literature (Li et al., 2018; Schatz and Kucharik, 2015) using the slope of the linear regression between temperature and impervious surface area. We calculated it separately for Tair measurements collected by Netatmo stations (canopy UHI - CUHI) and LST measurements from satellite data (surface UHI - SUHI) using the same method. Impervious surface area was derived from the 2015 edition of the Copernicus High Resolution Layer (http://land.copernicus.eu/pan-european) on imperviousness (Lefebvre et al., 2016) derived from 2 m resolution land cover data. We extracted impervious surface area within a circular buffer (100 m radius) of each weather station and, according to best practice for UHI calculation (Stewart, 2011), corrected for terrain effects by only including stations within 25 m of the mean city elevation. Excluding stations outside of this elevation bracket provided 340 stations to perform the UHI analysis.

2.5. LiDAR remote sensing data

Light detection and ranging (LiDAR) data are often used to characterize the 3D structure of the urban canopy which has significant effects on solar radiation trapping and consequent Tair dynamics (Masson et al., 2020). LiDAR data from airborne laser scanning missions have been processed by the Norwegian Mapping Authority to create a wall-to-wall 1 m digital terrain (DTM) and surface model (DSM) over Oslo. This data was used to derive a canopy height model (CHM) which is calculated by subtracting the DTM from the DSM resulting in a raster of all surface objects' heights above ground level including trees,

buildings, bridges and other city infrastructure. We calculated morphological characteristics for the entire CHM (all surface objects), tree CHM and building CHM. Tree canopies were isolated using orthophoto-derived greenness index and a watershed segmentation technique described in Hanssen et al. (2019). Building CHM objects were extracted using GIS derived geometries of building footprints curated and provided by the Norwegian Mapping Authority.

We calculated a proxy for the sky view factor (SVF) or hemispheric openness using a variant of the method proposed by Lindberg (2005). The method uses the 'hillshadow' GEE algorithm to cast hypothetical shadows using the DSM with the sun at a zenith angle or 45° and a pixel neighbourhood of 200 m. This is done for each 10° azimuth increment after which average shadow intensities are calculated (Fig. S4). Although the zenith of 45° angle does not match the sun angle at all times of day and year, we found that the added benefit from modelling all sun angles was not worth the computational power necessary. Morphological metrics included the mean object height, fractional cover, and neighbourhood heterogeneity in object height. These are all common metrics used to characterize urban canopies for modelling microscale temperature effects (Masson et al., 2020). The neighbourhood height heterogeneity or roughness was calculated at three spatial scales using resampling with a standard deviation reducer (Fig. S5). The reducer calculates the standard deviation for a pixel neighbourhood within a moving window (see Fig. S6 for illustration). Mean CHM rasters were resampled from 1, 4, and 20 m resolution up to 4, 20, and 100 m resolution, respectively using this reducer. This was done to capture the surface roughness of tree, building, and total CHM infrastructure in the city at varying spatial scales.

2.6. Tair modelling

To map spatially-explicit Tair, we chose to use the Random Forest (RF) machine learning algorithm which is a supervised regression model using remote sensing variables to predict Tair from private weather stations and then to extrapolate and interpolate over the entire study area. RF models account for nonlinear interactions between response and predictor variables, and build an ensemble of regression trees that have proven to be more accurate than ordinary least squares regression and support vector machine learning in mapping Tair (Ho et al., 2014). The satellite and LiDAR predictor variables were split into those that are open source and freely available at global scales, and those that are specific to Oslo and not freely available to the public (Table 1). This distinction is made because of its relevance to the transferability of this method anywhere on the globe.

We built separate RF models based on the input data used, and the scale of temporal, and spatial aggregation (Fig. 2). RF models were trained on a combination of Sentinel, Landsat and LiDAR-derived urban canopy metrics. Terrain variables were common to all models. The temporal aggregation involved merging satellite and Netatmo data at three levels of temporal synchrony (Fig. 2). First, Netatmo data were extracted for the exact hour at which the satellite image was taken. Second, daily minimum and maximum Tair were extracted for the day on which the satellite image was taken. Finally, the mean annual Tair for each station was extracted along with the mean value of the satellite data time-stack. Given that there were only 213 stations with 100% complete annual timeseries (Fig. S2), we tested whether we could use the daily differentials between Netatmo station and city-wide mean Tair to derive a complete set of mean annual Tair readings for all 1310 stations. We first calculated the daily city-wide Tair means using MET stations and Netatmo stations with > 90% availability. Then the daily Tair differential for each station relative to the city mean ($Tair_{station} - Tair_{city\_mean}$) was calculated and averaged over the year for all stations independent of their temporal coverage. Finally, the mean annual differential was corrected relative to the mean annual Tair for the city. This is based on the assumption that a station's Tair differential remains constant throughout the year. We tested this assumption using stations
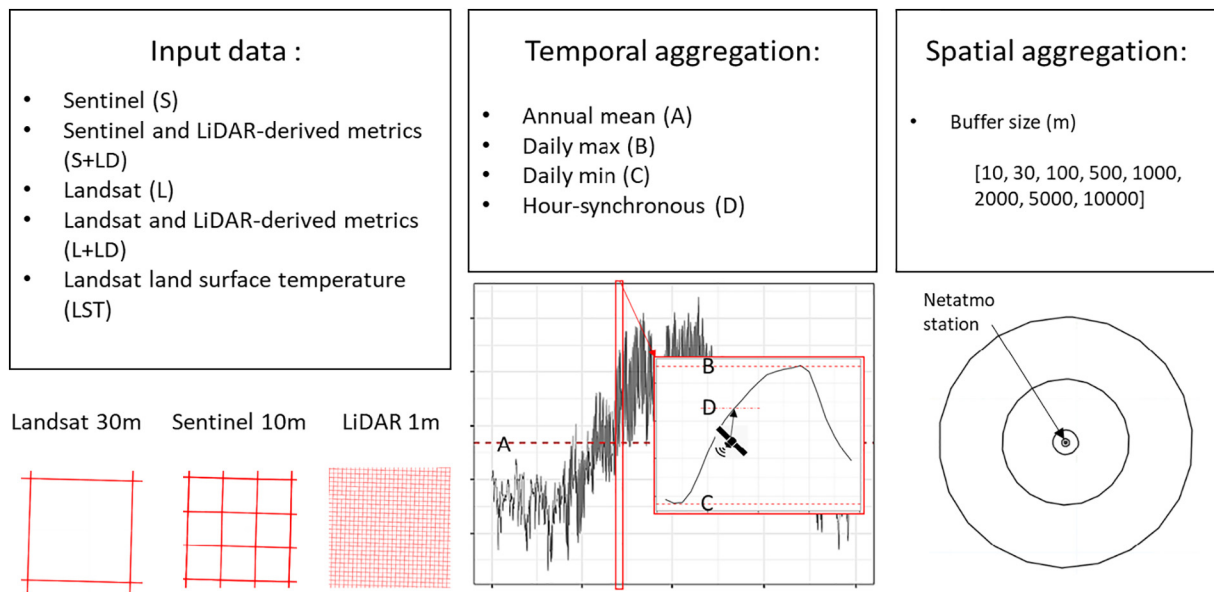
Fig. 2. Stratification of the Random Forest modelling procedure. Separate models were built for each combination of input data, temporal and spatial aggregation. Input data spatial resolutions are illustrated (lower left) with the exception of Landsat thermal infrared at 100 m resolution. The time series for one Netatmo station is plotted to illustrate the different temporal aggregation levels. The dashed red lines represent the annual mean (A), daily maximum (B) and minimum (C) and hour-synchronous (D) temperatures at the time of satellite fly-over. The buffer sizes around each weather station area also illustrated (lower right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with > 90% availability and found that the mean standard error in station Tair differential is 0.036 °C through the year. Thus, for instance, a station that gave daily Tair of 1 °C below the city mean during July is expected to also have a mean annual Tair of 1 ± 0.036 °C below the city average. We further tested the technique by iteratively removing a random fraction of data from stations with 100% temporal coverage and testing the interpolation performance. Even with 90% missing data, the interpolation is accurate with a RMSE of 0.1 °C (Fig. S9). Therefore, we found it reasonable to interpolate mean annual Tair for stations with missing data using this approach.

The spatial scale of data aggregation refers to the extraction of mean and standard deviation values for all input variables within a circular buffer zone of varying size around each Netatmo station (Fig. 2). The method of averaging predictor variables within a buffer zone has been used before (Ho et al., 2014) to account for the influence of neighbourhood surface and land cover on localized temperature measurements (Guo and Moore, 1998; Zakšek and Oštir, 2012). Although previous studies only extracted statistics within buffer zones of up to 1000 m radius (Ho et al., 2016), we did not want to preclude the possibility that model accuracies might be enhanced by gathering information at even larger spatial scales. Therefore, the range of buffer zone radii we used included 10, 30, 100, 500, 1000, 2000, 5000, and 10,000 m (Fig. S6). During the collection of training data, this equates to calculating zonal statistics for the raster stack (variables in Table 1), where the zones are vector objects defined by the buffered Netatmo stations. During the model prediction phase (i.e. extrapolating Tair over space using the RF model), we performed a pixel focal smoothing function which calculates the mean and standard deviation for all pixels within the circular neighbourhood of the given centroid pixel (see Fig. S6 for illustration). This is repeated for every single pixel in the raster stack. The rationale for calculating the standard deviation within the buffer zone is to account for the effect of spatial heterogeneity in urban morphology and its effect on Tair (Nordbo et al., 2013). This approach considers the surface-atmosphere exchange processes that effects Tair measured at one point in space.

