

RESEARCH ARTICLE

A spatiotemporal reconstruction of daily ambient temperature using satellite data in the Megalopolis of Central Mexico from 2003 to 2019

Iván Gutiérrez-Avila¹  | Kodi B. Arfer¹ | Sandy Wong² | Johnathan Rush¹ | Itai Kloog³ | Allan C. Just¹

¹Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York

²Department of Geography, Florida State University (FSU), Tallahassee, Florida

³Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Beersheba, Israel

Correspondence

Iván Gutiérrez-Avila, Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1057, New York, NY 10029, USA.
Email: ivan_2c@hotmail.com

Funding information

National Institutes of Health (NIH), Grant/Award Numbers: P30 ES023515, R00 ES023450, R01 ES013744, R01 ES014930, R01 ES021357, R24 ES028522, T32 HD049311

Abstract

While weather stations generally capture near-surface ambient air temperature (T_a) at a high temporal resolution to calculate daily values (i.e., daily minimum, mean, and maximum T_a), their fixed locations can limit their spatial coverage and resolution even in densely populated urban areas. As a result, data from weather stations alone may be inadequate for T_a -related epidemiology particularly when the stations are not located in the areas of interest for human exposure assessment. To address this limitation in the Megalopolis of Central Mexico (MCM), we developed the first spatiotemporally resolved hybrid satellite-based land use regression T_a model for the region, home to nearly 30 million people and includes Mexico City and seven more metropolitan areas. Our model predicted daily minimum, mean, and maximum T_a for the years 2003–2019. We used data from 120 weather stations and Land Surface Temperature (LST) data from NASA's MODIS instruments on the Aqua and Terra satellites on a 1×1 km grid. We generated a satellite-hybrid mixed-effects model for each year, regressing T_a measurements against land use terms, day-specific random intercepts, and fixed and random LST slopes. We assessed model performance using 10-fold cross-validation at withheld stations. Across all years, the root-mean-square error ranged from 0.92 to 1.92 K and the R^2 ranged from .78 to .95. To demonstrate the utility of our model for health research, we evaluated the total number of days in the year 2010 when residents ≥ 65 years old were exposed to T_a extremes (above 30°C or below 5°C). Our model provides much needed high-quality T_a estimates for epidemiology studies in the MCM region.

KEYWORDS

extreme air temperature, human exposure, land surface temperature, Megalopolis of Central Mexico, MODIS, remote sensing

Iván Gutiérrez-Avila and Kodi B. Arfer contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

1 | INTRODUCTION

Climate change has spurred worldwide efforts to model trends and fluctuations in near-surface (i.e., 2 m) ambient air temperature (T_a). Recent developments in climatology and related fields have yielded valuable datasets that quantify average and extreme temperatures at local and global spatial scales and at critical time scales that are relevant for exposure science and public health (Donat *et al.*, 2014; Oyler *et al.*, 2015; Behnke *et al.*, 2016). However, there is considerable heterogeneity in data sources, methods, and temperature products—as the process of producing T_a estimates with adequate spatial and temporal coverage is far from straightforward. Data have included in situ observations or reanalysis data from ground stations, radiosondes, satellites, and other sources (Donat *et al.*, 2014; Behnke *et al.*, 2016). Interpolation methods have included kriging, angular distance weighting, thin plate splines, and/or land use regression (Hofstra *et al.*, 2008). The heterogeneity among different products makes them difficult to compare. Estimates are less consistent and accurate for extremes than they are for averages in T_a . Furthermore, a major limitation even among widely used datasets is that they have a relatively coarse spatial resolution that spans many kilometres (Donat *et al.*, 2014; Behnke *et al.*, 2016).

There are ongoing efforts in urban climate research to improve the quantification of urban meteorological phenomena, including T_a (Šećerov *et al.*, 2019; Venter *et al.*, 2020). Urban areas are particularly important because they contain most of the world's population; thus, more accurate estimates of the urban environment would improve research efforts linking T_a to human health risks (Venter *et al.*, 2020). However, urban environments are complex and heterogeneous, and traditional meteorological monitoring stations, mostly focused on synoptic atmospheric conditions, are inadequate for capturing the spatiotemporal variability of many meteorological phenomena across urban landscapes (Meier *et al.*, 2017). One direction for moving urban climatology forward is to densify urban meteorological networks by increasing the density of various sensors beyond traditional monitoring stations to include open-source technology (Šećerov *et al.*, 2019) or crowdsourced data from private monitors (Muller *et al.*, 2013). There is also a trend toward integrating high-density networks with opportunistic sensing data to improve the quantification of urban T_a (Venter *et al.*, 2020). This includes remotely sensed data such as Land Surface Temperature (LST). LST is retrieved from the thermal infrared signal received by satellites and measures the thermal radiation emitted from the earth's surface as a result of the interaction between incoming solar energy and the ground, or the

top of the canopy in urban and vegetated areas. LST's spatially continuous and global coverage makes it possible to examine the thermal heterogeneity of the Earth's surface and changes over time in surface temperatures (Hulley *et al.*, 2019).

In epidemiology, there is a growing need for daily intracity T_a estimates (i.e., 1×1 km or higher resolution), as T_a effects on human health often occur at fine spatial and temporal scales (Mostofsky *et al.*, 2014; Phung *et al.*, 2016; Rowland *et al.*, 2020; Venter *et al.*, 2020). To date, many health studies only use T_a data from ground stations (Zanobetti and O'Neill, 2018). Data are often used from the airport weather stations closest to the study population or from a mix of local weather stations (Zanobetti and Schwartz, 2008; Ren *et al.*, 2011; Zhang *et al.*, 2017). The main problem with relying on a sparse network of ground stations or on stations located far from study populations is that it can introduce measurement bias and create exposure misclassification (Armstrong, 1998; Zeger *et al.*, 2000).

Remotely sensed LST has increasingly been used to refine gridded interpolations of T_a (Oyler *et al.*, 2016). LST and T_a are physically related; however, their correlation varies daily due to meteorological conditions, seasonality, soil moisture, land use, urban geometry, elevation, surface reflectance, and satellite-surface geometry (Oyler *et al.*, 2016; Shi *et al.*, 2016a; Pelta and Chudnovsky, 2017). Researchers have in recent years incorporated satellite data and land use regression to develop more accurate estimates of T_a for epidemiology (Kloog *et al.*, 2014; Ho *et al.*, 2016; Shi *et al.*, 2016b; Pelta and Chudnovsky, 2017; Rosenfeld *et al.*, 2017; Zhang *et al.*, 2017). These new models have been developed mostly for cities in northern latitudes, resulting in a geographic concentration of epidemiological research in these places (Shi *et al.*, 2015; Lee *et al.*, 2016; Shi *et al.*, 2016a, with very few studies carried out in other regions (Xu *et al.*, 2014).

In Mexico, epidemiological studies linking T_a exposure to health outcomes have mainly taken place in Mexico City and the larger Mexico City Metropolitan Area (MCMA; O'Neill *et al.*, 2005; Bell *et al.*, 2008; McMichael *et al.*, 2008; Hurtado-Díaz *et al.*, 2019). These studies reported associations with acute mortality in different age groups, including children (0–14 years old) and the elderly (≥ 65 years old), who are especially vulnerable to T_a exposure due to their physiological characteristics and their dependence on caregivers to regulate their body temperature. While these studies provide valuable findings, a major shortcoming is that they assigned citywide daily average temperatures using a few weather stations with limited spatiotemporal coverage. Moreover, there is limited research outside of the MCMA and within the greater Megalopolis of Central Mexico (MCM), which includes several major metropolitan areas neighbouring the MCMA.

Ta estimates with finer spatiotemporal resolutions are needed to move beyond the citywide approach to more accurately quantify individuals' exposures. In this paper, we integrated Ta records from ground meteorological networks, satellite remote sensing LST and other spatiotemporal predictors to reconstruct daily minimum, mean, and maximum Ta for the entire MCM for the 2003–2019 time period at a 1×1 km spatial resolution. As a case study, we provide estimates for the total number of days in the year 2010 when MCM residents ≥ 65 years old were exposed to Ta extremes. Our temperature model opens up opportunities to reconstruct historical Ta exposures for any MCM resident and to explore individual-level associations between Ta and health outcomes. These methods have yet to be applied to Mexico, a middle-income country with unique climate zones and sociodemographic characteristics.

2 | METHODS

2.1 | Study area

The central region of Mexico has experienced rapid urbanization leading to the development of the MCM,

which has a population of nearly 30 million inhabitants representing 25% of the country's total population. The MCM includes the MCMA in its centre and the metropolitan areas of Puebla-Tlaxcala, Cuernavaca, Cuautla, Toluca, Pachuca, Tula, and Tulancingo. A total of 184 municipalities integrate the MCM (USAID, 2014).

Plains, mountains, and hills cover 42, 34, and 24% of the MCM territory, respectively, and its climate is influenced by the humid tropical air of the Pacific Ocean, the Caribbean Sea, and the Gulf of Mexico as well as the polar air from the North American continent. There are three main seasons in the region: cold-dry, warm-dry, and rainy seasons and five thermal zones (% of the total MCM territory): cold (0.7%), semicold (10%), temperate (70%), semiwarm (19%), and warm (0.3%). The annual mean temperature for the region ranges from 4 to 26°C (USAID, 2014).

