

Journal of Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Ground-truth of a 1-km downscaled NLDAS air temperature product using the New York City Community Air Survey

Heather Eliezer
Sarah Johnson
William L. Crosson
Mohammad Z. Al-Hamdan
Tabassum Z. Insaf

SPIE.

Heather Eliezer, Sarah Johnson, William L. Crosson, Mohammad Z. Al-Hamdan, Tabassum Z. Insaf,
“Ground-truth of a 1-km downscaled NLDAS air temperature product using the New York City
Community Air Survey,” *J. Appl. Remote Sens.* **13**(2), 024516 (2019),
doi: 10.1117/1.JRS.13.024516.

Ground-truth of a 1-km downscaled NLDAS air temperature product using the New York City Community Air Survey

Heather Eliezer,^a Sarah Johnson,^{a,*} William L. Crosson,^b
Mohammad Z. Al-Hamdan,^b and Tabassum Z. Insaf^c

^aBureau of Environmental Surveillance and Policy, New York City Department of Health and Mental Hygiene, New York, New York, United States

^bUniversities Space Research Association, NASA Marshall Space Flight Center, Huntsville, Alabama, United States

^cCenter for Environmental Health, New York State Department of Health, New York and School of Public Health, Rensselaer, New York, United States

Abstract. Ground-truthing results are presented for a new 1-km air temperature product downscaled for New York City (NYC) from ~12 km North American Land Data Assimilation System (NLDAS) air temperature data using 1 km moderate resolution imaging spectroradiometer surface temperature data. The downscaled product was compared against a unique highly spatially resolved ground-level ambient air temperature dataset collected through the New York City Community Air Survey (NYCCAS), a neighborhood level air pollution and temperature monitoring network, for the years 2009 and 2010. This work focuses on the spatial variation in daily minimum temperatures within the five counties that comprise NYC (~784 km²). Overall, the downscaled daily minimum temperature was well correlated with ground station data, with NYCCAS minimum temperatures being slightly higher. Minimum temperature R^2 values were 0.9 and 0.92, and mean absolute errors were 0.69°C and 0.86°C for years 2009 and 2010, respectively. The smallest differences between NYCCAS and the downscaled data were seen at lower temperatures, in less densely urbanized areas, and in areas with higher vegetative cover, suggesting systematic bias in the downscaled data related to land-use. The 1-km dataset discerned neighborhood level temperature differences in high-density urban situations with heterogeneous land cover. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.13.024516](https://doi.org/10.1117/1.JRS.13.024516)]

Keywords: urban air temperature; down-scaling; temperature model; North American Land Data Assimilation System; New York City Community Air Survey; ground-truth.

Paper 190008 received Jan. 4, 2019; accepted for publication Apr. 16, 2019; published online May 13, 2019.

1 Introduction

As the frequency of extreme heat days has risen with the increase of anthropogenic global warming in recent years^{1,2} so has the availability of weather data with higher spatial and temporal resolutions. Making this rich data set available and relevant to the decisions that public health and policy professionals need to make is important for the development of interventions to reduce the health impacts and the frequency of extreme heat.³ In addition, the mapping of the physical determinants of heat risk allows for resources to be allocated efficiently and can be used for population-level assessment of human exposures for health studies.

The North American Land Data Assimilation System (NLDAS) derived from the North American Regional Reanalysis (NARR) datasets⁴ produces hourly meteorological forcing variables, such as air temperature, surface shortwave and longwave incident radiation, and other heat-related data at a 1/8 deg (~12 km) spatial resolution. Ground-truthing, i.e., validating modeled temperature data with on-the-ground temperature measurements, of the meteorological

*Address all correspondence to Sarah Johnson, E-mail: sjohnso5@health.nyc.gov

data products is a critical component in the development of accurate exposure metrics for health studies. Prior work in the validation of NLDAS components has relied on the Oklahoma Mesonet and the Atmospheric Radiation Measurement Program (ARM) as data sources for comparison. Luo et al.⁵ found that air temperature at 2 m from model simulations forced with NLDAS data compared well with those forced with station observations. Another retrospective forcing study of NLDAS radiation and precipitation data⁶ found the best results for radiation measures. Robock et al.⁶ found good agreement for soil temperature, but differences in surface energy partitioning between the model and station observed values. Properly estimating energy partitioning between latent and sensible heat is an ongoing challenge in many of these studies and may explain discrepancies in this study's outcomes as well. Energy absorbed at the surface and released as sensible heat will heat the surface and the air near the surface, while energy used to evaporate water will cool the same surface. The amount of surface cooling is related to both the vegetation coverage of the land surface and soil moisture content, both of which vary significantly across a coastal urban setting such as New York City (NYC).

Although performing well in ground-truthing analyses, one drawback of the NLDAS datasets is that the spatial resolution (~12 km) may not be appropriate for describing exposures in areas with large variation in land cover, such as complex coastlines, steep slopes, or urban areas. In order to address this gap, our project created a downscaled 1-km resolution air temperature dataset. Prior ground-truthing has not been done for a downscaled NLDAS dataset or for air temperature in a highly variable urban landscape such as NYC.

The New York City Community Air Survey (NYCCAS) has the most comprehensive geographic coverage of any urban air monitoring network in the United States, with up to 150 monitors in a 790-km² area. While developed to capture spatial variation in air pollution, it also records ground-level air temperature and relative humidity. Using NYCCAS' highly spatially resolved data, we assessed the extent to which downscaled NLDAS data products capture the intraurban patterns in warm season temperature and identified adjustments to the downscaling process that improved the accuracy of the data product.

Once the downscaled data products have been finalized, they will be incorporated into a collaborative heat stress vulnerability study involving the National Aeronautics and Space Administration (NASA), the New York State Department of Health (NYSDOH), and other agencies. The fine spatial scale of the final dataset will fill in spatial gaps in ground-based temperature sources. The accuracy of the 1-km downscaled data in highly variable landscapes, such as cities, will be evaluated and improved through the ground-truthing process in NYC. The final nationwide dataset will be disseminated via the Centers for Disease Control and Prevention (CDC) Environmental Public Health Tracking Network to be used as an alternative to ground-based or 12-km NLDAS surface temperature data in vulnerability mapping or potentially for the calibration or validation of other models.

