RESEARCH ARTICLE

Estimating near-surface air temperature across Israel using a machine learning based hybrid approach

Bin Zhou[1,2] | **Evyatar Erell[1,3]** | **Ian Hough[3,4]** | **Jonathan Rosenblatt[5]** | **Allan C. Just[6]** | **Victor Novack[7]** | **Itai Kloog[3]**

[1]Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Beer Sheva, Israel

[2]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany

[3]Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beer Sheva, Israel

[4]Univ. Grenoble Alpes, Inserm, CNRS, IAB, Grenoble, France

[5]Department of Industrial Engineering and Management, Ben Gurion University of the Negev, Beer Sheva, Israel

[6]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY

[7]Clinical Research Center, Soroka University Medical Center, Beer Sheva, Israel

Correspondence
Bin Zhou, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, 8499000 Midreshet Ben-Gurion, Israel.
Email: zhoub@post.bgu.ac.il

Abstract

Rising global temperatures over the last decades have increased heat exposure among populations worldwide. An accurate estimate of the resulting impacts on human health demands temporally explicit and spatially resolved monitoring of near-surface air temperature (T_a). Neither ground-based nor satellite-borne observations can achieve this individually, but the combination of the two provides synergistic opportunities. In this study, we propose a two-stage machine learning-based hybrid model to estimate 1×1 km^2 gridded intra-daily T_a from surface skin temperature (T_s) across the complex terrain of Israel during 2004–2016. We first applied a random forest (RF) regression model to impute missing T_s from the Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra satellites, integrating T_s from the geostationary Spinning Enhanced Visible and InfraRed Imager (SEVIRI) satellite and synoptic variables from European Centre for Medium-Range Weather Forecasts' (ECMWF) ERA5 reanalysis data sets. The imputed T_s are in turn fed into the Stage 2 RF-based model to estimate T_a at the satellite overpass hours of each day. We evaluated the model's performance applying out-of-sample fivefold cross validation. Both stages of the hybrid model perform very well with out-of-sample fivefold cross validated R^2 of 0.99 and 0.96, MAE of 0.42°C and 1.12°C, and RMSE of 0.65°C and 1.58°C (Stage 1: imputation of T_s, and Stage 2: estimation of T_a from T_s, respectively). The newly proposed model provides excellent computationally efficient estimation of near-surface air temperature at high resolution in both space and time, which helps further minimize exposure misclassification in epidemiological studies.

KEYWORDS
air temperature, health < 6. application/context, health exposure, MODIS, random forest, remote sensing < 1. tools and methods, statistical methods < 1. tools and methods, surface skin temperature

1 | INTRODUCTION

Numerous studies worldwide have demonstrated a robust relationship between outdoor temperature and mortality. Although temperature effects may vary geographically due to acclimatization, minimum mortality typically occurs at about the 76th percentile of the annual temperature (Guo *et al.*, 2014), increasing with both lower (Analitis *et al.*, 2008) and higher temperature. Exposure to extreme ambient temperatures during heat waves is increasingly associated with adverse health effects (Ye *et al.*, 2012; Gasparrini *et al.*, 2015; Shi *et al.*, 2015) such as adverse birth outcomes (Basu *et al.*, 2010; Kloog *et al.*, 2015); increasing cardiac output and peripheral blood circulation, which may promote dehydration and renal failure (Schwartz *et al.*, 2004); and increases in blood viscosity and cholesterol (Keatinge, 1997). The 2003 European heat wave, for example, was estimated to result in as many as 40,000 excess deaths that are attributed to heatstroke, hyperthermia, dehydration, cardiovascular, respiratory, and neurologic diseases (Fouillet *et al.*, 2006). Worse still, heat waves often coincide with extreme air pollution events at large scale (Schnell and Prather, 2017; Zhang *et al.*, 2017). These synergistic impact modifiers may further impair human health at a level greater than the sum of their individual parts (Stafoggia *et al.*, 2008; Chen *et al.*, 2018).

Most projections of global climate change predict more frequent and more intense extreme heat waves (Coumou and Rahmstorf, 2012; IPCC, 2014). Although it is hard to predict precise outcomes in terms of direct heat-related mortality (Guo *et al.*, 2018), it is clearly important to be able to model heat exposure and its health impacts, especially with regard to vulnerable populations, such as the elderly, the chronically ill, and people living alone (Rosenthal *et al.*, 2014). This requires accurate ambient near-surface air temperature (T_a) data that are temporally explicit and spatially resolved.

Regarding the latter challenge, previous epidemiological studies typically employed air temperature measurement at weather stations that are more densely installed in or close to urban areas rather than rural areas (Basu, 2009). This takes limited account of large variability of air temperature in space and time, not least because of the urban heat island effect, where the temperature in cities is higher relative to their rural surroundings (Arnfield, 2003; Oke *et al.*, 2017). Assigning a unique T_a for regions of varying geographical and socio-economic features could introduce exposure error, leading to a negative bias in results (Zeger *et al.*, 2000) and thus compromising the representativeness of those studies.

To minimize error in assessing human exposure to extreme ambient temperature, an increasing number of studies have started to employ *scattered* ground-based T_a and *spatially resolved* satellite-borne surface skin temperature (T_s) measurements to attain high-resolution continuous spatio-temporal T_a (Vancutsem *et al.*, 2010; Benali *et al.*, 2012; Kloog *et al.*, 2012, 2014; Zhu *et al.*, 2013, 2017; Kilibarda *et al.*, 2014; Weiss *et al.*, 2014; Oyler *et al.*, 2015). For a more detailed review on the methodologies used we refer to Zakšek and Schroedter-Homscheidt (2009); Lin *et al.* (2012); and Bechtel *et al.* (2014).

T_s and T_a differ in terms of measurement technique and medium sensed. T_s is the surface radiometric temperature confined to the instantaneous field-of-view of the sensor (Prata *et al.*, 1995), while T_a refers to the near-surface thermo- or aerodynamic air temperature measured by a thermometer (Li *et al.*, 2013), usually at a height of 1.5–2.0 m above the ground. The discrepancy between T_s and T_a is caused by the location-specific features of the surface energy balance, in particular the parcelling of the incoming radiant flux into sensible heat (which creates a difference in air temperature), and latent flux (which leads to a difference in the moisture content of the air). Changes in the near-surface air temperature (T_a) typically lag the development of the surface temperature (T_s) by several hours. The two temperature measures are further found to converge at vegetated places with high soil moisture, or on cloudy days (Prigent *et al.*, 2003). Several previous studies relied on calibrated T_a–T_s relationships to map daily minimum, mean, and maximum T_a, rather than T_a at hourly intervals, due to the temporal limitations of satellite data. The techniques applied include linear mixed effects models (Kloog *et al.*, 2017; Rosenfeld *et al.*, 2017), spatio-temporal regression-kriging (Kilibarda *et al.*, 2014), and advanced statistical regression (Benali *et al.*, 2012; Janatian *et al.*, 2017). Despite overall good performance [with root mean square error (RMSE) normally <2°C, and coefficient of determination (R^2) > 0.90], the inability of those models to capture the diurnal variation of exposure may to some extent restrict their application. At the other extreme, although geostationary satellites can provide temporally explicit T_s at 15 min intervals, their coarse resolution of normally > 3 km cannot capture the intra-urban variation of T_a, making them inadequate for health exposure analysis in cities (Freitas *et al.*, 2013).

Physically-based numerical models can also simulate air temperature from micro to regional scales, ranging from several meters to tens of kilometres (Mirzaei and Haghighat, 2010; Georgescu *et al.*, 2015). To attain a spatial resolution capable of resolving heterogeneous urban features, regional climate models (RCMs) are coupled with urban parameterization schemes for downscaling T_a (Chen *et al.*, 2011; Wang *et al.*, 2012; Hamdi *et al.*, 2014;

Grossman-Clarke *et al.*, 2016). Numerical models can predict T_a under different future climate and urban development scenarios. However, the urbanized meso-scale climate models are computation-intensive and require great expertise to implement. Thus, they have not yet been widely employed in epidemiological studies. In comparison, urban climate models coupled with simplified energy balance schemes constitute a computationally efficient and accurate alternative (Masson *et al.*, 2002; Erell and Williamson, 2006; Bueno *et al.*, 2013; De Ridder *et al.*, 2015). These models typically require a detailed description of urban morphology and apply forcing at the boundary level or from rural reference stations under the same meso-scale weather (Kaplan *et al.*, 2016; Zhou *et al.*, 2019). This may restrict their application to regions with high-quality geospatial databases and well-maintained weather stations.

To overcome limitations existing in previous studies (i.e., lack of a computation-efficient model capable of estimating diurnally explicit T_a), we propose a two-stage machine learning based hybrid approach to estimating high-resolution (1×1 km^2) intra-daily (in line with satellite overpasses) T_a from T_s across Israel for the period 2004–2016. The model incorporates remotely sensed T_s from polar-orbiting and geostationary satellites, synoptic meteorological variables from reanalysis data, and biophysical and socio-economic features accounting for the variation of T_a.

The machine learning technique underlying the proposed model is the random forest regression—an ensemble learning technique consisting of a large number of decision trees. Each tree is constructed using a subset of data that is independently sampled and identically distributed (Breiman, 2001; Liaw and Wiener, 2002). Its randomness stems from the number of trees grown and the number of features used in splitting at each node, making it robust against both outliers and overfitting (Breiman, 2001). Due to its high estimative performance and efficiency in implementation (Wright and Ziegler, 2017; Hengl *et al.*, 2018), it has already been applied in estimating particulate matter (PM) exposure (Brokamp *et al.*, 2017; Stafoggia *et al.*, 2019), but is rare in studying heat exposure (Li and Zha, 2018).

