



## A multi-resolution air temperature model for France from MODIS and Landsat thermal data

Ian Hough, Allan C Just, Bin Zhou, Michael Dorman, Johanna Lepeule, Itai  
Kloog

### ► To cite this version:

Ian Hough, Allan C Just, Bin Zhou, Michael Dorman, Johanna Lepeule, et al.. A multi-resolution air temperature model for France from MODIS and Landsat thermal data. *Environmental Research*, 2020, 183, pp.109244. 10.1016/j.envres.2020.109244 . inserm-03184627v2

**HAL Id: inserm-03184627**

**<https://inserm.hal.science/inserm-03184627v2>**

Submitted on 18 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A multi-resolution air temperature model for France from MODIS and Landsat thermal data

Ian Hough<sup>a,b,\*</sup>, Allan C. Just<sup>c</sup>, Bin Zhou<sup>b</sup>, Michael Dorman<sup>b</sup>, Johanna Lepeule<sup>a</sup>, and Itai Kloog<sup>b</sup>

<sup>a</sup> Univ. Grenoble Alpes, Inserm, CNRS, IAB, Site Sante – Allée des Alpes, 38700 La Tronche, France

<sup>b</sup> Department of Geography and Environmental Development, Ben-Gurion University of the Negev, P.O.B. 653, Be'er Sheva, Israel

<sup>c</sup> Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY 10029-5674, USA

\* Corresponding author. Institute for Advanced Biosciences (IAB), Site Santé – Allée des Alpes, 38700 La Tronche, France. E-mail: [ian.hough@univ-grenoble-alpes.fr](mailto:ian.hough@univ-grenoble-alpes.fr)

This is the author-produced version of an article accepted for publication in *Environmental Research* following peer review. The version of record is available online at: <https://doi.org/10.1016/j.envres.2020.109244>.

## Abstract

Understanding and managing the health effects of ambient temperature ( $T_a$ ) in a warming, urbanizing world requires spatially- and temporally-resolved  $T_a$  at high resolutions. This is challenging in a large area like France which includes highly variable topography, rural areas with few weather stations, and heterogeneous urban areas where  $T_a$  can vary at fine spatial scales. We have modeled daily  $T_a$  from 2000 – 2016 at a base resolution of 1 km<sup>2</sup> across continental France and at a 200 x 200 m<sup>2</sup> resolution over large urban areas. For each day we predict three  $T_a$  measures: minimum ( $T_{\min}$ ), mean ( $T_{\text{mean}}$ ), and maximum ( $T_{\max}$ ). We start by using linear mixed models to calibrate daily  $T_a$  observations from weather stations with remotely sensed MODIS land surface temperature (LST) and other spatial predictors (e.g. NDVI, elevation) on a 1 km<sup>2</sup> grid. We fill gaps where LST is missing (e.g. due to cloud cover) with additional mixed models that capture the relationship between predicted  $T_a$  at each location and observed  $T_a$  at nearby weather stations. The resulting 1 km  $T_a$  models perform very well, with ten-fold cross-validated  $R^2$  of 0.92, 0.97, and 0.95, mean absolute error (MAE) of 1.4 °C, 0.9 °C, and 1.4 °C, and root mean square error (RMSE) of 1.9 °C, 1.3 °C, and 1.8 °C ( $T_{\min}$ ,  $T_{\text{mean}}$ , and  $T_{\max}$ , respectively) for the initial calibration stage. To increase the spatial resolution over large urban areas, we train random forest and extreme gradient boosting models to predict the residuals ( $R$ ) of the 1 km  $T_a$  predictions on a 200 x 200 m<sup>2</sup> grid. In this stage we replace MODIS LST and NDVI with composited top-of-atmosphere brightness temperature and NDVI from the Landsat 5, 7, and 8 satellites. We use a generalized additive model to ensemble the random forest and extreme gradient boosting predictions with weights that vary spatially and by the magnitude of the predicted residual. The 200 m models also perform well, with ten-fold cross-validated  $R^2$  of 0.79, 0.79, and 0.85, MAE of 0.4, 0.3, and 0.3, and RMSE of 0.6, 0.4, and 0.5 ( $R_{\min}$ ,  $R_{\text{mean}}$ , and  $R_{\max}$ , respectively). Our model will reduce bias in epidemiological studies in France by improving  $T_a$  exposure assessment in both urban and rural areas, and our methodology demonstrates that MODIS and Landsat thermal data can be used to generate gap-free timeseries of daily minimum, maximum, and mean  $T_a$  at a 200 x 200 m<sup>2</sup> spatial resolution.

## Keywords

air temperature; land surface temperature; MODIS; Landsat; exposure error

## 1. Introduction

Ambient or near-surface air temperature ( $T_a$ ) is increasingly recognized as an important health risk. High or low  $T_a$  is associated with increased morbidity and mortality across regions and climates (Gasparrini et al., 2015; Guo et al., 2014; Song et al., 2017), and recent work suggests that high  $T_a$  may exacerbate the effect of exposure to particulate matter (PM), another major health hazard (Li et al., 2017).  $T_a$  exposure is a growing concern in cities, which are often warmer than the surrounding countryside due to increased heat accumulation and slower heat diffusion (Arnfield, 2003). Urban areas are now home to more than half the world's population, and this share is projected to increase to almost 70% by 2050 (United Nations, 2018). Health effects of  $T_a$  are also seen in rural populations (Lee et al., 2016), although fewer studies have examined these due to the difficulty of estimating  $T_a$  exposure. Meanwhile climate change is increasing  $T_a$  and the frequency of extreme events such as heat waves in both urban and rural areas (IPCC, 2013). The health burden of  $T_a$  exposure is expected to grow as climate change and urbanization continue (Gasparrini et al., 2017; Wang et al., 2018).

Understanding, monitoring, and managing  $T_a$  health effects requires spatiotemporally-resolved  $T_a$  at high resolutions. Weather station networks measure  $T_a$  at high temporal resolution, but rarely capture spatial variation at the scales needed for epidemiological studies (e.g. across a region, within a city). Failure to account for spatial variation in  $T_a$  can introduce error in exposure assessment, which tends to bias health effect estimates towards the null (Zeger et al., 2000). Some recent epidemiological studies have addressed this issue by using spatiotemporally-resolved  $T_a$  estimates from numerical weather prediction models such as WRF (Ha et al., 2017b, 2017a), but computational limitations currently restrict these models to medium spatial resolutions (e.g. 4 km) or small geographic areas (e.g. a single city). In urban areas, studies have used weather model  $T_a$  estimates or indicators such as sky view factor, vegetation abundance, and land surface temperature to create indexes that identify warmer and cooler areas within a city (Goggins et al., 2012; Ho et al., 2017; Laaidi et al., 2012; Milojevic et al., 2016; Smargiassi et al., 2009). Studies to date have focused on the typical spatial distribution of  $T_a$  during a specific time period (e.g. a single heat wave, the hot season) as the limited temporal variability of the indicator variables and cost of numerical weather prediction have precluded consideration of changes in the pattern of warmer and cooler areas over time.

Other recent studies have used  $T_a$  estimates from hybrid land use regression models that predict  $T_a$  based on remotely sensed 1 km land surface temperature (LST) and spatial and

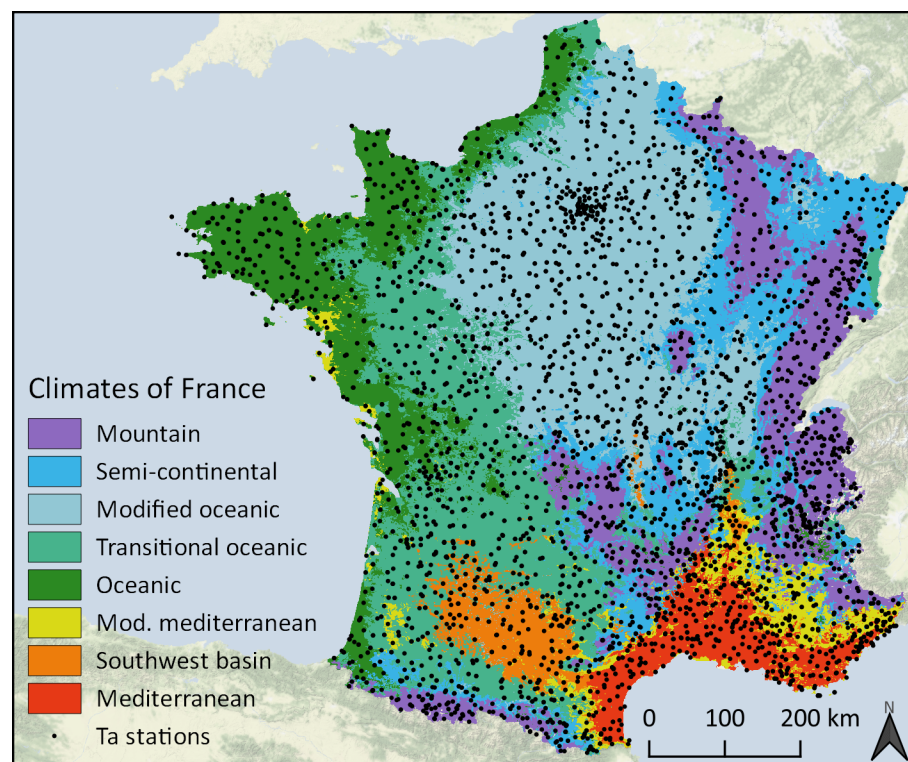
spatiotemporal variables such as elevation and normalized difference vegetation index (NDVI) (Kloog et al., 2015; Shi et al., 2016b, 2015). This approach takes advantage of the growing body of satellite earth observation data and the fact that LST is a good indicator of spatiotemporal variation in  $T_a$  (Oyler et al., 2016). In particular, a technique that uses linear mixed models to calibrate the relationship between daily 1 km LST from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument and  $T_a$  has been shown to perform well over large, heterogeneous areas including the northeastern USA (root mean square error [RMSE] 2.2 °C) (Kloog et al., 2014), the southeastern USA (RMSE 1.4 °C) (Shi et al., 2016a), France (RMSE 1.7 °C) (Kloog et al., 2017), and Israel (RMSE 1.2 °C) (Rosenfeld et al., 2017). These models are parsimonious compared to numerical weather prediction, which allows them to capture both spatial and temporal variation in  $T_a$  over large areas and long time periods. Their spatial resolution suffices for areas where  $T_a$  varies little at scales of less than 1 km and for studies where subjects' locations are only approximately known. But finer spatial resolution estimates are needed for studies with address-level location data, particularly in urban areas where  $T_a$  can vary markedly within a square kilometer. Very high spatiotemporal resolutions would also benefit studies that have time-location data (e.g. GPS tracks).