2 Data Sources and Methods

2.1 New York City Community Air Survey: Air Temperature Records

NYCCAS is a monitoring dataset that was developed primarily for the purposes of tracking street-level air pollutant distribution across the city. The spatial designation of the sensors was optimized to capture intraurban spatial variation in fine particulate pollutant concentrations from local sources in NYC. More details of the site selection process and monitoring strategy can be found in Ref. 7. In order to compute the temperature-corrected flow volume of the sensor air pump, temperature and relative humidity were also recorded, thereby providing more spatially resolved temperature data than the three National Weather Service locations in NYC. Most NYCCAS sites were monitored for 14 continuous days, once per season (four times per year). A small subset of sites (five) was continuously monitored, so as to track city-wide temporal variation in pollution. The monitors were mounted on utility poles 3 m above the street and recorded temperature and relative humidity every 15 min. Our study uses May to October data from the years 2009 to 2013, for which the NYCCAS and downscaled NLDAS datasets coincide. The number of NYCCAS sites has fluctuated over time due to funding availability,

with 155 sites monitored in 2009 and 2010, 100 sites in 2011 and 2012, and 60 sites in 2013. These air-sampling units were not standard weather stations and were variably subject to direct sunlight, which may have introduced additional errors in the monitored daytime maximum temperature across sites. Because of this, our study focused on minimum temperatures, which tended to occur during nighttime when the influence of shading across sites was presumed to be minimal.

To provide context to our temperature comparison, we retrieved NDVI data from the United States Geological Survey (Landsat 5, image taken April 22, 2010) and calculated the average value around each NYCCAS sampling location at 15 buffer distances from 100 to 1000 m. We also developed measures of the total interior square footage of all buildings within the 15 buffers around each location using extensive land use and geographic data at the tax lot level maintained by New York City Department of City Planning.⁸

2.2 NLDAS Meteorological Reanalysis $1 \times 1 \text{ km}^2$ Downscaled Temperature Data

A fine-scaled, near-surface (2 m above ground level) temperature dataset was derived from meteorological reanalysis and remote sensing data for the years 2005 to 2013. Historical NLDAS temperature data were derived from the NARR analysis fields, which have a 32-km spatial resolution and 3-hourly temporal frequency.⁴ The NARR fields that were used to generate NLDAS meteorological fields were spatially interpolated to the finer resolution of the NLDAS 1/8-deg grid ($\sim 12 \text{ km}$) and then temporally disaggregated to the NLDAS hourly frequency. From the hourly NLDAS data, we derived daily maximum and minimum air temperatures on the 1/8 deg grid for the months of May to October.

These daily temperatures, described in Ref. 9, were then downscaled to a 1-km product using the 1-km moderate resolution imaging spectroradiometer (MODIS) land surface temperature (LST) product “MYD11A2.” In this downscaling algorithm, LST data from the daytime and nighttime ($\sim 1:30 \text{ PM}$ and AM local standard time, respectively) aqua MODIS overpasses were used to provide 1-km spatial patterns of LST for the conterminous United States. The MODIS LST grids form the basis for the downscaled patterns of air temperatures, using the algorithm described below. Daytime LST data were used for downscaling daily maximum temperatures, and nighttime LST data were used for downscaling daily minimum temperatures.

We chose this remote sensing-based approach, as opposed to using available data sets that rely solely on ground station data, such as DayMet¹⁰ and PRISM (parameter-elevation relationships on independent slopes model),¹¹ in order to capture geographic temperature distributions, including intraurban variations, which are extremely difficult to represent using ground station data alone. Remotely sensed LST has been shown to be an excellent predictor of spatial patterns in near-surface air temperature over large expanses,^{12–14} which substantiates the use of LST for estimating air temperature at the MODIS scale.

The downscaling model is based on the assumption that in the absence of strong horizontal temperature advection, air temperature is driven by sensible heat flux from the surface, thus the spatial patterns of air temperature mimic the patterns of LST. Another assumption is that daily maximum air temperatures occur during the early-mid afternoon, near the time of the PM aqua overpass (1:30 PM local standard time), and the daily minimum air temperatures occur in the early morning, near the 1:30 AM aqua overpass. These two assumptions are generally appropriate for quiescent conditions associated with weak synoptic flow, typical of the warm season at mid-latitudes. A final assumption is that the spatial pattern (but not the magnitude) of temperatures is nearly constant from day to day within a 40-day period, so that use of MODIS LST data predating the current day by up to 40 days is still valid. This assumption can be violated by rapidly changing vegetation conditions or large changes in soil moisture, but neither of these conditions typically is relevant to an urban environment.^{15–17}

Air temperature variations tend to be much smaller in magnitude than corresponding LST variations. Therefore, our method computes and applies normalized MODIS LST spatial anomalies to disaggregate daily maximum NLDAS air temperature; the normalization procedure accounts for this difference in magnitude of variation. We first created daytime and nighttime

LST grids using the most recent previous 8-day composite MODIS LST products. If LST data are missing due to cloudiness in the most recent composite, we use data from prior periods, going back a maximum of five 8-day periods, or 40 days, from the day for which downscaling is being performed. By using up to 40 days of LST data, the missing data problem is virtually eliminated. After the composite LST grid is established, standardized LST departures, Z_{HR} , [Eq. (1)] are calculated from the composite LST grid, in which the spatial means and standard deviations are calculated within a local neighborhood or “moving window” (kernel). The size of the window can be varied; we consider it to be a “tunable parameter.” As a step toward optimizing the downscaling model, we tested two sizes of the spatial neighborhood, 3×3 NLDAS grid cells [version 1 (v1)] and 5×5 NLDAS grid cells [version 2 (v2)], or $\sim 36 \times 36 \text{ km}^2$ or $60 \times 60 \text{ km}^2$. The departures are calculated according to

$$Z_{HR} = (T_{HR} - T_{HR,mean})/\sigma_{HR}, \quad (1)$$

where T_{HR} is the high-resolution LST, and $T_{HR,mean}$ and σ_{HR} are the mean and standard deviation, respectively, of high-resolution LST over the neighborhood. Once the departures are calculated, we then compute the downscaled daily maximum air temperatures for each day, T_{DIS} , based on the standardized LST departures

$$T_{DIS} = T_{LR} + Z_{HR} \cdot \sigma_{LR} = T_{LR} + (T_{HR} - T_{HR,mean}) \cdot (\sigma_{LR}/\sigma_{HR}), \quad (2)$$

where T_{LR} and σ_{LR} are the mean and standard deviation, respectively, of low-resolution daily maximum or minimum air temperatures over the neighborhood. Through this algorithm, areas within the NLDAS grid cell for which the MODIS LST composite is warmer (colder) than the NLDAS grid cell mean will result in a maximum or minimum air temperature that is warmer (colder) than the NLDAS air temperature recorded for the respective day. More details of this procedure are provided in Ref. 18.