The objective of this research is to propose a computationally efficient model capable of seamless estimation of T_s and T_a at high resolution in both space and time. Such models are of great importance in research of heat-related impacts on health. Specifically, this study contributes to research aiming at minimizing the misclassification of heat exposure in Israel, where the intensity, length, and number of heat waves have increased by a factor of six to eight since the 1960s (Kuglitsch *et al.*, 2010), and are projected to increase still further (Hochman *et al.*, 2018a; Hochman *et al.*, 2018b).

2 | DATA AND METHODS

2.1 | Study area and meteorological data

The study area includes the entire land territory of the State of Israel, with a total area of about 21,670 km^2. Located at the eastern coast of the Mediterranean Sea, Israel is about 470 km from north to south and 135 km wide at the widest point. Despite its small size and elongated shape, Israel's terrain is characterized by extreme variations in elevation, with the lowest point in the Jordan Rift Valley (~430 m below sea level) and the highest point on Mount Hermon (+2,807 m above sea level), as shown in Figure 1.

As per altitude, latitude, and the dominant climate condition, Israel is divided into four geographical regions: the Mediterranean coastal plain, the Central Hills, the Jordan Rift Valley and the Negev Desert (Ochsenwald *et al.*, 2019).

Precipitation is unevenly distributed over time and across regions. The rainy cool winter lasts from November to March, while rainfall is fairly rare in the
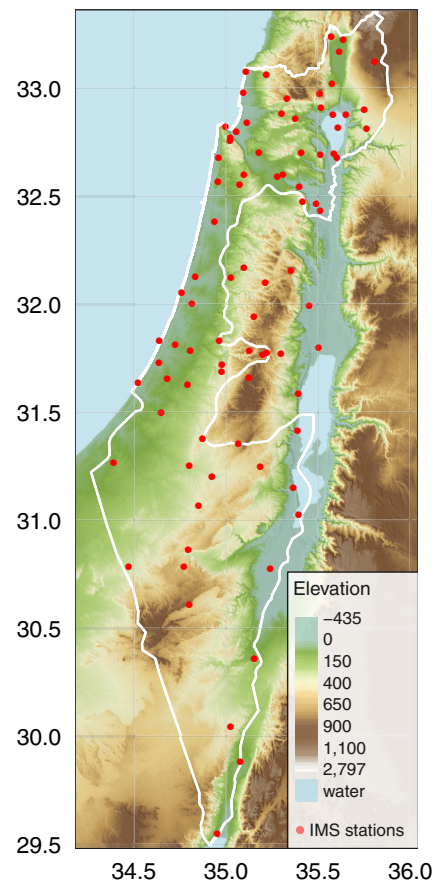
FIGURE 1 Digital elevation map (DEM) of Israel and the location of 85 IMS weather stations [Colour figure can be viewed at wileyonlinelibrary.com]

rest of the year. Annual precipitation ranges from 1,000 mm in the north Golan Heights to nearly 30 mm in the extreme south. The heterogeneity in precipitation has in turn resulted in a diversity of biomes (land cover types) across Israel.

Israel has a total population of about 8.9 million inhabitants in 2018 (Central Bureau of Statistics, 2018) exhibiting a pronounced heterogeneous spatial distribution: The urban agglomerations on the coastal plain accommodate more than half of state's population, whereas the Negev desert in the south is sparsely inhabited.

Figure 1 shows the location of 85 weather stations of the Israel Meteorological Service (IMS) employed in this study. We acquired hourly T_a from those stations covering the period 2004 to 2016.

2.2 | Remotely sensed surface skin temperature

The daily Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) products (version 6) are collected at about $1 \times 1 \text{ km}^2$ resolution from the MODIS sensors on board the Aqua (MYD11A1) and Terra (MOD11A1) satellites. The data used here cover the years 2004 to 2016 for the entire area of Israel corresponding to MODIS tiles h20v05, h20v06, h21v05, and h21v06. The Aqua and Terra satellites overpass the same location on Earth's surface twice per day, but with slightly varying times depending on their (polar) orbits. We retrieved the overpass time (in local solar time) for each MODIS LST value. When the overpass time for a pixel is not produced due to, for example, cloud contamination, the average overpass time in the scene (distinguished between day- and nighttime) is assigned.

Accounting for the slight variability in overpass times, the mean local solar time of four satellite overpasses during the period investigated is about 1,100 (Terra-Day), 1,300 (Aqua-Day), 2,200 (Terra-Night), and 0200 (Aqua-Night). We converted the local solar time to local time (LT) in Israeli Standard Time (IST, UTC + 2) using Equation (1) (Wan, 2013):

$$\text{LT (in IST)} = \text{Local Solar Time} - \text{longitude}/15 + 2 \quad (1)$$

We then rounded the local time to the nearest hour.

In contrast with polar-orbiting satellites (such as the aforementioned Aqua and Terra satellites), geostationary satellites such as the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) onboard the Meteosat Second Generation (MSG) satellites can provide continuous observation of T_s with high temporal resolution—but lower spatial detail. SEVIRI provides infra-red images every 15 minutes with a nadir spatial resolution of 3 km, which is geometrically degraded at large off-nadir view angles. In this study, we downloaded the SEVIRI LST data from LSA SAF (https://landsaf.ipma.pt/en/products/land-surface-temperature/lst, accessed on April 21, 2019) aggregated to hourly means. For Israel, the spatial resolution of the grid cell is about $4 \times 4 \text{ km}^2$.

2.3 | ERA5 reanalysis data

The ERA5 data are the fifth generation of the European Centre for Medium-Range Weather Forecasts' (ECMWF) atmospheric reanalysis of the global climate (Copernicus Climate Change Service [C3S], 2017). ERA5 is produced using data assimilation techniques based on ECMWF's latest Integrated Forecasting System (IFS). The IFS combines model data with all available historical in-situ and spaceborne observations with stringent quality control. ERA5 provides hourly estimates of atmospheric and surface parameters from 1979 onward at 0.28125° (31 km) worldwide, but can be bilinearly interpolated to any custom grid. Given the aim of this study, we integrated the following ERA5 parameters (summarized in Table 1) into the model: (a) skin temperature, (b) 2 m temperature (T_a at 2 m height), (c) boundary layer height, (d) 10 m wind speed [horizontal eastward (U) + northward (V) components], (e) soil temperature layer 1 (0–7 cm), and (f) total cloud cover. These parameters may be used to describe the surface energy balance, and thus are expected to account for the complex land surface-atmosphere interactions (Jin and Dickinson, 2010). The wind speed used in the model is the root of the sum of the squared U and V wind components. We downloaded the ERA5 data from 2004 to 2016 at 0.125° (~10 km) for Israel via the Copernicus Climate Change Service (C3S) Climate Data Store (https://cds.climate.copernicus.eu, accessed on April 21, 2019).

2.4 | Geospatial variables

Given the data availability and relevance, we incorporated the following geographical and socio-economical predictors into the model to account for the spatiotemporal variability of air temperature.

2.4.1 | NDVI

We used MODIS-derived monthly Normalized Difference Vegetation Index (NDVI) products (version 6) at $1 \times 1 \text{ km}^2$ resolution from Aqua (MYD13A3) and Terra (MOD13A3)

TABLE 1 ERA5 parameters used in this study

Name	Units	Short name in ERA5	Spatial resolution	Temporal resolution
Skin temperature	K	skt	0.125°(~10 km)	Hourly, 2004–2016
2 m air temperature	K	2t		
Boundary layer height	m	blh		
10 m U wind component (zonal velocity, horizontal eastward wind component)	m s^{-1}	10u		
10 m V wind component (meridional velocity, horizontal northward wind component)	m s^{-1}	10v		
Soil temperature level 1	K	stl1		
Total cloud cover	(0–1)	tcc		

satellites. The NDVI value for each grid cell is the mean of the two datasets. NDVI is computed as the ratio of the difference in spectral reflectance between near infrared and red wavebands to the sum of the reflectance of the same two bands (Tucker, 1979). It quantifies the relative abundance and health status of vegetation varying in time and space. NDVI values range from −1 to 1: a higher value indicates dense photosynthetically active vegetation, whereas senesced and dead plants, inorganic materials (e.g., rocks), and water courses normally have a low or even negative NDVI. For every timestep, each grid cell was assigned the mean NDVI of the month when the observation was taken.

2.4.2 | Road and population density

To calculate population density, we used the population estimates for the year of 2012 collected from 3,067 census units (as polygons) across Israel by the Israeli Central Bureau of Statistics (Central Bureau of Statistics, 2014). The population estimates are homogenously dis-aggregated into a grid of 10×10 m^2 raster cells and then re-allocated to the MODIS grid of 1×1 km^2 cells. Similarly, road density, defined as the total length of roads within a 1×1 km^2 cell, is generated based on the road vector data from the Israel survey bureau mapping service (Survey of Israel, 2013), using the line density tool in the ArcGIS 10.6 (Esri, 2018).

2.4.3 | Distance to large bodies of water

Water bodies have a higher specific heat capacity than land, and are expected to modify mesoscale climate conditions, and thus have to be taken into proper account in the model. We calculated the distance of each grid cell to the nearest water body by applying the distance function in R sf package (Pebesma, 2018) on the 1:10,000,000 coastline data version 4.1.0 from the Natural Earth portal (https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-coastline/, accessed on 01/04/2019).