The MCM is an irregularly shaped region that is 20,686 km² in area, and we used the MCM as the prediction area for which we generated and validated our Ta predictions. The longitude ranges from 99.9 to 97.8°W and the latitude ranges from 18.6 to 20.2°N (see Figure 1). To produce the study area for our Ta model, we expanded the spatial extent by 50 km outwards from each side of the bounding box of the MCM and rounded

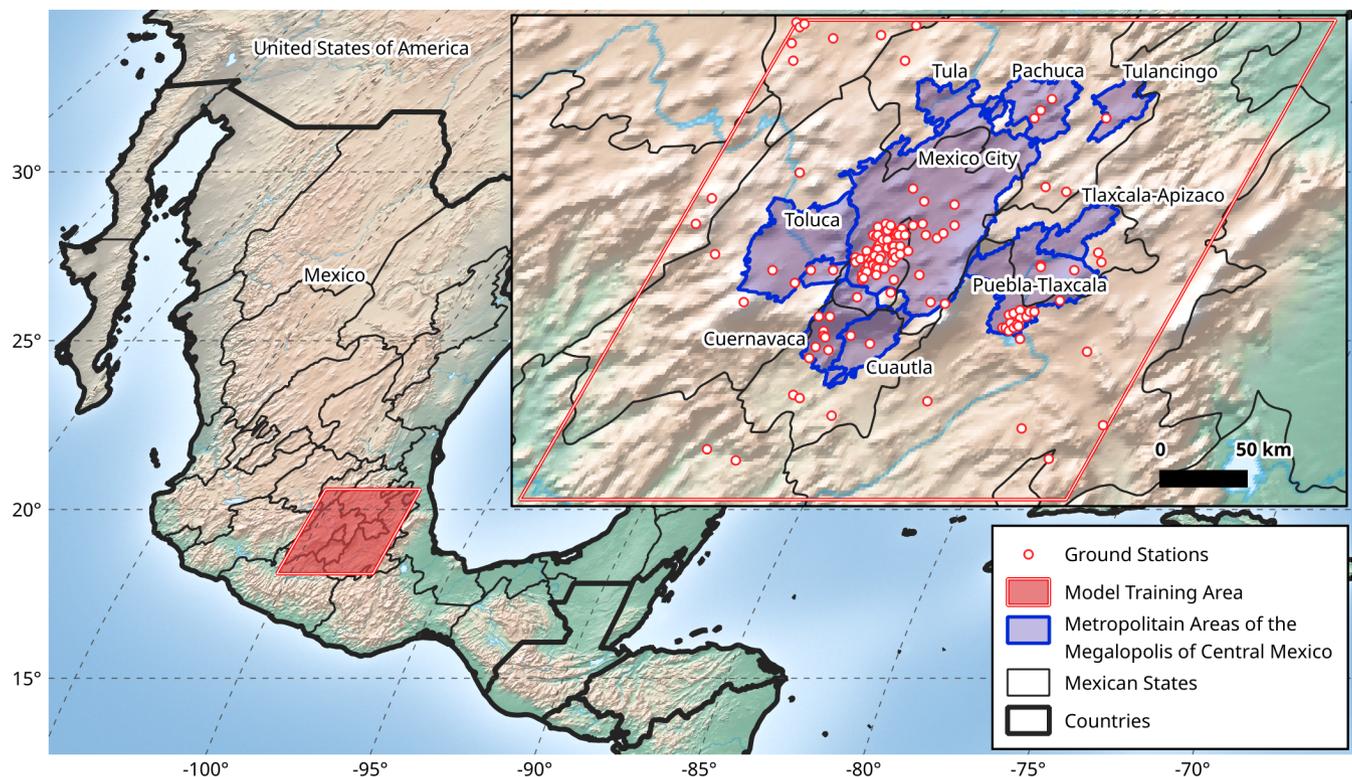


FIGURE 1 Study area showing all available ground meteorological stations ($n = 120$) used for our daily Ta predictions in the Megalopolis of Central Mexico (MCM, shown as indigo-coloured regions) from 2003 to 2019 [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

the limits to the nearest tenth of a degree of longitude or latitude. This allowed us to include additional stations that we used only as training data. The result was a rectangular region in the plate carrée projection that ranged from 100.4 to 97.4°W and 18.1 to 20.6°N. This was then reprojected to match the MODIS sinusoidal projection and divided into 101,892 total 1×1 km cells, and this output served as our master grid for modelling.

2.2 | Data sources

2.2.1 | Meteorological ground stations

We utilized weather station records from the year 2003 through 2019 inclusive, based on data availability from the different sources of information. By request, we obtained historical records of Ta and wind speed from the Servicio Meteorológico Nacional de México (SMN) that integrated three monitoring networks: the Estaciones Meteorológicas Automáticas network (EMAs; 30 stations), the Estaciones Sinópticas Meteorológicas network (ESIMES; 10 stations), and a network of observatories (9 stations). We also incorporated two networks for which data are publicly available on the Internet: the Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT; 34 stations; <http://www.aire.cdmx.gob.mx>) and the Programa de Estaciones Meteorológicas del Bachillerato Universitario at the Universidad Nacional Autónoma de México (PEMBU-UNAM; 13 stations; <https://www.ruoa.unam.mx/pembu>). Finally, we included observations from personal weather stations available from Weather Underground (70 stations; <http://wunderground.com>). Records for all networks are available in local time, except those from the EMAs and ESIMES networks which are available in UTC and were converted to local time.

Given the heterogeneous sources of information in our study which included crowdsourced data, we adapted methods developed for data quality assessment applied elsewhere to make them work in our study region (Meier *et al.*, 2017; Napoly *et al.*, 2018; Dirksen *et al.*, 2020). We discarded data that did not pass several checks. Each check, described below, is prepended with the number of observations that were deleted by the check. (a) (134) If a station had fewer than 20 observations in a given year, we dropped that station for that year. (b) (1,670) We dropped station-days with values that were impossible (e.g., a maximum Ta that is less than the minimum Ta or a negative mean wind speed) or implausible according to climatological records (i.e., colder than -30°C , hotter than 53°C , or having a mean wind speed greater than 114 m/s). (c) (674) We dropped runs of station-days

(ignoring unobserved days) in which one of the three daily Ta values (minimum, mean, and maximum), rounded to the nearest 0.01 K, was repeatedly equal. We allowed runs of up to two equal values for the mean Ta and three equal values for the maximum or minimum Ta. These limits were chosen based on the observation that longer runs are rare except in Weather Underground. (d) (16,297) We compared observed Ta to inverse-distance-weighted interpolations of Ta from other stations on the same day that were no more than 30 km away and 500 m different in elevation as a “buddy check” for spatial consistency. We term the squared differences between observations and these interpolations “deviations,” and for each Ta’s daily observation, we computed the 99th percentile of the deviation, excluding Weather Underground. We dropped station-days for which a deviation exceeded this percentile, and also dropped entire stations for which 20% or more of the temperature-days were dropped.

These checks removed 11% of Weather Underground observations, 6% of SMN-observatories observations, and less than 5% of observations from each of the other networks. Afterwards, 402,823 observations from 166 stations remained; the above per-network counts of stations are after data cleaning. Of these, there were 290,886 observations from 120 stations in the prediction area (i.e., the area of the MCM). We checked for possible duplicate stations (stations with different identifiers but similar locations and a substantial number of identical observations) but did not find any. A summary of the geographic location of stations from each network, number of records used in our analyses, land use/land cover, and local climate is shown in Table S1, Supporting Information. The geographic location of stations from the UNAM and REDMET networks can be found online in their websites (UNAM, <https://www.ruoa.unam.mx/pembu/index.php?page=map#>; REDMET, http://www.aire.cdmx.gob.mx/pendata/catalogos/cat_estacion.csv). A complete list with geographic coordinates for the EMAs and ESIMES networks of the SMN was downloaded from their website (<http://smn1.conagua.gob.mx/emas/catalogoa.html>). The same information for all the networks from the SMN was received in our data request. Information on land use/land cover (Instituto Nacional de Estadística y Geografía (INEGI), 2017) and local climate (Instituto Nacional de Estadística y Geografía (INEGI), 2008) was obtained from the Instituto Nacional de Estadística y Geografía de México (INEGI) from its website. In Data S1, we included general characteristics about mounting, location, and exposure of instruments and contact information for each network, based on guidance on metadata by Aguilar *et al.* (2003), when such information was available.

2.2.2 | Land surface temperature

LST records were extracted from the daily daytime and night-time MODIS LST products MOD11A1 (Terra satellite) and MYD11A1 (Aqua satellite) using the most recent reprocessing version (Collection 6). LST retrievals from both satellites were available by the first half of 2002, but we selected 2003 as the start year due to data completeness from both sensors (Tatem *et al.*, 2004). The spatial resolution was 1×1 km, and data were available for the entire study period. Local overpass times for Terra and Aqua were around 2230 and 0130 for night-time, and around 1030 and 1330 for daytime, respectively. MODIS LST products are derived from channels 31 (10.78–11.28 μm) and 32 (11.77–12.27 μm) in the thermal infrared band, and they are already corrected for emissivity and atmospheric effects using the split-window algorithm. These products have been used before in similar studies due to their high spatiotemporal resolution and free availability (Benali *et al.*, 2012; Kloog *et al.*, 2012; 2014; Shi *et al.*, 2016b; Rosenfeld *et al.*, 2017).

2.2.3 | Land use terms

Vegetation

To calculate a measure of vegetation density, the monthly Normalized Difference Vegetation Index (NDVI) from both Terra and Aqua MODIS instruments were averaged (Collection 6 MOD13_A3 and MYD13_A3) at the spatial resolution of 1×1 km.

Elevation

Elevation from the Shuttle Radar Topography Mission at a 30 m spatial resolution was aggregated by applying a Gaussian filter (150 m SD) and extracting data to the centroids of the MODIS 1×1 km products.

readings were available, the average was calculated. To impute cases where both satellite readings were missing, we used a linear-interpolation algorithm as follows. For a given day t and grid cell g , we found the closest days before and after t in the same year as t , t_0 , and t_1 , with a nonmissing value of the appropriate variable (daytime LST or night-time LST) at g , and we computed the imputed value as

$$y_{ig} = (y_0(t_1 - t) + y_1(t - t_0)) / (t_1 - t_0),$$

where y_{ig} is the LST at day t_i and place g . When no such t_0 exists for the given year, y_1 , unaltered, is used as the imputed value instead, and likewise y_0 substitutes when no t_1 exists. We also imputed missing wind speed. We did this by using the wind speed from the closest station with a value for the same day (or the previous day, if no other station was available that day, or the day before that if necessary, and so on).