To evaluate model performances, we used external validation by withholding 30% of the dataset from the model training stage and thereafter testing model predictions against it. By regressing observed

Tair on predicted Tair, we produced the root mean square error (*RMSE*) and adjusted R^2 as measures of model accuracy and fit (Willmott, 1981). To account for variation in model performance caused by the random partitioning of training and testing sets, we performed a bootstrapping approach to smooth over this variation. The bootstrapping involves running the RF model 25 times with a different randomly selected training and testing set. The mean of the resulting *RMSE* and R^2 values are then calculated. The RF algorithm also measures the relative importance of each predictor variable by quantifying the increase in prediction errors when a predictor is permuted in the validation data. We calculated the mean and standard error in permutation-based variable importance measures for each predictor variable across all models.

To assess the effect of Netatmo station number on Tair mapping accuracies we used a similar approach to the bootstrapping described above for mean annual Tair. We iteratively removed a random selection of Netatmo stations from the training dataset with increments of 10 stations from 1000 to 20. For each increment we performed the bootstrapping procedure of RF model accuracy assessment. We collected all the model residuals (difference between modelled and observed Tair in testing set) from the bootstrapping procedure and linked these to their corresponding Netatmo station ID and associated station locality. With this we could derive the spatial distribution of model error over the city. We then calculated mean annual Tair and model error by aggregating values for stations within Local Climate Zones (LCZ, Ching et al., 2018; Stewart and Oke, 2012) in Oslo using data from the LCZ map for Europe with pixels of 100 × 100 m (Demuzere et al., 2019). Each Netatmo station was assigned to a LCZ zone based on the station location and the intersecting LCZ map pixel value.

Statistical analyses of model accuracies were performed in R (RCoreTeam, 2017) using the 'stats' and 'car' packages for linear regression and significance testing. We performed a second order polynomial regression of model *RMSE* and R^2 values on log-transformed buffer size to identify an optimal buffer size for use in subsequent analyses. We used a polynomial regression for buffer size because it produced a stronger model fit than simple linear regression. Once the optimal buffer size was determined (lowest *RMSE* and highest R^2), we used multiple linear regression to regress model accuracies from these
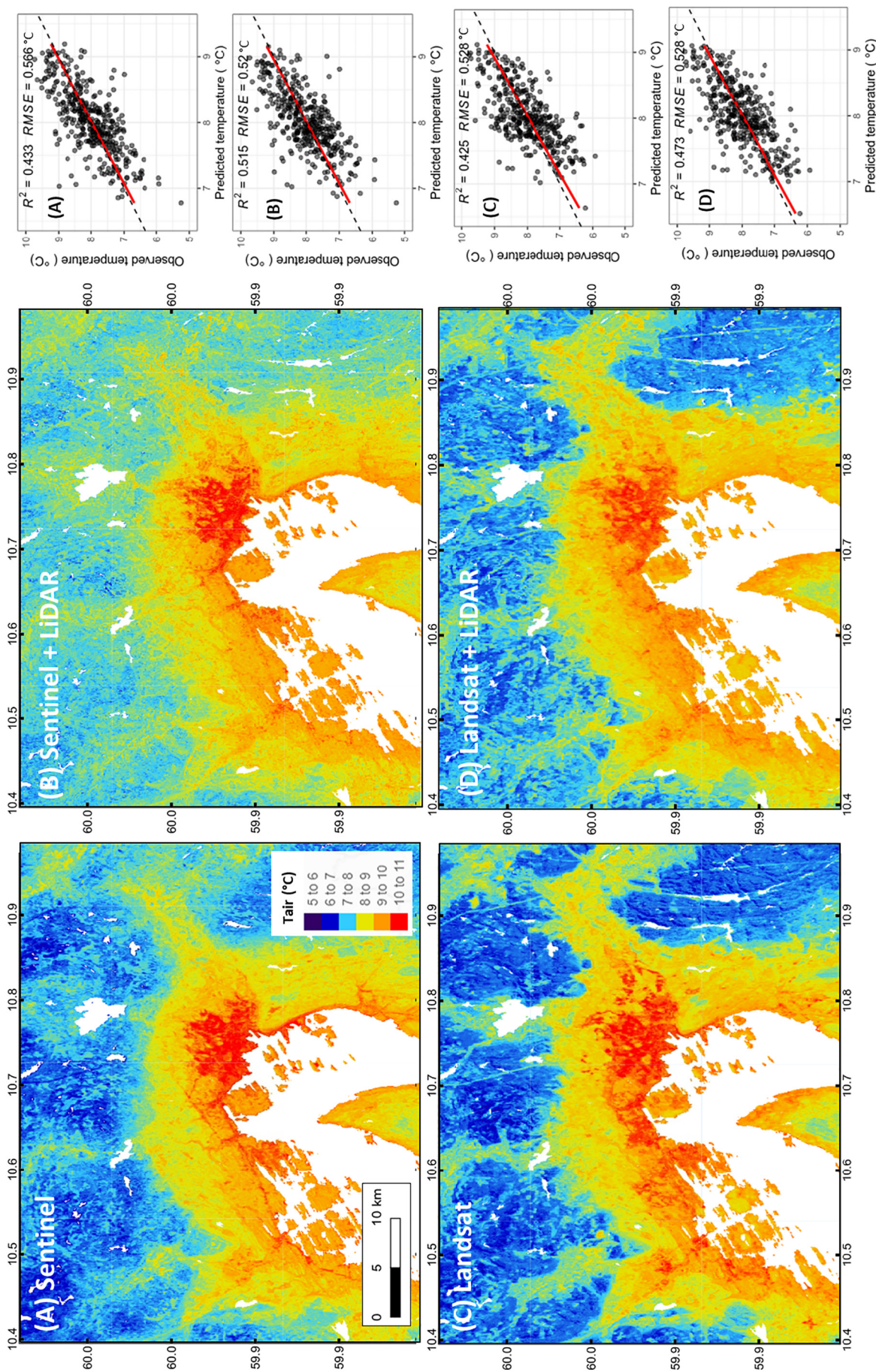
Fig. 3. Modelled mean annual Tair maps at 10–30 m resolution of Oslo using Sentinel (A, B), or Landsat (C, D) spectral responses, in combination with LiDAR-derived urban canopy parameters, and Landsat land surface temperature. Graphs indicating model accuracies are displayed in the right-hand panel. A with-held testing dataset (n = 392) of observed Tair from private weather stations are regressed on the Tair predicted by the model to produce a model r² value and root mean square error (*RMSE*). The linear regression line is plotted in red with a 1:1 line dashed line plotted in black for reference. White areas are water bodies and are masked from the analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

models on four climatic variables expected to explain the temporal variance in Tair mapping accuracy. Second order polynomial regression produced the strongest fit likely because it captured seasonal fluctuation in model results. These variables included the hour-synchronous city-wide Tair mean, Tair range, daily precipitation and wind velocity means. Precipitation and wind velocity data were derived from the ERA5 daily atmospheric reanalysis dataset at 0.25˚ × 0.25˚ resolution (Copernicus Climate Change Services, 2017). An ANOVA was run on the multiple regression output to test the significance of individual explanatory variables. We used second order polynomial terms for each explanatory variable where it reduced the model fit relative to a simple linear model.

3. Results

3.1. Annual mean temperature mapping

Mean annual temperature mapped using RF models was 8 ˚C, with a spatial standard deviation of 0.73 ˚C across the Netatmo stations. There was a steep temperature gradient from the center to periphery of the city (Fig. 3) defined by a mean annual CUHI (measured with Tair) and SUHI (measured with LST) of 1.5 and 6 ˚C, respectively. After accounting for the effect of terrain by only including stations within 25 m of the mean city elevation, the corrected CUHI and SUHI were found to be 0.6 and 3.5 ˚C, respectively.

Models of mean annual air temperature performed well, producing a mean $RMSE$ of 0.54 ˚C and R^2 of 0.46 across all combinations of predictor variables (Fig. 3). The addition of LiDAR to Sentinel and Landsat models reduced $RMSE$ by 0.044 and 0 ˚C, respectively, whereas R^2 was increased by 0.09 and 0.05. Landsat models performed best, reducing $RMSE$ by 0.025 ˚C relative to Sentinel models. At neighbourhood scales, the advantage of the high resolution models for identifying hyperlocal hotspots of Tair during heatwave events become apparent (Fig. 4). For example, in Fig. 4 one can observe a 5 ˚C range in night-time Tair at neighbourhood scales. The Landsat models appear to produce greater contrasting Tair between vegetation and impervious cover compared to Sentinel models. For example, in Fig. 3, the forested area surrounding Oslo appears cooler relative to the city center in the Landsat compared to the Sentinel maps. This subtle difference is also evident at the neighbourhood level (Fig. S7). Under the Local Climate Zone typology, dense trees produced the lowest mean annual Tair whereas compact lowrise to midrise zones produced the highest Tair (Fig. 5). $RMSE$ of mean annual Tair was lowest for bare rock or paved zones (mean of 0.16 ˚C) and highest for vegetated LCZs (mean of 0.31 ˚C, Fig. 5). Compact lowrise LCZ, which covers most residential areas, produced the second lowest $RMSE$ out of the LCZ classes.