In this study we extend the mixed modeling approach to predict daily minimum, mean, and maximum  $T_a$  ( $T_{\min}$ ,  $T_{\text{mean}}$ ,  $T_{\max}$ , respectively) at a 1 km resolution across continental France and at a 200 m resolution across 103 urban areas in continental France. We improve performance at the 1 km resolution by allowing the daily  $T_a \sim$  LST relationship to vary between climatic regions, and we consider both daytime and nighttime MODIS LST, which allows us to predict diurnal ( $T_{\max}$ ) and nocturnal ( $T_{\min}$ ) temperature in addition to  $T_{\text{mean}}$ . This is useful both for studies of urban heat islands, which exhibit different spatial patterns and intensities during day vs. night (Arnfield, 2003), and for studies of  $T_a$  variability, which recent work suggests may independently affect health (Guo et al., 2016; Molina and Saldarriaga, 2017; Shi et al., 2015). We also add a local stage that uses an ensemble of machine learning algorithms to predict the residuals of the 1 km model in urban areas based on higher spatial resolution predictors including thermal data from the Landsat 5, 7, and 8 satellites. This allows us to predict daily  $T_a$  over 17 years at a 200 m spatial resolution which better captures intra-urban  $T_a$  variation across 103 urban areas.

## 2. Data and methods

### 2.1. Study area and period

Our study area is continental France, comprising all French territory in Europe except Corsica. It covers 542,973 km<sup>2</sup> of topographically and climatically diverse terrain with elevations that range from -10 to 4,809 m. Joly et al. (2010) classify France into eight climatic regions based on the magnitude, variability, and seasonality of temperature and precipitation (**Fig. 1**). The north and west coasts have a wet, temperate oceanic climate, which transitions to a drier, cooler modified oceanic climate in the north center. The mountainous east, south center, and southwest have variable montane and semi-continental climates with cold winters. In the southeast, the Mediterranean coast has hot, dry summers with mild wet winters; the inland southeast and isolated segments of the west coast are similar but cooler. The southwest basin resembles the inland southeast but with drier winters.



**Fig. 1.** Climatic regions of France according to Joly et al. (2010) and METEO-FRANCE stations used in the current study.

The estimated population on January 1, 2018 was 64,388,583 (INSEE, 2018). About 80% of the population is urban, and this share is projected to grow to 88% by 2050 (United Nations, 2018). The largest urban area, Paris, has a population of 12.5 million (20% of the total) and the six next largest urban areas have a population of one to 2.3 million (combined 14% of total). A

further 10% of the population lives in cities of one half to one million, and 37% live in urban areas with fewer than half a million residents (**Fig. S1**). Our study period is January 1, 2000 through December 31, 2016.

## *2.2. Meteorological observations*

We use daily weather station observations from Météo France, the French national meteorological service. About 64% of the observations come from stations managed by Météo France; the remaining stations are managed by other entities. All observations are quality controlled by Météo France. We exclude stations with no metadata or that do not record hourly  $T_a$ , and for each month during the study period we exclude stations that were active for fewer than 21 days in the month. This leaves 1710 to 2314 stations on each day. The stations are distributed over the entire study region, but are denser in populous areas (e.g. Paris, the southeast) and the Alps (which has many ski resorts, hydroelectric dams, and avalanche monitors) (**Fig. 1**). Just 3% of the stations are located within large urban areas (as defined in section 2.7), 7% are in peri-urban areas (within 5 km of an urban area), and the remaining 90% are rural.

The stations calculate daily  $T_{\min}$  as the lowest  $T_a$  observed from 18 UTC the previous day until 18 UTC on the day; daily  $T_{\max}$  is the highest  $T_a$  observed from 6:00 UTC on the day until 6:00 UTC the following day. Most stations calculate  $T_{\text{mean}}$  as the mean of all (at least 24)  $T_a$  observations from 0 UTC on the day until 0 UTC the following day. However, about 40% of the  $T_{\text{mean}}$  observations were calculated as the average of  $T_{\min}$  and  $T_{\max}$ . We exclude these observations, meaning our final dataset has fewer observations for  $T_{\text{mean}}$  than for  $T_{\min}$  or  $T_{\max}$ . Daily  $T_a$  at the included stations during the study period ranged from  $T_{\min}$  of  $-31.2\text{ }^{\circ}\text{C}$  to  $T_{\max}$  of  $44.1\text{ }^{\circ}\text{C}$ ; mean  $T_{\text{mean}}$  was  $11.3\text{ }^{\circ}\text{C}$  with a standard deviation of  $7.1\text{ }^{\circ}\text{C}$  (**Table S1**).

## *2.3. Land surface temperature and emissivity*

We use version 6 of the widely-used MODIS daily 1 km land surface temperature and emissivity product from the Terra and Aqua satellites (MOD11A1 and MYD11A1, respectively) (**Table 1**). These products include daytime and nighttime LST derived using a split-window algorithm and land use classification-based emissivity and have been masked for clouds and validated to  $\pm 2\text{ K}$  in clear-sky conditions across 47 sites on all seven continents (Wan, 2014). We use the quality assessment band to exclude pixels with LST error  $> 2\text{ K}$ . As LST retrieval error increases over snow and water, we also exclude pixels with  $\text{NDVI} < 0$  or where the corresponding 1 km grid cell has land cover of  $> 33\%$  water.

**Table 1.** Satellite instruments used in this study.

Instrument	Satellite	Resolution	Revisit time	Overpass*	Time period
MODIS	Terra	1 km	12 hours	10:00 22:00	2000-02-02 – present
MODIS	Aqua	1 km	12 hours	13:00 01:00	2002-07-04 – present
TM	Landsat 5	120 m <sup>†</sup>	16 days	10:00	1984-03-01 – 2011-11-18
ETM+	Landsat 7	60 m <sup>†</sup>	16 days	10:00	1999-04-15 – present
TIRS	Landsat 8	100 m <sup>†</sup>	16 days	10:00	2013-02-11 – present

\*Approximate local solar time; <sup>†</sup>Resampled to 30 m

#### 2.4. Top-of-atmosphere brightness temperature

For large urban areas, we composite 30 m top-of-atmosphere brightness temperature ( $T_b$ ) from the Landsat 5, 7, and 8 satellites (**Table 1**).  $T_b$  is the kinetic temperature a perfect blackbody would have if it emitted the quantity of thermal radiation measured by the satellite instrument. Converting  $T_b$  to LST requires correcting for atmospheric effects and accounting for the emissivity of the earth's surface. This is difficult in the case of the Landsat satellites because Landsat 5 and 7 have only a single thermal band and the USGS Landsat 8's second thermal band is contaminated by stray light, precluding the use of the split-window algorithm (Li et al., 2013). A global Landsat LST product is under development but not yet available (Malakar et al., 2018), so for this study we use  $T_b$  from the USGS Landsat Collection 1 Level-2 surface reflectance products (USGS, 2018a, 2018b).

The 16-day revisit time of the Landsat satellites means that  $T_b$  is unavailable for many locations on many days. Cloud cover and sensor malfunctions also contribute to these data gaps and can increase error in  $T_b$  retrieval. To reduce error, we discard all scenes with cloud cover > 75%. We also discard all scenes captured during periods of instrument malfunction, which we identified by checking summary statistics of each scene for unrealistic values (e.g. mean  $T_b$  > 100 °C). We then trim the edges of Landsat 5 scenes by 2.5 km to remove abnormalities (Robinson et al., 2017) and mask pixels identified as high- or medium-confidence cloud in the pixel quality assessment band. We mask any remaining pixels where  $T_b \leq -25$  °C or  $T_b \geq 50$  °C. Finally, for each calendar month we composite all  $T_b$  retrievals during the entire study period (e.g. every January in 2000 – 2016). This yields 12 gap-free  $T_b$  datasets representing the 17-year mean  $T_b$  of each pixel in each calendar month.

#### 2.5. NDVI

We use version 6 of the MODIS monthly composite 1 km NDVI product from the Terra and Aqua satellites (MOD13A3 and MYD13A3, respectively). For large urban areas we also



composite 30 m NDVI from the Landsat 5, 7, and 8 Collection 1 Level-2 surface reflectance products. We use a similar quality assurance and compositing procedure as for  $T_b$ , first discarding all scenes with greater than 75% cloud cover or during periods of thermal sensor malfunction (as this results in unreliable cloud confidence scores in the pixel quality assessment band). We then trim the edges of Landsat 5 scenes by 2.5 km and adjust NDVI from Landsat 5 and Landsat 7 to match Landsat 8 using equation Eq. 1 (Robinson et al., 2017).

$$NDVI_{L8} = 0.0235 + 0.9723 \times NDVI_{L5,L7} \quad \text{Eq. 1}$$

Similar to Robinson et al. (2017), for each calendar month we create two 17-year mean composites, one using pixels marked as clear in the pixel quality assurance band (i.e. not cloud, cloud shadow, snow, or water) and a second using pixels marked as snow or water. Finally, we mosaic the two composites preferring the clear pixels composite.

## 2.6. Elevation, Population, Land Cover, and Climatic Regions

We use version 1.1 of the European Digital Elevation Model (EU-DEM) from the Copernicus Land Monitoring Service. These data have a 25 m spatial resolution and vertical RMSE of  $\pm 7$  m (Tøttrup, 2014). We also use 200 m gridded 2010 population from INSEE, the French national statistics agency (INSEE, 2017). We use the 100 m Corine Land Cover (CLC) inventory for 2000, 2006, and 2012. The 2000 edition has been validated to better than 85% thematic accuracy (Bossard et al., 2000). We aggregate the land cover classes into four groups: artificial, vegetation, bare, and water (**Table S2**). Finally, we use the eight climatic regions of Joly et al. (2010), which are based on temperature and precipitation patterns (**Fig. 1**).

## 2.7. Model grids

For the 1 km model, we create a grid covering continental France by making a 1 km square buffer around the centroid of each MODIS 1 km LST pixel in the ETRS89-LAEA Europe (EPSG:3035) equal-area projection. We associate each 1 km grid cell with the MODIS LST and NDVI pixel having the same centroid and calculate the mean elevation, total population, percent area of each land cover group, and climate region with greatest spatial overlap.