We calibrated daily T_a (minimum, mean, and maximum) on LST as follows: each T_a station was assigned the closest LST observation on a specific day (within each 1×1 km grid cell) using grid cells for which both T_a measurements and LST values (observed or imputed) were available. On each day we estimated a separate slope in the relationship between T_a and LST to capture the temporal variability in their relationship. The calibrations, fit separately per year and daily T_a outcome (minimum, mean, or maximum), were mixed-effects regression models implemented with the R package lme4 (Bates *et al.*, 2015). After calibration, we used the coefficients of the mixed-effects model to predict T_a in those grid cells without T_a information but with LST values. A generalization of the equations for the three models (for minimum, mean, and maximum T_a) is,

$$Ta_{ij} = (\alpha + \nu_j) + \beta_1 \text{day LST}_{ij} + \beta_2 \text{night LST}_{ij} + \beta_3 \text{imputed day LST}_{ij} + \beta_4 \text{imputed night LST}_{ij} + \beta_5 \text{NDVI}_{ij} + \beta_6 \sin 2\pi \text{time}_i + \beta_7 \cos 2\pi \text{time}_i + \beta_8 \text{elevation}_i + \beta_9 \text{mean wind speed}_{ij} + \beta_{10} \text{season}_j + \beta_{11} \text{season}_j * \text{day surface temperature}_{ij} + \beta_{12} \text{season}_j * \text{night surface temperature}_{ij} + \varepsilon_{ij},$$

2.3 | Statistical methods

Aqua and Terra readings were combined to form a daily LST variable for daytime and night-time. When one satellite's reading was missing for a given time and place, data from the other was used. When both

where Ta_{ij} is near-ground (2 m) air temperature (minimum, mean, or maximum) on location i on day j ; $(\alpha + \nu_j)$ are the fixed and random intercepts; day and night LST_{ij} are satellite day and satellite night LST; imputed day and night LST_{ij} are indicators of whether

satellite day and satellite night surface temperatures are imputed; $NDVI_{ij}$ is the mean NDVI of Aqua and Terra for grid cell i in month j ; $time_j$ is the time of year calculated as $(\text{day of the year}^{-1})/(\text{total days of the year}^{-1})$; $elevation_i$ is the mean elevation at site i ; and mean wind speed $_{ij}$ is daily mean wind speed from the nearest station. The season was defined as a categorical variable with three levels: cold-dry for November through February, warm-dry for March and April, and rainy for May to October. All continuous variables were centred and scaled before fitting. All analyses were conducted in R 4.0.2 (R Core Team, 2020).

2.3.1 | Assessment of model performance

Within each year, all available stations in the prediction area were randomly split into 10 cross-validation folds. Inside the cross-validation loop, models were trained with all observations (occurring in the year of interest) for stations in the training folds plus all stations outside the prediction area but within the study area (i.e., calibration folds with 90% of the data), and asked to predict the observations for the stations in the test fold (i.e., validation folds with 10% of the data). Mean wind speed for test stations was imputed as if all test stations were missing mean wind speed at all times. The primary measure of model performance was the root-mean-square error (RMSE) of cross-validated predictions. Tabular summaries of model results also report the SD and improvement ($SD - RMSE$) of each outcome (minimum, mean, or maximum T_a) in a particular year, season, or subregion. To evaluate prediction performance more evenly throughout the study region, our summary tables include spatially weighted versions of the SD and RMSE, for which each day and each 16th of a longitude–latitude grid cell (splitting each degree of longitude or latitude into four equal intervals) with at least one observation is given a total weight of 1. These metrics help to evaluate performance when we consider areas with few stations (e.g., the northern parts of the MCM) to be equally important to areas with lots of stations (e.g., Mexico City proper). In addition, we estimate the degree to which our predictions capture spatial and temporal patterns by computing for each T_a outcome y and prediction p , given per-station annual means of the outcome M_y and the predictions M_p :

- R^2 , the proportion of variance accounted for, as $1 - \text{mean}((y - p)^2)/\text{Var}(y)$.
- R^2_{spatial} , the squared correlation between M_y and M_p .
- R^2_{temporal} , the squared correlation between $(y - M_y)$ and $(p - M_p)$.

For summaries on the per-station annual means (R^2_{spatial}), the sample size is the number of stations instead of station-days (Kloog *et al.*, 2014).

To verify that including Weather Underground data did not impair prediction, we conducted variations of the cross-validation procedure in which we excluded all Weather Underground stations from testing. We computed the spatial RMSE from a cross-validation that tests and trains in non-Weather Underground stations, $RMSE_{\text{NWU}}$. Then Weather Underground stations were allowed in training and $RMSE_{\text{WU}}$ was computed in non-Weather Underground stations and finally subtracted from $RMSE_{\text{NWU}}$. Thus, a positive difference in $RMSE_{\text{NWU}} - RMSE_{\text{WU}}$ means an improvement in RMSE when Weather Underground was included in training.

We constructed a learning curve in order to illustrate how our model's predictive accuracy is influenced by the size of its training data. This analysis was conducted for mean daily temperature in 2018. We selected two folds to hold out for testing, while using various subsets of the remaining eight folds for training. The test folds were chosen to have the closest unweighted RMSE under the cross-validation as the overall RMSE for this year and dependent variable. These test folds ended up comprising 16 stations and 4,741 observations. There were seven rounds of analysis and 100 simulation replicates for each round. In each replicate of round 1, 10 stations were randomly selected, 2,500 observations were randomly selected from these 10 stations, and the model was trained on these 2,500 observations and tested on the test folds. Round 2 used 20 stations and 5,000 observations, round 3 used 30 stations and 7,500 observations, and so on up to round 7 with 70 stations and 17,500 observations.

2.3.2 | Estimation of at risk population

We obtained population density information at the AGEb level (equivalent to U.S. census tracts) from the 2010 Mexican Population Census that was carried out by the Instituto Nacional de Estadística y Geografía de México (Instituto Nacional de Estadística y Geografía (INEGI), 2016). AGEb data are available as a polygon layer with population data for all metropolitan areas in the MCM. From these records we calculated the number of at risk days (above 30°C and below 5°C) experienced by people ≥ 65 years old in the MCM during the year 2010.

3 | RESULTS

Approximately one quarter to one third of LST data were missing from both Aqua and Terra satellites and thus

had to be imputed for our temperature model. The proportions of missing LST data for both satellites were similar over time (23–34% missing daytime LST and 27–37% missing night-time LST across all years), except for a higher number of missing data in 2004 (52% daytime, 50% night-time).

For all years and dependent variables (i.e., minimum, mean, and maximum Ta), we observed substantially lower RMSE compared with the *SD*, indicating that our model was effective in predicting temperature. Table 1 presents our model's performance in which we conducted 10-fold cross-validation for mean Ta. Tables S2 and S3 show cross-validation results for minimum and maximum Ta, respectively. Across all results, the range for RMSEs was 0.9–1.9 K, with a mean of 1.5 K, whereas the range for *SD*s was 3.7–5.4 K. Values of out of sample R^2 ranged from 0.78 to 0.95 indicating good predictive abilities in our models, particularly for maximum Ta (R^2 0.80–0.92) and mean Ta (R^2 0.89–0.95), compared to minimum Ta (R^2 0.78–0.87). The spatial and temporal R^2 also showed good performance with average spatial R^2 values of 0.88, 0.94, and 0.90 for minimum, mean, and maximum Ta, respectively, and average temporal R^2 values of 0.80, 0.88, and 0.85 for minimum, mean, and maximum Ta, respectively. The averaged RMSEs for minimum, mean, and maximum Ta were 1.64, 1.14, and 1.58 K, respectively, and the spatially weighted RMSEs, in which

all monitored areas are equally important, were generally worse than the unweighted RMSEs (by 0.33 K on average), but still reasonably small.

Figure 2 provides an example of observations and predictions at two different monitoring stations, in June of 2010 and 2018, showing improvement in 2018 Ta predictions compared to 2010 for the southern region of Morelos as the number of stations in training increased between these years.

Figure 3 shows that the distribution of prediction errors in 2018 was similar by season in the MCM. Of the 23,558 predictions in the figure, 40 (1 in 589) have an error of -5 K or below. The corresponding observations are about evenly distributed by season. They come from 5 stations in the southern Valley of Mexico, 26 are from a single Weather Underground station, and 10 are from a single EMAS station.

We examined model performance by metropolitan area in order to compare how our model performed in areas with fewer stations relative to those with more stations. Table 2 summarizes cross-validated model performance within each metropolitan area of the MCM with >5 stations for 2018. By comparing the RMSEs to the *SD*s of the Ta responses we can see how much more Ta variation is explained with our model. For results for all metropolitan areas, see Table S4. There were large differences in both *SD* and RMSE across municipalities.

TABLE 1 Prediction accuracy for the Megalopolis of Central Mexico: 10-fold cross-validation (CV) results for daily mean Ta predictions from 2003 to 2019

Year	Station-days (<i>N</i>)	Number of stations	<i>SD</i>	RMSE	R^2	<i>SD</i> _{weighted}	RMSE _{weighted}	R^2 _{spatial}	R^2 _{temporal}
2003	9,622	32	3.94	0.92	.95	5.02	1.21	.97	.92
2004	10,453	35	3.80	1.04	.92	5.20	1.37	.93	.89
2005	11,489	36	4.16	1.09	.93	5.55	1.40	.95	.91
2006	10,882	36	3.94	1.11	.92	5.17	1.40	.95	.87
2007	9,854	39	3.95	1.04	.93	5.21	1.29	.94	.87
2008	11,430	41	4.05	1.11	.92	5.52	1.44	.96	.89
2009	13,114	48	4.13	1.21	.91	5.89	1.48	.93	.90
2010	13,980	51	4.50	1.26	.92	6.35	1.71	.95	.91
2011	14,036	46	4.25	1.16	.93	5.84	1.46	.95	.89
2012	15,161	53	3.93	1.06	.93	5.38	1.35	.96	.87
2013	17,317	59	4.21	1.14	.93	5.23	1.32	.96	.86
2014	18,685	62	4.02	1.10	.92	5.21	1.29	.96	.86
2015	20,712	69	3.92	1.09	.92	5.38	1.23	.95	.84
2016	23,716	74	4.18	1.24	.91	5.37	1.33	.94	.88
2017	23,915	80	4.15	1.30	.90	5.45	1.47	.91	.87
2018	23,558	91	3.77	1.26	.89	5.09	1.29	.90	.88
2019	29,093	99	3.68	1.22	.89	5.27	1.31	.92	.85

Note: *SD* and RMSE are in K.