Model $RMSE$ decreases and R^2 increases with the number of Netatmo stations available (Fig. 6). The model accuracies decline gradually with reducing station availability from 1000 to approx. 250 stations or 0.87 stations km^{-2} (Fig. 6). After the 250 station threshold, model $RMSE$ reduced by 0.04 ˚C for every 10 stations removed. The highest model fits and lowest model errors in the mapping of mean annual temperatures were achieved with a buffer radius of between 100 and 500 m (Fig. 7). Buffer sizes that were small (< 100 m) or very large (> 1000 m) produced lower model accuracies. Smaller buffer sizes performed better when LiDAR data was included.

3.2. Daily temperature mapping

When mapping Tair at each satellite fly-over through the year (Fig. 8), model errors were lowest for the mapping of daily minimum Tair (RMSE = 1.44 ± 0.03 ˚C, R² = 0.33 ± 0.007), and highest for the mapping of daily maximum Tair ($RMSE$ = 1.86 ± 0.04 ˚C, R² = 0.05 ± 0.004). The mapping of hour synchronous Tair produced a $RMSE$ of 1.62 ± 0.03 ˚C. The change in model fits and error for daily minimum Tair did not vary over the year, however, model errors were

lower in winter relative to summer months for mapping of hour-synchronous and daily maximum Tair (Fig. 8). The temporal variation in the hour-synchronous model fit (R^2) was partially explained by variation in daily mean temperatures, temperature range and daily precipitation (Fig. 9). Model fits were greatest at low temperature means and ranges and high precipitation, although there is a large variation in this response. However, these atmospheric conditions do not foster the development of pronounced thermal differences at the local scale. Hyperlocal variation of Tair is very limited under these atmospheric conditions. The temporal variance in model $RMSE$ was partially explained by the spatial range in temperature and wind velocity with greatest model accuracies at low temperature ranges and high wind velocities (Fig. 9).

3.3. Predictor variable importance

Out of the 20 most important predictors, 6 were from closed-source and 14 were from open-source datasets (Fig. 10). The most important satellite-derived predictors were LST, IBI and NDVI. Mean annual Tair increases with LST and IBI and decreases with NDVI (Fig. 10). Elevation and distance to coastline were the most important terrain/landscape predictors and both showed a negative linear relationship to Tair. The two most important closed-source predictors were related to LiDAR-derived building canopy characteristics (Fig. 10). These included building height, and the standard deviation in building height for an aggregation scale from 20 to 100 m (see Fig. S5). Tair increases linearly with the standard deviation in building height (a measure of building canopy roughness).

4. Discussion

Up until now the limited availability of Tair measurements has hindered the mapping of thermal environment as a hazard for heat stress in urban contexts (Wicki et al., 2018; Zhou et al., 2018). Using a machine learning framework with crowd-sourced temperature data, we were able to map mean annual, daily maximum and minimum Tair over Oslo at 10-30 m resolution with a RMSE of 0.54, 1.86 and 1.46 ˚C, respectively. Ho et al. (2014), using Landsat data in a RF model were able to map daily maximum Tair in Vancouver with a comparable $RMSE$ of 2.31 ˚C. Other attempts to map Tair using machine learning and statistical models have all relied on MODIS satellite data at 1 km resolution. For example, Yoo et al. (2018) mapped daily maximum Tair in Los Angeles with an RMSE of 1.7 ˚C, while others have mapped daily mean Tair outside of urban areas at national scales with an $RMSE$ of 2.3 ˚C and 0.5 ˚C in Iran (Janatian et al., 2017), and China (Zhu et al., 2019), respectively. No studies have utilized Sentinel surface reflectance data in combination with crowdsourced data from private weather stations for Tair mapping. Although Sentinel does not measure thermal infrared reflectance, we find that it produces comparable mapping accuracies to Landsat-driven models. This appears to be largely because of the importance of near infrared spectral indices (NDVI and IBI) and ancillary landscape predictors (elevation and distance to coast) which were included in both the Landsat and Sentinel models.

4.1. Relevance of Tair mapping

The Tair maps produced using the approach presented here are relevant to urban planning in several ways. They provide a high resolution map of urban hotspots and coldspots that can inform health risk mitigation strategies such as planting vegetation of increasing surface albedo. The hyperlocal effects of mitigation strategies can be monitored over time. This is directly relevant to heat exposure epidemiology and indirectly related to air pollution epidemiology, particularly in northern European cities where air temperatures can significantly reduce city ventilation (Wolf-Grosse et al., 2017). The Tair maps also offer a powerful tool for validation of climate models (Chen
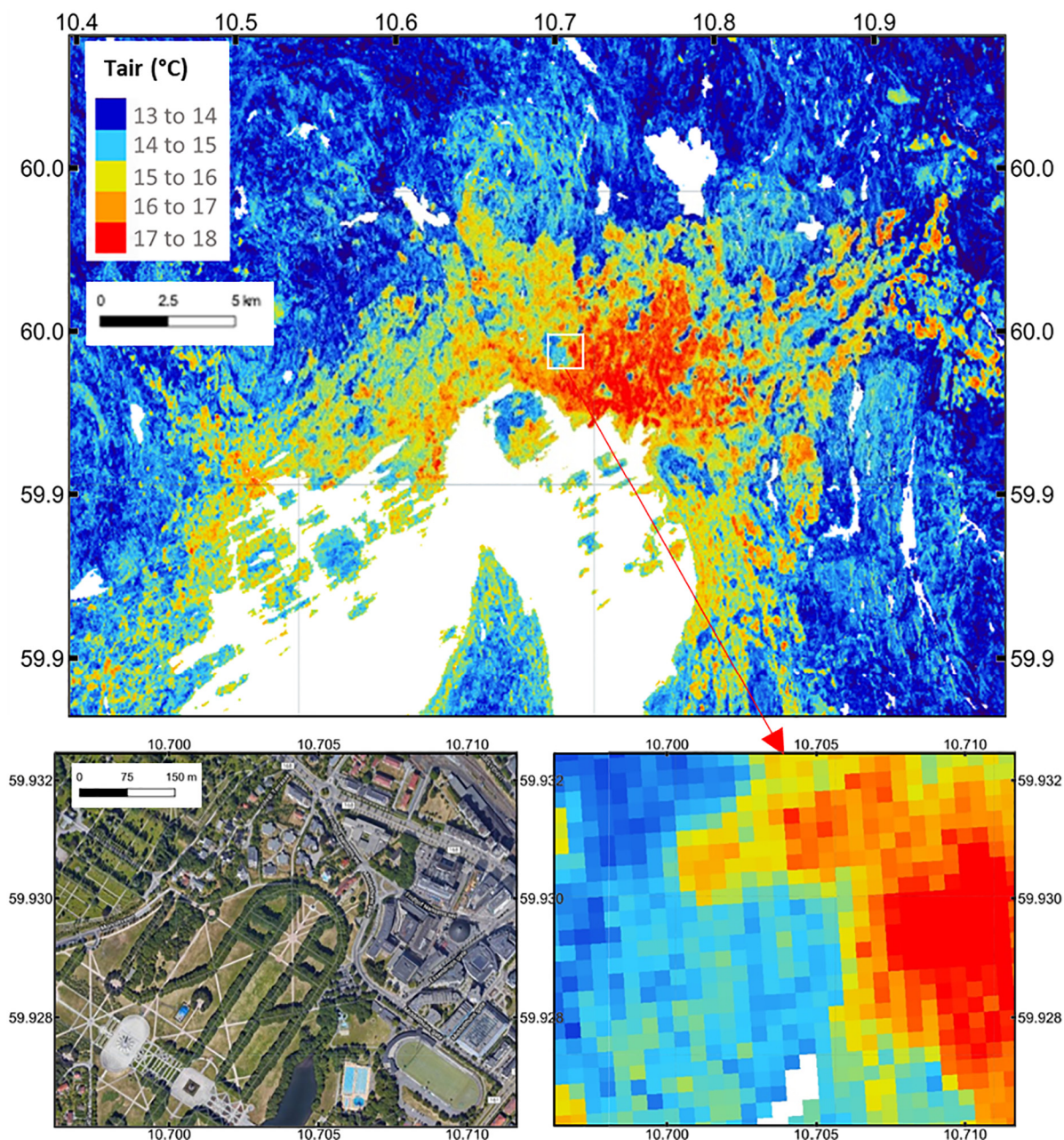
Fig. 4. An example of a relevant hyperlocal heat-hazard situation over Oslo on 1 June, one of the hottest days during the summer heat wave of 2018. Night-time (daily minimum) Tair is mapped at 30 m resolution using the Landsat and LiDAR Random Forest model over the city with a zoomed extent of Vigeland Park (bottom right) with a high resolution satellite image for reference (bottom left). The model had an *RMSE* of 1.8 °C at this time point.

et al., 2011) that include urban canopy models such as the building effect parameterization-building energy model (BEP–BEM, Martilli, 2002; Salamanca et al., 2009). Such models are often used to test the effect of urban planning scenarios on air temperature and rely on assumptions about relationships between urban canopy parameters and air temperatures that are seldom tested in both micro- and mesoscale applications. The RF modelling approach builds predictions based on observed linear and non-linear relationships between explanatory variables and Tair (Breiman, 2001). Therefore RF models could be used to estimate the hyperlocal response in Tair to the hypothetical alteration of urban canopy parameters (e.g. tree removal or building development). However, this would be a retrospective analysis based on recent climate observations and cannot replace the forecasting ability of urban canopy models. Finally, mapping Tair using RF modelling gives insight into determinants of urban cooling. The relationships between

land cover characteristics and Tair corroborate those observed in the broader literature on both LST (Zhou et al., 2018 and studies therein) and Tair (e.g. Oliveira et al., 2011; Yan et al., 2014; Zhang et al., 2014). It is clear that Tair increases with impervious surface area and a decline in vegetation greenness (Fig. 10), and after accounting for the effect of terrain, the urban core is 0.6 °C warmer than the city's surroundings (CUHI intensity). This aligns with the broader European Commission agenda for nature-based solutions to climate change adaptation (European Commission, 2013).

