

## Title of project or data set

Code associated with: Effects of urbanization on cloud-to-ground lightning strike frequency: a global perspective

## Abstract

Urbanization tends to increase local lightning frequency (i.e., the "lightning enhancement" effect). Despite many urban areas showing lightning enhancement, the prevalence of these effects is unknown, and the drivers underlying these patterns are poorly quantified. We conducted a global assessment of cloud-to-ground lightning flashes (lightning strikes) across 349 cities to evaluate how the likelihood and magnitude of lightning enhancement vary with geography, climate, air pollution, topography, and urban development. The likelihood of exhibiting lightning enhancement increased with higher temperature and precipitation in urban areas relative to their natural surroundings (i.e., urban heat islands and elevated urban precipitation), higher regional lightning strike frequency, greater distance to water bodies, and lower elevations. Lightning enhancement was stronger in cities with conspicuous heat island and elevated urban precipitation effects, higher lightning strike frequency, larger urban areas, and lower latitudes. The particularly strong effects of elevated urban temperature and precipitation indicate that these are dominant mechanisms by which cities cause local lightning enhancement.

## Creators

First Name	Middle Initial	Last Name	Organization	e-mail address	ORCID ID (optional)
J.	Pablo	Narvaez	Smithsonian Tropical Research Institute Marquette University	pablo.narvaez@marquette.edu	0009-0004-3094-4309
Evan	M.	Gora	Smithsonian Tropical Research Institute Cary Institute	gorae@caryinstitute.org	0000-0002-0537-5835
Steven	P.	Yanoviak	Smithsonian Tropical Research Institute University of Louisville	steve.yanoviak@louisville.edu	0000-0001-6425-1413
Phillip	M.	Bitzer	The University of Alabama in Huntsville	pm.bitzer@uah.edu	0000-0002-6665-9778
Jeffrey	C.	Burchfield	The University of Alabama in Huntsville	jcb0003@uah.edu	0000-0003-2858-053X

## Keywords

Lightning, urban, remote sensing, human disturbance, modeling, data

## Funding of this work:

Add rows to table if several grants were involved, list only the main PI, start with main grant first:

PI First Name	PI Middle Initial	PI Last Name	PI ORCID ID (optional)	Title of Grant	Funding Agency	Funding Identification Number
Evan	M.	Gora		Earl S. Tupper Fellowship	Smithsonian Tropical Research Institute	
Evan	M.	Gora		DEB	Natural Science Foundation	2213245
Phillip	M.	Bitzer		DEB	Natural Science Foundation	2213247
Steven	P.	Yanoviak		DEB	Natural Science Foundation	2213246

## License

The code for this project is licensed via the MIT License. For full licensing and attribution information see <https://opensource.org/license/MIT>. See the references below under “Data Provenance” for the licensed data used in this global assessment.

## Timeframe

- Begin date: January 2013
- End date: December 2020
- Data collection ongoing/completed? Completed

## Geographic location

349 cities worldwide

## Methods

### Lightning strike data

We quantified the urban lightning enhancement effect using Earth Networks Total Lightning Network (ENTLN) data. ENTLN continuously detects and locates lightning using each discharge's time and signal amplitude (Liu & Heckman, 2012); here, we focus on the ENTLN-classified cloud-to-ground flashes, which we call lightning strikes. We omitted lightning strikes <10 kA in magnitude to avoid

misclassification with in-cloud lightning (Cummins et al., 1998). We calculated monthly mean lightning strike frequency (lightning strikes  $\text{km}^{-2} \text{yr}^{-1}$ ) on a  $0.05 \times 0.05$ -degree grid (ca.  $5 \times 5$  km) extending from  $60^\circ\text{N}$  to  $60^\circ\text{S}$  latitude for 2013–2020.