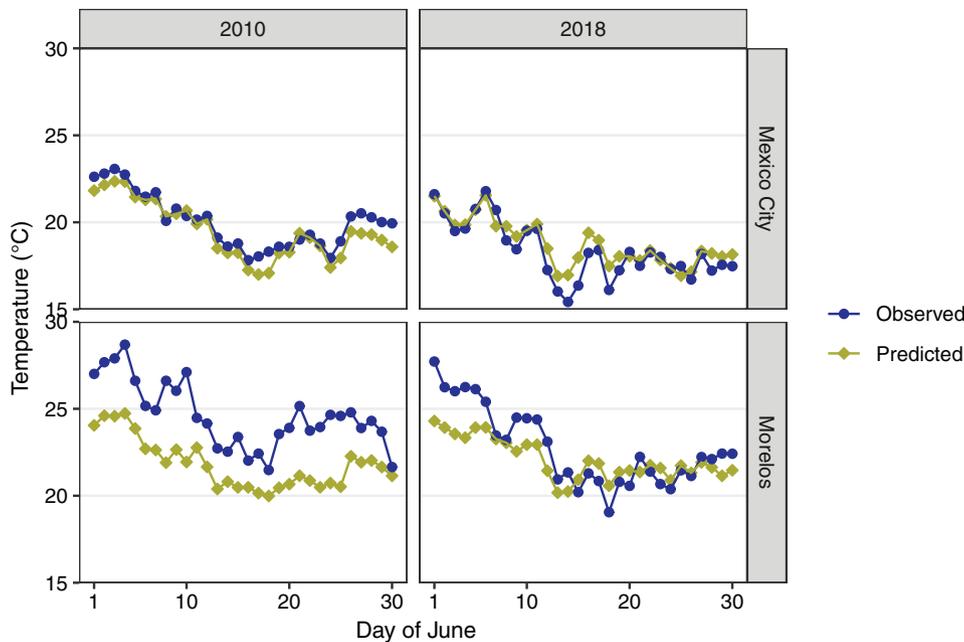


FIGURE 2 Observed and predicted T_a from CV, for station 8 (in Mexico City proper) and station 24 (in the southern region of the study area, in the state of Morelos) in two different years [Colour figure can be viewed at wileyonlinelibrary.com]

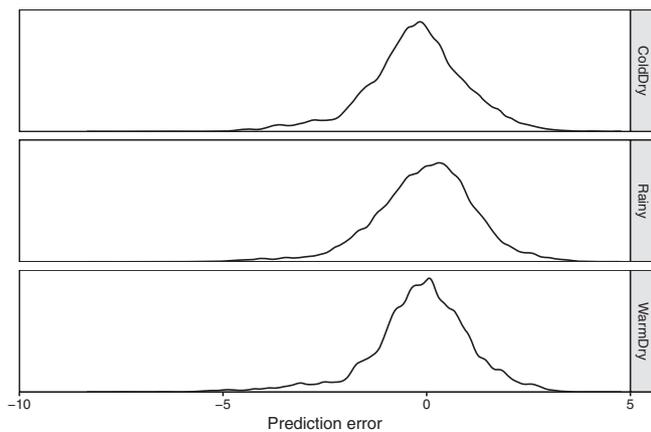


FIGURE 3 Density plots of the CV-predicted T_a minus observed T_a in K for 2018, aggregated across stations but stratified by season

Still, the improvement was at least 1 K in most cases. On average, our models yield better predictions than the naïve estimate of the mean for all T_a outcomes (i.e., our predictions' measure of variation RMSE reduces the randomness in all observed T_a outcomes better than the SD).

We evaluated model performance for each ground monitoring network from which we obtained temperature records. Table 3 shows the accuracy in cross-validated T_a predictions by type of network. The lowest RMSE values were obtained for UNAM records for minimum, mean, and maximum T_a predictions. The highest RMSE values came from ESIMEs for minimum and maximum T_a , and Weather Underground for mean T_a . One reason why UNAM had the lowest RMSEs for all T_a outcomes despite having fewer stations in 2018 compared to

most networks is because UNAM stations were evenly distributed across a smaller and more climatically homogeneous area, and potentially received higher maintenance compared to the other networks.

Model performance was compared by season type: cold-dry, warm-dry, and rainy. Table 4 presents the average cross-validated prediction accuracy for each T_a outcome by season for the entire MCM from 2003 to 2019. The lowest average RMSE values were for mean T_a for all seasons. The highest RMSE values were observed for minimum and maximum T_a during the cold-dry and rainy seasons, respectively, and for mean T_a in the cold-dry season. The highest precision improvements were observed during the cold-dry season for minimum T_a , and during the rainy season for mean and maximum T_a .

Results from our test for impaired prediction from allowing data from Weather Underground stations in training, $RMSE_{WU}$, compared to excluding them from training, $RMSE_{NWU}$, showed that while the mean difference of -0.04 K in $RMSE_{NWU} - RMSE_{WU}$ favours training without Weather Underground stations, there was considerable variation in $RMSE_{NWU} - RMSE_{WU}$ by dependent variable and year. For instance, the inclusion of Weather Underground stations improved the RMSE for minimum T_a for 7 years. Table S5 shows $RMSE_{NWU} - RMSE_{WU}$ for all years and dependent variables in the MCM.

Figure 4 shows the RMSE in two held-out folds that can be achieved when training our model on various subsets of the remaining data. The mean RMSE across simulation replicates decreases rapidly as the training set

TABLE 2 Prediction accuracy by metropolitan area in the Megalopolis of Central Mexico: 10-fold cross-validation (CV) results for Ta predictions for 2018

Metropolitan area	Number of stations	<i>N</i>	Temperature	<i>SD</i>	RMSE	<i>SD</i> – RMSE
Cuernavaca	7	1,717	Minimum	4.21	1.70	2.51
			Mean	4.31	1.22	3.09
			Maximum	4.41	1.93	2.47
Puebla-Tlaxcala	16	3,553	Minimum	3.22	1.61	1.61
			Mean	2.75	1.15	1.61
			Maximum	2.99	1.57	1.42
Mexico City	65	17,092	Minimum	3.56	1.76	1.80
			Mean	3.18	1.27	1.91
			Maximum	3.60	1.45	2.14

Note: *SD*, RMSE, and *SD* – RMSE are in K.

TABLE 3 Prediction accuracy by ground monitoring network in the Megalopolis of Central Mexico for 2018

Ta	Network	Station-days (<i>N</i>)	Number of stations	<i>SD</i>	RMSE	<i>SD</i> – RMSE
Minimum	EMAs	2,740	14	5.66	1.92	3.75
	ESIMES	823	4	3.72	2.47	1.25
	SIMAT	7,787	25	3.34	1.75	1.59
	UNAM	3,187	12	2.76	1.04	1.72
	Weather Underground	8,736	35	3.66	1.89	1.77
Mean	EMAs	2,740	14	6.53	1.50	5.02
	ESIMES	823	4	2.93	1.10	1.82
	SIMAT	7,787	25	2.77	1.04	1.73
	UNAM	3,187	12	2.47	0.72	1.74
	Weather Underground	8,736	35	3.20	1.51	1.68
Maximum	EMAs	2,740	14	7.50	2.10	5.40
	ESIMES	823	4	3.31	2.15	1.16
	SIMAT	7,787	25	3.22	1.26	1.96
	UNAM	3,187	12	2.82	0.94	1.88
	Weather Underground	8,736	35	3.54	1.77	1.76

Note: *SD*, RMSE, and *SD* – RMSE are in K.

TABLE 4 Prediction accuracy by season in the Megalopolis of Central Mexico: Average *SD*, RMSE, and *SD* – RMSE for minimum, mean, and maximum Ta predictions from 2003 to 2019

Ta	Cold dry			Warm dry			Rainy		
	<i>SD</i>	RMSE	<i>SD</i> – RMSE	<i>SD</i>	RMSE	<i>SD</i> – RMSE	<i>SD</i>	RMSE	<i>SD</i> – RMSE
Minimum	3.65	1.82	1.84	3.85	1.76	2.09	3.35	1.47	1.88
Mean	3.70	1.16	2.54	4.08	1.13	2.95	3.66	1.12	2.53
Maximum	4.49	1.54	2.95	4.71	1.56	3.14	4.55	1.60	2.94

Note: *SD*, RMSE, and *SD* – RMSE are in K.

grows from 2,500 to 7,500 observations, then levels off around 1.33 K. This example suggests that our sample is

more than large enough to achieve the best accuracy possible with this model in this region.

Model predictions were made for each 1×1 km grid cell in the entire prediction area for summarization and Figure 5 shows the 95th percentile of the minimum and maximum temperatures across days for each grid cell in 2018. In general, the inhabited areas were hotter than the mountainous borders between municipalities. The southernmost region of the MCM, comprising the metropolitan areas of Cuernavaca and Cuautla, was substantially hotter than the rest of the MCM.

Finally, we used population data to quantify human exposure to extreme ambient temperatures for people aged ≥ 65 years old in the urban AGEBs of the MCM in

2010 (Figure 6a). There were over 51 and 18 million person-days of exposure to extreme low and high temperatures, respectively, in 2010. The highest number of person-days of exposure to daily minimum $T_a \leq 5^\circ\text{C}$ was concentrated in the metropolitan areas of Toluca, MCMA, Puebla-Tlaxcala, and Pachuca (Figure 6b). As for exposure to daily maximum $T_a \geq 30^\circ\text{C}$, Figure 6c shows that the metropolitan areas with a higher number of person-days above this point were Cuernavaca, Cuautla, MCMA, and Puebla-Tlaxcala.