2.4.4 | Elevation and slope aspect

The elevation assigned to each MODIS grid cell (1 km resolution) is the mean value calculated from 30 m raster data provided by the Advanced Spaceborne Thermal Emissions and Reflection Radiometer (ASTER) global digital elevation model version 2 (GDEM V2). As the compass direction a slope faces (aspect) can influence the insolation during the daytime, we applied the aspect function in ArcGIS 10.6 (Esri, 2018) on the elevation data to calculate the slope aspect of each grid cell. Analogously, we aggregated the output to 1 km in line with the MODIS grid.

2.4.5 | Urban and vegetation fractions

The urban and vegetation fractions are derived from the 100 m land use and land cover data for the year 2014 from the Israeli Central bureau of Statistics (Central Bureau of Statistics, 2015). The land use and land cover data have 39 categories, which were regrouped into five classes: urban built-up, mining, water courses, vegetation (forest, orchard, and shrubland), and desert (barren).

2.5 | Statistical methods

We propose a two-stage hybrid model approach based on random forest regression to estimate T_a from T_s, as shown in Figure 2.
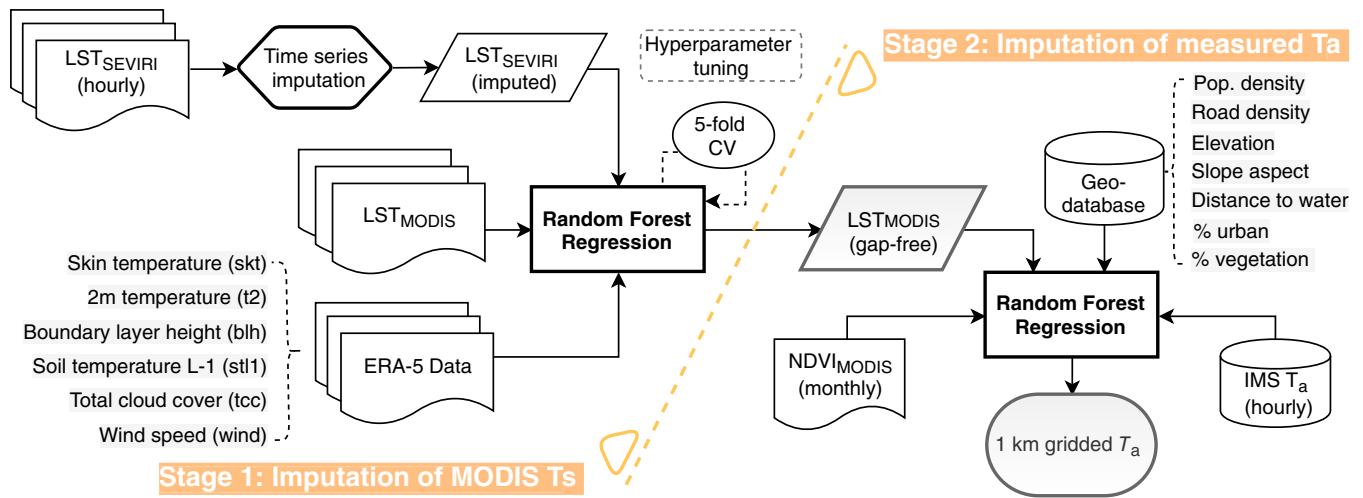
FIGURE 2 Schematic diagram of the two-stage model of estimating 1 km gridded T_a [Colour figure can be viewed at wileyonlinelibrary.com]

First, we imputed T_s values for missing T_s grid cells (obscured by cloud cover, etc.). Then we fed the imputed T_s data for each grid cell into the Stage 2 model to estimate T_a at the same spatial and temporal resolution. We performed the random forest regression using the R *ranger* package (Wright and Ziegler, 2017) wrapped in *mlr* package (Bischl et al., 2016).

2.5.1 | Stage 1 model: Imputation of MODIS T_s

Given the large volume of the MODIS T_s data available for training (on average > 33 million instances each year), the random forest regression was applied on data of each month individually to the missing values in the MODIS T_s. Predictor variables for the Stage 1 model are T_s from the SEVIRI data set, hour of overpass, and spatial coordinates (latitude and longitude), as well as synoptic variables from ECMWF's ERA5 reanalysis datasets including skin temperature (skt), 2 m air temperature (2 t), level 1 soil temperature (stl1), boundary layer height (blh), total cloud cover (tcc), and wind speed. The acronyms in parentheses refer to the short names of parameters in the ECMWF nomenclature. The Stage 1 model is expressed below as Equation 2.

$$T_s^{MODIS} \tilde{R}F \begin{pmatrix} T_s^{SEVIRI}, hour, long., lat., skt^{ERA5}, 2t^{ERA5}, \\ blh^{ERA5}, stl1^{ERA5}, tcc^{ERA5}, wind^{ERA5} \end{pmatrix}$$

$$(2)$$

Like MODIS, SEVIRI data are also limited to clear-sky conditions. However, the hourly SEVIRI data, in contrast with four daily overpasses of the MODIS, may incorporate more information pertaining to the diurnal and seasonal patterns. The additional information could facilitate imputing

data gaps in the SEVIRI data. We interpolated the SEVIRI T_s based on the Seasonal Trend Decomposition using the Loess (STL) algorithm (Cleveland *et al.*, 1990) employed in the R *forecast* package (Hyndman and Khandakar, 2008). The STL divides a time series into the underlying trend, seasonality and remainder. We linearly interpolated the seasonally adjusted data and added back the seasonal component.

We applied a fivefold cross-validation (CV) resampling scheme to tune the Stage 1 model and to estimate its performance. The tuning seeks to identify the optimum hyperparameters resulting in the best model performance, which we evaluate with the overall mean squared error (MSE). As the MSE decreases monotonously and converges asymptotically with growing num.trees (Oshiro *et al.*, 2012; Probst and Boulesteix, 2017), num.trees was set to 300. In the Stage 1 model, we only tuned the number of features (mtry).

Validation of the model was performed using a fivefold CV resampling scheme, which partitions the dataset into five approximately equal-sized subsets. Each subset is iteratively used for testing model performance, while the remaining 4 subsets comprise the training set. Overall model performance was estimated by calculating the "out-of-sample" Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), maintaining a complete separation between data used to train the model and data used to test its performance.

2.5.2 | Stage 2 model: Imputation of T_a from T_s

After obtaining the imputed gap-free T_s for all grid elements, we performed the random forest regression again to calibrate T_a measurements from the IMS stations and the corresponding T_s. The model is based on the data collected from the entire period of investigation, 2004–2016.
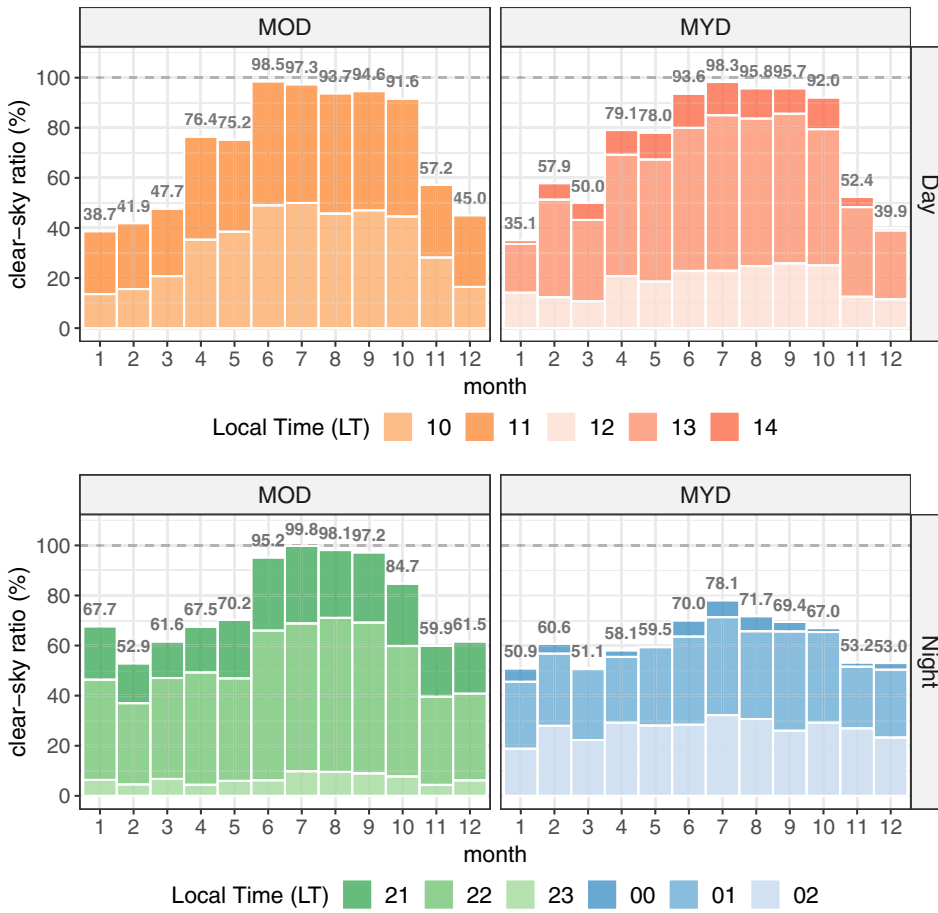
FIGURE 3 The monthly clear-sky ratio (number of overpasses with a clear-sky condition/total number of overpasses) and the contribution of individual overpass hours for each MODIS overpass mode (MOD-day, MYD-night, MYD-day, and MYD-night) in a selected 1-year period (2016) [Colour figure can be viewed at wileyonlinelibrary.com]

Stage 2 of the model is described by Equation (3):

$$T_a \tilde{R} F\left(T_S^{MODIS*}, DoY, Year, NDVI^{MODIS}, Population\ Density, \right.$$

$$Road\ Density, Slope\ Aspect, Elevation,$$

$$\left. Water\ Distance, Urban\ Fraction, Vegetation\ Fraction\right)$$

$$(3)$$

To evaluate the model's capacity to capture the spatio-temporal variability of T_a, we disaggregated the overall performance into spatial and temporal components year by year. The spatial component contrasts the station-wise difference between the annual mean of observed and estimated T_a in the 85 ground stations, whereas the temporal one contrasts at each station the difference in daily T_a between the observation and the estimation, detrended by subtracting its actual value from the respective annual mean (Kloog *et al.*, 2017; Stafoggia *et al.*, 2017).