4.2. Tair mapping accuracy

It is important to communicate the accuracy of Tair mapping to urban planners, and understanding the factors that influence the spatial and temporal variation in accuracy is helpful in this regard. Mapping
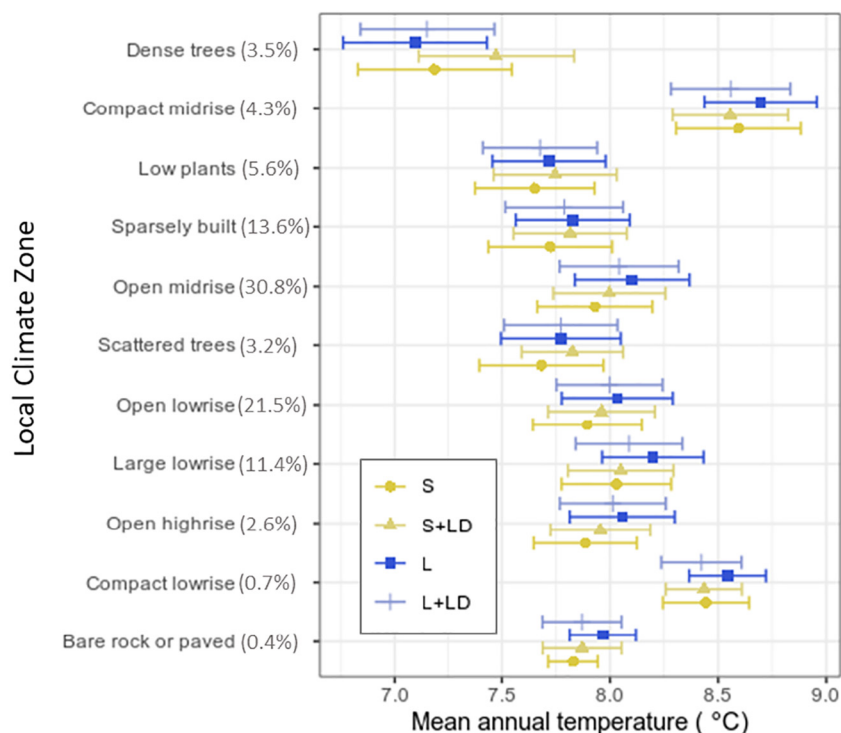
Fig. 5. Modelled mean annual Tair statistics for Local Climate Zones (LCZ) over Oslo during 2018. Mean and associated error margins are indicated with points and error bars for each LCZ and Random Forest model type. Error margins are derived from bootstrapping Random Forest model validation and assigning average *RMSE* values for Netatmo stations within each LCZ. LCZ classes are ordered from top to bottom in order of decreasing mean *RMSE*. The percentage coverage of each LCZ is indicated in parentheses. Please refer to Fig. 2 for input data code definitions.

errors were largest in LCZs with vegetation cover (e.g. dense trees and low plants) and lowest in those without (e.g. compact lowrise and bare ground, Fig. 5). Additionally, we found that it was more difficult to map Tair accurately in summer than in winter and that on days with high wind speeds and low temperature ranges, mapping errors were lowest. A potential explanation is that on hot, calm days, hyperlocal temperatures are very variable and thus difficult to map whereas on windy days, canopy turbulence homogenizes Tair and makes it easier to map (Choi et al., 2018). Few studies have quantified the factors that influence Tair mapping accuracy, however Ho et al. (2014) have inferred from their data that wind variability may reduce mapping accuracies especially in areas close to the ocean. We did not have data on wind gusts or hourly variations in wind velocity and future studies may be improved by adding this data to models of urban Tair. Indeed, some Netatmo weather stations do measure hourly wind velocity and direction, however quality assurance protocols like those for Tair (Napoly et al.,

2018) do not currently exist. In addition, the Netatmo data on humidity holds potential for mapping apparent temperature, which is a measure of air temperature closer to that perceived by humans, often quantified with the Humidex (Gosling et al., 2014; Masterton and Richardson, 1979).

Hyperlocal Tair can be significantly influenced by the surrounding land cover and urban canopy characteristics (Ho et al., 2016; Massetti et al., 2014). This is supported by our finding that aggregating predictor data within 100-500 m of each weather station resulted in higher Tair mapping accuracies than only including data within 30-100 m of each weather station (Fig. 7). The buffer size may differ significantly relative to city size depending on the topographic characteristics influencing temperature station fetch. However, Ho et al. (2014) found that predictors averaged over comparable buffer sizes to ours, (between 200 and 800 m) contributed the most to reducing model errors when mapping Tair in Vancouver. An alternative explanation is that the
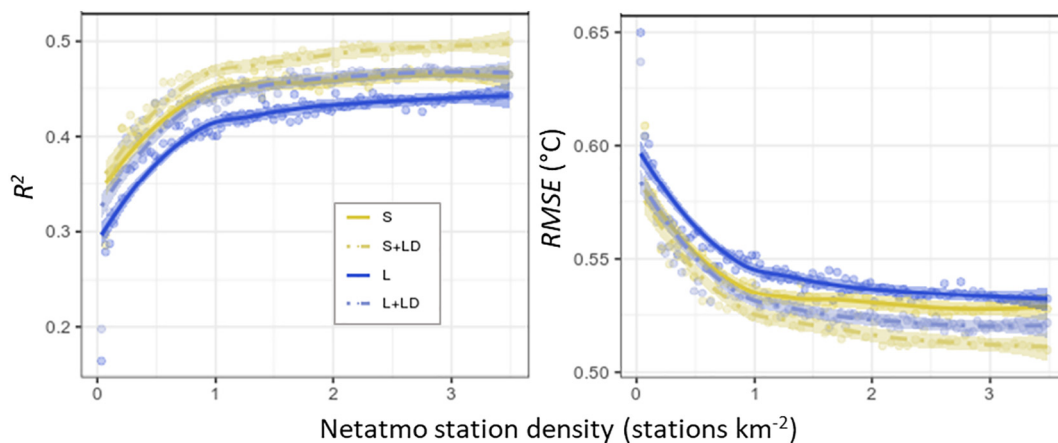

Fig. 6. Model performance statistics for the mapping of mean annual Tair over Oslo after iteratively removing a random selection of Netatmo stations from the original training dataset. For each training dataset size (station density), the observed Tair from private weather stations are regressed on the Tair predicted by the model to produce an adjusted model R^2 value and root mean square error (*RMSE*). Loess regression lines are fitted with 95% confidence interval ribbons. Please refer to Fig. 2 for input data code definitions.
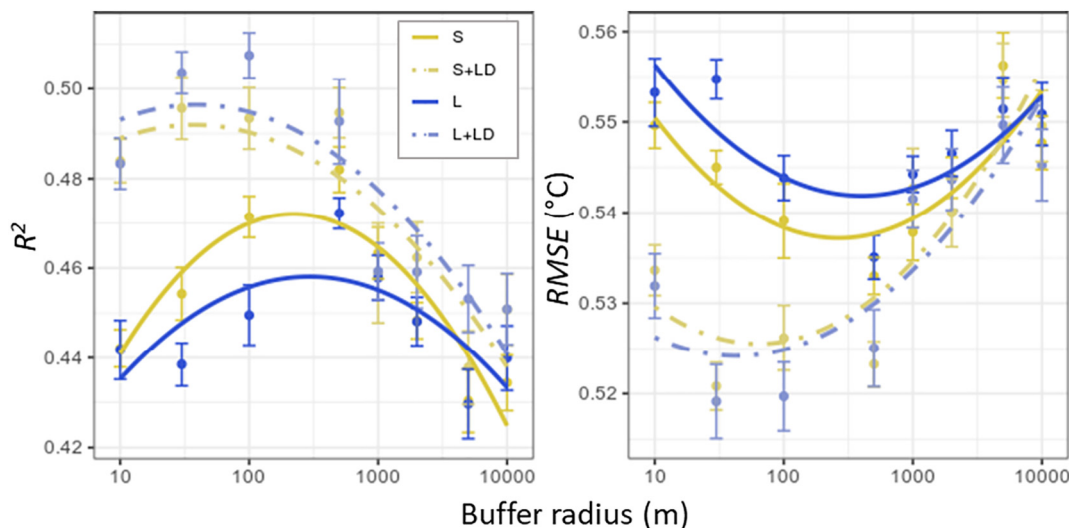
Fig. 7. Model performance statistics for the mapping of mean annual Tair over Oslo at different spatial aggregation scales. Aggregation scale refers to the radius of the circular buffer around each weather station within which the mean and standard deviation of predictor variables were calculated. For each buffer radius and input data combination, the observed Tair from private weather stations are regressed on the Tair predicted by the model to produce an adjusted model R^2 value and root mean square error (*RMSE*). Second order polynomial regression lines are fitted. Please refer to Fig. 2 for input data code definitions.