Raw data, processed data, predictions, and code are archived in the Zenodo open-access digital repository (doi: <http://doi.org/10.5281/zenodo.3362523>).

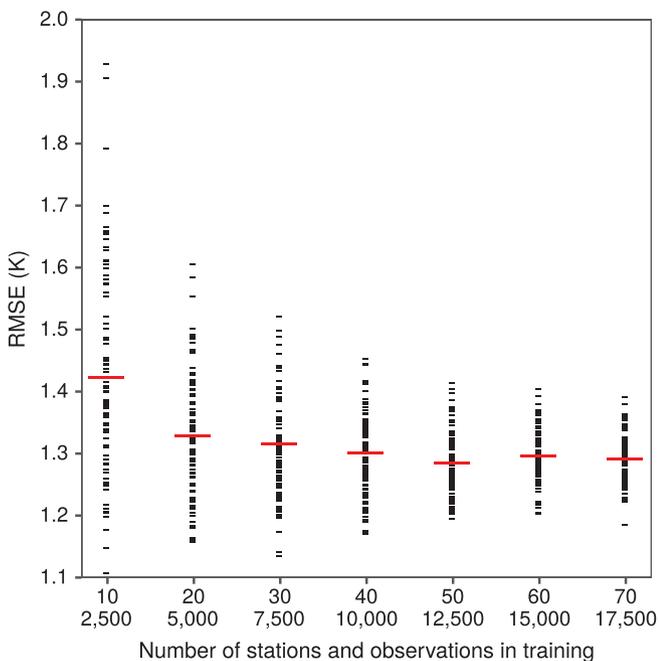


FIGURE 4 Model performance learning curve for minimum T_a in 2018 as a function of the size of its training data. Horizontal bars represent the RMSE mean [Colour figure can be viewed at wileyonlinelibrary.com]

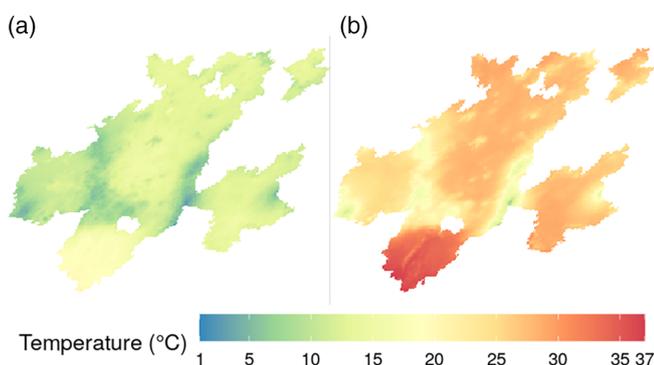


FIGURE 5 Spatial pattern of the 95th percentiles of minimum (a) and maximum (b) temperature across days for each 1 km^2 grid cell in the Megalopolis of Central Mexico for 2018 [Colour figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

In this paper we present the performance of our spatio-temporally resolved hybrid satellite-based land use regression model for predicting T_a across the MCM, the most populated megacity in North America. Our daily predictions of minimum, mean, and maximum T_a are the first highly resolved 1×1 km air temperature estimates in the MCM. Our temperature models generated T_a predictions for an extensive region spanning several large metropolitan areas in central Mexico with diverse geographic characteristics. We addressed multiple data quality issues and incorporated many relevant predictor terms, including LST with daily varying slopes, in our models. Our high-quality T_a predictions for the MCM have potential applications in a variety of settings such as public health, urban planning, and climatology, among others.

Consistent with results of similar studies carried out in other regions with the aim of predicting daily T_a using LST from MODIS (Benali *et al.*, 2012; Kilibarda *et al.*, 2014; Rosenfeld *et al.*, 2017), our model predicted mean T_a more accurately than minimum or maximum T_a . Compared to previous works, the average RMSE in our models (1.64, 1.14, and 1.58 K for minimum, mean, and maximum T_a , respectively) was similar to the RMSE reported by Kloog *et al.*, 2014 for the states from the northeast U.S. region (RMSE ranged from 1.43 to 2.91) for mean T_a , and to the RMSE reported by Shi *et al.* (2016b) (average RMSE of 1.66 K) for mean T_a in the southeast U.S. region. Results from Rosenfeld *et al.*, 2017 for Israel were also similar to ours in terms of RMSE for all T_a outcomes (1.65, 0.97, 1.53 K, for minimum, mean, and maximum T_a , respectively) using LST from the MODIS instrument on the Terra satellite alone and were also similar to the RMSEs (1.65, 1.00, and 1.52 K for minimum, mean, and maximum T_a , respectively) using LST data from MODIS on the Aqua satellite alone. Previous

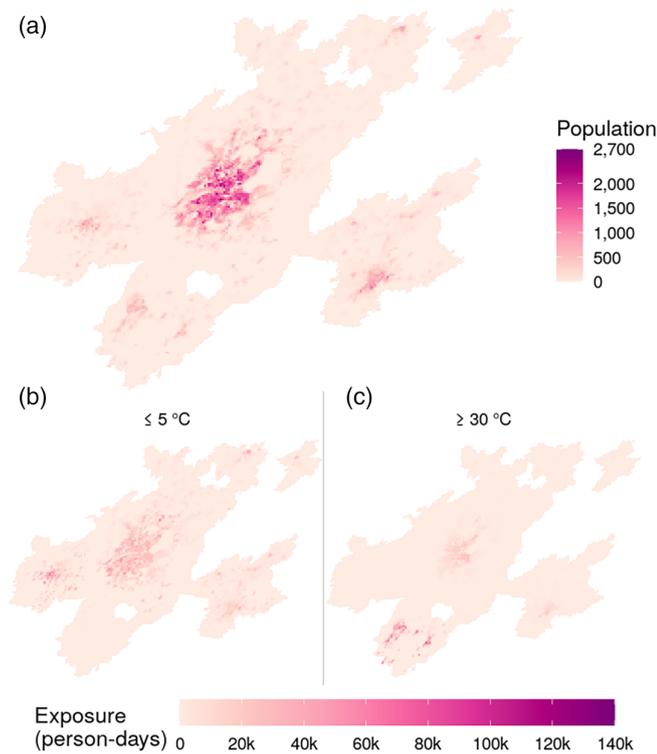


FIGURE 6 Spatial distribution of population density (a), person-days exposure to $\leq 5^{\circ}\text{C}$ (b) and $\geq 30^{\circ}\text{C}$ (c) for each 1 km^2 grid cell in the Megalopolis of Central Mexico for people ≥ 65 years old for 2010 [Colour figure can be viewed at wileyonlinelibrary.com]

studies agree on how often LST data are missing and that the main causes of missing LST records were related to cloud cover, weather conditions, snow, and retrieval errors. Despite the high amount of missing LST data from both Terra and Aqua satellites in 2004 ($\sim 50\%$) in our study region, compared to the rest of the years in the period of study, we did not observe a substantially worse performance in our models for this year (Kloog *et al.*, 2014; Shi *et al.*, 2016b; Rosenfeld *et al.*, 2017). Our spatially weighted RMSEs were higher compared to those reported in previous studies which could be related to a better distribution in ground stations in other study regions (Kloog *et al.*, 2014; Shi *et al.*, 2016b; Rosenfeld *et al.*, 2017). In other words, spatially weighted evaluation in our models' performance suggests worse results for subregions of the MCM with just a few ground monitoring sites such as the metropolitan areas of Toluca and Cuernavaca. In general, our model performance was similar across seasons for all daily Ta predictions.

The utility of daily Ta predictions in epidemiology from models with similar performance (i.e., RMSE) to the ones reported in our study has been demonstrated by Kloog *et al.* (2015) and Lee *et al.* (2016); both were able to assign Ta exposure with less error compared with using a

closest-monitor approach, regardless of the distance between a participant's residence and the closest ground monitor. The lower spatial resolution of Ta from ground stations increased exposure measurement error, to the extent of not finding significant associations.

As shown in Figure 6, within the MCM there are important contrasts in the spatial distribution of the at-risk populations from exposure to extreme Ta, which justify temperature-related epidemiological research in this region. The regions within the MCM with the greatest exposure to low temperatures correspond to the metropolitan areas of Mexico City, Pachuca, Toluca, and Puebla-Tlaxcala. In turn, the areas with the highest concentration of hot temperatures are located in the metropolitan areas of Mexico City, Cuernavaca, Cuautla, and Puebla-Tlaxcala. This spatial distribution of temperatures can be used for planning interventions aimed at mitigating the adverse effects from Ta on the population's health within the MCM.

One of our goals was to use the best available Ta records from ground stations in the MCM, so it is difficult to assess if our model performance is similar across the years. The underlying set of monitors that we used changed substantially with an expansion of ground monitoring from 2003 to 2019 in the MCM. A model primarily designed for long-term climatic trends might be fit to the smaller subset of long-running stations (Oyler *et al.*, 2015). It is also possible that measurement error that is inherent to each monitoring network affected performance in our models. To address this limitation, we applied multiple criteria for quality assessment of weather records from all data sources included in our analyses. However, applying the same criteria to heterogeneous data sources to the same extent can be challenging, especially when including crowdsourced information. Unlike monitoring sites operated under international standards, private stations may not necessarily adhere to the same operation principles, with the possibility of generating data with a quality that is not suitable for research purposes (Bell *et al.*, 2015). In this regard, specific quality control methods for crowdsourced data have been proposed (Droste *et al.*, 2020); most of them have included statistically based steps (Chapman *et al.*, 2017; Napoly *et al.*, 2018) and comparisons to reference networks (Meier *et al.*, 2017). Nonetheless, these methods might not be directly transferred to all locations and crowdsourced datasets, because they are only suitable for places with specific climatic characteristics and specific crowdsourced techniques, may require high-quality reference data from dedicated urban climate observational networks, and are also dependent on records from other physical variables, which are not always available in all weather stations (Meier *et al.*, 2017; Droste *et al.*, 2020). Thus, replication of such methods in different settings such as the MCM may require some adaptations.