### **Urban and natural areas**

We used the 2018 Moderate Resolution Imaging Spectroradiometer (MODIS) land cover data (MCD12C1 Version 6; Friedl & Sulla-Menashe, 2019) on a  $0.05$ -degree grid to identify urbanized land and its surrounding natural areas. Spatially, we defined cities as clusters of  $0.05 \times 0.05$ -degree cells with more than 50% urbanization overlapping the city center, defined by the United Nations World's Cities in 2018–Data Booklet (UN, 2018), or contiguous with other urban cells. Because the urbanization footprint of a city is often a mosaic of developed and undeveloped space (e.g., water bodies), we also included any cell with >50% urban area that was within two cells of the city center or contiguous city area (no cells of <50% urban area were included in a city). This process collapsed 22 pairs of cities into a single urban center (e.g., Dallas/Ft. Worth). We identified 884 cities with >300,000 inhabitants and at least one cell comprising >50% urban area.

We used MODIS to identify natural areas surrounding each city. Specifically, we defined natural areas as any combination of non-modified MODIS terrestrial layers (excluding water bodies, urban area, and croplands) within 150 km of the boundaries of a city. When a cell was within 150 km of multiple cities, we associated that natural area cell with the closest city. To limit edge effects, we removed all natural areas within two cells (ca. 10 km) of any cell with >50% urban area. We only retained cities in our dataset if they had at least  $100 \text{ km}^2$  of associated natural area (691 cities qualified). The natural areas capture typical lightning frequency of each region with limited direct influence of urbanization, functioning as a reference point for evaluating the effect of each urban area.

### **Calculating lightning frequency**

We calculated each pixel's average annual lightning frequency using only months with meaningful lightning activity. We removed all cities with < 1 lightning strike  $\text{km}^{-2} \text{yr}^{-1}$  in their associated natural areas. We also removed months from individual cities if their natural areas exhibited < 1 lightning strike  $\text{km}^{-2} \text{yr}^{-1}$  in those months (328 cities removed). Additionally, we removed months and cities (8 cities in total) lacking data for their covariates (e.g., precipitation data was not available for 2019 and 2020, and the aerosol optical depth sensor could not make its measurement in certain areas during the study). Following these criteria, we ultimately included 349 cities in the analyses.

We used Glass's delta effect size and a simulation approach to evaluate whether each city exhibited unambiguous lightning enhancement. To calculate Glass's delta, we divided the mean difference in lightning strike frequency (lightning strikes  $\text{km}^{-2} \text{yr}^{-1}$ ) between urban and natural areas by the standard deviation of lightning strike frequency of the associated natural area. Glass's delta was preferable to other effect size metrics because the much larger sample size of the natural areas, relative to the cities, results in a more precise estimate of standard deviation. Effect sizes  $\geq 0.5$  were considered significant (Cohen, 1992). We confirmed that 218 of the 228 cities with effect sizes  $> 0.5$  were also identified as significant using a simulation test based on random pulls from the natural area associated with each

city. We considered the 218 cities identified with both approaches as those exhibiting unambiguous lightning enhancement. This conservative approach likely eliminated false positives while potentially producing some false negatives.

The detection efficiency of ENTLN likely exhibits unquantifiable spatial biases. However, the spatial grain of these biases is much larger than that of our city and natural area measurements (150 km radius) because individual ENTLN sensors detect lightning over distances >1000 km. We measured the strength of urban enhancement by dividing a city's average lightning strike frequency by the average lightning strike frequency in its associated natural area (hereafter, urban-natural strike ratio). Accordingly, this approach is insensitive to possible differences in detection efficiency among cities or over time.

### **Climatological, topographical, and geographic covariates**

We used spatially explicit, gridded data products to aggregate climatological, topographical, and geographic covariates for each 0.05 x 0.05-degree cell. We assigned each 0.05 x 0.05 cell the proportional average of overlapping sulfur dioxide (SO<sub>2</sub>) values because of the mismatch in resolution. All other data were downscaled or upscaled to the same spatial grain as the lightning data. Climate and pollution data were aggregated monthly. The temperature metrics captured monthly averages of daily trends and were advantageous because of their broad spatial coverage and fine resolution, but they did not capture detailed within-day variation, which could influence both rainfall and lightning activity (Sheperd et al., 2015). All other variables had a single value because they did not change during the study period (e.g., topography) or data were limited (e.g., population).