For the 200 m model, we create a grid covering large urban areas. Starting from a 200 m grid in the ETRS89-LAEA Europe (EPSG:3035) equal-area projection, we select all cells in continental France containing “Urban fabric” or “Industrial or commercial units” in the 2012 CLC inventory. We associate each cell with the corresponding INSEE gridded population and

select cells with 50 or more inhabitants as well as the eight surrounding cells (i.e. including diagonal neighbors). We define urban areas as four-wise contiguous (i.e. excluding diagonal neighbors) groups of cells and sum the population of all cells in each urban area. Finally, we eliminate urban areas with population  $< 50,000$ . This leaves 103 large urban areas ranging from greater Paris (9.4 million inhabitants) to Armentières (50,260 inhabitants). For each 200 m grid cell in a large urban area or that contains a weather station we calculate the mean 17-year composite Landsat  $T_b$  and NDVI for each calendar month, mean elevation, and percent area of each land cover group.

## 2.8. Statistical methods

We use a four-stage approach to predict  $T_a$ : stages 1 and 2 predict daily 1 km  $T_a$  across continental France and stages 3 and 4 predict daily 200 m  $T_a$  within large urban areas. We consider each year during the study period (2000 – 2016) and each  $T_a$  measure ( $T_{min}$ ,  $T_{max}$ , and  $T_{mean}$ ) separately. Stages 1 and 2 are an extension of the method used in (Kloog et al., 2017) and are detailed in Appendix A. Sections 2.8.1 to 2.8.2 detail stages 3 and 4; the following is a brief overview of all stages.

In stage 1 we calibrate  $T_a$  at each station as a function of daily 1 km LST and emissivity, monthly 1 km NDVI, and 1 km elevation, population, and land cover. We use a linear mixed model to allow the  $T_a \sim \text{LST}$  relationship to vary by day within each climatic region. We use this calibrated relationship to predict 1 km  $T_a$  ( $T_{ap\_s1}$ ) for all cell-days where LST is available.

In stage 2, we fill gaps in  $T_{ap\_s1}$  where 1 km LST was not available by calibrating  $T_{ap\_s1}$  as a function of daily 1 km inverse distance weighting interpolated observed  $T_a$  ( $T_{IDW}$ ). We use a linear mixed model to allow the  $T_{ap\_s1} \sim T_{IDW}$  relationship to vary by location. We use this calibrated relationship to fill gaps in  $T_{ap\_s1}$ , producing gap-free daily 1 km predicted  $T_a$  ( $T_{ap\_1km}$ ). This is the 1 km  $T_a$  model.

In stage 3, we calculate the daily 200 m residuals of the 1 km  $T_a$  model ( $R$ ) and train random forest (RF) and extreme gradient boosting (GB) models to predict  $R$  based on latitude, longitude, Julian day, climatic region, 200 m composite  $T_b$  and NDVI, and 200 m elevation, population, and land cover. We use each of these models predict the residual for all 200 m cell-days ( $R_{p\_rf}$  and  $R_{p\_gb}$ , respectively).

In stage 4, we calibrate a generalized additive model that ensembles  $R_{p\_rf}$  and  $R_{p\_gb}$ . We use a tensor product smooth with interaction to allow the relative performance of the RF and GB

models to vary by location and with the magnitude of the predicted residual. Finally, we add the ensemble predictions to  $T_{ap\_1km}$  to get daily 200 m predicted  $T_a$  for large urban areas ( $T_{ap\_200m}$ ). This is the 200 m  $T_a$  model.

#### 2.8.1. Stage 3: increasing spatial resolution to 200 m across large urban areas

In stage 3 we increase the spatial resolution of our predictions over large urban areas. We start by associating each 200 m grid cell with  $T_{ap\_1km}$  ( $T_a$  predicted in stage 2 by the final 1 km model) from the 1 km grid cell that contains the 200 m grid cell. Next, we calculate the residuals ( $R$ ) for all 200 m grid cell-days with a weather station  $T_a$  observation by subtracting observed  $T_a$  from  $T_{ap\_1km}$ . The number of cell-days with a weather station observation varies by year; on average there are about 462 thousand for  $T_{mean}$  and 789 thousand for each of  $T_{min}$  and  $T_{max}$ . We use these cell-days to train a random forest and an extreme gradient boosting (XGBoost) model with the equation:

$$R_{ij} = f \left( \begin{matrix} T_{ap\_1kmi j}, T_{bim}, NDVI_{im}, Land\ Cover_{ily}, \\ Climate_i, Elevation_i, Population_i, x_i, y_i, j \end{matrix} \right) + \varepsilon_{ij} \quad \text{Eq. 2}$$

where  $R_{ij}$  is the residual of the 1 km  $T_a$  model associated with 200 m grid cell  $i$  on day  $j$ ;  $f$  designates the random forest or extreme gradient boosting function;  $T_{ap\_1kmi j}$  is the 1 km  $T_a$  model prediction associated with 200 m grid cell  $i$  on day  $j$ ;  $T_{bim}$  is the Landsat top-of-atmosphere brightness temperature of cell  $i$  for the calendar month  $m$  in which day  $j$  falls;  $NDVI_{im}$  is the Landsat NDVI of cell  $i$  for the calendar month  $m$  in which day  $j$  falls;  $Land\ Cover_{ily}$  is the fraction of cell  $i$  occupied by each land cover group  $l$  in the CLC inventory year  $y$  closest to day  $j$ ;  $Climate_i$  is the climatic region of cell  $i$ ;  $Elevation_i$  is the elevation of cell  $i$ ;  $Population_i$  is the population of cell  $i$ ;  $x_i$  and  $y_i$  are the geographical coordinates of cell  $i$ ;  $j$  is the Julian day; and  $\varepsilon_{ij}$  is the error for cell  $i$  on day  $j$ .

We use the R packages ranger (Wright and Ziegler, 2017), XGBoost (Chen and Guestrin, 2016), and mlr (Bischl et al., 2016) to train the random forest and XGBoost models. We tune the models using the sequential model-based optimization of package mlrMBO (Bischl et al., 2017). Briefly, mlrMBO estimates optimal hyperparameter values by iteratively training and evaluating a model using hyperparameter values that are chosen based on the performance of previous iterations. We use a fixed number of iterations and evaluate performance as the mean RMSE of two random 80% holdouts (i.e. we train the model on a 20% random sample of the data, predict and calculate RMSE for the held-out 80%, repeat, and take the mean of the two

276 RMSEs). Initial exploration showed that this resampling approach produced stable estimates of  
277 RMSE at a lower computational cost than cross-validation.

278 For the random forest, we use 400 trees and a minimum of 5 observations per node, and tune  
279 mtry (the number of variables to consider for each split) from 3 to 12 (25% to 100% of the  
280 explanatory variables) using 6 mlrMBO iterations. Initial exploration showed that using more  
281 than 400 trees only marginally increased performance and had a high computational cost. For  
282 the XGBoost model, we use the gbtrees booster with 100 rounds and set gamma (the minimum  
283 loss reduction for a split) to 5. We use 24 mlrMBO iterations to tune eta (the learning rate) from  
284 0.1 to 0.3, the maximum tree depth from 5 to 20, the minimum number of observations per node  
285 from 3 to 30, and the fraction of features used in each tree from 0.75 to 1.

286 We evaluate the performance of the stage 3 models using 5-fold cross-validation with nested  
287 tuning. We use the final stage 3 random forest and XGBoost models to predict the residual of  
288 the 1 km  $T_a$  model ( $R_{p\_rf}$  and  $R_{p\_xgb}$ , respectively) for all 200 m cell-days.

#### 289 2.8.2. Stage 4: improving 200 m predictions

290 In stage 4 we improve the stage 3 predictions by ensembling. We use all 200 m grid cell-days  
291 with a weather station  $T_a$  observation to calibrate a generalized additive model (GAM) with the  
292 formula:

$$R_{ij} = t(x_i, y_i) \times R_{p\_rfij} + t(x_i, y_i) \times R_{p\_gbij} + \varepsilon_{ij} \quad \text{Eq. 3}$$

293 where  $R_{ij}$  is the residual of the 1 km  $T_a$  model associated with 200 m grid cell  $i$  on day  $j$ ;  $t(x_i,$   
294  $y_i)$  is a tensor product smooth of the  $x$  and  $y$  coordinates of cell  $i$ ;  $R_{p\_rfij}$  and  $R_{p\_gbij}$  are the  
295 predicted residuals of the 1 km  $T_a$  model from the stage 3 random forest and XGBoost model,  
296 respectively, for cell  $i$  on day  $j$ ; and  $\varepsilon_{ij}$  is the error for cell  $i$  on day  $j$ . The GAM averages the  
297 random forest and XGBoost predicted residuals using weights that vary both by location and  
298 with the magnitude of each model's predicted residual. Finally, we add the ensemble-predicted  
299 residuals for all 200 m grid cells to  $T_{ap\_1km}$  ( $T_a$  predicted in stage 2 by the final 1 km model) to  
300 obtain daily 200 m predicted  $T_a$  ( $T_{ap\_200m}$ ) across large urban areas.

#### 301 2.8.3. Performance assessment

302 We use 10-fold out-of-sample cross-validation to assess the overall performance of the models.  
303 For the random forest and XGBoost model we use nested tuning (i.e. within each cross-  
304 validation fold we tune the model as described in section 2.8.1). To evaluate the models' ability

to capture both spatial and temporal patterns in  $T_a$ , we also calculate the spatial and temporal components of the errors. The spatial component is the difference at each station between the annual mean of daily observed  $T_a$  ( $\overline{T_a}$ ), and the annual mean of daily predicted  $T_a$  ( $\overline{T_{ap}}$ ). The temporal component is the difference at each station between  $\Delta T_a$  and  $\Delta T_{ap}$  where  $\Delta T_a$  is the difference between daily observed  $T_a$  and  $\overline{T_a}$  and  $\Delta T_{ap}$  is the difference between daily predicted  $T_a$  and  $\overline{T_{ap}}$ . We use Google Earth Engine (Gorelick et al., 2017) to quality assure and composite Landsat  $T_b$  and NDVI and aggregate them to the 200 m grid cells. For all other data processing and analyses we use R version 3.4.4 (R Core Team, 2018).