smallest buffer size we used (10 m) was not small enough to capture micro-scale urban canopy metrics that might explain more of the variance in Netatmo Tair readings. Netatmo sensors are exposed to small-scale turbulent eddies and radiation reflection/trapping which are affected by different surface types (grass, pavement, walls) that can cause unpredictable variations in Tair due to distinct micro-scale and intra-LCZ variability of Tair (Choi et al., 2018; Fenner et al., 2017; Meier et al., 2017; Napoly et al., 2018). For instance, a station placed close to a wall might produce temperature readings that are distorted by the wall's effect on the local radiation budget. These factors cannot be captured by satellite data and only partly by LiDAR data as illustrated by the unexplained variance in Tair for important predictor variables in Fig. 10. Possible solutions might be to screen out spurious Tair measurements relative to surrounding station averages or for private weather station users and companies to report more detailed metadata on station siting. Nevertheless, our findings imply that urban heat mitigation strategies need to take broader spatial context into account and that small building canopy alterations are unlikely to have a dominant influence on local ambient Tair.

4.3. Limitations and transferability

A major limitation to mapping Tair using satellite remote sensing is that satellite predictor variables cannot be collected on cloudy days. In this way satellite-derived Tair maps will be biased toward clear-sky weather conditions even in the case when the cloud mask gaps could be filled in through one or another form of interpolation. Others have found that cloud cover can be as important as other meteorological variables (including wind speed, humidity and atmospheric pressure) in determining UHI intensity (Kolokotroni and Giridharan, 2008; Morris et al., 2001) as well as changes of UHI intensity during heat waves (Fenner et al., 2019). For example, in Canada, Stewart (2000), found that daytime and post-sunset cloud cover explain 20% more variance in the ensuing nocturnal UHI intensities than do daytime and post-sunset wind speeds. Research in higher latitude cities shows that clear sky conditions are necessary for establishing a strong UHI whereas cloud cover homogenizes temperature observations during wintertime (Konstantinov et al., 2018; Varentsov et al., 2018). Given that Oslo is a cloudy city, it may be more appropriate to map urban Tair using other
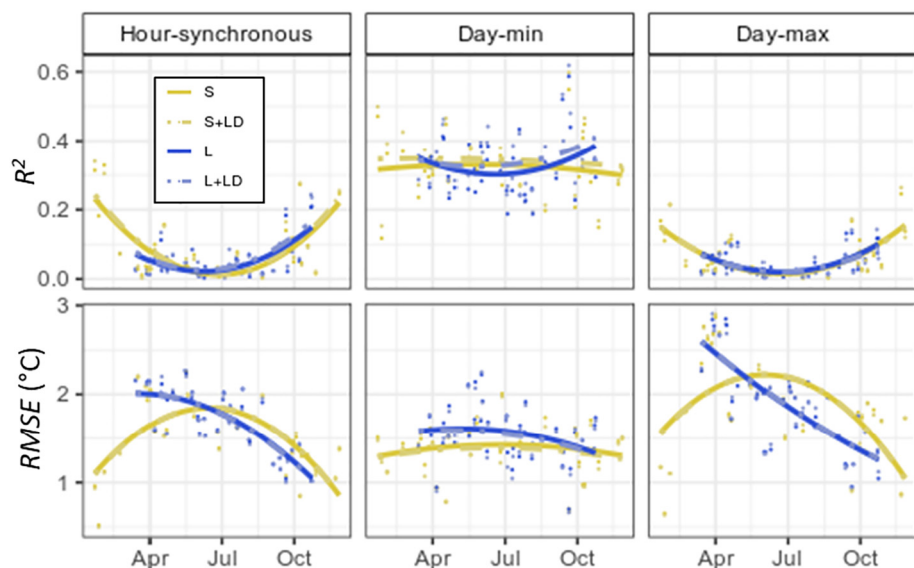

Fig. 8. Model performance statistics for the mapping of hour-synchronous, daily minimum and daily maximum Tair over Oslo during 2018 at 500 m spatial aggregation scale. For each time point and input data combination, the observed Tair from private weather stations are regressed on the Tair predicted by the model to produce a model R^2 value and root mean square error (*RMSE*). Second order polynomial regression lines are fitted. Please refer to Fig. 2 for input data code definitions.
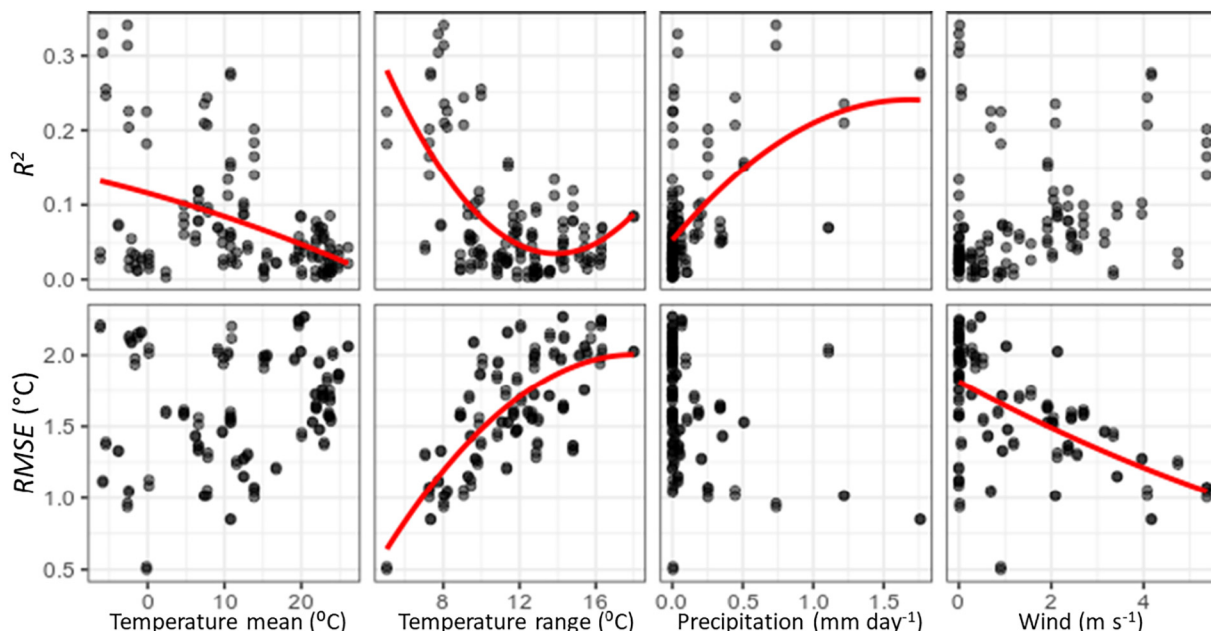
Fig. 9. Relationship between climatic covariates and Tair mapping accuracies over Oslo during 2018 using models with 500 m spatial aggregation scale. To assess model accuracy, observed Tair from private weather stations are regressed on the Tair predicted by the model to produce a model R^2 value and root mean square error (*RMSE*). Red lines represent second order polynomial regressions. Plots without lines represent non-significant relationships ($p > 0.05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
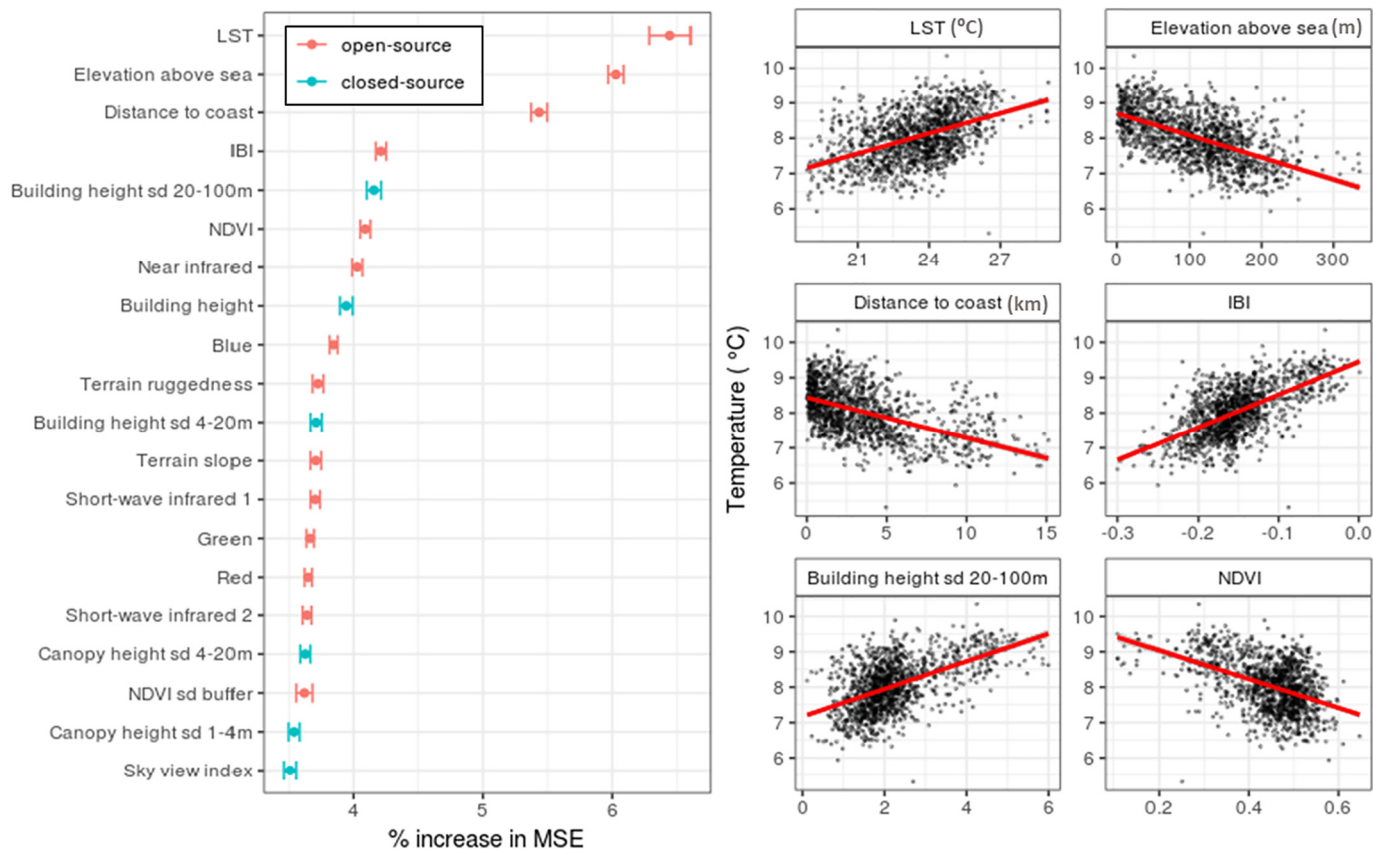