We used these spatially explicit datasets to calculate potential predictors of variation in the lightning enhancement effect. For each variable described in Table 2, we extracted its average value for each urban area during the months retained in the dataset (i.e., months with > 1 lightning strike km<sup>-2</sup> yr<sup>-1</sup>). We calculated annual and cumulative averages of those values from 2013-2020. The only exception was regional lightning frequency, which equaled the mean lightning strike frequency across all natural and urban cells (i.e., the region). To assess the density of urbanization within each city, we calculated the percentage of land covered by natural areas within the urban cells of each city (hereafter, greenspace). We also calculated the local effect of urbanization on temperature, precipitation, and all aerosol variables. Specifically, we divided the average values of these predictors in the urban areas by their average across all cells in the associated natural areas, and we referred to these variables as the “variable” ratio (e.g., temperature ratio or precipitation ratio). This allowed us to determine if lightning enhancement was directly associated with the effect of urbanization on local climate and pollution, such as the urban heat island effect (i.e., urban temperature divided by natural area temperature). We log-transformed overdispersed variables before analysis (12 of the 17 fixed-effect predictors were transformed; average temperature, local precipitation, total elevation, absolute latitude, and greenspace were not transformed). Because the annual data for aerosol depth and urban-natural strike ratio included 3 and 4 zero values, respectively, we added half the smallest positive value (0.0020 for aerosol optical depth and 0.0446 for urban-natural strike ratio) to each variable before transformation.

### **Model averaging**

We used Akaike Information Criterion (AIC) model averaging to explore spatiotemporal variation in the likelihood and magnitude of the lightning enhancement effect. To evaluate the probability of enhancement, we constructed a generalized linear model with a binary response variable indicating whether there was lightning enhancement (determined by a threshold of Glass' delta  $\geq 0.5$ ). This model included a single value for each city with 17 predictors averaged across all years (349 observations). To explore spatiotemporal variation in enhancement strength, we assessed how the urban-natural strike ratio varied among cities with unambiguous enhancement using annual data from 2013-2018 (218 cities with 1,217 city-year observations). Specifically, we constructed a mixed-effect linear model (fitted with the lmer function of the lme4 package; Bates et al., 2015) with an urban-natural strike ratio as the response variable, a random effect for the city (accounting for the annual lightning variation of each city), and the same collection of 17 fixed-effect predictors (Table 2; representing the linear relationships between these predictors and the response variable). We used unique annual values for all variables with yearly data (i.e., all lightning, climate, and pollution variables). We note that some variables were omitted from this final set of predictors (i.e., mean maximum temperature, mean minimum temperature, the ratios between urban and natural areas for these two variables, and the total concentration of NO<sub>2</sub>) because of collinearity, as determined by Pearson correlations ( $R > 0.7$ ) and variance inflation factors ( $VIF > 5$ ).

We fitted models for every possible combination of these terms (function dredge). Then, we averaged all models with AICc values within 4 of the lowest AICc values (function model.avg in package MuMIn; Barton, 2010). We scaled all variables (Z-transformation) to allow direct comparison of coefficients, and we identified significant predictors as model-averaged coefficients with 95% confidence intervals that did not overlap with zero. Additionally, we performed forward model selection and assessed whether including pairwise interaction terms between the significant predictors decreased model AIC. We verified the appropriate model fit and the need for all transformations by evaluating model residuals (e.g., Q-Q plots).

## Data Table

**Table name(s):** Pixels\_lightning\_month\_year\_v2.csv

**Table description(s):** We used this table to create the two datasets needed to run the model and estimate the strength and magnitude of the lightning enhancement over urban areas. Based on this dataset, we aggregated the data by city by year (to measure the strength of the lightning enhancement) and only by city (to measure the magnitude of the lightning enhancement).