### 3. Results

**Table 2** presents the mean 10-fold cross-validated performance of the stage 1 models (predicting daily 1 km  $T_a$  from LST) across all years. The models perform very well, with  $R^2$  of 0.92 or higher, RMSE of less than 2 °C, and mean absolute error (MAE) of less than 1.5 °C. All models have very low bias: the slope of observed vs. predicted  $T_a$  is 1.00 while the intercept ranges from 0.01 to 0.02. The  $T_{mean}$  models perform best overall (MAE 0.94), followed by the  $T_{max}$  (MAE 1.35) and  $T_{min}$  (MAE 1.43) models. The models capture both spatial and temporal variation in  $T_a$  and show little variation in performance between years, although overall  $T_{mean}$  performance decreases slightly after 2010, possibly reflecting degradation of the Terra MODIS instrument (**Table S4**). Consistent with previous studies, nighttime LST is the best predictor of  $T_{min}$  and  $T_{mean}$  while daytime LST is the best predictor of  $T_{max}$  (Oyler et al., 2016; Rosenfeld et al., 2017; Yoo et al., 2018). Aqua LST is a better predictor of  $T_{min}$  and  $T_{max}$  while Terra LST is a better predictor of  $T_{mean}$ . This is expected as the Aqua overpasses (approximately 1:30 and 13:30 local solar time) are closer to the time at which  $T_{min}$  and  $T_{max}$  typically occur in France. However, Aqua LST is only available since July 2002, so we use Terra LST for all models prior to 2003.

**Table 2.** Stage 1 model (predicting daily 1 km  $T_a$  from LST): 10-fold cross-validated performance across all years (2000 – 2016), overall, spatial, and temporal components.

	N*	Overall			Spatial			Temporal		
		$R^2$	RMSE	MAE	$R^2$	RMSE	MAE	$R^2$	RMSE	MAE
$T_{min}$	354	0.92	1.89	1.43	0.89	1.10	0.80	0.94	1.61	1.19
$T_{mean}$	205	0.97	1.29	0.94	0.95	0.83	0.57	0.97	1.15	0.84
$T_{max}$	324	0.95	1.81	1.35	0.88	1.23	0.89	0.96	1.52	1.12

\* N = mean thousands of observations used to fit each annual model

**Table 3** presents the 10-fold cross-validated performance of the stage 1 models across all years by calendar month and season and **Table 4** presents the performance by climatic region and urban vs. rural locations. The  $T_{\min}$  and  $T_{\text{mean}}$  models perform slightly less well in winter months, possibly due to higher LST missingness from more frequent cloud cover. The  $T_{\max}$  model performs best in late winter, early spring, and fall. The models perform less well in the mountain, semi-continental, and modified Mediterranean climates. These climates occur in mountainous areas where large contrasts in topography and land cover make modelling particularly challenging; other factors not included in the model may also reduce performance in these areas. The models perform slightly better in peri-urban areas than in urban and rural areas, possibly due to the higher density of weather stations (peri-urban areas have the most stations per km<sup>2</sup>).

**Fig. 2** shows the spatial pattern of the daily 1 km  $T_a$  predictions of the stage 2 model on selected winter and summer days. On the cold winter day of Feb 18, 2003, predictions range from  $T_{\min}$  of -17 °C in parts of the Alps, the Massif Central, and the Pyrenees to  $T_{\max}$  of 11 °C on the Mediterranean coast. The urban heat island of Paris is faintly visible in the north center of the  $T_{\min}$  and  $T_{\text{mean}}$  maps but disappears on the  $T_{\max}$  map. Spatial contrasts corresponding to terrain features are well resolved, and the spatial pattern of  $T_{\min}$  vs.  $T_{\text{mean}}$  vs.  $T_{\max}$  varies most in the north, northeast, and southwest.

**Table 3.** Stage 1 model performance (predicting daily 1 km  $T_a$  from LST): 10-fold cross-validated performance across all years (2000 – 2016), by month and season.

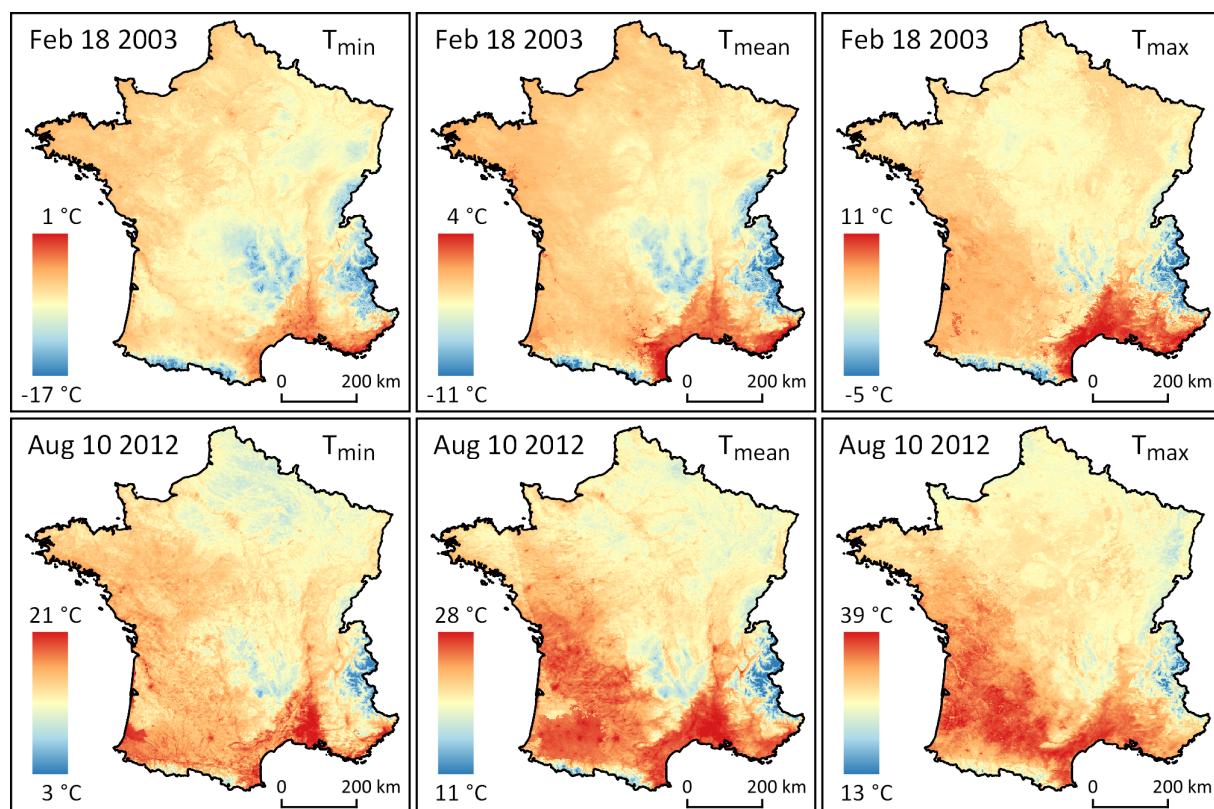
	$T_{\min}$			$T_{\text{mean}}$			$T_{\max}$		
	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
Jan	0.83	2.16	1.60	0.89	1.54	1.11	0.86	1.87	1.37
Feb	0.84	2.03	1.51	0.91	1.37	0.99	0.89	1.74	1.28
Mar	0.80	1.92	1.46	0.91	1.22	0.91	0.89	1.72	1.28
Apr	0.77	1.82	1.39	0.91	1.17	0.85	0.87	1.75	1.32
May	0.80	1.75	1.33	0.92	1.20	0.86	0.85	1.85	1.39
Jun	0.81	1.74	1.32	0.92	1.23	0.90	0.84	1.94	1.46
Jul	0.79	1.71	1.30	0.92	1.19	0.88	0.84	1.90	1.44
Aug	0.78	1.77	1.35	0.92	1.18	0.88	0.87	1.89	1.43
Sep	0.79	1.83	1.40	0.92	1.12	0.84	0.87	1.70	1.29
Oct	0.83	1.94	1.47	0.91	1.26	0.93	0.88	1.67	1.25
Nov	0.82	2.02	1.52	0.89	1.42	1.03	0.88	1.69	1.25
Dec	0.82	2.17	1.61	0.86	1.69	1.21	0.84	1.94	1.39
Winter	0.83	2.12	1.57	0.89	1.55	1.11	0.86	1.86	1.35
Spring	0.86	1.83	1.40	0.94	1.20	0.87	0.91	1.77	1.33
Summer	0.80	1.74	1.32	0.92	1.20	0.89	0.86	1.91	1.44
Fall	0.87	1.92	1.46	0.95	1.26	0.92	0.93	1.69	1.27

**Table 4.** Stage 1 model performance (predicting daily 1 km  $T_a$  from LST): 10-fold cross-validated performance across all years (2000 – 2016), by climatic region and urban vs. rural locations.

	$T_{\min}$			$T_{\text{mean}}$			$T_{\max}$		
	$R^2$	RMSE	MAE	$R^2$	RMSE	MAE	$R^2$	RMSE	MAE
Mountain	0.90	2.22	1.71	0.95	1.69	1.25	0.93	2.26	1.73
Semi-continental	0.91	2.11	1.61	0.96	1.44	1.07	0.95	2.00	1.52
Modified oceanic	0.94	1.53	1.16	0.98	0.98	0.73	0.98	1.33	1.01
Transitional oceanic	0.92	1.81	1.37	0.97	1.20	0.88	0.95	1.74	1.31
Oceanic	0.90	1.79	1.33	0.96	1.20	0.88	0.94	1.83	1.36
Mod. Mediterranean	0.90	2.22	1.71	0.96	1.43	1.07	0.94	2.03	1.55
Southwest basin	0.94	1.60	1.22	0.98	1.04	0.76	0.97	1.40	1.04
Mediterranean	0.93	1.81	1.40	0.98	1.11	0.84	0.96	1.62	1.25
Urban	0.93	1.85	1.35	0.97	1.32	0.96	0.95	1.79	1.35
Peri-urban*	0.93	1.71	1.29	0.97	1.18	0.87	0.96	1.71	1.27
Rural	0.92	1.90	1.44	0.97	1.30	0.94	0.95	1.82	1.36

\* non-urban locations within 5 km of a large urban area

On the hot summer day of Aug 10, 2012, predictions ranged from a  $T_{\min}$  of 3 °C in parts of the Alps to a  $T_{\max}$  of 39 °C in the southeast and southwest. On the  $T_{\min}$  map, the southwestern cities of Toulouse and Bordeaux stand out as hotspots, while Paris and Rouen are faintly visible as warm spots in the north. The north is colder than the Vosges mountains in the northeast and the Pyrenees in the southwest are warmer than the alps. The warmest areas are the southern Rhone river valley in the southeast and a patch of the southwestern Atlantic coast. On the  $T_{\text{mean}}$  map, Paris and Rouen are still visible, Lyon stands out in the east, and a few northwestern cities appear. Much of the southwest is as warm as the southeast, and the southwestern cities are harder to distinguish from the countryside. On the  $T_{\max}$  map, Lyon, Rouen, and some northwestern cities remain faintly visible, Pau and Tarbes appear in the southwest, and the north is warmer than the Vosges.