Fig. 10. Variable importance plot for the 20 most important predictor variables included in models of Tair over Oslo during 2018 (left panel). Variable importance is measured as the percentage decrease in model performance (mean standard error) when the given variable is left out of the model. Points and error bars represent means and standard errors across model accuracy from hour-synchronous and daily minimum and maximum temperature mapping. The relationship between mean annual Tair (y-axis) and the six most important variables (x-axis) are plotted along with red linear regression lines (right panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

statistical climate downscaling methods such as kriging or simple regressions with land cover information (Smid and Costa, 2018).

Despite the limitations on cloudy days, the approach presented here is technically transferable to any city in the world containing enough private weather stations to train and validate a machine learning algorithm like RF, however it is unknown whether the mapping will produce adequate accuracies. We found that mapping accuracies drop significantly with less than approx. 250 stations or 1 station km^{-2}. The current density of available Netatmo stations over Europe suggests that many cities will meet this criteria (Fig. S8). In cities without privately-owned Netatmo stations, it may be feasible for research institutes or similar to deploy their own high-density network of Netatmo stations (e.g. Chapman et al., 2015). At a per unit cost of 160 EUR, and a minimum density of approx. 0.87 stations km^{-2} this would translate to 140 EUR km^{-2}. The stations would need to leverage open-access WIFI networks in order to sync and store measurements in the cloud.

We also found that the difference in accuracies between models with open-source and closed-source predictors was minimal. The *RMSE* difference between models with different input data was in fact below the sensitivity of the temperature sensors of the stations and therefore the preference for one or another input dataset is not decisive. Therefore, using Sentinel, Landsat, and terrain data that are globally-available is potentially adequate for Tair mapping with comparable accuracies to those presented for Oslo. New methods for using Sentinel, Landsat and open source radar data to derive more detailed urban canopy characteristics like building height (Misra et al., 2018) and SVF (Hodul et al., 2016) may also act as substitutes for closed-source LiDAR-derived data. The broad applicability of the approach presents an opportunity for standardized assessments of heat risk exposure in cities which can be adopted in policy instruments such as national (e.g. Heatwave plan for England, Public Health England, 2018) or European-wide heat risk governance plans (Lass et al., 2011). Similarly, the derived relationships between green cover and Tair can be used in ecosystem service accounting frameworks to assess and monitor potential nature-based solutions to mitigating climate change in cities. Finally, recent advances in integrating Netatmo weather data into short-term Tair forecasts (Nipen et al., 2019) may benefit from the hyperlocal mean annual Tair maps produced here to aid city planners and public health programmes to target heat wave adaptation strategies.

5. Conclusion

The combined study of the crowd-sourced and satellite temperature at very fine spatial resolution reveals reliability and relevance of the hyperlocal air temperature mapping to urban environmental management. We demonstrated that the crowd-sourced Tair data have a significant utility for urban temperature analysis despite their irregularity and varying quality. The crowd-sourced data indicate warmer temperatures toward the more dense city center and in the areas collocated with more significant surface roughness. There are two major advantages of this data source, namely: the temperature readings are available in any weather conditions; and the readings are concentrated in the populated residential areas where the observed temperature anomalies are likely to have the major impact on population health.

Satellite data provide estimation of the LST, which can be used to characterize the surface UHI. However, the satellite temperature readings are sporadic and biased toward clear-sky weather conditions and are not necessarily relevant to human thermal comfort. This study showed that the combination of satellite and crowd-sourced Tair data is beneficial. We used satellite, LiDAR and terrain data within a Random Forest regression modelling framework to map Tair over Oslo and our findings support the conclusion that this method is broadly transferable to cities with private weather station data. The methods presented here are made available in simplified scripts stored in a GitHub repository (https://github.com/NINAnor/cityTairMapping). We can recommend that future studies aggregate explanatory variables within a 100-500 m

buffer of each station, and that a minimum of approx. 1 stations km^{-2} is necessary to achieve enhanced modelling accuracies (explore city-level station density here: https://weathermap.netatmo.com/). Furthermore, high resolution LiDAR data is not necessary as it does not produce a substantial reduction in mapping error compared to models built with purely open-source terrain and satellite data. Important variables to include in Tair models are LST, distance to coast and elevation, as well as NDVI and IBI which capture the warming effect of urban impervious surfaces. Tair mapping errors should be investigated and quantified in local contexts, however we found the highest confidence in model outputs in the compact lowrise LCZ and on days with low temperature ranges and higher wind speeds. Future research may produce improvements on the relatively low accuracies we found during times of year (summer) and areas (compact midrise) with highest absolute temperatures – conditions during which accurate Tair maps are presumably in highest demand. Mapping hour-synchronous Tair did not offer significant improvements in accuracy relative to daily minimum and maximums. These methods and the maps of hyperlocal urban air temperatures are particularly relevant to stakeholders in urban planning and human health and provide scope for better climate change adaptation in a rapidly urbanizing world.

CRediT authorship contribution statement

Zander S. **Venter:**Conceptualization, Methodology, Writing - original draft.**Oscar Brousse:**Methodology, Writing - review & editing.**Igor Esau:**Methodology, Writing - review & editing.**Fred Meier:**Methodology, Writing - review & editing.

Declaration of competing interest

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

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2020.111791.

References

As-syakur, A., Adnyana, I., Arthana, I.W., Nuarsa, I.W., 2012. Enhanced built-up and bareness index (EBBI) for mapping built-up and bare land in an urban area. Remote Sens. 4, 2957–2970.

Barton, D.N., Stange, E., Blumentrath, S., Traaholt, N.V., 2015. Economic Valuation of Ecosystem Services for Policy. A Pilot Study on Green Infrastructure. Oslo. NINA Rep. 1114. pp. 77.

Barton, D.N., Obst, C., Day, D., Caparrós, A., Dadvand, P., Fenichel, E., Havinga, I., Hein, L., McPherson, T., Randrup, T., Zulian, G., 2019. Recreation Services from Ecosystems. Discussion Paper 10. Paper Submitted to the Expert Meeting on Advancing the Measurement of Ecosystem Services for Ecosystem Accounting, New York, 22–24 January 2019 and Subsequently Revised. Version of 25 March 2019. New York, USA.

Bizjak, M., Žalik, B., Štumberger, G., Lukač, N., 2018. Estimation and optimisation of buildings' thermal load using LiDAR data. Build. Environ. 128, 12–21. https://doi.org/10.1016/j.buildenv.2017.11.016.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brousse, O., Georganos, S., Demuzere, M., Vanhuysse, S., Wouters, H., Wolff, E., Linard, C., Nicole, P.-M., Dujardin, S., 2019. Using local climate zones in sub-Saharan Africa to tackle urban health issues. Urban Clim. 27, 227–242.