Once calibrated and validated using fivefold cross-validation, we applied this model to estimate T_a for all grid cells throughout the study area and period.

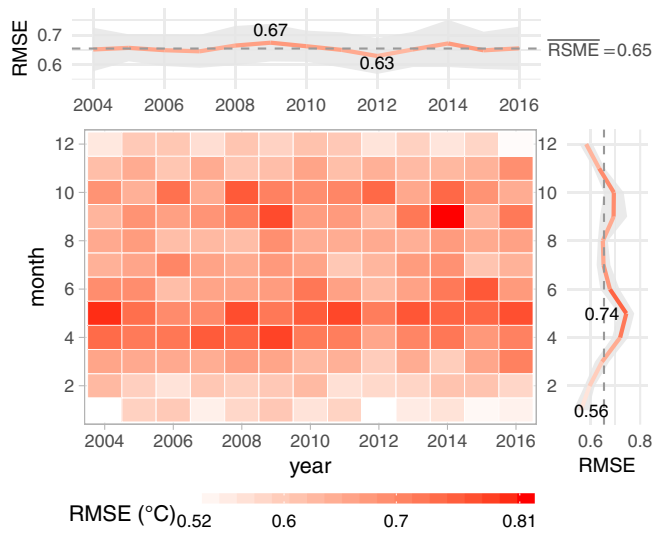# 3 | RESULTS

3.1 | MODIS overpass time

Figure 3 illustrates the temporal distribution and the percentage of clear-sky conditions during the MODIS satellite overpasses for each month in 2016. The MODIS images captured from June to October are mostly cloudless, except for the Aqua-Night overpass, since summer weather in Israel is characterized by the occasional formation of low-level clouds at night, which dissipate soon after sunrise. For Terra-Night and Aqua-Day, the predominant overpass time is LT2200 and LT1300, respectively, whereas there is not a major overpass time for Terra-Day and Aqua-Night.

3.2 | Performance of the stage 1 model

Figure 4 illustrates an imputed image of MODIS Terra-Day T_s at LT1100, January 20, 2016 using the Stage 1 model, in contrast with the original image with a total cloud cover of 55%. Model performance is evaluated for all visible pixels (i.e., pixels not obscured by cloud).
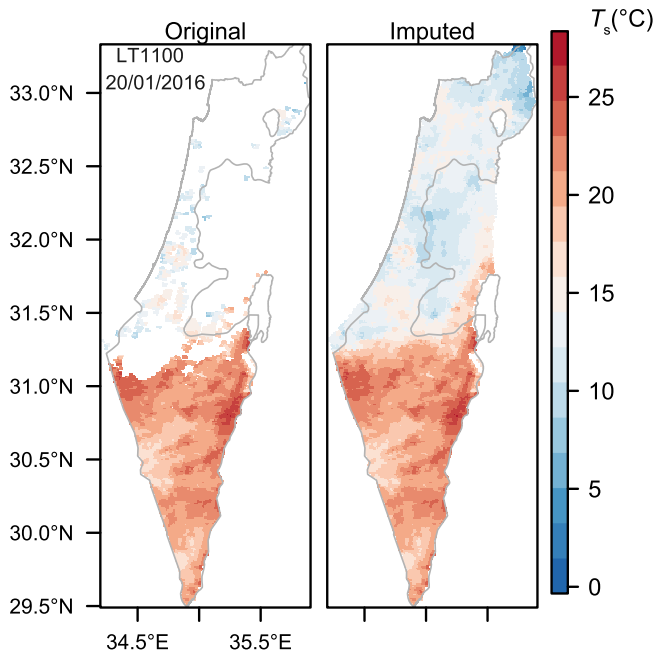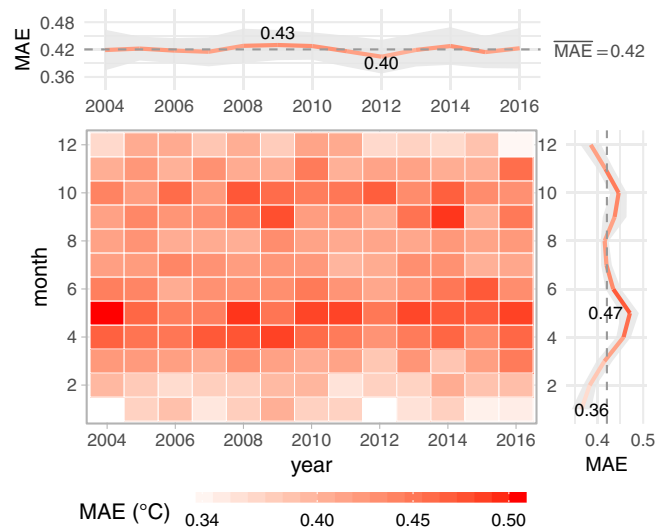
FIGURE 4 An example of imputing MODIS Terra LST image at LT1100 on January 20, 2016 with a total cloud cover of 55% [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 6 Mean absolute error (MAE) of the Stage 1 model using fivefold cross validation (CV5) for each month from 2004 to 2016. The model achieved an overall MAE of 0.42°C, with its annual and monthly means given in the top and right margins, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5 Root mean square error (RMSE) of the Stage 1 model using fivefold cross validation (CV5) for each month from 2004 to 2016. The model achieved an overall RMSE of 0.65°C, with its annual and monthly means given in the top and right margins, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 6, respectively. The Stage 1 model achieved an overall RMSE of 0.65°C and MAE of 0.42°C, with small annual fluctuations. However, both performance measures exhibit pronounced bimodal monthly variability: the model performs relatively well during the winter months (December–January–February, hereafter DJF), whereas the performance declines in May and October. In cloudless months (June–July–August, hereafter JJA), the model's performance returns to an average level.
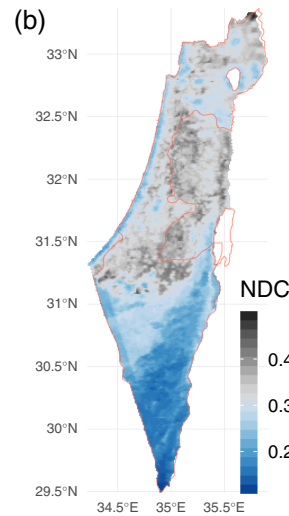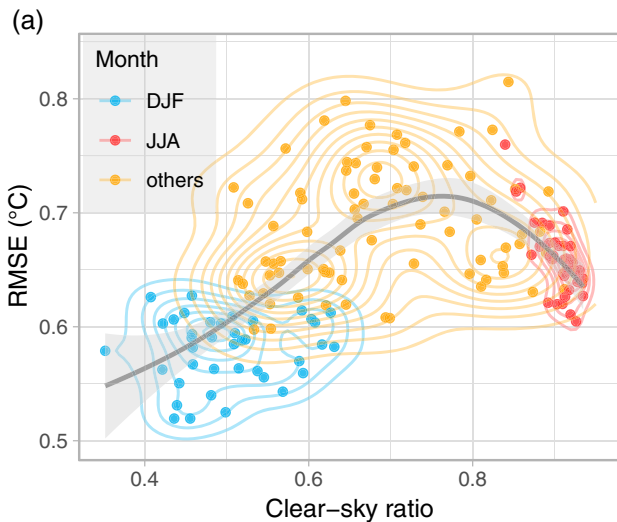
We attribute this seasonality to the relative abundance of clear-sky data in each month that exhibits remarkable spatial and temporal (Figure 3) fluctuation. The spatio-temporal variation of multi-annual monthly mean clear-sky ratio (CSR) across Israel is provided in the Supporting Information (SI) Figure S1.

To quantify the fluctuation of cloudiness across seasons for each grid cell, we proposed a measure denoted as Normalized Difference Cloudiness (NDC) that is defined as follows:

$$NDC = \frac{CSR_{JJA} - CSR_{DJF}}{CSR_{JJA} + CSR_{DJF}}, \quad (4)$$

where CSR_{JJA}, CSR_{DJF} represent the respective clear-sky ratios, defined as the ratio of clear-sky observations to the total number of overpasses, for the months JJA (the driest months dominant by cloudless skies), and for DJF where the precipitation is most likely to happen in Israel, that is, with the highest cloud cover. As CSR_{JJA} is always

Though it was mostly overcast in the North of Israel during this overpass, the model was still able to generate sound clear-sky T_s estimation over the region with an RMSE of 0.54°C, and an MAE of 0.35°C.

The RMSE and MAE of the Stage 1 model for each month from 2004 to 2016 are shown in Figure 5 and

(a)

(b)

FIGURE 7 (a) Relationship between RMSE and clear-sky ratio (CSR) in the Stage 1 model for each month in the study period; and (b) the mean spatial pattern of the normalized difference cloudiness [NDC = (CSR$_{JJA}$ − CSR$_{DJF}$)/ (CSR$_{JJA}$ + CSR$_{DJF}$)] across Israel for all years. The NDC quantifies the fluctuation of clear-sky conditions between JJA (June–July–August, relatively cloudless) and DJF (December–January–February, the main rain season) [Colour figure can be viewed at wileyonlinelibrary.com]

larger than CSR$_{DJF}$, the NDC ranges theoretically between 0 and 1, whereby smaller values indicate less fluctuation in clear-sky conditions throughout the year.