**Fig. 2.** Predicted 1 km  $T_a$  from the stage 2 model on selected days: Feb 18, 2003 (top row) and Aug 10, 2012 (bottom row).

**Table 5** presents the 10-fold cross-validated performance of the stage 4 models (predicting daily 200 m residuals of the 1 km model using an ensemble) across all years and by month and season; **Table 6** presents the performance by climatic region and urban *vs.* rural locations. These models also perform well, with overall  $R^2$  of 0.79 to 0.85, RMSE of 0.41 to 0.63, and MAE of 0.26 to 0.39 (residual scale). As with the stage 1 models, the  $R_{T_{mean}}$  predictions are slightly better than the  $R_{T_{min}}$  or  $R_{T_{max}}$  predictions and the models perform least well in the mountain, semi-continental, and modified Mediterranean climates. The  $R_{T_{min}}$  model performs slightly worse in late summer; otherwise performance is quite consistent across months and seasons. The models have low bias, with a slope of observed *vs.* predicted of 1.00 and intercept of zero for every year. Performance is consistent across years except for the  $R_{T_{min}}$  model, which performs slightly better in 2000 – 2002, and the  $R_{T_{mean}}$  model, which performs best in 2004 (**Table S6**).



**Table 5.** Stage 4 model performance (predicting daily 200 m residuals with an ensemble): 10-fold cross-validated performance across years (2000 – 2016), overall and by month and season (residual scale).

	R <sub>Tmin</sub>			R <sub>Tmean</sub>			R <sub>Tmax</sub>		
	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
Overall	0.79	0.63	0.39	0.79	0.41	0.26	0.85	0.51	0.31
Jan	0.84	0.56	0.34	0.82	0.40	0.24	0.85	0.48	0.27
Feb	0.82	0.59	0.36	0.81	0.39	0.24	0.84	0.49	0.28
Mar	0.80	0.63	0.39	0.79	0.40	0.26	0.83	0.50	0.30
Apr	0.77	0.63	0.40	0.76	0.39	0.25	0.83	0.51	0.31
May	0.77	0.60	0.37	0.76	0.38	0.24	0.84	0.51	0.31
Jun	0.77	0.62	0.40	0.79	0.39	0.25	0.87	0.52	0.33
Jul	0.76	0.66	0.43	0.77	0.42	0.28	0.86	0.55	0.35
Aug	0.77	0.67	0.44	0.78	0.41	0.28	0.87	0.54	0.34
Sep	0.77	0.69	0.46	0.75	0.42	0.29	0.84	0.54	0.34
Oct	0.78	0.65	0.41	0.76	0.42	0.27	0.82	0.52	0.32
Nov	0.80	0.61	0.37	0.79	0.41	0.25	0.81	0.50	0.29
Dec	0.83	0.60	0.37	0.84	0.43	0.27	0.84	0.52	0.31
Winter	0.83	0.58	0.36	0.83	0.41	0.25	0.84	0.50	0.28
Spring	0.78	0.62	0.39	0.77	0.39	0.25	0.84	0.51	0.31
Summer	0.76	0.65	0.42	0.78	0.41	0.27	0.86	0.54	0.34
Fall	0.78	0.65	0.41	0.77	0.42	0.27	0.82	0.52	0.32

**Table 6.** Stage 4 model performance (predicting daily 200 m residuals with an ensemble): 10-fold cross-validated performance across all years (2000 – 2016), by climatic region and urban vs. rural locations (residual scale).

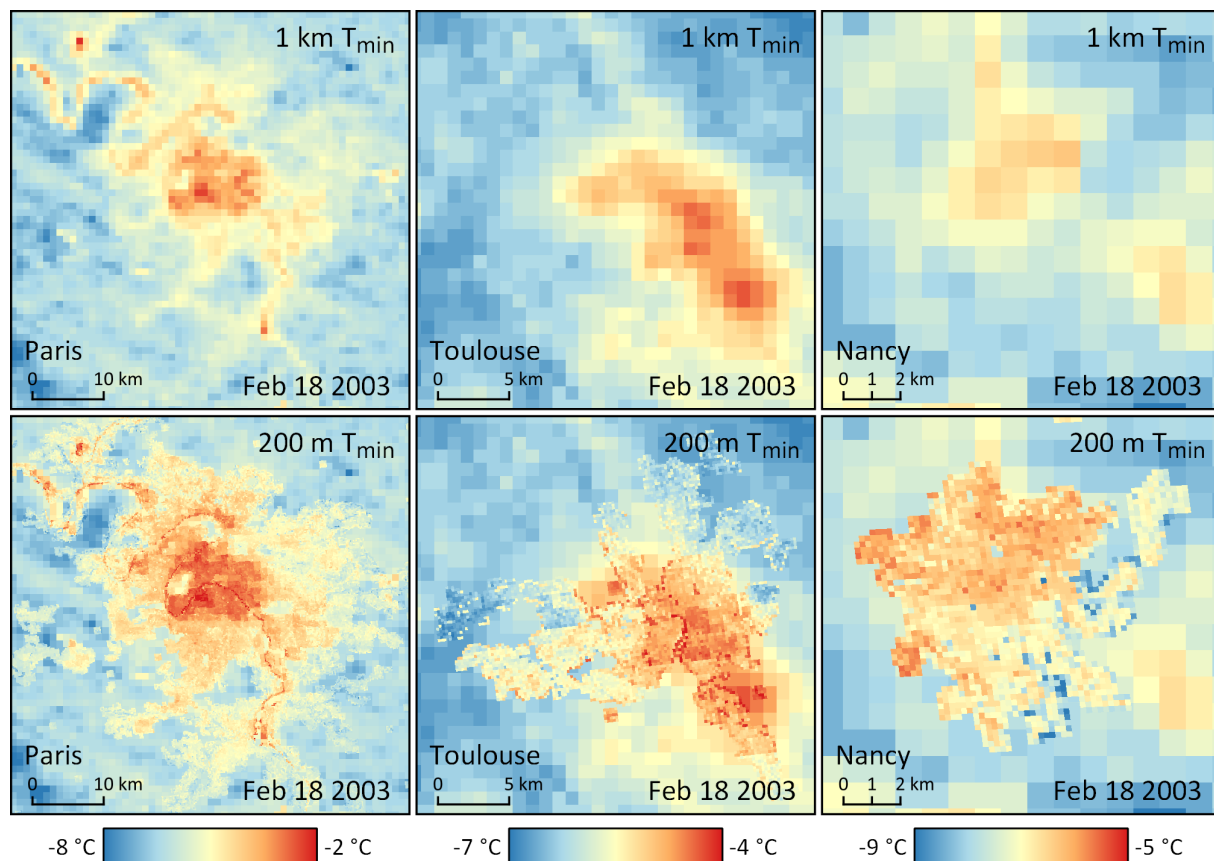
	R <sub>Tmin</sub>			R <sub>Tmean</sub>			R <sub>Tmax</sub>		
	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>	RMSE	MAE
Mountain	0.83	0.67	0.42	0.83	0.46	0.30	0.88	0.58	0.36
Semi-continental	0.81	0.66	0.42	0.79	0.43	0.28	0.86	0.55	0.34
Modified oceanic	0.75	0.54	0.33	0.76	0.33	0.21	0.81	0.40	0.23
Transitional oceanic	0.77	0.62	0.39	0.78	0.39	0.25	0.84	0.50	0.30
Oceanic	0.75	0.62	0.40	0.77	0.39	0.26	0.83	0.50	0.30
Mod. Mediterranean	0.82	0.73	0.47	0.78	0.47	0.31	0.84	0.62	0.41
Southwest basin	0.75	0.59	0.36	0.69	0.38	0.24	0.78	0.48	0.29
Mediterranean	0.77	0.67	0.44	0.73	0.42	0.28	0.80	0.57	0.39
Urban	0.79	0.53	0.32	0.82	0.37	0.23	0.84	0.46	0.27
Peri-urban*	0.76	0.58	0.36	0.78	0.37	0.24	0.83	0.47	0.28
Rural	0.79	0.63	0.40	0.79	0.41	0.26	0.85	0.52	0.32

\* non-urban locations within 5 km of a large urban area

Spatial location and elevation are generally the most important features in the random forest and XGBoost models (**Fig. S2 – S3**). Day of year and predicted 1 km T<sub>a</sub> were equally or even more important in some models but less important in others. Landsat T<sub>b</sub> and NDVI and

population also contributed to the models, particularly for  $R_{T_{\text{mean}}}$ . The land cover and climatic region variables were the least important.

**Fig. 3** shows the spatial pattern of predicted 1 km  $T_{\text{min}}$  from the stage 2 model and predicted 200 m  $T_{\text{min}}$  from the stage 4 model for the Paris metropolitan area (northern France, population 12.5 million), the Toulouse metropolitan area (southwestern France, Population 1.3 million), and the Nancy metropolitan area (northeastern France, population 250,000) on the cold winter day of Feb 18, 2003. In the large city of Paris, an urban heat island is clearly visible centered over the large urban core where  $T_{\text{min}}$  is about 5 °C warmer than the rural surroundings. The 200 m predictions are slightly higher than the 1 km predictions in the peripheral built-up areas and capture fine details such as the warmer Seine river and cooler parks. In the midsize city of Toulouse, the 1 km predictions capture an urban heat island over the dense city center and the suburbs to the northwest and southeast, with  $T_{\text{min}}$  about 3 °C warmer than the rural surroundings. The 200 m predictions show warm  $T_{\text{min}}$  in the southwestern suburbs where 1 km  $T_{\text{min}}$  was cool and capture the Garonne river in the center. The northwestern and northeastern suburbs have greater contrast with some areas slightly cooler than in the 1 km predictions and others slightly warmer. In the small city of Nancy, at 1 km both the city center and an area of ponds to the southeast have  $T_{\text{min}}$  about 2 °C warmer than the surroundings. The 200 m predictions show warmer  $T_{\text{min}}$  throughout most of the built-up area with sharp contrasts between built and open areas: compared to the 1 km predictions,  $T_{\text{min}}$  is up to 2 °C higher in the center, north, and west of the built-up area and up to 2 °C lower over parks and over fields abutting the eastern edges of the city.