Chakraborty, T., Lee, X., 2019. A simplified urban-extent algorithm to characterize surface urban heat islands on a global scale and examine vegetation control on their spatiotemporal variability. Int. J. Appl. Earth Obs. Geoinf. 74, 269–280. https://doi.org/10.1016/j.jag.2018.09.015.

Chapman, L., Muller, C.L., Young, D.T., Warren, E.L., Grimmond, C.S.B., Cai, X.-M., Ferranti, E.J.S., 2015. The Birmingham urban climate laboratory: an open meteorological test bed and challenges of the Smart City. Bull. Am. Meteorol. Soc. 96, 1545–1560.

Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C.S.B., Grossman-Clarke, S., Loridan, T., Manning, K.W., Martilli, A., Miao, S., Sailor, D., Salamanca, F.P., Taha, H., Tewari, M., Wang, X., Wyszogrodzki, A.A., Zhang, C., 2011. The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. Int. J. Climatol. 31, 273–288. https://doi.org/10.1002/joc.2158.

Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X., Ren, C., Brousse, O., Martilli, A., Neophytou, M., Mouzourides, P., 2018. WUDAPT: an urban weather, climate, and environmental modeling infrastructure for the anthropocene. Bull. Am. Meteorol. Soc. 99, 1907–1924.

Choi, Y., Lee, S., Moon, H., 2018. Urban physical environments and the duration of high air temperature: focusing on solar radiation trapping effects. Sustainability 10, 4837.

Copernicus Climate Change Service (C3S), 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS). https://cds.climate.copernicus.eu/cdsapp#!/home, Accessed date: 10 January 2019.

Demuzere, M., Bechtel, B., Middel, A., Mills, G., 2019. Mapping Europe into local climate zones. PLoS One 14, e0214474.

Esau, I., Miles, V., Varentsov, M., Konstantinov, P., Melnikov, V., 2019. Spatial structure and temporal variability of a surface urban heat island in cold continental climate. Theor. Appl. Climatol. 1–16.

European Commission, 2013. Green Infrastructure (GI)—Enhancing Europe's Natural Capital. European Commission, Brussels.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography mission. Rev. Geophys. 45. https://doi.org/10.1029/2005RG000183.

Fenner, D., Meier, F., Bechtel, B., Otto, M., Scherer, D., 2017. Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. Meteorol. Z. 26, 525–547. https://doi.org/10.1127/metz/2017/0861.

Fenner, D., Holtmann, A., Meier, F., Langer, I., Scherer, D., 2019. Contrasting changes of urban heat island intensity during hot weather episodes. Environ. Res. Lett. 14. https://doi.org/10.1088/1748-9326/ab506b.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Gosling, S.N., Bryce, E.K., Dixon, P.G., Gabriel, K.M.A., Gosling, E.Y., Hanes, J.M., Hondula, D.M., Liang, L., Bustos Mac Lean, P.A., Muthers, S., Nascimento, S.T., Petralli, M., Vanos, J.K., Wanka, E.R., 2014. A glossary for biometeorology. Int. J. Biometeorol. 58, 277–308. https://doi.org/10.1007/s00484-013-0729-9.

Grassmann, T., Napoly, A., Meier, F., Fenner, D., 2018. CrowdQC v1.2.0 - Quality Control for Crowdsourced Data From CWS. Technische Universität Berlin, Berlin. https://doi.org/10.14279/depositonce-6740.3.

Guo, L.J., Moore, J.M., 1998. Pixel block intensity modulation: adding spatial detail to TM band 6 thermal imagery. Int. J. Remote Sens. 19, 2477–2491.

Hafner, J., Kidder, S.Q., 1999. Urban heat island modeling in conjunction with satellite-derived surface/soil parameters. J. Appl. Meteorol. 38, 448–465.

Hajat, S., Sheridan, S.C., Allen, M.J., Pascal, M., Laaidi, K., Yagouti, A., Bickis, U., Tobias, A., Bourque, D., Armstrong, B.G., 2010. Heat–health warning systems: a comparison of the predictive capacity of different approaches to identifying dangerously hot days. Am. J. Public Health 100, 1137–1144.

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, 1241–1257. https://doi.org/10.1002/joc.5447.

Hanssen, F., Barton, D., Cimburova, Z., 2019. Mapping Urban Tree Canopy Cover Using Airborne Laser Scanning –Applications to Urban Ecosystem Accounting for Oslo NINA Report.

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

Ho, H.C., Knudby, A., Xu, Y., Hodul, M., Aminipouri, M., 2016. A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area. Sci. Total Environ. 544, 929–938.

Hodul, M., Knudby, A., Ho, C.H., 2016. Estimation of continuous urban sky view factor from Landsat data using shadow detection. Remote Sens. 4, 2957–2970. https://doi.org/10.3390/rs8070568.

Huang, W., Kan, H., Kovats, S., 2010. The impact of the 2003 heat wave on mortality in Shanghai, China. Sci. Total Environ. 408, 2418–2420.

Janatian, N., Sadeghi, M., Sanaeinejad, S.H., Bakhshian, E., Farid, A., Hasheminia, S.M., Ghazanfari, S., 2017. A statistical framework for estimating air temperature using MODIS land surface temperature data. Int. J. Climatol. 37, 1181–1194. https://doi.org/10.1002/joc.4766.

Jiménez-Muñoz, J.C., Sobrino, J.A., Skoković, D., Mattar, C., Cristóbal, J., 2014. Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. IEEE Geosci. Remote. Sens. Lett. 11, 1840–1843.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77, 37–47.

Katsouyanni, K., Pantazopoulou, A., Touloumi, G., Tselepidaki, I., Moustris, K., Asimakopoulos, D., Poulopoulou, G., Trichopoulos, D., 1993. Evidence for interaction between air pollution and high temperature in the causation of excess mortality. Arch. Environ. Heal. An Int. J. 48, 235–242.

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

Kolokotroni, M., Giridharan, R., 2008. Urban heat island intensity in London: an investigation of the impact of physical characteristics on changes in outdoor air temperature during summer. Sol. Energy 82, 986–998. https://doi.org/10.1016/j.solener.2008.05.004.

Konstantinov, P., Varentsov, M., Esau, I., 2018. A high density urban temperature network deployed in several cities of Eurasian Arctic. Environ. Res. Lett. 13, 75007.

Kosatsky, T., Henderson, S.B., Pollock, S.L., 2012. Shifts in mortality during a hot weather event in Vancouver, British Columbia: rapid assessment with case-only analysis. Am. J. Public Health 102, 2367–2371.

Lass, W., Haas, A., Hinkel, J., Jaeger, C., 2011. Avoiding the avoidable: towards a European heat waves risk governance. Int. J. Disaster Risk Sci. 2, 1–14.

Lefebvre, A., Sannier, C., Corpetti, T., 2016. Monitoring urban areas with sentinel-2A data: application to the update of the Copernicus high resolution layer imperviousness degree. Remote Sens. 8, 606.

Li, H., Zhou, Y., Li, X., Meng, L., Wang, X., Wu, S., Sodoudi, S., 2018. A new method to quantify surface urban heat island intensity. Sci. Total Environ. 624, 262–272. https://doi.org/10.1016/j.scitotenv.2017.11.360.

Lindberg, F., 2005. Towards the use of local governmental 3-D data within urban climatology studies. Mapp. Image Sci. 2, 32–37.

Marble, S., 2018. Everything that can be measured will be measured. Technol. + Des. 2, 127–129. https://doi.org/10.1080/24751448.2018.1497355.

Martilli, A., 2002. Numerical study of urban impact on boundary layer structure: sensitivity to wind speed, urban morphology, and rural soil moisture. J. Appl. Meteorol. 41, 1247–1266. https://doi.org/10.1175/1520-0450(2002)041<1247:NSOUIO>2.0.CO;2.

Massetti, L., Petralli, M., Brandani, G., Orlandini, S., 2014. An approach to evaluate the intra-urban thermal variability in summer using an urban indicator. Environ. Pollut. 192, 259–265. https://doi.org/10.1016/j.envpol.2014.04.026.

Masson, V., Heldens, W., Bocher, E., Bonhomme, M., Bucher, B., Burmeister, C., de Munck, C., Esch, T., Hidalgo, J., Kanani-Sühring, F., Kwok, Y.T., 2020. City-descriptive input data for urban climate models: model requirements, data sources and challenges. Urban Clim. 31, 100536.

Masterton, J.M., Richardson, F.A., 1979. Humidex: A Method of Quantifying Human Discomfort Due to Excessive Heat and Humidity. Environment Canada, Atmospheric Environment.

Meier, F., Fenner, D., Grassmann, T., Otto, M., Scherer, D., 2017. Crowdsourcing air temperature from citizen weather stations for urban climate research. Urban Clim. 19, 170–191.

Miles, V., Esau, I., 2017. Seasonal and spatial characteristics of urban heat islands (UHIs) in northern west Siberian cities. Remote Sens. 9, 989.