2.3 Statistical Analysis

In order to assess the performance of the downscaling algorithm, we intersected NYCCAS monitor locations with the native NLDAS grid ($\sim 12 \text{ km}$) and downscaled (1 km) NLDAS grid (see Fig. 1 for graphical explanation) and matched the minimum temperatures by day for all days present in both datasets during the months of May to October. We then averaged the values by year and NLDAS grid cell (see Table 1 for sample sizes). We characterized the relationship between native and downscaled NLDAS data and NYCCAS monitor data using the following metrics: coefficient of determination (R^2) from a linear model with the spatial distribution of the NYCCAS minimum temperatures as the dependent variable and the downscaled NLDAS minimum temperatures as the explanatory variable, root-mean square error (RMSE) for information on bias and precision, mean absolute error (MAE) to describe the accuracy of the modeled data, and the slope of the regression equation to determine any bias in the linear relationship. Initial comparisons included all 5 years (2009 to 2013); however, the analyses presented here focused on years with the most spatial coverage in the NYCCAS program, 2009 and 2010, a cooler and a typical summer, respectively. We tested for spatial autocorrelation in the residuals from this model using Moran’s I using the R package *ape*.¹⁹ In order to account for spatial autocorrelation in the residuals, we included a thin-plate smooth function of the XY coordinates in a generalized additive model. We tested various degrees of freedom in the spatial smooth term and selected the minimum number necessary to remove spatial autocorrelation in the residuals. We explored the spatial pattern of residuals from the model without the smooth term by correlating them with indicators of the intensity of development and greenness (NDVI, density of interior built space). All analyses were conducted using statistical package R version 3.3.2 (R Core Team, Vienna).

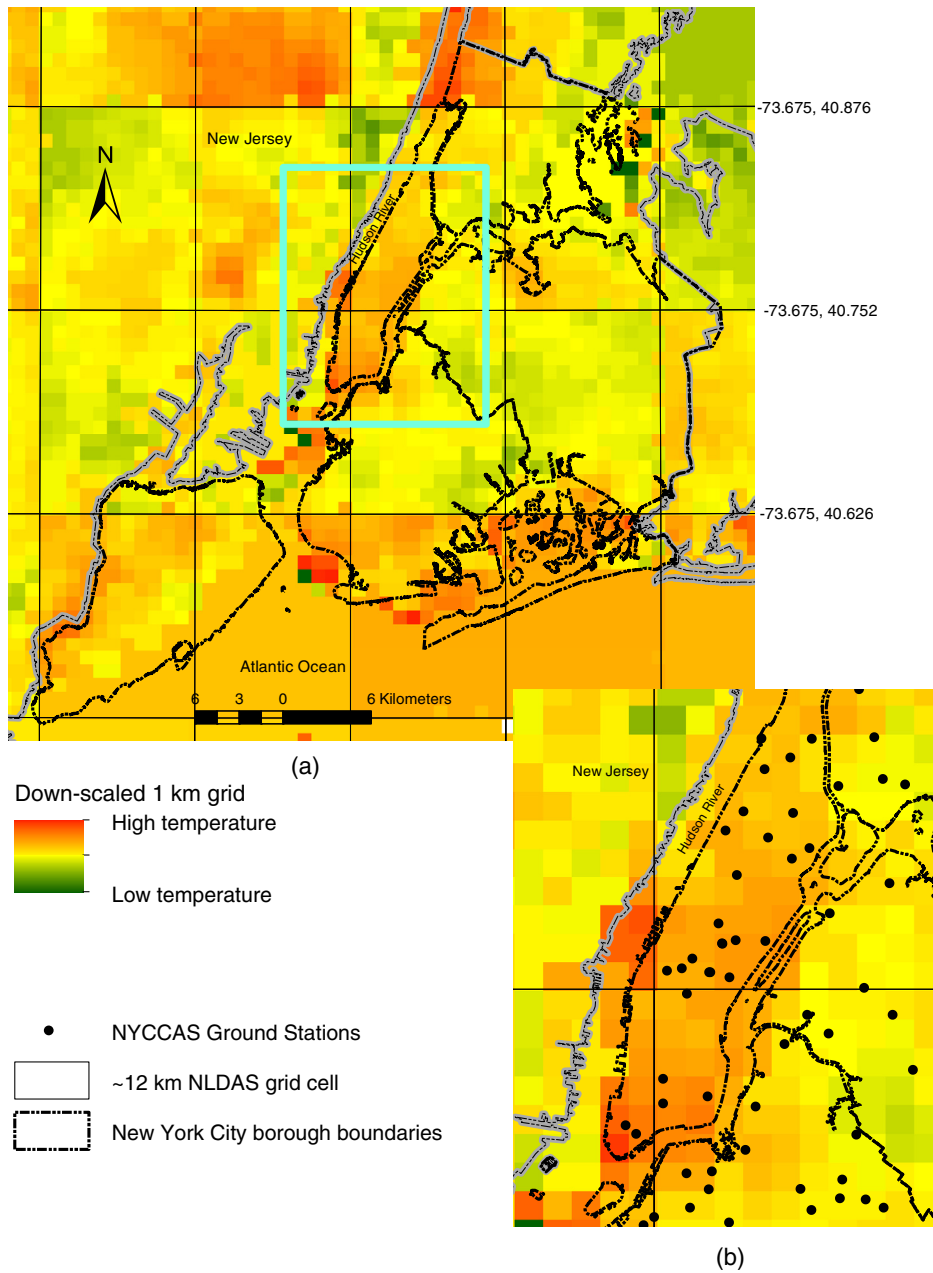


Fig. 1 Illustration of how the NYCCAS data are matched geographically to the NLDAS temperature data. (a) Citywide extent of the (~12 km) and 1 km NLDAS grid cells. (b) Detailed inset with example of the spatial distribution of NYCCAS ground sites.

3 Results

3.1 12 km NLDAS-NYCCAS Minimum Temperature Comparisons

The results from the comparison of the native NLDAS data with average NYCCAS temperature was based on only the five cells whose centroids fall within NYC boundaries (Table 1 and Fig. 2). These results provided context for the 1-km outcomes, demonstrating that the temperature variability found within NYC boundaries was not well described at the ~12 km resolution.

Table 1 Performance of downscaled NLDAS data in describing summertime average minimum temperatures from NYCCAS. Comparison of regression metrics from separate models for the years 2009 and 2010, for the native NLDAS grid (~12 km), downscaled data using a 3 × 3 and a 5 × 5 kernel (versions 1 and 2, respectively), and with a smooth term to account for spatial autocorrelation of the residuals (indicated by “+”). Metrics include coefficient of determination (R^2), RMSE, MAE, the slope and intercept and their standard errors (SE), and the sample size (n).