As shown in Figure 7b, the South of Israel is generally less cloudy and presents less variability in the clear-sky ratio across seasons, whereas the clear-sky ratio in the region of Judaean Mountains (central) and Northern Mountains varies significantly with seasons. This spatio-temporal heterogeneity resulted in varying data abundance/availability of each region, and thus in varying regional representativeness in the model. In the cloudless period of the year, for example, during the summer, each region is equally represented in the training datasets, whereas in the rainy season the data from the South prevail. In comparison, data in the transition period between the rainy and sunniest seasons are well mixed but with a high degree of fragmentation in the central part of the country.

Given the spatial autocorrelation that underlies numerous methods for the spatial interpolation (Cressie, 1993), data predominantly collected from contiguous geographical areas (in the case of both dry and rainy seasons) exhibit a higher degree of correlation and thus enable good predictability from the training to the testing datasets. This explains the inverted "U" curve by displaying the relationship between the RMSE and the clear-sky ratio for each month in the study period (Figure 7a). The model's performance is the highest in the rainy winter, as the model is mostly evaluated in the south. Its performance falls in the transition period and then rises again in the summer, as in the summer the model deals with the entire country.

3.3 | Performance of the stage 2 model

The Stage 2 model achieved very good CV performance with RMSE of 1.58°C, MAE of 1.12°C and coefficient of determination (R^2) of 0.96 (a scatterplot is given in the SI Figure S10). The linear regression of the observed versus the estimated T_a across all hours results in a slope of 1.01 and an intercept of −0.27°C, implying a tendency of the model to slightly underestimate large T_a (over 27°C) and over-estimate small T_a for the left-out stations.

The variation of performance across weather stations is provided in the SI (Figures S2–S4). The model performs equally well for each station based on the out-of-sample fivefold cross-validation. No spatial pattern exists across stations, except that a station at the southern end of the Dead Sea exhibits a significantly lower R^2, which we attribute to the systematic or random measurement error in the weather station data or satellite data. As this station is located on an earth levee in the shallow south basin of the dwindling Dead Sea, the MODIS LST algorithm typically applied for homogenous land surfaces may become less effective for mudflats and result in dubious estimation of LST (Wan, 2014).

However, the performance varies slightly between hours. For scatterplots of observed and estimated T_a for each overpass hour across all years, we refer to Figure S11 in the SI. Figure 8 presents the variation of performance measures (overall and disaggregated into spatial and temporal components) for each overpass hour across all years based on the fivefold cross-validation. The model captures both spatial and temporal variation of T_a with little fluctuation both among hours and years, except for the Aqua nighttime hours LT0000 – LT0200. The relatively lower performance with larger inter-annual variance could be ascribed to the relatively lower clear-sky data coverage in those hours than other satellite modes throughout the year (Figure 3). The detailed annual values of performance measures are provided in SI, Figures S5–S9.

FIGURE 8 Variation of performance measures by year across all years: Overall performance with spatial and temporal components. The number of observations for each hour varies annually, dependent on the number of weather stations in operation. On average, there are annually more than 25,000 observations for each hour [Colour figure can be viewed at wileyonlinelibrary.com]

Variation of performance measures by year
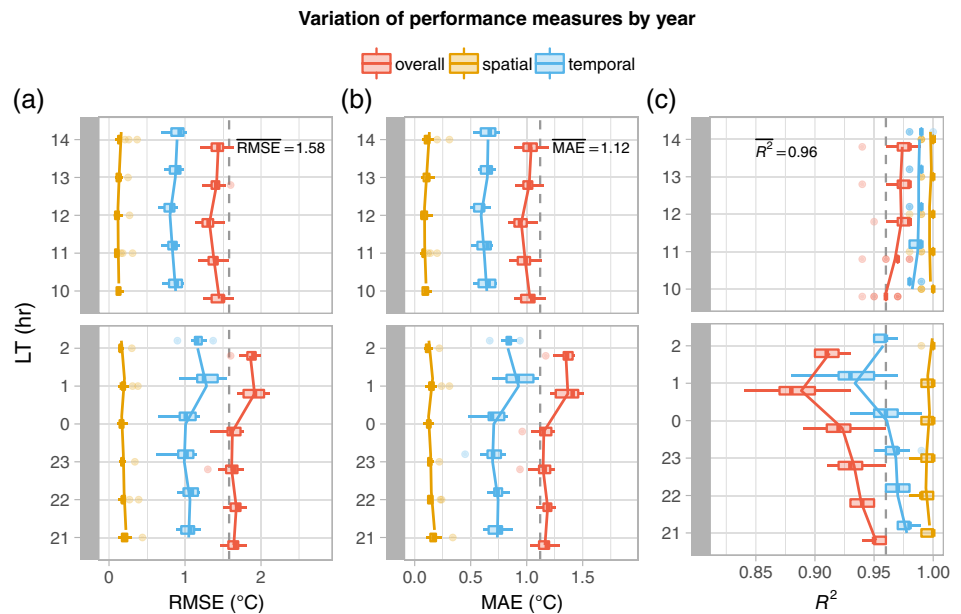
For all measures, the model's performance reaches a local extremum (either minimum or maximum) in 2010 (SI, Figures S5–S9), that is, the model becomes less capable in estimating T_a in that year. We hypothesized that the heat wave that swept across the Northern Hemisphere in 2010 could account for the relatively poor performance (Coumou and Rahmstorf, 2012). In Israel, 2010 was the warmest year ever recorded, with an annual average temperature 2–3°C above the long-term average of 1981–2000 (IMS, 2015). Despite the challenges posed by the unusual weather, the model attained a mean RMSE of 1.66°C and mean MAE of 1.2°C, which are smaller than the anomaly in the annual average temperature.

Figure 9 presents the spatial pattern of multi-annual mean estimated T_a at each overpass hour from the Stage 2 model. The estimations ranged from 11.4°C to 26.3°C with notable diurnal variation. The daytime and nighttime mean T_a are 20.5°C and 17.7°C, respectively. The Jordan Rift Valley and the Hermon mountain on the Golan Heights present constant hot and cold spots throughout day- and nighttime, respectively. This may highlight the remarkable relevance of elevation in accounting for the variability of temperature. Other terrain features such as the Judaean Mountains in central Israel and the Negev dune field (marked with dashed magenta triangle in Figure 10) are also clearly discernible.

Figure 10 shows the daytime and night multi-annual mean estimated T_a. The major metropolitan areas in Israel, for example, Tel Aviv, Haifa, and Beer Sheva, constitute constant hot spots with more pronounced contrast during nighttime than daytime. This can be ascribed to the urban heat island effect arising from the anthropogenic modification of natural landscapes and the consequent atmospheric and thermophysical changes in the urban boundary layer (Arnfield, 2003; Oke et al., 2017).

The only exception is Jerusalem, where its hilly terrain and relatively high altitude dampen the amplitude of temperature variation within the city. The sharp topographical gradient around Jerusalem from the Mediterranean Coast, via the Judaean Plateau, and further to the Dead Sea Depression, also obscures the contrast of urban temperature to its surroundings.

4 | DISCUSSION

This paper presents a two-stage machine learning based hybrid approach to estimating seamless intra-daily $1 \times 1 \text{ km}^2$ T_a from multi-sourced T_s across Israel for 2004–2016. The model demonstrates a good capacity for estimating spatial and temporal variation of T_a at high resolution. To our best knowledge, this study constitutes one of the first efforts towards an hourly estimate of heat exposure across a large geographical area based on a data-driven hybrid approach.

In terms of conventional error statistics (e.g., RMSE and R^2), our model performs comparably to a previous study by Rosenfeld et al. (2017) and outperforms others (Kloog et al., 2012, 2014, 2017). These studies use linear mixed effects models to regress daily extrema/mean T_a on T_s for a given MODIS overpass, whereby the regression coefficients can vary from day to day. As each satellite overpass fluctuates periodically with orbits, ignoring such dynamics may result in a certain degree of uncertainty.
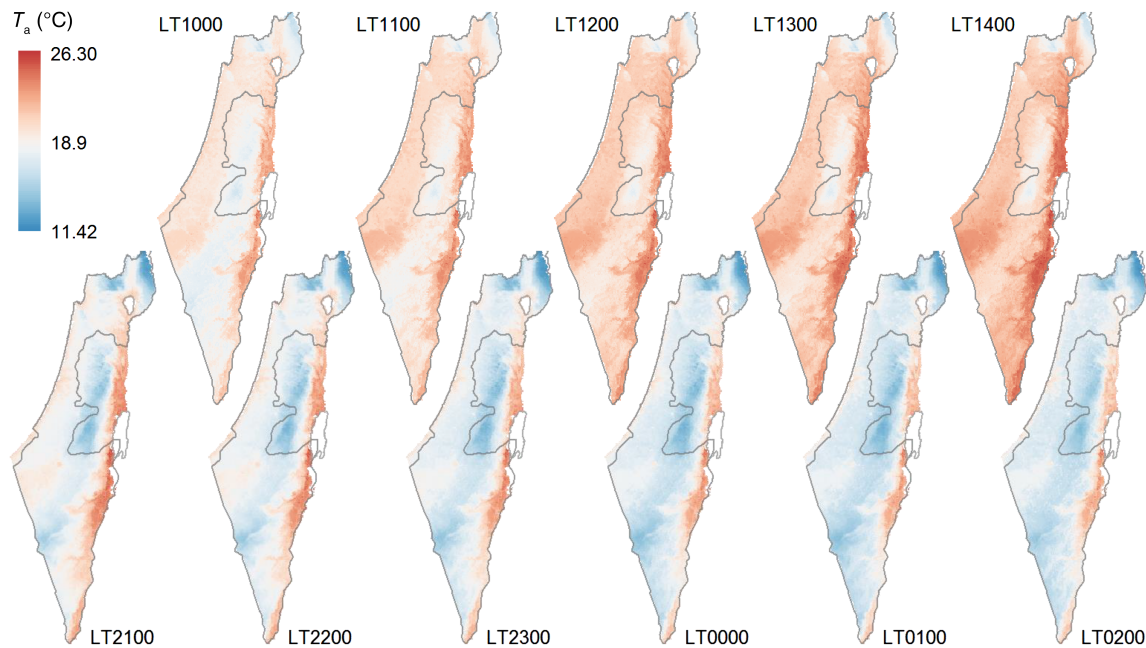
FIGURE 9 Spatial pattern of multi-annual (2004–2016) mean of the estimated T_a at each MODIS overpass hour, obtained from the Stage 2 model. Daytime hours are shown in the upper panel, and nighttime hours in the lower one [Colour figure can be viewed at wileyonlinelibrary.com]