Because no reference urban-climatic networks exist within the MCM, implementation of this approach for quality assessment of Weather Underground records in the metropolitan areas that make up the MCM is not possible. Also, the Weather Underground network in our study region includes multiple types of weather stations (see Table S1); then, quality control methods developed for specific types of weather stations such as Netatmo must be adapted to work in our study region. We implemented a quality assessment approach that was not dependent on a dedicated reference network and was not intended to be used just on a specific type of data source. In addition, we proposed and implemented a geostatistical filter to remove stations-days considered as “deviations” by comparing observations from all networks to inverse-distance-weighted interpolations of temperature from neighbouring monitoring sites (excluding Weather Underground), within similar ranges of elevation. In the end, our quality assessment approach filtered out more observations from Weather Underground than from any other network, emphasizing the importance of quality assessment of crowdsourced information. We report that 26 of the 40 errors of -5 K or below in 2018 mean Ta were from a single Weather Underground station which could be caused by the inappropriate microscale siting of that respective site (e.g., next to a wall), resulting in biased measurements from that station which are small enough in magnitude to avoid flagging by our deviance criteria. Overall, the rarity of errors of this magnitude (1-in-589) and the robustness of model results to the exclusion of the Weather Underground network give us confidence that the trade-offs of stringency and data availability are appropriate for our goals.

For estimating daily Ta, our model improves on previously used methods by incorporating satellite data to produce more accurate predictions for any given location within the MCM. Studies conducted in this region that aimed to estimate temperature variations beyond the spatial coverage of meteorological ground stations are scarce and not suited to generate Ta estimates with adequate spatiotemporal resolution (Carrera-Hernández and Gaskin, 2007). The use of LST data from MODIS (alone and limited to cloud-free conditions) in studies performed in subregions of the MCM have sought to examine seasonal variations in the urban heat island (UHI) in Mexico City using 8-day LST averages in 2006 (Cui and de Foy, 2012) and to correlate the abundance of the dengue virus mosquito vector with 28–29-day LST average in Puebla City (Moreno-Madriñán *et al.*, 2014). None of the above studies generated daily predictions of Ta with a potential use in intra-urban assessment of human exposure to extreme temperatures.

Thus far, exposure assessment of Ta in epidemiological research in Mexico has relied heavily on measurements from a sparse network of ground stations that are not always located close to the populations under study (O'Neill *et al.*, 2005; Bell *et al.*, 2008; McMichael *et al.*, 2008; Hurtado-Díaz *et al.*, 2019). The data from different monitoring networks can be challenging to work with because they are often organized heterogeneously (i.e., different file formats, temporal resolutions, time zones, differences and changes in the organization of records over time, and often coding errors), which may explain why only some and not all ground station data in Mexico have been used in previous epidemiological studies. The trade-off is a very coarse spatial resolution, and these studies assigned daily minimum, mean, and maximum Ta exposures at spatial scales ranging from tens to a hundred kilometres.

While ground stations collect vital meteorological information, their locations constrain the utility of their data for epidemiological research particularly in urban areas. When placing ground stations, most networks follow international standards that avoid interference from large buildings, reflective surfaces, and other sources of heat or radiation (Llansó, 2003). This approach may preclude meteorological stations from being located in areas of interest for Ta-related spatial epidemiology. For these reasons, it is important to consider the use of different quality-controlled data sources and data fusion techniques to better reflect Ta variability within large urban areas. The inclusion of spatiotemporal covariates to predict Ta should improve on ground station data alone and would more accurately estimate the population's actual exposures (Pelta and Chudnovsky, 2017; Dirksen *et al.*, 2020).

Biometeorological indices have been alternatively used to assess heat stress in epidemiology research (Basu, 2009). These biometeorological indices include apparent temperature, humidex, heat index, and net effective temperature, among others. The aim of these indices has been to reflect the actual ambient heat perceived by humans as a function of air temperature and different metrics of humidity (McGregor *et al.*, 2015). However, such metrics are also more difficult to map at highly resolved spatial scales from remote sensing compared to LST. Also, the relationship between LST and biometeorological indices is complex because of the influence of factors such as wind speed and wind direction in the mass of air above the surface, land surface heat capacity, and near-surface water available for evapotranspiration, and these data are not always available in all weather networks (Ho *et al.*, 2016).

Although our temperature model offers robust predictions compared to more commonly used citywide Ta

exposure estimates, our model has limitations related to its spatial and temporal resolution. It is possible that the 1×1 km grid cell size in our model may be too coarse of a spatial resolution for capturing biologically relevant exposures for specific types of health conditions, in steep areas with large changes in elevation, and also for exposure assessment in the different microenvironments where individuals spend most of their time. There is evidence that spatial resolutions lower than 50 m may underestimate the UHI (Sobrino *et al.*, 2012), a phenomena that has been linked to different mortality and morbidity outcomes (Heaviside *et al.*, 2017). Also, Ta in urban areas can change rapidly in short distances given the influence of specific city characteristics such as proximity to green areas and water bodies, surface albedo, sky view factor, and construction materials, which can induce significant variations in the actual temperature experienced by individuals within 1×1 km grid cells in large metropolitan areas (Schinasi *et al.*, 2018). On the other hand, there is limited availability of remote sensing products with finer spatially resolved temperature data, such as the satellite data derived from USGS Landsat satellites. However, while Landsat has a higher spatial resolution, its temporal resolution is limited: once every 16 days under cloud-free conditions (Tomlinson *et al.*, 2011; Ho *et al.*, 2014). Additionally, our daily temporal resolution may lead to exposure misclassification for specific acute health outcomes occurring at finer time scales. For instance, Rowland *et al.* (2020) assessed the impact of hourly temperature predictions on myocardial infarctions incidence in New York State, finding critical windows of exposure in the hours leading up to the onset of the event (Rowland *et al.*, 2020). With the development of computational resources, numerical models like the Weather Research and Forecasting (WRF) model offer an alternative for capturing high-resolution characteristics of urban climate and have been successful at simulating 2 m Ta and extreme Ta episodes such as UHIs (Yang *et al.*, 2012; Jandaghian *et al.*, 2018; Li *et al.*, 2019). High-resolution numerical models for studying urban climate at subkilometre scales are now possible (Jandaghian and Berardi, 2020). Recently, numerical models have been successfully employed for human exposure assessment to extreme temperatures and the projection of climate change scenarios in public health research (Ha *et al.*, 2017; 2018; José *et al.*, 2017; Lou *et al.*, 2019). Potential limitations documented for these models are related to the available options in the WRF solver, exclusion of anthropogenic heat emissions leading to underestimation in Ta, the extensive analysis of data needed for daily long-term simulations over years (including land use and/or urban canopy data), and comparison of area-averaged data from simulations to observational data

(point Ta observations) from reference networks (Yang *et al.*, 2012; Jandaghian *et al.*, 2018).

There are also potential limitations pertaining to the replicability of our methods in other places and to the comparability of our results to other studies using MODIS products to predict Ta. Previous studies in other regions have utilized different modelling techniques, have climates that are not similar to the MCM, and they have different data availability and relevance of spatial predictors such as detailed land use predictors, road density, and distance from water bodies (Kloog *et al.*, 2014; Rosenfeld *et al.*, 2017). Nonetheless, similar modelling approaches can generally be adopted elsewhere to improve existing Ta estimates based on limited data sources (e.g., ground stations or satellite data only).

We expect that the three different daily Ta predictions from our model can be used to improve results from prior epidemiologic research on well-defined Ta-related pathologies in the MCM such as cardio-respiratory diseases and emergent vector-borne diseases (dengue, zika, and chikungunya). They can also be used to explore new hypotheses linking Ta with birth outcomes, cardiometabolic outcomes, cognitive function, and mental health disorders (Kloog *et al.*, 2015; Dai *et al.*, 2016; Wallwork *et al.*, 2017; Zanobetti *et al.*, 2017). Our Ta predictions will facilitate investigations of the health impacts from UHI in the different metropolitan areas of the MCM and have the potential to be harnessed for early warning systems, which are increasingly important as we experience climate changes and extreme variations in temperature. Our model may assist decision-makers in public health and meteorology in the MCM to design interventions aimed to reduce population health risks from exposure to dangerous temperature levels. While it is not a forecast system, our temperature model has the capacity to augment existing early warning systems by pinpointing more hotspots across the entire MCM and aiding the people who live there. Our exposure maps on people aged ≥ 65 years old exposed to extreme cold ($\leq 5^\circ\text{C}$) and hot ($\geq 30^\circ\text{C}$) temperatures support the idea that alert systems should consider differences in local meteorology, demographics, and urban structure to address health impacts from extreme temperatures in a megalopolitan context (McGregor *et al.*, 2015).

Finally, to increase the utility of our prediction model and adhere to best practices for open and transparent research Wilson *et al.*, 2017, we have archived the raw and cleaned station data, our temperature predictions, metadata, and all of the R code used in this project in a citable and open research repository (doi: <http://doi.org/10.5281/zenodo.3362523>). This allows others studying different parts of the world to reproduce, modify, and build on our work for the advancement of exposure science.

5 | CONCLUSION

Evidence about the intra-urban health effects from extreme temperatures are much needed in the Megalopolis of Central Mexico (MCM), which includes the Mexico City Metropolitan Area and several neighbouring major metropolitan areas. The aim of our study was to generate daily predictions of T_a with a spatial resolution of 1×1 km from 2003 to 2019, with application in public health in this region. For this we used T_a records from ground monitors and LST data from the MODIS instrument on the Terra and Aqua satellites as our main predictor of daily T_a . We calibrated LST to T_a using mixed-effect models, land use predictors, and separate slopes for each day. Performance of our T_a models showed that daily minimum, mean, and maximum T_a can be reliably predicted using daily LST data even across the heterogeneous geography of the MCM with high accuracy. To illustrate the utility of our T_a models for public health, we calculated the number of days in the year 2010 that people ≥ 65 years old were at risk of exposure to extremely low or high temperatures. We estimated over 51 million person-days of exposure to extreme cold and 18 million person-days of exposure to extreme heat, and these exposures were concentrated in particular metropolitan areas where critical public health efforts may be most needed. Our findings reveal the potential for our daily T_a predictions to improve epidemiological research in this region.