Year	Version	R^2	RMSE (°C)	MAE (°C)	Slope (SE)	Intercept (SE) (°C)	n
2009	1	0.83	0.93	0.75	0.94 (0.04)	1.95 (0.59)	146
2009	1+	0.9	0.69	0.54	0.93 (0.03)	2.12 (0.46)	146
2009	2	0.85	0.87	0.69	0.92 (0.03)	1.94 (0.54)	146
2009	2+	0.9	0.67	0.51	0.91 (0.03)	2.11 (0.44)	146
2010	1	0.85	1.1	0.88	1.03 (0.04)	1.73 (0.64)	141
2010	1+	0.92	0.75	0.58	1.01 (0.03)	2.11 (0.46)	141
2010	2	0.85	1.07	0.86	1.01 (0.04)	1.82 (0.61)	141
2010	2+	0.92	0.79	0.62	0.98 (0.03)	2.28 (0.48)	141
2009	12 km	0.05	0.6	0.52	-0.44 (1.14)	23.81 (18.27)	5
2010	12 km	0.01	0.92	0.78	0.24 (1.16)	14.54 (19.22)	5

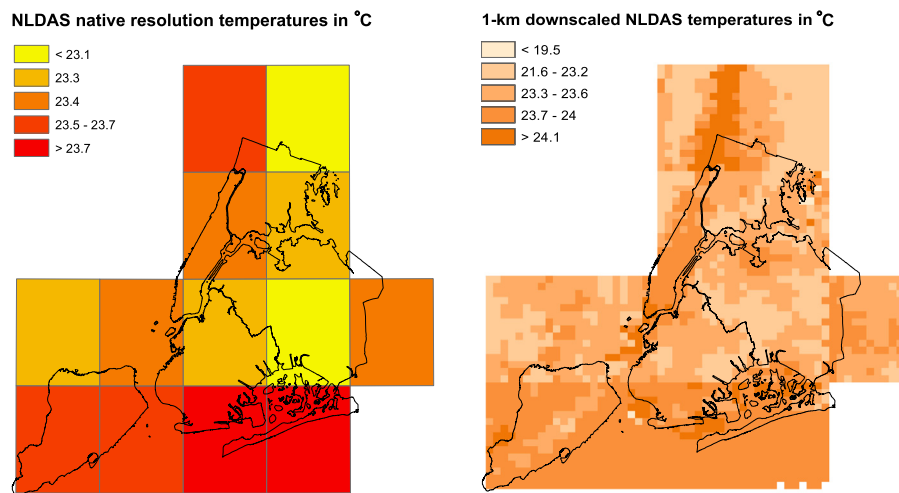


Fig. 2 Example of the native and downscaled NLDAS minimum temperatures for July 7, 2010; the high reported at JFK weather station on this day was 38°C, and the low was 25°C.

3.2 Downscaled NLDAS-NYCCAS Minimum Temperature Comparison

Unique locations in the downscaled NLDAS–NYCCAS comparison totaled 146 in 2009 and 141 in 2010. The number of matching days at each location ranged from 15 to 153 with a median of number of days of 26 by year of analysis (2009 and 2010). NLDAS minimum temperatures explained the spatial variation in NYCCAS data well with a maximum R^2 without spatial smooth of 0.85 in both 2009 and 2010 [v2, Figs. 3(b) and 3(d), Table 1]. 2009 was a relatively cool year compared to those before and after, with May to October average minimum temperature of 17.4°C as measured by NYCCAS as compared to 19.1°C in 2010. Both v1 (3 × 3 kernel) and v2 (5 × 5 kernel) of the downscaled data were highly correlated with NYCCAS data (R^2 ranging from 0.83 to 0.92, Table 1). Error was smaller in 2009 (v2 RMSE = 0.87) than in 2010 (v2 RMSE = 1.07) and for v2 (2009 RMSE = 0.87) compared to v1 (2009

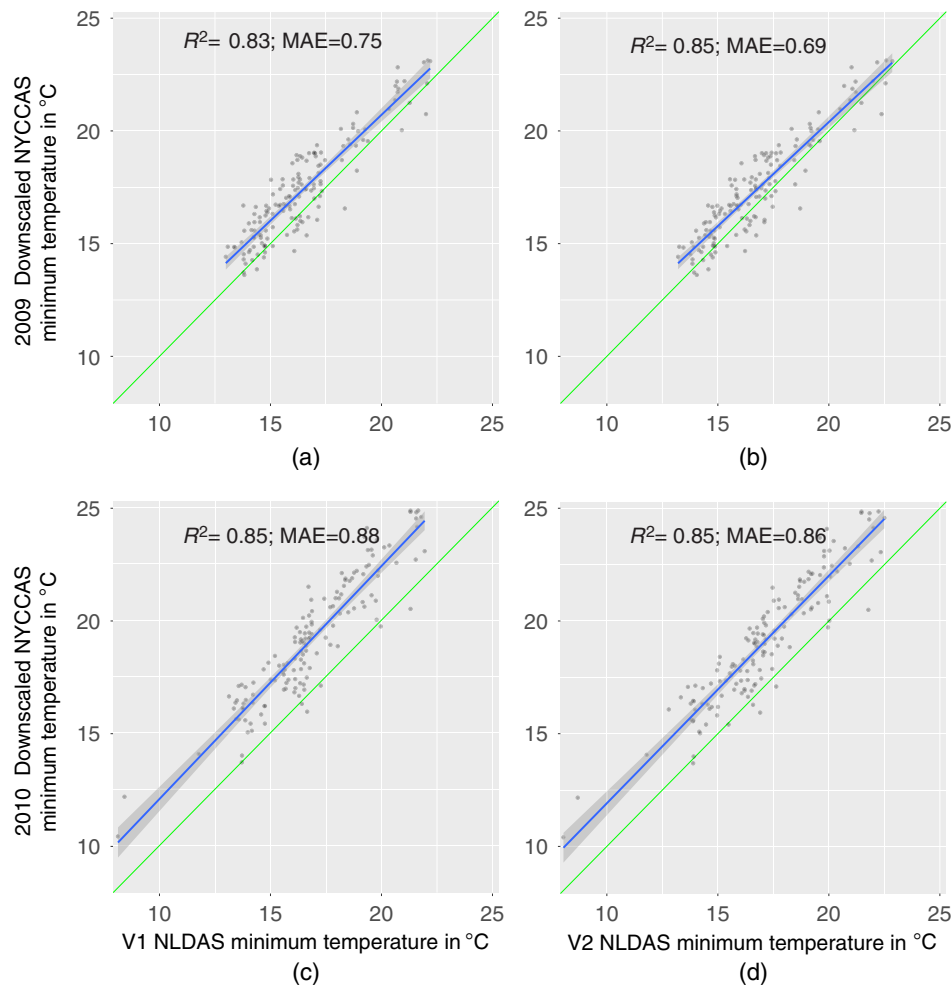


Fig. 3 Scatterplots of (a) 2009 NYCCAS versus NLDAS downscaled 1 km averaged values for the version 1 model (3×3 kernel), and (b) the same for version 2 (5×5 kernel). Plots (c) and (d) are the same scatterplots for versions 1 and 2, respectively, for 2010. The blue line shows values fitted to a linear model; gray shading shows the 95% confidence interval limits. The green line shows a 1:1 fit.

RMSE = 0.93). The better agreement in 2009, the cooler year, is consistent with the observation that NLDAS underestimates higher temperatures on average across both versions and all years. There was also a narrower range of temperatures in 2009 compared with 2010, which included two much cooler outliers [Figs. 3(c) and 3(d)]. Examination of the residuals did not reveal violations of linear model assumptions; however, we did see a significant spatial pattern in the residuals (Moran's I p -value < 0.000001). The xy smooth function with seven degrees of freedom was sufficient to control spatial autocorrelation in the residuals, which increased the R^2 from 0.85 to 0.9, decreased the error by over 20% while changing the slope by less than 3% and the intercept by on average 15% (Table 1).