Our model differs from previous ones in the estimated temperature measures (quasi-hourly versus daily extrema, averages), region of interest (Israel versus France, the US, etc.), and time span of the period of investigation (multiple years versus single year). All these factors impede an effective communication and intercomparison of results. It is therefore difficult to conclude that our model is superior to the previous ones. Nevertheless, a random forest-based model does manifest marked advantages over linear mixed effects models.

First, a two-stage RF-based model achieves greater ease of operation than multi-stage linear mixed effects models requiring to temporally and spatially smooth T_a with a thin-plate spline (Kloog *et al.*, 2012; Rosenfeld *et al.*, 2017). The smoothing procedure in particular is found to result in overfitting (Stafoggia *et al.*, 2019). Although linear mixed effects models allow regression coefficients between T_a and T_s to vary by day, the assumption that T_a varies linearly with T_s is in any sense questionable, especially in countries with extremely diverse landscape and climate like Israel. Rather, the linear mixed effects model is regarded as a compromise between computational efficiency and estimation accuracy. In contrast, random forests do not assume a linear relationship. A random forest consisting of a number of independent trees can easily be parallelized, endowing it with excellent performance in speed.

Second, a random forest model does not entail a priori filtering of features (removal of features assumed

to be irrelevant), and the ensuing transformation. A random forest model can reveal feature importance, that is, how each feature contributes to outcomes estimation (Breiman, 2001). Features accounting less for the variability of T_s/T_a are less frequently selected in splitting nodes in trees. Nevertheless, we do consider the relevance of each feature fed into our model viewing the underlying physics, as a recent study suggested that inclusion of counterproductive features could result in over-fitting and worsen estimation performance (Meyer *et al.*, 2018).

Moreover, because random forests can analyse nonlinear relationships and high-order interactions between variables (Basu *et al.*, 2018), it is not required to transform features using for example, standard score normalization or box cox transformation to meet the normality assumption of the linear regression (and its variants).

One unanticipated finding of this study was an inverted "U" shaped performance curve as a function of clear-sky ratio in the Stage 1 model, which we attributed to the uneven distribution of cloud in space and time. This is of crucial importance in any data-driven modelling schemes, as it implies that varying representation of groups in training/testing data sets may result in a model more adapted to the over-representative groups.

Autocorrelation, commonly present in spatial data (Tobler, 1970), might also affect the performance estimate. It is prone to overestimate the performance if training/testing data sets are predominated by spatial
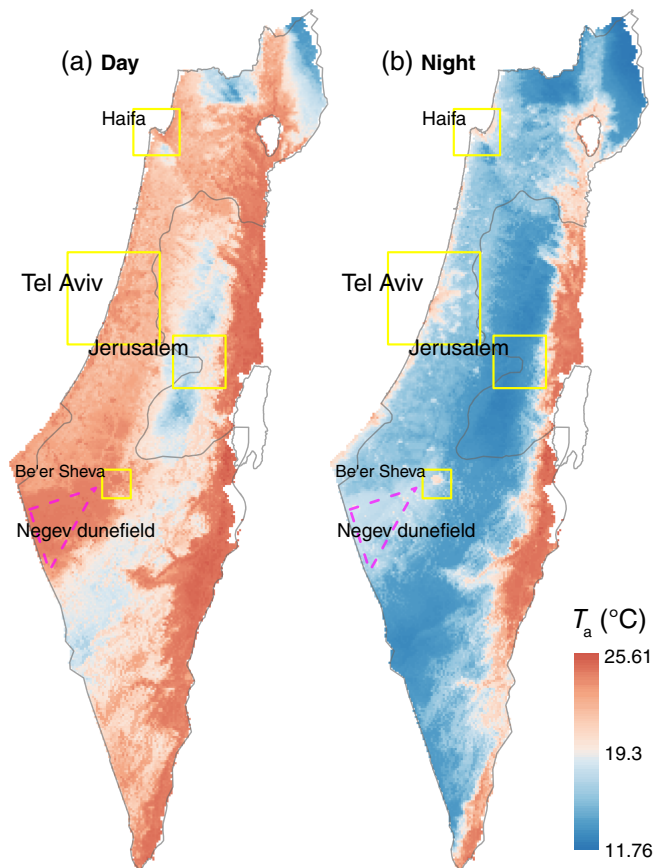
FIGURE 10 Spatial pattern of multi-annul daytime and nighttime mean T_a, which are averaged over LT1000–LT1400 and LT2100–LT0200, respectively. The major metropolitan areas in Israel are marked in yellow [Colour figure can be viewed at wileyonlinelibrary.com]

continua. To what extent the dynamics of spatio-temporal arrangement of data could influence the estimated outcomes remains a topic of debate for future work.

Kloog *et al.* (2014) reported a slightly better model performance in urban areas than rural ones, whereas Benali *et al.* (2012) indicated the opposite, that is, homogeneous surfaces like forests are more predictable. These contrasting findings suggest de facto the complexity between competing factors determining performance, that is, the density of monitoring sensors, and the heterogeneity of respective land forms.

On the one hand, weather stations and monitoring sensors are much more densely installed close to human settlements (e.g., cities, and villages) than in remote rural areas, resulting in a better estimation for urban areas over the rural. On the other hand, urban landscape is characterized by enormous heterogeneity in morphology and socio-economic properties, resulting in notable micro-climate variability within short distances (tens of meters). It is difficult to estimate urbanized T_a because it is the result of an ensemble of effects from different influencing factors.

In this study, we used 85 weather stations scattered across Israel, which are more densely installed in the populous North and Central Districts than in the southern Negev Desert. Meanwhile, we see a weaker tendency of further densifying the current monitoring network, especially in the sparsely populated South of Israel. The adverse conditions associated with the insufficient monitoring sensors may still hold in the near future. To attain an accurate estimate of heat exposure for all locations, we suggest a stratified modelling strategy—a random forest model for homogenous land use and land covers of large spatial extent, enhanced with established computation-efficient urban canopy climate models for heterogenous urban areas (e.g., Masson, 2000; Erell and Williamson, 2006; Bueno *et al.*, 2013; De Ridder *et al.*, 2015). This goes beyond the scope of this study but constitutes a basis for future work.

Despite an overall good performance of our model, we acknowledge several limitations of this study.

Though we have increased the number of daily T_a estimations by decomposing the satellite overpass time, there are still several hours missing. A truly hourly estimation of T_a is still in demand. A model incorporating the diurnal cycle of both T_a and T_s could pave the way (Zakšek and Oštir, 2012).

We imputed T_s for cloud-contaminated grid cells at Stage 1 using models trained mostly by clear-sky T_s. The imputed T_s may neither reflect the real T_s under clouds nor serve as a good proxy to T_a, as T_s and T_a converge in highly cloudy locations (Prigent *et al.*, 2003). Irrespective of this discrepancy, the imputed T_s is indiscriminately (without introducing weighting) fed into the Stage 2 model. An unanswered question is to what extent error propagates through stages, and whether a random forest model trained with weighted data samples could improve the estimation performance.

5 | CONCLUSION

In this study, we proposed a two-stage machine learning based approach to estimating near-surface air temperature (T_a) from remotely sensed surface skin temperature (T_s). It enables a seamless estimation of T_s and T_a at high resolution in both time and space, while achieving good performance and computational efficiency. The novel method takes advantage of an extensive catalogue of geospatial data sets that are becoming increasingly available at a global scale, endowing it with excellent scalability and transferability. Though focusing specifically on

Israel, the model presented in this study can be easily replicated to any other regions where comparable geospatial databases are available. The outcomes help further minimize exposure misclassification in epidemiological studies and may also benefit a wide range of stakeholders including policy-makers, urban-planers, and insurance firms. As urbanization and climate change proceed, incorporating their impacts into daily routine has gradually become a cross-sectoral modus operandi. This is exactly where a high-resolution continuous spatio-temporal T_a database could come into full play.

ORCID

Bin Zhou https://orcid.org/0000-0001-8853-0724
Ian Hough https://orcid.org/0000-0001-7948-6995

REFERENCES

Analitis, A., Katsouyanni, K., Biggeri, A., Baccini, M., Forsberg, B., Bisanti, L., Kirchmayer, U., Ballester, F., Cadum, E., Goodman, P.G., Hojs, A., Sunyer, J., Tiittanen, P. and Michelozzi, P. (2008) Effects of cold weather on mortality: results from 15 European cities within the PHEWE project. *American Journal of Epidemiology*, 168(12), 1397–1408. https://doi.org/10.1093/aje/kwn266.