ACKNOWLEDGEMENTS

Data used in the preparation of this article were obtained from the Programa de Estaciones Meteorológicas del Bachillerato Universitario (PEMBU) database (www.ruoa.unam.mx/pembu). The PEMBU project is led by Prof. Michel Grutter de la Mora from the Centro de Ciencias de la Atmósfera at the Universidad Nacional Autónoma de México (CCA-UNAM). We also obtained historical meteorological records from the Coordinación General del Servicio Meteorológico Nacional (CGSMN) and advice from Adolfo Portocarrero Reséndiz, Project Manager.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

Raw and processed data, including daily temperature predictions for every grid cell, as well as a research notebook with reproducible R code and instructions, are archived in Zenodo at: <http://doi.org/10.5281/zenodo.3362523>. Other investigators may use the raw data, cleaned meteorological data, or predictions directly, or they may rerun or modify the data-processing and analysis code.

ORCID

Iván Gutiérrez-Avila  <https://orcid.org/0000-0001-8040-2007>

REFERENCES

- Aguilar, E., Auer, I., Brunet, M., Peterson, T.C. and Wieringa, J. (2003) Guidance on metadata and homogenization. *WMO TD*, 1186, 1–53.
- Armstrong, B.G. (1998) Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occupational and Environmental Medicine*, 55(10), 651–656. <https://doi.org/10.1136/oem.55.10.651>.
- Basu, R. (2009) High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environmental Health*, 8, 40. <https://doi.org/10.1186/1476-069X-8-40>.
- Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <http://doi.org/10.18637/jss.v067.i01>.
- Behnke, R., Vavrus, S., Allstadt, A., Albright, T., Thogmartin, W.E. and Radeloff, V.C. (2016) Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecological Applications*, 26(5), 1338–1351. <http://doi.org/10.1002/15-1061>.
- Bell, M.L., O'Neill, M.S., Ranjit, N., Borja-Aburto, V.H., Cifuentes, L. A. and Gouveia, N.C. (2008) Vulnerability to heat-related mortality in Latin America: a case-crossover study in São Paulo, Brazil, Santiago, Chile and Mexico City, Mexico. *International Journal of Epidemiology*, 37(4), 796–804. <http://doi.org/10.1093/ije/dyn094>.
- Bell, S., Cornford, D. and Bastin, L. (2015) How good are citizen weather stations? Addressing a biased opinion. *Weather*, 70, 75–84. <https://doi.org/10.1002/wea.2316>.
- 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. <http://doi.org/10.1016/j.rse.2012.04.024>.
- Carrera-Hernández, J.J. and Gaskin, S.J. (2007) Spatio temporal analysis of daily precipitation and temperature in the basin of Mexico. *Journal of Hydrology*, 336(3), 231–249. <https://doi.org/10.1016/j.jhydrol.2006.12.021>.
- Chapman, L., Bell, C. and Bell, S. (2017) Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Net-atmo weather stations. *International Journal of Climatology*, 37(9), 3597–3605. <https://doi.org/10.1002/joc.4940>.
- Cui, Y.Y. and de Foy, B. (2012) Seasonal variations of the urban heat island at the surface and the near-surface and reductions due to urban vegetation in Mexico City. *Journal of Applied Meteorology and Climatology*, 51(5), 855–868. <https://doi.org/10.1175/JAMC-D-11-0104.1>.
- Dai, L., Kloog, I., Coull, B.A., Sparrow, D., Spiro, A., Vokonas, P.S. and Schwartz, J.D. (2016) Cognitive function and short-term exposure to residential air temperature: A repeated measures study based on spatiotemporal estimates of temperature. *Environmental Research*, 150, 446–451. <http://doi.org/10.1016/j.envres.2016.06.036>.
- Dirksen, M., Knap, W.H., Steeneveld, G., Holtslag, A.A.M. and Klein Tank, A.M.G. (2020) Downscaling daily air-temperature measurements in the Netherlands. *Theoretical and Applied Climatology*, 142(1-2), 751–767. <http://doi.org/10.1007/s00704-020-03313-1>.

- Donat, M.G., Sillmann, J., Wild, S., Alexander, L.V., Lippmann, T. and Zwiers, F.W. (2014) Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, 27(13), 5019–5035. <http://doi.org/10.1175/jcli-d-13-00405.1>.
- Droste, A.M., Heusinkveld, B.G., Fenner, D. and Steeneveld, G. (2020) Assessing the potential and application of crowdsourced urban wind data. *Quarterly Journal of the Royal Meteorological Society*, 146(731), 2671–2688. <http://doi.org/10.1002/qj.3811>.
- Ha, S., Liu, D., Zhu, Y., Kim, S.S., Sherman, S. and Mendola, P. (2017) Ambient temperature and early delivery of singleton pregnancies. *Environmental Health Perspectives*, 125(3), 453–459. <http://doi.org/10.1289/ehp97>.
- Ha, S., Liu, D., Zhu, Y., Sherman, S. and Mendola, P. (2018) Acute associations between outdoor temperature and premature rupture of membranes. *Epidemiology*, 29(2), 175–182. <http://doi.org/10.1097/ede.0000000000000779>.
- Heaviside, C., Macintyre, H. and Vardoulakis, S. (2017) The urban heat island: implications for health in a changing environment. *Current Environmental Health Reports*, 4(3), 296–305. <https://doi.org/10.1007/s40572-017-0150-3>.
- Ho, H.C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M. and Henderson, S.B. (2014) Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, 154, 38–45. <http://doi.org/10.1016/j.rse.2014.08.012>.
- Ho, H.C., Knudby, A., Xu, Y., Hodul, M. and Aminipouri, M. (2016) A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area. *Science of The Total Environment*, 544, 929–938. <http://doi.org/10.1016/j.scitotenv.2015.12.021>.
- Hofstra, N., Haylock, M., New, M., Jones, P. and Frei, C. (2008) Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research*, 113 (D21), D05109. <http://doi.org/10.1029/2008jd010100>.
- Hulley, G.C., Ghent, D., Götsche, F.M., Guillevic, P.C., Mildrexler, D.J. and Coll, C. (2019) Land surface temperature. In: Hulley, G.C. and Ghent, D. (Eds.) *Taking the Temperature of the Earth*. Amsterdam, Netherlands: Elsevier, pp. 57–127. <https://doi.org/10.1016/B978-0-12-814458-9.00003-4>.
- Hurtado-Díaz, M., Cruz, J.C., Texcalac-Sangrador, J.L., Félix-Arellano, E.E., Gutiérrez-Avila, I., Briseño-Pérez, A.A., Saavedra-Lara, N., Tobías, A. and Riojas-Rodríguez, H. (2019) Short-term effects of ambient temperature on non-external and cardiovascular mortality among older adults of metropolitan areas of Mexico. *International Journal of Biometeorology*, 63 (12), 1641–1650. <http://doi.org/10.1007/s00484-019-01778-y>.
- Instituto Nacional de Estadística y Geografía (INEGI). (2008) Conjunto de datos vectoriales escala 1:1 000 000. Unidades climáticas. Available at: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825267568> [Accessed 25th January 2021].
- Instituto Nacional de Estadística y Geografía (INEGI). (2016) Censo de Población y Vivienda 2010. Available at: <https://www.inegi.org.mx/programas/ccpv/2010/> [Accessed 15th July 2019].
- Instituto Nacional de Estadística y Geografía (INEGI). (2017) Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación. Escala 1:250 000. Serie VI. Conjunto Nacional Información de Interés Nacional. Available at: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463598459> [Accessed 24th January 2021].
- Jandaghian, Z. and Berardi, U. (2020) Comparing urban canopy models for microclimate simulations in weather research and forecasting models. *Sustainable Cities and Society*, 55, 102025. <https://doi.org/10.1016/j.scs.2020.102025>.
- Jandaghian, Z., Touchaei, A.G. and Akbari, H. (2018) Sensitivity analysis of physical parameterizations in WRF for urban climate simulations and heat island mitigation in Montreal. *Urban Climate*, 24, 577–599. <http://doi.org/10.1016/j.uclim.2017.10.004>.
- San José, R., Pérez, J.L., Pérez, L. and Gonzalez Barras, R.M. (2018) Effects of climate change on the health of citizens modelling urban weather and air pollution. *Energy*, 165, 53–62. <http://doi.org/10.1016/j.energy.2018.09.088>.
- 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. <http://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. <http://doi.org/10.1016/j.scitotenv.2012.05.095>.
- 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. <http://doi.org/10.1016/j.rse.2014.04.024>.
- 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. <http://doi.org/10.1289/ehp.1308075>.
- Lee, M., Shi, L., Zanobetti, A. and Schwartz, J.D. (2016) Study on the association between ambient temperature and mortality using spatially resolved exposure data. *Environmental Research*, 151, 610–617. <http://doi.org/10.1016/j.envres.2016.08.029>.
- Li, H., Zhou, Y., Wang, X., Zhou, X., Zhang, H. and Sodoudi, S. (2019) Quantifying urban heat island intensity and its physical mechanism using WRF/UCM. *Science of The Total Environment*, 650, 3110–3119. <http://doi.org/10.1016/j.scitotenv.2018.10.025>.
- Lou, J., Wu, Y., Liu, P., Kota, S.H. and Huang, L. (2019) Health effects of climate change through temperature and air pollution. *Current Pollution Reports*, 5(3), 144–158. <http://doi.org/10.1007/s40726-019-00112-9>.
- McGregor, G.R., Bessemoulin, P., Ebi, K. and Menne, B. (2015) *Heatwaves and Health: Guidance on Warning-System Development*. Geneva: World Meteorological Organization and World Health Organization Available at: https://www.who.int/globalchange/publications/WMO_WHO_Heat_Health_Guidance_2015.pdf?ua=1. [Accessed 25th January 2021].