NYCCAS-NLDAS temperature differences are mapped in Figs. 4(a) and 4(b). (See Fig. 5 for histograms). A comparison of these two plots shows the improvement that the 5×5 kernel makes in the downscaling, with average difference between NYCCAS and NLDAS decreasing from 0.93°C with a 3×3 kernel (v1) to 0.64°C with a 5×5 (v2) kernel for 2009. In 2010, while the average differences are higher, the 5×5 kernel improves the performance of the downscaled data to the same extent as is 2009 [2.29°C for v1 and 1.98°C for v2, Figs. 4(c) and 4(d)]. Orange points in Fig. 4(c), representing a 4°C to 5°C difference between NYCCAS and downscaled NLDAS are not present in Fig. 4(d). Figures 4(a) and 4(b) also show greater divergence between the datasets in the higher density areas of mid-town Manhattan and Brooklyn. NYCCAS

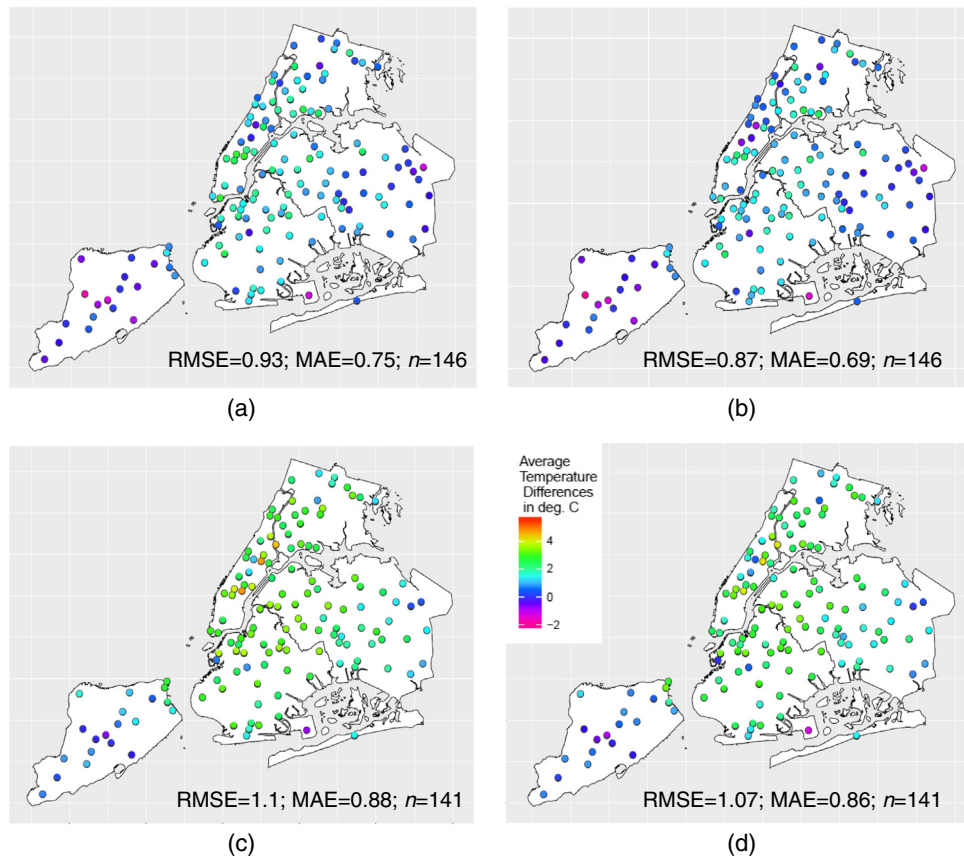


Fig. 4 NYCCAS-downscaled NLDAS 2009 1-km downscaled average minimum temperature differences for the (a) version 1 model (3×3 kernel) and (b) version 2 (5×5 kernel). Plots (c) and (d) are the same NYCCAS-downscaled NLDAS differences for versions 1 and 2, respectively, for 2010.

minimum temperatures tend to be higher than downscaled estimates in most of NYC except Staten Island, the least built-up borough, as well as in less densely urbanized parts of the other boroughs. We found the spatial pattern in the residuals from the model without the spatial smooth to be correlated most strongly with NDVI at a buffer size of 300 m ($r = -0.65$ for 2010 v2), with lower NDVI values indicating more urbanization and less green cover corresponding to areas where the downscaled dataset underestimated NYCCAS temperatures the most [Fig. 6(b)] while temperatures in greener areas were overestimated NLDAS.

4 Discussion and Conclusion

4.1 Overall Results

This study takes advantage of a unique high-density monitoring network in NYC to ground-truth remotely sensed and modeled temperature estimates, specifically in terms of the spatial variability in temperature within a complex urban and coastal environment. The native NLDAS (~ 12 km) resolution performs well in regions that are flat and landlocked with comparatively low population density, such as the Southern Great Plains, as demonstrated in the many studies that use Oklahoma Mesonet data.^{20,21} NYC has, in common with many other cities, a high population density and complex coastal features, which are among the contributors to surface temperature heterogeneity that are not well captured at a ~ 12 -km scale. However, the spatial variability in warm season daily minimum temperatures in NYC is well described by downscaling the native NLDAS air temperature estimates to a 1-km product using an algorithm that