**Fig. 3.** Predicted 1 km  $T_{\min}$  from the stage 2 model alone (top row) and with predicted 200 m  $T_{\min}$  from the stage 4 model overlaid (bottom row) on Feb 18, 2003 over the Paris, Toulouse, and Nancy metropolitan areas.

#### 4. Discussion

Spatiotemporally-resolved  $T_a$  at high resolutions is essential to understanding, monitoring, and managing the health effects of  $T_a$ , a pressing issue in a warming, urbanizing world. We have developed the longest (2000 – 2016), highest spatial resolution (1 km) model of daily  $T_a$  available for continental France aimed at public health research. Furthermore, our model provides an unprecedented spatial resolution of 200 m over large urban areas.

A key feature of our model is its ability to capture spatial variation in  $T_a$ . Previous epidemiological research in France linked geographical variation in mortality risk to both typical (Laaïdi et al., 2006) and extreme  $T_a$  (Le Tertre et al., 2006) using weather stations. Recent studies in the USA showed that a daily 1 km  $T_a$  dataset similar to ours was needed to detect associations with low birth weight (Kloog et al., 2015) and mortality (Shi et al., 2015). Our model will allow future studies in France to include participants in rural areas far from weather stations and will also improve exposure estimates in urban areas.

Another key feature is our model's 200 m spatial resolution over urban areas. Estimating  $T_a$  exposure in cities is particularly challenging due to complex built environments and the scarcity of representative  $T_a$  measurements, as weather stations tend to be located outside cities (e.g. at airports) or in large parks. Consequently, few epidemiological studies have examined intra-urban variation in  $T_a$ . In Milan, Italy, de'Donato et al. (2008) found that on hot summer days temperature measured at a nearby airport tended to be higher and more strongly associated with mortality than temperature measured in the city center, but in Turin and Rome there was little difference in temperature or its association with mortality between the city center and a nearby airport. In Paris, France, Laaidi et al. (2012) used 1 km LST as a proxy for  $T_a$  and found an association between minimum LST and mortality during the August 2003 heatwave. In Brisbane, Australia, Guo et al. (2013) found no significant difference in the mortality  $\sim T_a$  relationship when estimating  $T_a$  exposure using a central weather station vs. kriging, but noted that there was little spatial variation in temperature across the city. In Seattle, USA, Ho et al. (2017) found a significant association between spatial variation in mortality on extremely hot days and modeled humidex (a measure of both  $T_a$  and humidity). Our model will help future studies clarify the health effects of intra-urban  $T_a$  variation.

Our model's unique combination of lower spatial resolution (1 km) predictions over a large geographical extent and higher spatial resolution (200 m) predictions over more densely populated areas will be particularly helpful for epidemiological studies. Broad geographical coverage is essential to including rural residents which have often been excluded from epidemiological studies, especially in France where the 103 largest urban areas covered by our 200 m  $T_a$  model contain less than half of the population. At the same time, high spatial resolution is important in dense urban areas where  $T_a$  can vary at fine spatial scales and the effect of spatial  $T_a$  variation is less well understood. Limiting the 200 m resolution predictions to large urban areas reduces computational effort while still covering a large portion of the population.

A fourth feature of our model is its ability to predict daily  $T_{\min}$ ,  $T_{\text{mean}}$ , and  $T_{\max}$ . While  $T_{\text{mean}}$  suffices for many health studies (Barnett et al., 2010), certain research questions may benefit from having  $T_{\min}$  and  $T_{\max}$ . For example, heatwave studies may wish to use heatwave definitions that refer to  $T_{\min}$  or  $T_{\max}$  (Xu et al., 2016) or explore whether certain populations are sensitive to  $T_{\min}$  or nighttime  $T_a$  (Laaidi et al., 2012; Murage et al., 2017).  $T_{\max}$  might also be of interest because it tends to occur in the afternoon when people are more likely to be outside and active

(Guo et al., 2017).  $T_{\min}$  and  $T_{\max}$  also allow calculating diurnal  $T_a$  range for studies of  $T_a$  variability and delineating diurnal and nocturnal urban heat islands for urban climate studies.

We demonstrate that allowing the relationship between 1 km LST and  $T_a$  to vary by climatic region as well as by day slightly improves performance: our stage 1  $T_{\text{mean}}$  model achieves overall  $R^2$  of 0.97 with RMSE of 1.29 whereas an initial version achieved  $R^2$  of 0.96 with RMSE of 1.52 (Kloog et al., 2017). We also demonstrate that a GAM ensemble of machine learning models can use higher spatial resolution predictors including Landsat thermal data to account for some of the residual error in our daily 1 km  $T_a$  predictions. Adding this local stage both increases the spatial resolution of our model and improves performance.

One limitation of our method is its reliance on historical satellite thermal data. Our model is restricted to the MODIS period of record, which starts in 2000. Older thermal data is available from other satellites (e.g. Landsat), but not with a twice-daily revisit time. In the USA, Oyler et al. (2015) showed that an anomaly-climatology approach could model daily  $T_{\min}$  and  $T_{\max}$  since 1948 from 8-day composite MODIS LST, although their approach may have smoothed spatiotemporal  $T_a$  trends.

Our model can estimate past  $T_a$  but, unlike numerical weather prediction models, cannot forecast future  $T_a$ . However, our model is much simpler, which allows us to run it at relatively high spatial resolutions (1 km and 200 m). In comparison, Météo France's weather prediction model has run at a spatial resolution of 1.3 km only since 2015, and the ECMWF's most recent ERA5 reanalysis has a spatial resolution of just 30 km. And recent studies suggest that incorporating LST from geostationary satellites might allow us to estimate close to real-time  $T_a$  (Bechtel et al., 2017; Keramitsoglou et al., 2016), or possibly forecast next-day  $T_a$  from present-day MODIS LST (Yoo et al., 2018).

Another limitation of our approach is the temporal misalignment between observations of LST and  $T_a$  in the stage 1 model: the satellite overpass does not always coincide with the time that  $T_{\min}$  or  $T_{\max}$  occurs. Our model's low MAE (typically less than 1.5 °C) suggests that it produces good  $T_a$  estimates despite this; incorporating high temporal-resolution (e.g. hourly) LST from geostationary satellites might improve performance.

A fourth limitation of our model is the need to fill gaps in satellite thermal data. This can introduce error and may make modelling impossible in some areas or time periods. Landsat data is particularly challenging due to the satellites' 16-day revisit time; parts of France have

no usable Landsat observations during some winters. The few previous studies that used Landsat thermal data to model  $T_a$  limited their analysis to days and locations where Landsat data was available (Pelta and Chudnovsky, 2017) or used a few scenes that were deemed typical of hot summer days (Ho et al., 2016, 2014; Wicki et al., 2018). We fill gaps in Landsat  $T_b$  by compositing all scenes for each calendar month across 17 years. This smooths spatial patterns and means we rely entirely on MODIS to capture short-term temporal variation in LST. Combining data from Landsat 5, 7, and 8 may also introduce error as the sensors operate at different wavelengths and spatial resolutions (**Table 1**). Future studies may benefit from the forthcoming Landsat Surface Temperature product (Malakar et al., 2018) which might be more consistent, and would allow using LST as a predictor rather than brightness temperature.

Future studies could also make use of high spatial-resolution LST from forthcoming satellites. Landsat 9 will have a spatial resolution and revisit time similar to the previous Landsat satellites, but should offer better LST retrieval thanks to the correction of the stray light contamination that affects Landsat 8 (Hair et al., 2018). HypsIRI aims to provide a 60 m spatial resolution with a revisit time of 5 days (Lee et al., 2015), while MISTIGRI aims for 50 m spatial resolution with a daily revisit, but with coverage only within 15 ground tracks (Lagouarde et al., 2013). If these satellites improve LST retrieval and reduce missingness then they could improve our method's ability to capture  $T_a$  over urban areas.

MODIS LST also contains gaps, which we do not fill. Rather, we predict daily 1 km  $T_a$  only where MODIS LST is available and fill gaps in the predictions based on nearby  $T_a$  observations. Li et al. (2018) achieved similar performance (RMSE 2.1 °C  $T_{\min}$ , 1.9 °C  $T_{\max}$ ) for urban and surrounding areas in the USA by first filling gaps in MODIS LST using spatiotemporally nearby LST observations and then predicting daily  $T_a$  using geographically weighted regression. These approaches both assume that the spatial distribution of  $T_a$  or LST is similar on clear and cloudy days. Zhu et al. (2017) used the MODIS atmospheric profile and cloud cover products to estimate instantaneous  $T_a$  in parts of China and the USA. Their approach had the additional advantage of not requiring any weather station  $T_a$  observations to calibrate the model, but it produced larger errors (RMSE 3.4 °C China, 2.9, USA).

Despite these limitations, our model provides very good predictions of historical daily  $T_a$  for continental France at a 1 km or finer spatial resolution. These predictions may help compare rural and urban populations, identify and monitor urban heat islands, and better understand

health effects. More broadly, our methodology and predictions may be useful in other geographical areas and for any application where  $T_a$  is a key variable.

## Declarations of interest

None.

## Acknowledgements

This study was funded by the Climate Health Effects In Pregnant Women And Children-A Multi-Cohort Study In France And Israel grant (CNRS PRC 2018-2020), the Fondation de France (n° 00081169), and the Effects of Urban Microclimate Variability and Global Climate Change on Heat-Related Cardiovascular Outcomes in the Semi-Arid Environment of Southern Israel grant (MOST PRC 2018-2020). Ian Hough is supported by a grant from the French National Research Agency in the framework of the “Investissements d’avenir” program (ANR-15-IDEX-02) and Ben Gurion University of the Negev. Allan C. Just is supported by NIH grants P30ES023515 and R00ES023450. Some data processing and analyses were performed on the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07\_13 CIRA: <http://www.ci-ra.org>). We thank Météo France for providing data from the weather monitoring network.