Miljødepartementet, N., 2009. Lov om naturområder i Oslo og nærliggende kommuner (markaloven). (Norway).

Misra, P., Avtar, R., Takeuchi, W., 2018. Comparison of digital building height models extracted from AW3D, TanDEM-X, ASTER, and SRTM digital surface models over Yangon City. Remote Sens. 10, 2008.

Mora, C., Dousset, B., Caldwell, I.R., Powell, F.E., Geronimo, R.C., Bielecki, C.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., Trauernicht, C., 2017. Global risk of deadly heat. Nat. Clim. Chang. 7, 501–506. https://doi.org/10.1038/nclimate3322.

Mora, S., Anjomshoaa, A., Benson, T., Duarte, F., Ratti, C., 2019. Towards large-scale drive-by sensing with Multi-Purpose City scanner nodes. In: 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), pp. 743–748. https://doi.org/10.1109/WF-IoT.2019.8767186.

Morris, C.J.G., Simmonds, I., Plummer, N., 2001. Quantification of the influences of wind and cloud on the nocturnal urban heat island of a large city. J. Appl. Meteorol. 40, 169–182.

Muller, C.L., Chapman, L., Grimmond, C.S.B., Young, D.T., Cai, X., 2013. Sensors and the city: a review of urban meteorological networks. Int. J. Climatol. 33, 1585–1600.

Muller, C.L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., Leigh, R.R., 2015. Crowdsourcing for climate and atmospheric sciences: current status and future potential. Int. J. Climatol. 35, 3185–3203.

Napoly, A., Grassmann, T., Meier, F., Fenner, D., 2018. Development and application of a statistically-based quality control for crowdsourced air temperature data. Front. Earth Sci. 6, 118.

Nipen, T.N., Seierstad, I.A., Lussana, C., Kristiansen, J., Hov, Ø., 2019. Adopting citizen observations in operational weather prediction. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/BAMS-D-18-0237.1.

Nordbo, A., Järvi, L., Haapanala, S., Moilanen, J., Vesala, T., 2013. Intra-city variation in urban morphology and turbulence structure in Helsinki, Finland. Bound.-Layer Meteorol. 146, 469–496. https://doi.org/10.1007/s10546-012-9773-y.

Nyelele, C., Kroll, C.N., Nowak, D.J., 2019. Present and future ecosystem services of trees in the Bronx, NY. Urban For. Urban Green. 42, 10–20.

Oke, T.R., 1988. The urban energy balance. Prog. Phys. Geogr. 12, 471–508.

Oke, T.R., 2007. Siting and exposure of meteorological instruments at urban sites. In: Air Pollution Modeling and Its Application XVII. Springer, pp. 615–631.

Oke, T.R., Maxwell, G.B., 1975. Urban heat island dynamics in Montreal and Vancouver. Atmos. Environ. 9, 191–200. https://doi.org/10.1016/0004-6981(75)90067-0.

Oke, T.R., Mills, G., Christen, A., Voogt, J.A., 2017. Urban Climates. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781139016476.

Oliveira, S., Andrade, H., Vaz, T., 2011. The cooling effect of green spaces as a contribution to the mitigation of urban heat: a case study in Lisbon. Build. Environ. 46, 2186–2194. https://doi.org/10.1016/j.buildenv.2011.04.034.

Oslo kommune, 2017. Statistikkbanken.

Parastatidis, D., Mitraka, Z., Chrysoulakis, N., Abrams, M., 2017. Online global land surface temperature estimation from Landsat. Remote Sens. https://doi.org/10.3390/rs9121208.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418. https://doi.org/10.1038/nature20584.

Public Health England, 2018. Heatwave Plan for England – Protecting Health and Reducing Harm From Severe Heat and Heat Waves.

RCoreTeam, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Saaroni, H., Ziv, B., 2010. Estimating the urban heat island contribution to urban and rural air temperature differences over complex terrain: application to an arid city. J. Appl. Meteorol. Climatol. 49, 2159–2166.

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, 331. https://doi.org/10.1007/s00704-009-0142-9.

Schatz, J., Kucharik, C.J., 2015. Urban climate effects on extreme temperatures in Madison, Wisconsin, USA. Environ. Res. Lett. 10, 94024. https://doi.org/10.1088/1748-9326/10/9/094024.

Schifano, P., Cappai, G., De Sario, M., Michelozzi, P., Marino, C., Bargagli, A.M., Perucci, C.A., 2009. Susceptibility to heat wave-related mortality: a follow-up study of a cohort of elderly in Rome. Environ. Health 8, 50.

Smid, M., Costa, A.C., 2018. Climate projections and downscaling techniques: a discussion for impact studies in urban systems. Int. J. Urban Sci. 22, 277–307. https://doi.org/10.1080/12265934.2017.1409132.

Statistikkbanken, 2018. Population Statistics: Annually, Estimated Figures. [WWW Document]. URL https://www.ssb.no/en/befolkning/statistikker/folkemengde/aar-berekna, Accessed date: 7 September 2019.

Stewart, I.D., 2000. Influence of meteorological conditions on the intensity and form of the urban heat island effect in Regina. Can. Geogr. 44, 271–285. https://doi.org/10.1111/j.1541-0064.2000.tb00709.x.

Stewart, I.D., 2011. A systematic review and scientific critique of methodology in modern urban heat island literature. Int. J. Climatol. 31, 200–217. https://doi.org/10.1002/joc.2141.

Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. Bull. Am. Meteorol. Soc. 93, 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1.

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., Eklundh, L., 2007. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sens. Environ. 110, 262–274.

Sun, Y.-J., Wang, J.-F., Zhang, R.-H., Gillies, R.R., Xue, Y., Bo, Y.-C., 2005. Air temperature retrieval from remote sensing data based on thermodynamics. Theor. Appl. Climatol. 80, 37–48.

Tooke, T.R., van der Laan, M., Coops, N.C., 2014. Mapping demand for residential building thermal energy services using airborne LiDAR. Appl. Energy 127, 125–134.

https://doi.org/10.1016/j.apenergy.2014.03.035.

Trenberth, K.E., Fasullo, J.T., 2012. Climate extremes and climate change: the Russian heat wave and other climate extremes of 2010. J. Geophys. Res. Atmos. 117.

Tsin, P.K., Knudby, A., Krayenhoff, E.S., Ho, H.C., Brauer, M., Henderson, S.B., 2016. Microscale mobile monitoring of urban air temperature. Urban Clim. 18, 58–72. https://doi.org/10.1016/j.uclim.2016.10.001.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

Turrini, T., Knop, E., 2015. A landscape ecology approach identifies important drivers of urban biodiversity. Glob. Chang. Biol. 21, 1652–1667. https://doi.org/10.1111/gcb.12825.

Varentsov, M., Konstantinov, P., Baklanov, A., Esau, I., Miles, V., Davy, R., 2018. Anthropogenic and natural drivers of a strong winter urban heat island in a typical Arctic city. Atmos. Chem. Phys. 18, 17573–17587.

Voogt, J.A., Oke, T.R., 2003. Thermal remote sensing of urban climates. Remote Sens. Environ. 86, 370–384. https://doi.org/10.1016/S0034-4257(03)00079-8.

Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: methods, applications, and trends. ISPRS J. Photogramm. Remote Sens. 64, 335–344.

Wicki, A., Parlow, E., Feigenwinter, C., 2018. Evaluation and modeling of urban Heat Island intensity in Basel, Switzerland. Climate. https://doi.org/10.3390/cli6030055.

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194.

Wolf-Grosse, T., Esau, I., Reuder, J., 2017. Sensitivity of local air quality to the interplay between small- and large-scale circulations: a large Eddy simulation study. Atmos. Chem. Phys. 17, 7261–7276. https://doi.org/10.5194/acp-17-7261-2017.

Xu, H., 2008. A new index for delineating built-up land features in satellite imagery. Int. J. Remote Sens. 29, 4269–4276. https://doi.org/10.1080/01431160802039957.

Yan, H., Fan, S., Guo, C., Hu, J., Dong, L., 2014. Quantifying the impact of land cover composition on intra-urban air temperature variations at a mid-latitude city. PLoS One 9, e102124.

Yang, J., Bou-Zeid, E., 2019. Designing sensor networks to resolve spatio-temporal urban temperature variations: fixed, mobile or hybrid? Environ. Res. Lett. 14 (7), 074022.

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

Zakšek, K., Oštir, K., 2012. Downscaling land surface temperature for urban heat island diurnal cycle analysis. Remote Sens. Environ. 117, 114–124. https://doi.org/10.1016/j.rse.2011.05.027.

Zhang, B., Xie, G., Gao, J., Yang, Y., 2014. The cooling effect of urban green spaces as a contribution to energy-saving and emission-reduction: a case study in Beijing, China. Build. Environ. 76, 37–43. https://doi.org/10.1016/j.buildenv.2014.03.003.

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

Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Frolking, S., Yao, R., Qiao, Z., Sobrino, A.J., 2018. Satellite remote sensing of surface urban Heat Islands: Progress, challenges, and perspectives. Remote Sens. https://doi.org/10.3390/rs11010048.

Zhu, W., Lü, A., Jia, S., 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sens. Environ. 130, 62–73.

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