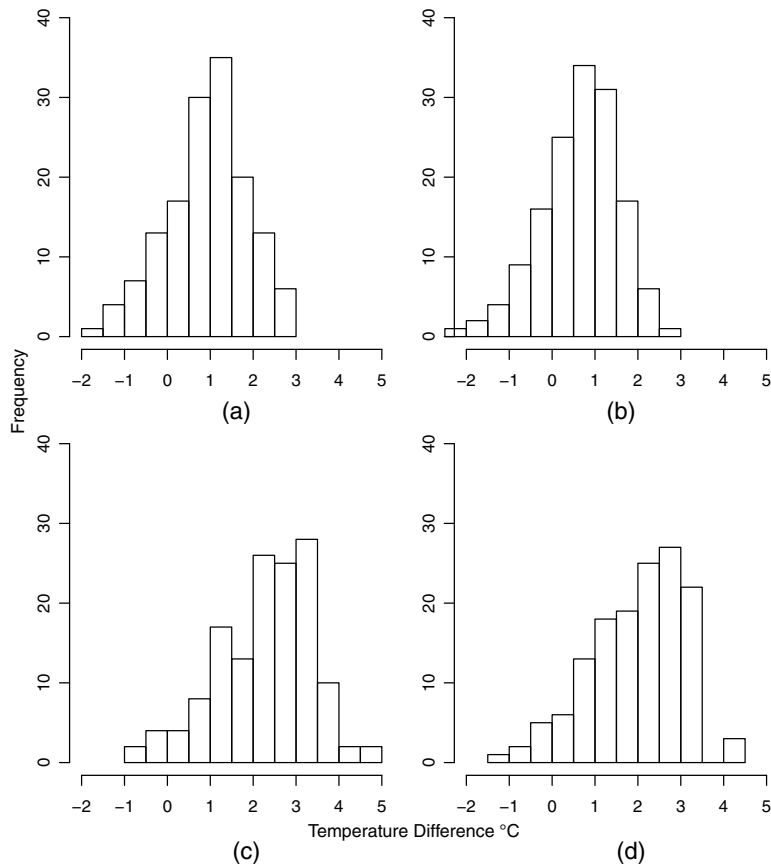


Fig. 5 NYCCAS-downscaled NLDAS 2009 1 km downscaled average minimum temperature differences for the (a) version 1 model (3×3 kernel) and (b) version 2 (5×5 kernel). Plots (c) and (d) are the same NYCCAS-downscaled NLDAS differences for versions 1 and 2, respectively, for 2010.

employs a 5×5 kernel neighborhood to impose the spatial temperature pattern from MODIS LST data. While the datasets were highly correlated ($R^2 = 0.85$), NLDAS downscaled data consistently underestimated the warmest minimum temperatures measured by NYCCAS. Using the relationship of the downscaled NLDAS data to ground station monitoring that we have determined for NYC, other cities can make use of the regression parameters from our adjusted model and apply them to the downscaled data now available for the contiguous United States. The US Centers for Disease Control (CDC) recommends that health departments use data at the finest geographic resolution available when assessing vulnerability to climate change²² and many local health departments express the desire for finer resolution exposure data.

Similar measures of R^2 , MAE, and RMSE have been used in prior studies that compare remote sensing sources with monitored air temperature.^{23–25} A good correspondence (i.e., a high R^2) between satellite-based and ground-measured data is possible only when the observations contain sufficient variation in measured values, which is more often typical of country-scale studies. The very small spatial extent of cities make a good fit less likely; Ho et al.²⁵ found that the spatial pattern of Landsat-derived daytime surface temperature correlated poorly ($R^2 = 0.39$) with maximum air temperature measured at 39 weather stations for 6 hot summer days in the greater Vancouver area. In Birmingham, United Kingdom, nighttime MODIS-based surface temperature averaged over June to August was moderately correlated ($R^2 = 0.6$) with minimum air temperatures measured at 107 locations.²³ In comparison, we find good correspondence in the spatial patterns between the temperature measures in our study.

Other data sets exist that provide high-resolution daily meteorological variables. For our specific application, the primary shortcoming of some of these, such as DayMet¹⁰ and PRISM¹¹ (parameter-elevation relationships on independent slopes model), is that their temperatures are

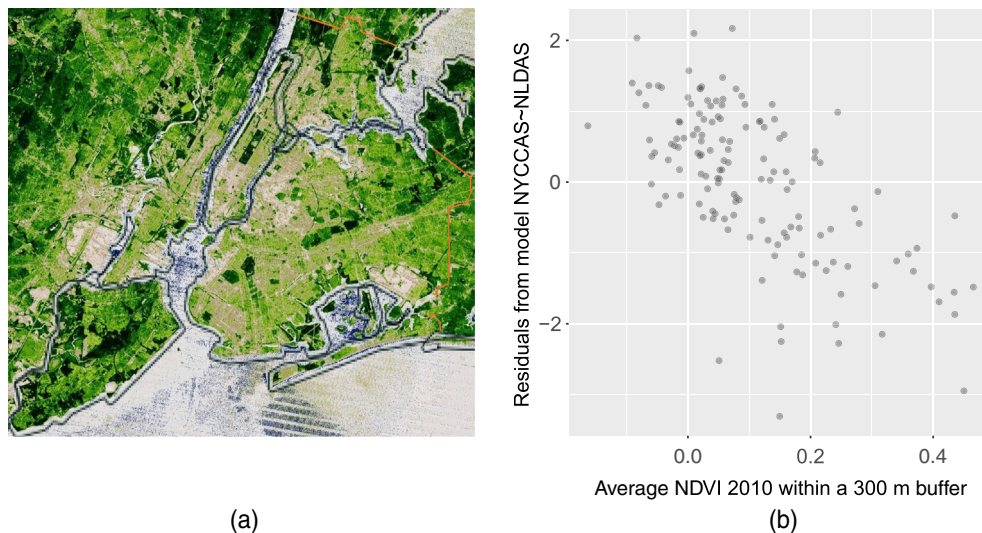


Fig. 6 (a) Surface NDVI, V3.0, from Landsat/WELD for July 21, 2010, NYC borders added in orange (Roy 2010). (b) Scatterplot of residuals from the linear regression of NYCCAS-downscaled NLDAS (v2) and NDVI averaged within a 300-m buffer for 2010.

based on station data alone, with no remotely sensed inputs. From the station data, DayMet and PRISM interpolate to a 1-km or 800-m grid through a sophisticated spatial algorithm, which also adjusts temperatures based on surface elevation in the case of DayMet, and surface elevation and other physiographic factors such as coastal proximity, topographic facet orientation, and vertical atmospheric layer in the case of PRISM. However, ground stations are rarely available at anything close to 1 km or 800 m spacing. Thus, even in urban areas where surface observations are relatively dense, the estimated temperatures rely heavily on spatial interpolation. In more rural areas, ground observations are very sparse and thus the DayMet temperatures are even more heavily interpolated. In our approach, use of remote sensing data at the desired spatial resolution of the resulting data set allows us to capture the thermal details of, for example, an urban environment. Even in rural areas, temperature variations that occur due to agricultural patterns, soil moisture variations, and riparian areas are not well captured by ground observations and thus are mostly absent in the DayMet and PRISM temperature data sets. These features are observed by MODIS LST and are therefore represented in our downscaled temperature data. Admittedly, the MODIS data are not of air temperature, but they at least provide the fine-scale structure of temperature, and we argue that our downscaling model provides a bridge between LST and air temperature.