Arnfield, A.J. (2003) Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat Island. *International Journal of Climatology*, 23(1), 1–26. https://doi.org/10.1002/joc.859.

Basu, R. (2009) High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environmental Health: A Global Access Science Source*, 8(1), 190–202. https://doi.org/10.1186/1476-069X-8-40.

Basu, R., Malig, B. and Ostro, B. (2010) High ambient temperature and the risk of preterm delivery. *American Journal of Epidemiology*, 172(10), 1108–1117. https://doi.org/10.1093/aje/kwq170.

Basu, S., Kumbier, K., Brown, J.B. and Yu, B. (2018) Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences*, 115(8), 1943–1948. https://doi.org/10.1073/pnas.1711236115.

Bechtel, B., Wiesner, S. and Zakšek, K. (2014) Estimation of dense time series of urban air temperatures from multitemporal geostationary satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(10), 4129–4137. https://doi.org/10.1109/JSTARS.2014.2322449.

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

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E,, Casalicchio, G. and Jones, Z. (2016) mlr: Machine Learning in R. Journal of Machine Learning Research, 17(170), 1–5. Available at: http://jmlr.org/papers/v17/15-066.html.

Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brokamp, C., Jandarov, R., Rao, M.B., LeMasters, G. and Ryan, P. (2017) Exposure assessment models for elemental components of particulate matter in an urban environment: a comparison of regression and random forest approaches. *Atmospheric Environment*, 151, 1–11. https://doi.org/10.1016/j.atmosenv.2016.11.066.

Bueno, B., Norford, L., Hidalgo, J. and Pigeon, G. (2013) The urban weather generator. *Journal of Building Performance Simulation*, 6(4), 269–281. https://doi.org/10.1080/19401493.2012.718797.

Central Bureau of Statistics (2014) *Population Estimates by Updated Statistical Regions for the End of 2012*.

Central Bureau of Statistics (2015) *Land Use 2014*.

Central Bureau of Statistics (2018) *Statistical Abstract of Israel 2018*.

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. and Zhang, C. (2011) The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *International Journal of Climatology*, 31(2), 273–288. https://doi.org/10.1002/joc.2158.

Chen, K., Wolf, K., Breitner, S., Gasparrini, A., Stafoggia, M., Samoli, E., Andersen, Z.J., Bero-Bedada, G., Bellander, T., Hennig, F., Jacquemin, B., Pekkanen, J., Hampel, R., Cyrys, J., Peters, A. and Schneider, A. (2018) Two-way effect modifications of air pollution and air temperature on total natural and cardiovascular mortality in eight European urban areas. *Environment International*, 116(April), 186–196. https://doi.org/10.1016/j.envint.2018.04.021.

Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I. (1990) STL: a seasonal-trend decomposition procedure based on loess (with discussion). *Journal of Official Statistics*, 6, 3–73.

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)*.

Coumou, D. and Rahmstorf, S. (2012) A decade of weather extremes. *Nature Climate Change*, 2(7), 1–6. https://doi.org/10.1038/nclimate1452.

Cressie, N.A.C. (1993) *Statistics for Spatial Data. Terra Nova.* Hoboken, NJ: John Wiley & Sons, Inc..

De Ridder, K., Lauwaet, D. and Maiheu, B. (2015) UrbClim – a fast urban boundary layer climate model. *Urban Climate*, 12, 21–48. https://doi.org/10.1016/j.uclim.2015.01.001.

Erell, E. and Williamson, T. (2006) Simulating air temperature in an urban street canyon in all weather conditions using measured data at a reference meteorological station. *International Journal of Climatology*, 26(12), 1671–1694. https://doi.org/10.1002/joc.1328.

Esri (2018) *ArcGIS Desktop: Release 10.6.* Redlands, CA.

Fouillet, A., Rey, G., Laurent, F., Pavillon, G., Bellec, S., Guihenneuc-Jouyaux, C., Clavel, J., Jougla, E. and Hémon, D. (2006) Excess mortality related to the August 2003 heat wave in France. *International Archives of Occupational and Environmental Health*, 80(1), 16–24.

Freitas, S.C., Trigo, I.F., Macedo, J., Barroso, C., Silva, R. and Perdigão, R. (2013) Land surface temperature from multiple geostationary satellites. *International Journal of Remote Sensing*, 34(9–10), 3051–3068. https://doi.org/10.1080/01431161.2012.716925.

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., Forsberg, B., Leone, M., De Sario, M., Bell, M.L., Guo, Y.L.L., Wu, C.F., Kan, H., Yi, S.M., De Sousa Zanotti Stagliorio Coelho, M., Saldiva, P.H.N., Honda, Y., Kim, H. and Armstrong, B. (2015) Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386(9991), 369–375. https://doi.org/10.1016/S0140-6736(14)62114-0.

Georgescu, M., Chow, W.T.L., Wang, Z.H., Brazel, A., Trapido-Lurie, B., Roth, M. and Benson-Lira, V. (2015) Prioritizing urban sustainability solutions: coordinated approaches must incorporate scale-dependent built environment induced effects. *Environmental Research Letters*, 10(6), 61001. https://doi.org/10.1088/1748-9326/10/6/061001.

Grossman-Clarke, S., Schubert, S. and Fenner, D. (2016) Urban effects on summertime air temperature in Germany under climate change. *International Journal of Climatology*, 37(2), 905–917. https://doi.org/10.1002/joc.4748.

Guo, Y., Gasparrini, A., Li, S., Sera, F., Vicedo-Cabrera, A.M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P.H.N., Lavigne, E., Tawatsupa, B., Punnasiri, K., Overcenco, A., Correa, P.M., Ortega, N.V., Kan, H., Osorio, S., Jaakkola, J.J.K., Ryti, N.R.I., Goodman, P.G., Zeka, A., Michelozzi, P., Scortichini, M., Hashizume, M., Honda, Y., Seposo, X., Kim, H., Tobias, A., Íñiguez, C., Forsberg, B., Åström, D.O., Guo, Y.L., Chen, B.-Y., Zanobetti, A., Schwartz, J., Dang, T.N., Van DD, B.M.L., Armstrong, B., Ebi, K.L. and Tong, S. (2018) Quantifying excess deaths related to heatwaves under climate change scenarios: a multicountry time series modelling study. *PLoS Medicine*, 15(7), e1002629. https://doi.org/10.1371/journal.pmed.1002629.

Hamdi, R., Van de Vyver, H., De Troch, R. and Termonia, P. (2014) Assessment of three dynamical urban climate downscaling methods: Brussels's future urban heat Island under an A1B emission scenario. *International Journal of Climatology*, 34(4), 978–999. https://doi.org/10.1002/joc.3734.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M. and Gräler, B. (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e26693v3. https://doi.org/10.7287/peerj.preprints.26693v3.

Hochman, A., Harpaz, T., Saaroni, H. and Alpert, P. (2018a) The seasons' length in 21st century CMIP5 projections over the eastern Mediterranean. *International Journal of Climatology*, 38(6), 2627–2637. https://doi.org/10.1002/joc.5448.

Hochman, A., Mercogliano, P., Alpert, P., Saaroni, H. and Bucchignani, E. (2018b) High-resolution projection of climate change and extremity over Israel using COSMO-CLM. *International Journal of Climatology*, 38(14), 5095–5106. https://doi.org/10.1002/joc.5714.

Hyndman, R.J. and Khandakar, Y. (2008) Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03.

IMS (2015) *Significant climatic events in Israel during 2010.*

IPCC. (2014) *Climate Change 2013 – The Physical Science Basis. 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: Cambridge University Press.

Janatian, N., Sadeghi, M., Sanaeinejad, S.H., Bakhshian, E., Farid, A., Hasheminia, S.M. and Ghazanfari, S. (2017) A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, 37(3), 1181–1194. https://doi.org/10.1002/joc.4766.

Jin, M. and Dickinson, R.E. (2010) Land surface skin temperature climatology: benefitting from the strengths of satellite observations. *Environmental Research Letters*, 5(4), 044004. https://doi.org/10.1088/1748-9326/5/4/044004.

Kaplan, S., Peeters, A. and Erell, E. (2016) Predicting air temperature simultaneously for multiple locations in an urban environment: a bottom up approach. *Applied Geography*, 76, 62–74. https://doi.org/10.1016/j.apgeog.2016.09.015.

Keatinge, W.R. (1997) Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *The Lancet*, 349(9062), 1341–1346. https://doi.org/10.1016/S0140-6736(96)12338-2.

Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Perčec Tadić, M. and Bajat, B. (2014) Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres*, 119(5), 2294–2313. https://doi.org/10.1002/2013JD020803.

Kloog, I., Chudnovsky, A., Koutrakis, P. and Schwartz, J. (2012) Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA. *Science of the Total Environment*, 432, 85–92. https://doi.org/10.1016/j.scitotenv.2012.05.095.

Kloog, I., Melly, S.J., Coull, B.A., Nordio, F. and Schwartz, J.D. (2015) Using satellite-based spatiotemporal resolved air temperature exposure to study the association between ambient air temperature and birth outcomes in Massachusetts. *Environmental Health Perspectives*, 123(10), 1053–1058. https://doi.org/10.1289/ehp.1308075.

Kloog, I., Nordio, F., Coull, B.A. and Schwartz, J. (2014) Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern

USA. *Remote Sensing of Environment*, 150, 132–139. https://doi.org/10.1016/j.rse.2014.04.024.

Kloog, I., Nordio, F., Lepeule, J., Padoan, A., Lee, M., Auffray, A. and Schwartz, J. (2017) Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data. *International Journal of Climatology*, 37(1), 296–304. https://doi.org/10.1002/joc.4705.