## References

- Arnfield, A.J., 2003. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* 23, 1–26.  
<https://doi.org/10.1002/joc.859>
- Barnett, A.G., Tong, S., Clements, A.C.A., 2010. What measure of temperature is the best predictor of mortality? *Environ. Res.* 110, 604–611. <https://doi.org/10.1016/j.envres.2010.05.006>
- Bechtel, B., Zakšek, K., Oßenbrügge, J., Kaveckis, G., Böhner, J., 2017. Towards a satellite based monitoring of urban air temperatures. *Sustain. Cities Soc.* 34, 22–31.  
<https://doi.org/10.1016/j.scs.2017.05.018>
- Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., Jones, Z.M., 2016. Mlr: Machine learning in R. *J. Mach. Learn. Res.* 17, 1–5.
- Bischl, B., Richter, J., Bossek, J., Horn, D., Thomas, J., Lang, M., 2017. mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions.  
<https://doi.org/10.13140/RG.2.2.11865.31849>
- Bossard, M., Feranec, J., Otahel, J., 2000. CORINE land cover technical guide - Addendum 2000. Copenhagen.
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: KDD '16 Proceedings

558 of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data  
559 Mining. San Francisco, CA USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>

560 De’Donato, F.K., Stafoggia, M., Rognoni, M., Poncino, S., Caranci, N., Bisanti, L., Demaria, M.,  
561 Forastiere, F., Michelozzi, P., Pelosini, R., Perucci, C.A., 2008. Airport and city-centre  
562 temperatures in the evaluation of the association between heat and mortality. *Int. J. Biometeorol.*  
563 <https://doi.org/10.1007/s00484-007-0124-5>

564 Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S.,  
565 Rocklöv, J., Forsberg, B., Leone, M., De Sario, M., Bell, M.L., Guo, Y.-L.L., Wu, C.-F., Kan,  
566 H., Yi, S.M., De Sousa Zanotti Stagliorio Coelho, M., Saldiva, P.H.N., Honda, Y., Kim, H.,  
567 Armstrong, B.G., 2015. Mortality risk attributable to high and low ambient temperature: A  
568 multicountry observational study. *Lancet* 386, 369–375. [https://doi.org/10.1016/S0140-](https://doi.org/10.1016/S0140-6736(14)62114-0)  
569 [6736\(14\)62114-0](https://doi.org/10.1016/S0140-6736(14)62114-0)

570 Gasparrini, A., Guo, Y., Sera, F., Vicedo-Cabrera, A.M., Huber, V., Tong, S., de Sousa Zanotti  
571 Stagliorio Coelho, M., Nascimento Saldiva, P.H., Lavigne, E., Matus Correa, P., Valdes Ortega,  
572 N., Kan, H., Osorio, S., Kyselý, J., Urban, A., Jaakkola, J.J.K., Rytí, N.R.I., Pascal, M.,  
573 Goodman, P.G., Zeka, A., Michelozzi, P., Scortichini, M., Hashizume, M., Honda, Y., Hurtado-  
574 Diaz, M., Cesar Cruz, J., Seposo, X., Kim, H., Tobias, A., Iñiguez, C., Forsberg, B., Åström,  
575 D.O., Ragettli, M.S., Guo, Y.-L.L., Wu, C.-F., Zanobetti, A., Schwartz, J., Bell, M.L., Dang,  
576 T.N., Van, D. Do, Heaviside, C., Vardoulakis, S., Hajat, S., Haines, A., Armstrong, B.G., 2017.  
577 Projections of temperature-related excess mortality under climate change scenarios. *Lancet*  
578 *Planet. Heal.* 360–367. [https://doi.org/10.1016/S2542-5196\(17\)30156-0](https://doi.org/10.1016/S2542-5196(17)30156-0)

579 Goggins, W.B., Chan, E.Y.Y., Ng, E., Ren, C., Chen, L., 2012. Effect modification of the association  
580 between short-term meteorological factors and mortality by urban heat islands in Hong Kong.  
581 *PLoS One* 7, 9–14. <https://doi.org/10.1371/journal.pone.0038551>

582 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth  
583 Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27.  
584 <https://doi.org/10.1016/j.rse.2017.06.031>

585 Guo, Y., Barnett, A.G., Tong, S., 2013. Spatiotemporal model or time series model for assessing city-  
586 wide temperature effects on mortality? *Environ. Res.* 120, 55–62.  
587 <https://doi.org/10.1016/j.envres.2012.09.001>

588 Guo, Y., Gasparrini, A., Armstrong, B.G., Li, S., Tawatsupa, B., Tobias, A., Lavigne, E., De Sousa  
589 Zanotti Stagliorio Coelho, M., Leone, M., Pan, X., Tong, S., Tian, L., Kim, H., Hashizume, M.,  
590 Honda, Y., Guo, Y.-L.L., Wu, C.-F., Punnasiri, K., Yi, S.M., Michelozzi, P., Saldiva, P.H.N.,  
591 Williams, G.M., 2014. Global variation in the effects of ambient temperature on mortality: A  
592 systematic evaluation. *Epidemiology* 25, 781–789.  
593 <https://doi.org/10.1097/EDE.0000000000000165>

594 Guo, Y., Gasparrini, A., Armstrong, B.G., Tawatsupa, B., Tobias, A., Lavigne, E., De Sousa Zanotti  
595 Stagliorio Coelho, M., Pan, X., Kim, H., Hashizume, M., Honda, Y., Leon Guo, Y.L., Wu, C.-F.,  
596 Zanobetti, A., Schwartz, J.D., Bell, M.L., Scortichini, M., Michelozzi, P., Punnasiri, K., Li, S.,  
597 Tian, L., Garcia, S.D.O., Seposo, X., Overcenco, A., Zeka, A., Goodman, P.G., Dang, T.N., Van  
598 Dung, D., Mayvaneh, F., Saldiva, P.H.N., Williams, G., Tong, S., 2017. Heat wave and  
599 mortality: A multicountry, multicomunity study. *Environ. Health Perspect.* 125, 1–11.  
600 <https://doi.org/10.1289/EHP1026>

601 Guo, Y., Gasparrini, A., Armstrong, B.G., Tawatsupa, B., Tobias, A., Lavigne, E., Zanotti Stagliorio



Coelho, M. de S., Pan, X., Kim, H., Hashizume, M., Honda, Y., Guo, Y.-L.L., Wu, C.-F.,  
 Zanobetti, A., Schwartz, J., Bell, M.L., Overcenco, A., Punnasiri, K., Li, S., Tian, L., Saldiva, P.,  
 Williams, G., Tong, S., 2016. Temperature variability and mortality: A multi-country study.  
 Environ. Health Perspect. 124, 1554–1559. <https://doi.org/10.1289/EHP149>

Ha, S., Liu, D., Zhu, Y., Sherman, S., Mendola, P., 2017a. Acute Associations between Outdoor  
 Temperature and Premature Rupture of Membranes. Epidemiology 29, 1.  
<https://doi.org/10.1097/EDE.0000000000000779>

Ha, S., Zhu, Y., Liu, D., Sherman, S., Mendola, P., 2017b. Ambient temperature and air quality in  
 relation to small for gestational age and term low birthweight. Environ. Res. 155, 394–400.  
<https://doi.org/10.1016/j.envres.2017.02.021>

Hair, J.H., Reuter, D.C., Tonn, S.L., McCorkel, J., Amy, A.S., Djam, M., Alexander, D., Ballou, K.,  
 Barclay, R., Coulter, P., Edick, M., Efremova, B., Finneran, P., Florez, J., Graham, S., Harbert,  
 K., Hewitt, D., Hickey, M., Hicks, S., Hoge, W., Jhabvala, M., Lilly, C., Lunsford, A., Mann, L.,  
 Masters, C., Montanaro, M., Muench, T., Otero, V., Parong, F., Pearlman, A., Penn, J., Vigneau,  
 D., Wenny, B., 2018. Landsat 9 thermal infrared sensor 2 architecture and design. Int. Geosci.  
 Remote Sens. Symp. 2018-July, 8841–8844. <https://doi.org/10.1109/IGARSS.2018.8518269>

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., Walker, B.B., Henderson, S.B., 2017. Delineation of spatial variability in the  
 temperature-mortality relationship on extremely hot days in greater Vancouver, Canada. Environ.  
 Health Perspect. 125, 66–75. <https://doi.org/10.1289/EHP224>

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.  
<https://doi.org/10.1016/j.scitotenv.2015.12.021>

INSEE, 2018. Estimation de la population au 1er janvier 2018 [WWW Document]. URL  
<https://www.insee.fr/fr/statistiques/1893198>

INSEE, 2017. Estimation de la population au 1er janvier 2016 [WWW Document]. URL  
<https://www.insee.fr/fr/statistiques/1893198>

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to  
 the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge  
 University Press, Cambridge, United Kingdom. <https://doi.org/10.1017/CBO9781107415324>

Joly, D., Brossard, T., Cardot, H., Cavailles, J., Hilal, M., Wavresky, P., 2010. Les types de climats en  
 France, une construction spatiale. Cybergeog 501. <https://doi.org/10.4000/cybergeog.23155>

Keramitsoglou, I., Kiranoudis, C.T., Sismanidis, P., Zakšek, K., 2016. An online system for  
 nowcasting satellite derived temperatures for urban areas. Remote Sens. 8, 1–17.  
<https://doi.org/10.3390/rs8040306>

Kloog, I., Melly, S.J., Coull, B.A., Nordio, F., Schwartz, J., 2015. Using Satellite-Based  
 Spatiotemporal Resolved Air Temperature Exposure to Study the Association between Ambient  
 Air Temperature and Birth Outcomes in Massachusetts. Environ. Health Perspect. 123, 1053–  
 1058. <https://doi.org/10.1289/ehp.1308075>

644 Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2014. Predicting spatiotemporal mean air temperature  
 645 using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote*  
 646 *Sens. Environ.* 150, 132–139. <https://doi.org/10.1016/j.rse.2014.04.024>