In NYC, there is variability in built density as measured by NDVI, and the fact that we found monitored values to exceed remotely sensed modeled estimates in areas of high built density is consistent with other studies. Bechtel compared temporal trends in air temperature with remotely sensed LST at six ground-station sites in Hamburg, Germany, which ranged from suburban neighborhoods and urban parks to very compact buildings and highly sealed areas; the remotely sensed data underestimated air temperature in sites with extensive impervious surface.²⁴ This inverse relationship of modeled versus observed temperature differences and NDVI could have several explanations.^{26,27} MODIS surface temperature used in the downscaling process tends to amplify temperature extremes in both directions, especially over built or sparsely vegetated surfaces.^{9,28} Satellite-derived surface temperature spatial patterns have been found to diverge from the spatial pattern of air temperature in urban settings, hypothesized to be due to the advection of heat produced in the city center.²³ Perhaps the warm season synoptic flow in NYC is not as weak as is assumed in the downscaling model. The summertime prevailing winds from the southwest in NYC may push heat created in Midtown Manhattan and the heavily industrialized area of New Jersey into residential neighborhoods to the north and east, exacerbating air temperatures already impacted by high surface temperatures as seen in Birmingham, United Kingdom.²⁹ Weather patterns in NYC, as with many other coastal cities, include complex land-water boundaries with sharp changes in energy and moisture exchanges. These drive a host

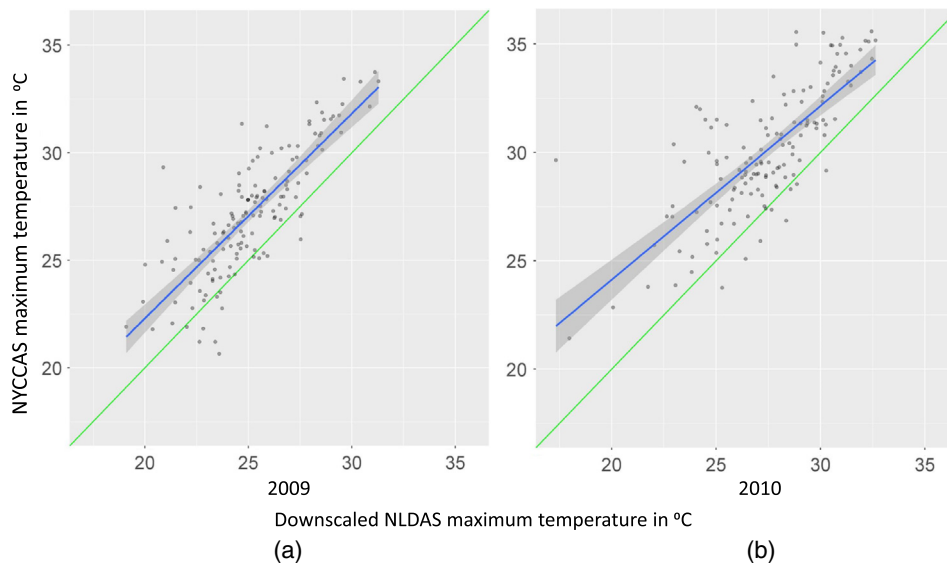


Fig. 7 Scatterplots of (a) 2009 and (b) 2010 NYCCAS versus downscaled NLDAS 1 km averaged maximum temperatures for the version 2 model. The R^2 is 0.80 for 2009 and 0.76 for 2010. The blue and gray as above show the linear model and the 95% CI.

of meteorological effects, such as the pattern of land-sea breezes, which are distinct from inland areas and can vary in effect on the urban climate depending on the weather pattern.^{26,27} While the combination of the NLDAS-based air temperatures and the MODIS surface temperatures did well in describing air temperatures in NYC, further steps could be taken to improve the fit such as using land-use regression modeling to combine the downscaled data with NDVI. Evrendilek et al.³⁰ used satellite-based and mesoscale regression modeling of air temperature with other variables including NDVI to compare against 83 ground-based stations in Turkey. The adjusted R^2 of the best minimum air temperature model was 0.65 and the RMSE 5.9°C and included NDVI in addition to the remotely sensed surface temperature. The downscaled NLDAS dataset is available for the continental United States as is NDVI (USGS) which would allow other cities to produce 1-km resolution temperature surfaces for exposure assessment in health studies and evaluation of urban heat mitigation activities.

4.2 Limitations

The entire area of NYC is coastal and urban, and weather patterns evident here may not apply to inland locations. We do not have access to information regarding when monitors are in sunlight or shade, thus daytime NYCCAS temperatures are likely more subject to errors than recordings at night. Daytime comparisons (Fig. 7) should be interpreted with caution. While we limited this analysis to ground-truthing overnight minimum temperatures due to daytime temperature limitations, minimum temperatures are highly correlated with maximum temperatures and heat index in NYC and have similar association with warm-season natural cause mortality.³¹ The frequency of extreme warm overnight temperatures is increasing more rapidly than extreme daytime temperatures in the United States (graph^{32,33}).

4.3 Summary

Our study has demonstrated the improvement of the downscaled NLDAS model in capturing the spatial variability of temperature across NYC neighborhoods over the native ~12-km resolution. Comparisons of warm-season (May to October) downscaled NLDAS modeled temperatures with NYCCAS ground station measurements showed strong agreement overall, with better agreement for v2, which uses a 5×5 kernel. NYCCAS measured minimum temperatures were warmer on average than the downscaled NLDAS temperatures, with the best agreement occurring for areas

and years with cooler minimum temperatures. Improvements to the downscaling algorithm in v2 reduced these differences further for both normal and cooler summers and improved model fit in the more densely urbanized areas.

This study demonstrates that the 1-km downscaled NLDAS dataset provides high accuracy temperature data at a temporal and spatial resolution that is desired by end-users in the climate and health community for research and policy development and evaluation purposes.³⁴ While the ~12-km data are useful in places with less heterogeneous land-cover and lower population density, the higher resolution is needed for coastal and/or urban applications. In cities without a high resolution network of temperature monitors such as NYCCAS, this dataset could provide exposure estimates useful in the understanding of heat-related health effects, as heat-health vulnerability exhibits high rates of geographic variability.³⁵

Acknowledgments

We would like to thank several colleagues who have helped to improve this manuscript: Muhammad Barik helped by suggesting some supporting references, along with colleagues at NYC DOHMH who provided invaluable guidance in statistical analysis and editing. This research was made possible by NASA Grant No. ROSES 2013-NNX15AN77G- Using Remote Sensing and Environmental Data to Quantify Social Vulnerabilities to Heat Stress and Strengthen Environmental Public Health Tracking and Heat Mitigation Efforts.