Kuglitsch, F.G., Toreti, A., Xoplaki, E., Della-Marta, P.M., Zerefos, C.S., Türkeş, M. and Luterbacher, J. (2010) Heat wave changes in the eastern Mediterranean since 1960. *Geophysical Research Letters*, 37(4), L04802. https://doi.org/10.1029/2009GL041841.

Li, L. and Zha, Y. (2018) Mapping relative humidity, average and extreme temperature in hot summer over China. *Science of the Total Environment*, 615, 875–881. https://doi.org/10.1016/j.scitotenv.2017.10.022.

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F. and Sobrino, J.A. (2013) Satellite-derived land surface temperature: current status and perspectives. *Remote Sensing of Environment*, 131, 14–37. https://doi.org/10.1016/j.rse.2012.12.008.

Liaw, A. and Wiener, M. (2002) Classification and regression by randomForest. *R News*, 2, 18–22.

Lin, S., Moore, N.J., Messina, J.P., DeVisser, M.H. and Wu, J. (2012) Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa. *International Journal of Applied Earth Observation and Geoinformation*, 18(1), 128–140. https://doi.org/10.1016/j.jag.2012.01.004.

Masson, V. (2000) A physically-based scheme for the urban energy budget in atmospheric models. *Boundary-Layer Meteorology*, 94(3), 357–397. https://doi.org/10.1023/A:1002463829265.

Masson, V., Grimmond, C.S.B. and Oke, T.R. (2002) Evaluation of the town energy balance (TEB) scheme with direct measurements from dry districts in two cities. *Journal of Applied Meteorology*, 41(10), 1011–1026. https://doi.org/10.1175/1520-0450(2002)041<1011:EOTTEB>2.0.CO;2.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. and Nauss, T. (2018) Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling and Software*, 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

Mirzaei, P.A. and Haghighat, F. (2010) Approaches to study urban heat Island – abilities and limitations. *Building and Environment*, 45(10), 2192–2201. https://doi.org/10.1016/j.buildenv.2010.04.001.

Ochsenwald, W.L., Sicherman, H., et al. (2019) *Israel*. Encyclopædia Britannica: Encyclopædia Britannica, Inc..

Oke, T.R., Mills, G., Christen, A. and Voogt, J.A. (2017) *Urban Climates*. Cambridge: Cambridge University Press.

Oshiro, T.M., Perez, P.S. and Baranauskas, J.A. (2012) How many trees in a random forest? In: Perner, P. (Ed.) *Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Lecture Notes in Computer Science, Vol 7376*. Berlin, Heidelberg: Springer, pp. 154–168.

Oyler, J.W., Ballantyne, A., Jencso, K., Sweet, M. and Running, S.W. (2015) Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *International Journal of Climatology*, 35(9), 2258–2279. https://doi.org/10.1002/joc.4127.

Pebesma, E. (2018) Simple features for R: standardized support for spatial vector data. *The R Journal*, 10(1), 439–446. https://doi.org/10.32614/RJ-2018-009.

Prata, A.J., Caselles, V., Coll, C., Sobrino, J.A. and Ottle, C. (1995) Thermal remote sensing of land surface temperature from satellites: current status and future prospects. *Remote Sensing Reviews*, 12(3–4), 175–224. https://doi.org/10.1080/02757259509532285.

Prigent, C., Aires, F. and Rossow, W.B. (2003) Land surface skin temperatures from a combined analysis of microwave and infrared satellite observations for an all-weather evaluation of the differences between air and skin temperatures. *Journal of Geophysical Research*, 108(D10), 1–14. https://doi.org/10.1029/2002JD002301.

Probst, P. and Boulesteix, A.-L. (2017) To tune or not to tune the number of trees in random forest? *Journal of Machine Learning Research*, 18, 1–18.

Rosenfeld, A., Dorman, M., Schwartz, J., Novack, V., Just, A.C. and Kloog, I. (2017) Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel. *Environmental Research*, 159(March), 297–312. https://doi.org/10.1016/j.envres.2017.08.017.

Rosenthal, J.K., Kinney, P.L. and Metzger, K.B. (2014) Intra-urban vulnerability to heat-related mortality in New York City, 1997–2006. *Health & Place*, 30, 45–60. https://doi.org/10.1016/j.healthplace.2014.07.014.

Schnell, J.L. and Prather, M.J. (2017) Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America. *Proceedings of the National Academy of Sciences*, 114(11), 2854–2859. https://doi.org/10.1073/pnas.1614453114.

Schwartz, J., Samet, J.M. and Patz, J.A. (2004) Hospital admissions for heart disease. *Epidemiology*, 15(6), 755–761. https://doi.org/10.1097/01.ede.0000134875.15919.0f.

Shi, L., Kloog, I., Zanobetti, A., Liu, P. and Schwartz, J.D. (2015) Impacts of temperature and its variability on mortality in New England. *Nature Climate Change*, 5(11), 988–991. https://doi.org/10.1038/nclimate2704.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., De, H.K., De, D.F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., Schwartz, J., de Hoogh, K., de' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I. and Schwartz, J. (2019) Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environment International*, 124(November), 170–179. https://doi.org/10.1016/j.envint.2019.01.016.

Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., de' Donato, F., Gaeta, A., Leone, G., Lyapustin, A., Sorek-Hamer, M., de Hoogh, K., Di, Q., Forastiere, F. and Kloog, I. (2017) Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environment International*, 99(2017), 234–244. https://doi.org/10.1016/j.envint.2016.11.024.

Stafoggia, M., Schwartz, J., Forastiere, F. and Perucci, C.A. (2008) Does temperature modify the association between air pollution and mortality? A multicity case-crossover analysis in Italy. *American Journal of Epidemiology*, 167(12), 1476–1485. https://doi.org/10.1093/aje/kwn074.

Survey of Israel (2013) *Survey of Israel, 2013*.

Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234. https://doi.org/10.2307/143141.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

Vancutsem, C., Ceccato, P., Dinku, T. and Connor, S.J. (2010) Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, 114(2), 449–465. https://doi.org/10.1016/j.rse.2009.10.002.

Wan, Z. (2013) *MODIS Land Surface Temperature Products Users' Guide*. Available at: https://lpdaac.usgs.gov/documents/118/MOD11_User_Guide_V6.pdf [Accessed 07th April 2020].

Wan, Z. (2014) New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, 140, 36–45. https://doi.org/10.1016/j.rse.2013.08.027.

Wang, J., Feng, J., Yan, Z., Hu, Y. and Jia, G. (2012) Nested high-resolution modeling of the impact of urbanization on regional climate in three vast urban agglomerations in China. *Journal of Geophysical Research*, 117(D21), 1–18. https://doi.org/10.1029/2012JD018226.

Weiss, D.J., Atkinson, P.M., Bhatt, S., Mappin, B., Hay, S.I. and Gething, P.W. (2014) An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98, 106–118. https://doi.org/10.1016/j.isprsjprs.2014.10.001.

Wright, M.N. and Ziegler, A. (2017) Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17. https://doi.org/10.18637/jss.v077.i01.

Ye, X., Wolff, R., Yu, W., Vaneckova, P., Pan, X. and Tong, S. (2012) Ambient temperature and morbidity: a review of epidemiological evidence. *Environmental Health Perspectives*, 120(1), 19–28. https://doi.org/10.1289/ehp.1003198.

Guo, Y.Y.L., Gasparrini, A., Armstrong, B., Li, S., Tawatsupa, B., Tobias, A., Lavigne, E., de Sousa Zanotti Stagliorio Coelho, M., Leone, M., Pan, X., Tong, S., Tian, L., Kim, H., Hashizume, M., Honda, Y., YYL, G., Wu, C.-F., Punnasiri, K., Yi, S., Michelozzi, P., Saldiva, P.H.N. and Williams, G. (2014) Global variation in the effects of ambient temperature on mortality. *Epidemiology*, 25(6), 781–789. https://doi.org/10.1097/EDE.0000000000000165.

Zakšek, K. and Oštir, K. (2012) Downscaling land surface temperature for urban heat Island diurnal cycle analysis. *Remote Sensing of Environment*, 117, 114–124. https://doi.org/10.1016/j.rse.2011.05.027.

Zakšek, K. and Schroedter-Homscheidt, M. (2009) Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4), 414–421. https://doi.org/10.1016/j.isprsjprs.2009.02.006.

Zeger, S.L., Thomas, D., Dominici, F., Samet, J.M., Schwartz, J., Dockery, D. and Cohen, A. (2000) Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental Health Perspectives*, 108(5), 419–426. https://doi.org/10.1289/ehp.00108419.

Zhang, H., Wang, Y., Park, T.W. and Deng, Y. (2017) Quantifying the relationship between extreme air pollution events and extreme weather events. *Atmospheric Research.*, 188, 64–79. https://doi.org/10.1016/j.atmosres.2016.11.010.

Zhou, B., Kaplan, S., Peeters, A., Kloog, I. and Erell, E. (2019) "Surface", "satellite" or "simulation": mapping intra-urban microclimate variability in a desert city. *International Journal of Climatology*, 2019, 1–19. https://doi.org/10.1002/joc.6385.

Zhu, W., Lű, A. and Jia, S. (2013) Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, 130, 62–73. https://doi.org/10.1016/j.rse.2012.10.034.

Zhu, W., Lű, A., Jia, S., Yan, J. and Mahmood, R. (2017) Retrievals of all-weather daytime air temperature from MODIS products. *Remote Sensing of Environment*, 189, 152–163. https://doi.org/10.1016/j.rse.2016.11.011.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.