647 Kloog, I., Nordio, F., Lepeule, J., Padoan, A., Lee, M., Auffray, A., Schwartz, J., 2017. Modelling  
 648 spatio-temporally resolved air temperature across the complex geo-climate area of France using  
 649 satellite-derived land surface temperature data. *Int. J. Climatol.* 37, 296–304.  
 650 <https://doi.org/10.1002/joc.4705>

651 Laaidi, K., Zeghnoun, A., Dousset, B., Bretin, P., Vandentorren, S., Giraudet, E., Beaudeau, P., 2012.  
 652 The impact of heat islands on mortality in Paris during the August 2003 heat wave. *Environ.*  
 653 *Health Perspect.* 120, 254–259. <https://doi.org/10.1289/ehp.1103532>

654 Laaidi, M., Laaidi, K., Besancenot, J.P., 2006. Temperature-related mortality in France, a comparison  
 655 between regions with different climates from the perspective of global warming. *Int. J.*  
 656 *Biometeorol.* 51, 145–153. <https://doi.org/10.1007/s00484-006-0045-8>

657 Lagouarde, J.P., Bach, M., Sobrino, J.A., Boulet, G., Briottet, X., Cherchali, S., Coudert, B., Dadou, I.,  
 658 Dedieu, G., Gamet, P., Hagolle, O., Jacob, F., Nerry, F., Oliso, A., Ottlé, C., Roujean, J. louis,  
 659 Fargant, G., 2013. The MISTIGRI thermal infrared project: Scientific objectives and mission  
 660 specifications. *Int. J. Remote Sens.* 34, 3437–3466.  
 661 <https://doi.org/10.1080/01431161.2012.716921>

662 Le Tertre, A., Lefranc, A., Eilstein, D., Declercq, C., Medina, S., Blanchard, M., Chardon, B., Fabre,  
 663 P., Filleul, L., Jusot, J.F., Pascal, L., Prouvost, H., Cassadou, S., Ledrans, M., 2006. Impact of  
 664 the 2003 heatwave on all-cause mortality in 9 French cities. *Epidemiology* 17, 75–79.  
 665 <https://doi.org/10.1097/01.ede.0000187650.36636.1f>

666 Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D.J., Middleton, E.M., 2015. An  
 667 introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory  
 668 activities. *Remote Sens. Environ.* 167, 6–19. <https://doi.org/10.1016/j.rse.2015.06.012>

669 Lee, M., Shi, L., Zanobetti, A., Schwartz, J., 2016. Study on the association between ambient  
 670 temperature and mortality using spatially resolved exposure data. *Environ. Res.* 151, 610–617.  
 671 <https://doi.org/10.1016/j.envres.2016.08.029>

672 Li, J., Woodward, A., Hou, X.Y., Zhu, T., Zhang, J., Brown, H., Yang, J., Qin, R., Gao, J., Gu, S., Li,  
 673 J., Xu, L., Liu, X., Liu, Q., 2017. Modification of the effects of air pollutants on mortality by  
 674 temperature: A systematic review and meta-analysis. *Sci. Total Environ.* 575, 1556–1570.  
 675 <https://doi.org/10.1016/j.scitotenv.2016.10.070>

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

679 Li, Z.-L., Tang, B.H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013. Satellite-  
 680 derived land surface temperature: Current status and perspectives. *Remote Sens. Environ.* 131,  
 681 14–37. <https://doi.org/10.1016/j.rse.2012.12.008>

682 Malakar, N.K., Hulley, G.C., Hook, S.J., Laraby, K.G., Cook, M., Schott, J.R., 2018. An Operational  
 683 Land Surface Temperature Product for Landsat Thermal Data: Methodology and Validation.  
 684 *IEEE Trans. Geosci. Remote Sens.* 1–19. <https://doi.org/10.1109/TGRS.2018.2824828>

685 Milojevic, A., Armstrong, B.G., Gasparrini, A., Bohnenstengel, S.I., Barratt, B., Wilkinson, P., 2016.  
 686 Methods to estimate acclimatization to urban heat island effects on heat-and cold-related

687 mortality. *Environ. Health Perspect.* 124, 1016–1022. <https://doi.org/10.1289/ehp.1510109>

688 Molina, O., Saldarriaga, V., 2017. The perils of climate change: In utero exposure to temperature  
689 variability and birth outcomes in the Andean region. *Econ. Hum. Biol.* 24, 111–124.  
690 <https://doi.org/10.1016/j.ehb.2016.11.009>

691 Murage, P., Hajat, S., Kovats, R.S., 2017. Effect of night-time temperatures on cause and age-specific  
692 mortality in London. *Environ. Epidemiol.* 1. <https://doi.org/10.1097/ee9.0000000000000005>

693 Oyler, J.W., Ballantyne, A., Jencso, K., Sweet, M., Running, S.W., 2015. Creating a topoclimatic daily  
694 air temperature dataset for the conterminous United States using homogenized station data and  
695 remotely sensed land skin temperature. *Int. J. Climatol.* 35, 2258–2279.  
696 <https://doi.org/10.1002/joc.4127>

697 Oyler, J.W., Dobrowski, S.Z., Holden, Z.A., Running, S.W., 2016. Remotely sensed land skin  
698 temperature as a spatial predictor of air temperature across the conterminous United States. *J.*  
699 *Appl. Meteorol. Climatol.* 55, 1441–1457. <https://doi.org/10.1175/JAMC-D-15-0276.1>

700 Pelta, R., Chudnovsky, A.A., 2017. Spatiotemporal estimation of air temperature patterns at the street  
701 level using high resolution satellite imagery. *Sci. Total Environ.* 579, 675–684.  
702 <https://doi.org/10.1016/j.scitotenv.2016.11.042>

703 R Core Team, 2018. R: A language and environment for statistical computing.

704 Robinson, N.P., Allred, B.W., Jones, M.O., Moreno, A., Kimball, J.S., Naugle, D.E., Erickson, T.A.,  
705 Richardson, A.D., 2017. A dynamic landsat derived normalized difference vegetation index  
706 (NDVI) product for the conterminous United States. *Remote Sens.* 9, 1–14.  
707 <https://doi.org/10.3390/rs9080863>

708 Rosenfeld, A., Dorman, M., Schwartz, J., Novack, V., Just, A.C., Kloog, I., 2017. Estimating daily  
709 minimum, maximum, and mean near surface air temperature using hybrid satellite models across  
710 Israel. *Environ. Res.* 159, 297–312. <https://doi.org/10.1016/j.envres.2017.08.017>

711 Shi, L., Kloog, I., Zanobetti, A., Liu, P., Schwartz, J., 2015. Impacts of temperature and its variability  
712 on mortality in New England. *Nat. Clim. Chang.* 5, 988–991.  
713 <https://doi.org/10.1038/nclimate2704>

714 Shi, L., Liu, P., Kloog, I., Lee, M., Kosheleva, A., Schwartz, J., 2016a. Estimating daily air  
715 temperature across the Southeastern United States using high-resolution satellite data: A  
716 statistical modeling study. *Environ. Res.* 146, 51–58.  
717 <https://doi.org/10.1016/j.envres.2015.12.006>

718 Shi, L., Liu, P., Wang, Y., Zanobetti, A., Kosheleva, A., Koutrakis, P., Schwartz, J., 2016b. Chronic  
719 effects of temperature on mortality in the Southeastern USA using satellite-based exposure  
720 metrics. *Sci. Rep.* 6, 1–8. <https://doi.org/10.1038/srep30161>

721 Smargiassi, A., Goldberg, M.S., Plante, C., Fournier, M., Baudouin, Y., Kosatsky, T., 2009. Variation  
722 of daily warm season mortality as a function of micro-urban heat islands. *J. Epidemiol.*  
723 *Community Health* 63, 659–664. <https://doi.org/10.1136/jech.2008.078147>

724 Song, X., Wang, S., Hu, Y., Yue, M., Zhang, T., Liu, Y., Tian, J., Shang, K., 2017. Impact of ambient  
725 temperature on morbidity and mortality: An overview of reviews. *Sci. Total Environ.* 586, 241–  
726 254. <https://doi.org/10.1016/j.scitotenv.2017.01.212>

727 Tøttrup, C., 2014. EU-DEM Statistical Validation Report.

728 United Nations, D. of E. and S.A.-P.D., 2018. World Urbanization Prospects: The 2018 Revision.

729 USGS, 2018a. Landsat 4-7 Surface Reflectance (LEDAPS) Product Guide.

730 USGS, 2018b. Landsat 8 Surface Reflectance Code (LASRC) Product Guide.

731 Wan, Z., 2014. New refinements and validation of the collection-6 MODIS land-surface  
 732 temperature/emissivity product. *Remote Sens. Environ.* 140, 36–45.  
 733 <https://doi.org/10.1016/j.rse.2013.08.027>

734 Wang, Y., Nordio, F., Nairn, J., Zanobetti, A., Schwartz, J., 2018. Accounting for adaptation and  
 735 intensity in projecting heat wave-related mortality. *Environ. Res.* 161, 464–471.  
 736 <https://doi.org/10.1016/j.envres.2017.11.049>

737 Wicki, A., Parlow, E., Feigenwinter, C., 2018. Evaluation and Modeling of Urban Heat Island  
 738 Intensity in Basel, Switzerland. *Climate* 6, 1–25. <https://doi.org/10.3390/cli6030055>

739 Wright, M.N., Ziegler, A., 2017. Ranger: A fast implementation of random forests for high  
 740 dimensional data in C++ and R. *J. Stat. Softw.* 77. <https://doi.org/10.18637/jss.v077.i01>

741 Xu, Z., FitzGerald, G., Guo, Y., Jalaludin, B., Tong, S., 2016. Impact of heatwave on mortality under  
 742 different heatwave definitions: A systematic review and meta-analysis. *Environ. Int.* 89–90, 193–  
 743 203. <https://doi.org/10.1016/j.envint.2016.02.007>

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

747 Zeger, S.L., Thomas, D., Dominici, F., Samet, J.M., Schwartz, J., Dockery, D., Cohen, A., 2000.  
 748 Exposure measurement error in time-series studies of air pollution: Concepts and consequences.  
 749 *Environ. Health Perspect.* 108, 419–426. <https://doi.org/10.1289/ehp.00108419>

750 Zhu, W., Lü, A., Jia, S., Yan, J., Mahmood, R., 2017. Retrievals of all-weather daytime air  
 751 temperature from MODIS products. *Remote Sens. Environ.* 189, 152–163.  
 752 <https://doi.org/10.1016/j.rse.2016.11.011>

753