References

1. J. Hansen, M. Sato, and R. Ruedy, "Perception of climate change," *Proc. Natl. Acad. Sci. U. S. A.* **109**(37), E2415–E2423 (2012).
2. G. Luber and M. McGeehin, "Climate change and extreme heat events," *Am. J. Prev. Med.* **35**(5), 429–435 (2008).
3. J. K. Vanos, L. S. Kalkstein, and T. J. Sanford, "Detecting synoptic warming trends across the US Midwest and implications to human health and heat-related mortality," *Int. J. Climatol.* **35**, 85–96 (2015).
4. F. Mesinger et al., "North American regional reanalysis," *Bull. Am. Meteorol. Soc.* **87**, 343–360 (2006).
5. L. Luo et al., "Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains," *J. Geophys. Res. Atmos.* **108**(d22), 8843 (2003).
6. A. L. Robock et al., "Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season," *J. Geophys. Res. Atmos.* **108**(d22), 8846 (2003).
7. T. D. Matte et al., "Monitoring intraurban spatial patterns of multiple combustion air pollutants in New York City: design and implementation," *J. Exposure Sci. Environ. Epidemiol.* **23**(3), 223–231 (2013).
8. "BYTES of the BIG APPLE," New York City Department of City Planning (NYCDCP) (2009).
9. M. Z. Al-Hamdan et al., "Environmental public health applications using remotely sensed data," *Geocarto Int.* **29**(1), 85–98 (2014).
10. P. E. Thornton et al., "Daymet: daily surface weather data on a 1-km grid for North America, Version 3," ORNL DAAC (2018).
11. C. Daly et al., "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States," *Int. J. Climatol.* **28**(15), 2031–2064 (2008).
12. S. Shen and G. G. Leptoukh, "Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature," *Environ. Res. Lett.* **6**, 045206 (2011).
13. C. Vancutsem et al., "Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa," *Remote Sens. Environ.* **114**(2), 449–465 (2010).

14. W. Zhu, A. Lü, and S. Jia, “Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products,” *Remote Sens. Environ.* **130**, 62–73 (2013).
15. Y. Feng et al., “Spatial patterns of land surface temperature and their influencing factors: a case study in Suzhou, China,” *Remote Sens.* **11**(2), 182 (2019).
16. M. Z. Al-Hamdan et al., “Evaluating land cover changes in eastern and Southern Africa from 2000 to 2010 using validated Landsat and MODIS data,” *Int. J. Appl. Earth Obs. Geoinf.* **62**, 8–26 (2017).
17. M. Z. Al-Hamdan et al., “Using Landsat, MODIS, and a biophysical model to evaluate LST in urban centers,” *Remote Sens.* **8**(11), 952 (2016).
18. W. L. Crosson, M. Z. Al-Hamdan, and T. Z. Insaf, “Downscaling NLDAS daily maximum air temperatures using MODIS land surface temperatures,” *PLoS One*, in review.
19. E. Paradis, J. Claude, and K. Strimmer, “APE: analyses of phylogenetics and evolution in R language,” *Bioinformatics* **20**(2), 289–290 (2004).
20. B. A. Cosgrove et al., “Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project,” *J. Geophys. Res. Atmos.* **108**(D22), 8842 (2003).
21. K. Nemunaitis-Berry et al., *Validation of the North American Land Data Assimilation System (NLDAS) using Data from Oklahoma Mesonet Oasis Sites* (2004).
22. A. P. Manangan et al., *Assessing Health Vulnerability to Climate Change: a Guide for Health Departments U.S. Climate Resilience Toolkit*, Climate and Health Technical Report Series, Centers for Disease Control and Prevention (2014).
23. J. A. Azevedo, L. Chapman, and C. L. Muller, “Quantifying the daytime and nighttime urban heat Island in Birmingham, UK: a comparison of satellite derived land surface temperature and high resolution air temperature observations,” *Remote Sens.* **8**(2), 153 (2016).
24. B. Bechtel, S. Wiesner, and K. Zakšek, “Estimation of dense time series of urban air temperatures from multitemporal geostationary satellite data,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**(10), 4129–4137 (2014).
25. H. C. Ho et al., “A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area,” *Sci. Total Environ.* **544**, 929–938 (2016).
26. S. D. Gedzelman et al., “Mesoscale aspects of the Urban Heat Island around New York City,” *Theor. Appl. Climatol.* **75**(1), 29–42 (2003).
27. P. Ramamurthy et al., “Impact of heatwave on a megacity: an observational analysis of New York City during July 2016,” *Environ. Res. Lett.* **12**(5), 054011 (2017).
28. A. Wang, “Comparison of land skin temperature from a land model, remote sensing, and in situ measurement,” *J. Geophys. Res. Atmos.* **119**(6), 3093–3106 (2014).
29. R. Bassett et al., “Observations of urban heat island advection from a high-density monitoring network,” *Q. J. R. Meteorol. Soc.* **142**(699), 2434–2441 (2016).
30. F. Evrendilek et al., “Satellite-based and mesoscale regression modeling of monthly air and soil temperatures over complex terrain in Turkey,” *Expert Syst. Appl.* **39**(2), 2059–2066 (2012).
31. K. B. Metzger, K. Ito, and T. D. Matte, “Summer heat and mortality in New York City: how hot is too hot?” *Environ. Health Perspect.* **118**(1), 80–86 (2010).
32. K. L. Gleason et al., “A revised U.S. Climate Extremes Index,” *J. Clim.* **21**(10), 2124–2137 (2008).
33. National Center for Environmental Information (NCEI), “U.S. Climate Extremes Index (CEI),” <https://www.ncdc.noaa.gov/extremes/cei/graph/us/2/06-08> (Accessed 21 December 2018).
34. Z. A. Hamstead et al., “Classification of the heterogeneous structure of urban landscapes (STURLA) as an indicator of landscape function applied to surface temperature in New York City,” *Ecol. Indic.* **70**, 574–585 (2016).
35. J. Madrigano et al., “A case-only study of vulnerability to heat wave-related mortality in New York City (2000-2011),” *Environ. Health Perspect.* **123**(7), 672–678 (2015).

Heather Eliezer worked as a research scientist at New York City's Department of Health & Mental Hygiene. She received her PhD from the CUNY Graduate Center in Earth & Environmental Science and a BA degree from Binghamton University in Biology and Environmental Studies. Her interests also include green infrastructure and multimodal planning. She currently cochairs the Sustainability Advisory Board for the Village of Pelham, New York.

Sarah Johnson is the director of the Air Quality Program at the New York City Department of Health. She has a MS degree in ecology from University of California, Davis, and a MPH from UC Berkeley. She has worked for the Health Department since 2005 and focuses on understanding spatial variation in air pollution, temperature and vectors in NYC and measuring their health effects with the goal of informing program planning and evaluation.

William L. Crosson recently retired from his position as a senior research scientist with the Universities Space Research Association at NASA Marshall Space Flight Center in Huntsville, Alabama, USA. He received his PhD and MS degrees in meteorology from Florida State University and performed basic and applied research in the areas of hydrology, meteorology, remote sensing, air quality and public health for over three decades.

Mohammad Z. Al-Hamdan is a senior research scientist with the Universities Space Research Association at NASA Marshall Space Flight Center in Huntsville, Alabama, USA. He received his PhD and MS degrees in Civil and Environmental Engineering from the University of Alabama in Huntsville. His research expertise includes remote sensing and GIS applications to environmental modeling and assessment for air quality, water, urban heat island, ecological and public health applications.

Tabassum Z. Insaf is the research director for the Bureau of Environmental and Occupational Epidemiology at the New York State Department of Health. She also serves as assistant professor at the School of Public Health, University at Albany. She is interested in geo-spatial epidemiology and its applications into assessment of social and environmental determinants of health including health effects of climate change.