- McMichael, A.J., Wilkinson, P., Kovats, R.S., Pattenden, S., Hajat, S., Armstrong, B., Vajanapoom, N., Niciu, E.M., Mahomed, H., Kingkeow, C., Kosnik, M., O'Neill, M.S., Romieu, I., Ramirez-Aguilar, M., Barreto, M.L., Gouveia, N. and Nikiforov, B. (2008) International study of temperature, heat and urban mortality: the 'ISOTHURM' project. *International Journal of Epidemiology*, 37(5), 1121–1131. <http://doi.org/10.1093/ije/dyn086>.
- Meier, F., Fenner, D., Grassmann, T., Otto, M. and Scherer, D. (2017) Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19, 170–191. <http://doi.org/10.1016/j.uclim.2017.01.006>.
- Moreno-Madriñán, M., Crosson, W., Eisen, L., Estes, S., Estes Jr. M., Hayden, M., Hemmings, S., Irwin, D., Lozano-Fuentes, S., Monaghan, A., Quattrochi, D., Welsh-Rodriguez, C. and Zielinski-Gutierrez, E. (2014) Correlating Remote Sensing Data with the Abundance of Pupae of the Dengue Virus Mosquito Vector, *Aedes aegypti*, in Central Mexico. *ISPRS International Journal of Geo-Information*, 3(2), 732–749. <http://doi.org/10.3390/ijgi3020732>.
- Mostofsky, E., Wilker, E.H., Schwartz, J., Zanobetti, A., Gold, D.R., Wellenius, G.A. and Mittleman, M.A. (2014) Short-Term changes in ambient temperature and risk of ischemic stroke. *Cerebrovascular Diseases Extra*, 4(1), 9–18. <http://doi.org/10.1159/000357352>.
- Muller, C.L., Chapman, L., Grimmond, C.S.B., Young, D.T. and Cai, X. (2013) Sensors and the city: a review of urban meteorological networks. *International Journal of Climatology*, 33(7), 1585–1600. <http://doi.org/10.1002/joc.3678>.
- Napoly, A., Grassmann, T., Meier, F. and Fenner, D. (2018) Development and application of a statistically-based quality control for crowdsourced air temperature data. *Frontiers in Earth Science*, 6 <http://doi.org/10.3389/feart.2018.00118>.
- O'Neill, M.S., Hajat, S., Zanobetti, A., Ramirez-Aguilar, M. and Schwartz, J. (2005) Impact of control for air pollution and respiratory epidemics on the estimated associations of temperature and daily mortality. *International Journal of Biometeorology*, 50(2), 121–129. <http://doi.org/10.1007/s00484-005-0269-z>.
- 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. <http://doi.org/10.1002/joc.4127>.
- Oyler, J.W., Dobrowski, S.Z., Holden, Z.A. and Running, S.W. (2016) Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. *Journal of Applied Meteorology and Climatology*, 55(7), 1441–1457. <http://doi.org/10.1175/jamc-d-15-0276.1>.
- Pelta, R. and Chudnovsky, A.A. (2017) Spatiotemporal estimation of air temperature patterns at the street level using high resolution satellite imagery. *Science of the Total Environment*, 579, 675–684. <https://doi.org/10.1016/j.scitotenv.2016.11.042>.
- Phung, D., Thai, P.K., Guo, Y., Morawska, L., Rutherford, S. and Chu, C. (2016) Ambient temperature and risk of cardiovascular hospitalization: An updated systematic review and meta-analysis. *Science of The Total Environment*, 550, 1084–1102. <http://doi.org/10.1016/j.scitotenv.2016.01.154>.
- Plummer, N., Allsopp, T., Lopez, J.A. and Llansó, P. (2003) *Guidelines on climate observation networks and systems*. Geneva: World Meteorological Organization WMO/TD No. 1185, WCDMP No. 52. https://library.wmo.int/index.php?lvl=notice_display&id=11634#.YDarEWhKhPY.
- R Core Team. (2020) *R: a language and environment for statistical computing*. Vienna: The R Foundation. Available at: <https://www.R-project.org>. [Accessed 25th January 2021].
- Ren, C., O'Neill, M.S., Park, S.K., Sparrow, D., Vokonas, P. and Schwartz, J. (2011) Ambient temperature, air pollution, and heart rate variability in an aging population. *American Journal of Epidemiology*, 173(9), 1013–1021. <http://doi.org/10.1093/aje/kwq477>.
- 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, 297–312. <http://doi.org/10.1016/j.envres.2017.08.017>.
- Rowland, S.T., Boehme, A.K., Rush, J., Just, A.C. and Kioumourtzoglou, M. (2020) Can ultra short-term changes in ambient temperature trigger myocardial infarction?. *Environment International*, 143, 105910. <http://doi.org/10.1016/j.envint.2020.105910>.
- Schinasi, L.H., Benmarhnia, T. and De Roos, A.J. (2018) Modification of the association between high ambient temperature and health by urban microclimate indicators: a systematic review and meta-analysis. *Environmental Research*, 161, 168–180. <https://doi.org/10.1016/j.envres.2017.11.004>.
- Šećerov, I.B., Savić, S.M., Milošević, D.D., Arsenović, D.M., Dolinaj, D.M. and Popov, S.B. (2019) Progressing urban climate research using a high-density monitoring network system. *Environmental Monitoring and Assessment*, 191(2), <http://doi.org/10.1007/s10661-019-7210-0>.
- Shi, L., Liu, P., Kloog, I., Lee, M., Kosheleva, A. and Schwartz, J. (2016b) Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study. *Environmental Research*, 146, 51–58. <http://doi.org/10.1016/j.envres.2015.12.006>.
- Shi, L., Liu, P., Wang, Y., Zanobetti, A., Kosheleva, A., Koutrakis, P. and Schwartz, J. (2016a) Chronic effects of temperature on mortality in the Southeastern USA using satellite-based exposure metrics. *Scientific Reports*, 6(1), <http://doi.org/10.1038/srep30161>.
- 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. <http://doi.org/10.1038/nclimate2704>.
- Sobrino, J.A., Oltra-Carrió, R., Sòria, G., Bianchi, R. and Paganini, M. (2012) Impact of spatial resolution and satellite overpass time on evaluation of the surface urban heat island effects. *Remote Sensing of Environment*, 117, 50–56. <http://doi.org/10.1016/j.rse.2011.04.042>.
- Tatem, A.J., Goetz, S.J. and Hay, S.I. (2004) Terra and Aqua: new data for epidemiology and public health. *International Journal of Applied Earth Observation and Geoinformation*, 6(1), 33–46. <https://doi.org/10.1016/j.jag.2004.07.001>.
- Tomlinson, C.J., Chapman, L., Thornes, J.E. and Baker, C. (2011) Remote sensing land surface temperature for meteorology and climatology: a review. *Meteorological Applications*, 18(3), 296–306. <http://doi.org/10.1002/met.287>.

- USAID. (2014) Actualización al diagnóstico de la Megalópolis del centro de México. Mexico Low Emissions Development Program (MLED). Available at: <http://www.plataformaeds.org/productos-programa-mled.php> [Accessed 9th June 2019].
- Venter, Z.S., Brousse, O., Esau, I. and Meier, F. (2020) Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. *Remote Sensing of Environment*, 242, 111791. <http://doi.org/10.1016/j.rse.2020.111791>.
- Wallwork, R.S., Colicino, E., Zhong, J., Kloog, I., Coull, B.A., Vokonas, P., Schwartz, J.D. and Baccarelli, A.A. (2017) Ambient fine particulate matter, outdoor temperature, and risk of metabolic syndrome. *American Journal of Epidemiology*, 185(1), 30–39. <http://doi.org/10.1093/aje/kww157>.
- Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L. and Teal, T.K. (2017) Good enough practices in scientific computing. *PLOS Computational Biology*, 13(6), e1005510. <http://doi.org/10.1371/journal.pcbi.1005510>.
- Xu, Z., Liu, Y., Ma, Z., Li, S., Hu, W. and Tong, S. (2014) Impact of temperature on childhood pneumonia estimated from satellite remote sensing. *Environmental Research*, 132, 334–341. <http://doi.org/10.1016/j.envres.2014.04.021>.
- Yang, B., Zhang, Y. and Qian, Y. (2012) Simulation of urban climate with high-resolution WRF model: a case study in Nanjing, China. *Asia-Pacific Journal of Atmospheric Sciences*, 48, 227–241. <https://doi.org/10.1007/s13143-012-0023-5>.
- Zanobetti, A. and O'Neill, M.S. (2018) Longer-term outdoor temperatures and health effects: a review. *Current Epidemiology Reports*, 5(2), 125–139. <https://doi.org/10.1007/s40471-018-0150-3>.
- Zanobetti, A. and Schwartz, J. (2008) Temperature and mortality in nine US cities. *Epidemiology*, 19(4), 563–570. <https://doi.org/10.1097/EDE.0b013e31816d652d>.
- Zanobetti, A., Coull, B.A., Kloog, I., Sparrow, D., Vokonas, P.S., Gold, D.R. and Schwartz, J. (2017) Fine-scale spatial and temporal variation in temperature and arrhythmia episodes in the VA Normative Aging Study. *Journal of the Air & Waste Management Association*, 67(1), 96–104. <http://doi.org/10.1080/10962247.2016.1252808>.
- 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. <http://doi.org/10.1289/ehp.00108419>.
- Zhang, Y., Yu, C. and Wang, L. (2017) Temperature exposure during pregnancy and birth outcomes: an updated systematic review of epidemiological evidence. *Environmental Pollution*, 225, 700–712. <https://doi.org/10.1016/j.envpol.2017.02.066>.

SUPPORTING INFORMATION

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

How to cite this article: Gutiérrez-Avila I, Arfer KB, Wong S, Rush J, Kloog I, Just AC. A spatiotemporal reconstruction of daily ambient temperature using satellite data in the Megalopolis of Central Mexico from 2003 to 2019. *Int J Climatol*. 2021;41:4095–4111. <https://doi.org/10.1002/joc.7060>