Column	Description	Units	Code explanation or date format	Empty values code
city	City			
country	Country			

month	Month of analysis		Jan: January, Feb: February, Mar: March, Apr: April, May: May, Jun: June, Jul: July, Aug: August, Sep: September, Oct: October, Nov: November, Dec: December	
year	Year of analysis			
Area_urban	Area of the cities (aggregation of cells with more than 50% of urban coverage).	Km <sup>2</sup>		
Lat_city and lon_city	Latitude and longitude of the city center are defined in the United Nations, Department of Economic and Social Affairs, Population Division (2018). World Urbanization Prospects: The 2018 Revision, Online Edition, respectively.	degrees		
Urban_area	Area of the city defined by the United Nations, Department of Economic and Social Affairs, Population Division (2018). World Urbanization Prospects: The 2018 Revision, Online Edition.	Km <sup>2</sup>		
Sd_CG_urban	The standard deviation of the Cloud-to-Ground (CG) lightning on each urban cell aggregated by city, month, and year.	number of CG lightning km <sup>2</sup> year <sup>-1</sup>		NA
Mean_CG_urban	Mean of the CG lightning on each urban cell aggregated by city, month, and year.	number of CG lightning km <sup>2</sup> year <sup>-1</sup>		NA
Sd_total_urban	The standard deviation of the total lightning (CG and in-cloud lightning) on each urban cell aggregated by city, month, and year.	number of total lightning km <sup>2</sup> year <sup>-1</sup>		NA
Mean_total_urban	Mean of the total lightning on each urban cell aggregated by city, month, and year.	number of total lightning km <sup>2</sup> year <sup>-1</sup>		NA
Per_urban	Average percentage of urban coverage in the urban cells or	%		NA

	pixels (>50%) that make up the area of the city.			
Per_natar_urban	Average percentage of natural coverage in the urban cells or pixels (>50%) that make up the area of the city.	%		NA
Pop_2018	Population of 2018 defined in the United Nations, Department of Economic and Social Affairs, Population Division (2018). World Urbanization Prospects: The 2018 Revision, Online Edition.	Thousands of people		NA
Mean_elev_urban	Average elevation of the urban cells aggregated by city, month, and year.	masl		NA
Mean_prec_urban	Average precipitation of the urban cells aggregated by city, month, and year.	kg m <sup>-2</sup> month <sup>-1</sup>		NA
Mean_tas_urban, mean_tmax_urban, mean_tmin_urban	Average mean, minimum and maximum temperature of the urban cells aggregated by city, month, and year.	K		NA
Mean_aer_urban	Average aerosol optical depth of the urban cells aggregated by city, month, and year.	µm of particulates scaled from 0 to 1		NA
Mean_sulf_urban	Average sulfate dioxide of the urban cells aggregated by city, month, and year.	µg m <sup>-3</sup>		NA
Mean_nitr_urban	Average nitrate dioxide of the urban cells aggregated by city, month, and year.	billion molecules mm <sup>-2</sup>		NA
Area_natar	Area of the natural ecosystems (cells of non-human ecosystems; aggregation of cells with more than 90% of natural coverage).			NA
Sd_CG_natar	Standard deviation of the Cloud-to-Ground (CG) lightning on each natural cell aggregated by city, month, and year.	number of CG lightning km <sup>2</sup> year <sup>-1</sup>		NA
Mean_CG_natar	Mean of the CG lightning on each natural cell aggregated by city, month, and year.	number of CG lightning km <sup>2</sup> year <sup>-1</sup>		NA

Sd_total_natar	Standard deviation of the total lightning (CG and in-cloud lightning) on each natural cell aggregated by city, month, and year.	number of total lightning km <sup>2</sup> year <sup>-1</sup>		NA
Mean_total_natar	Mean of the total lightning on each natural cell aggregated by city, month, and year.	number of total lightning km <sup>2</sup> year <sup>-1</sup>		NA
Per_natar	Average percentage of natural coverage in the natar cells or pixels (>90%) that make up the area of the natural ecosystem associated with each city.	%		NA
Mean_elev_natar	Average elevation of the natural cells aggregated by city, month, and year.	masl		NA
Mean_prec_natar	Average precipitation of the natural cells aggregated by city, month, and year.	kg m <sup>-2</sup> month <sup>-1</sup>		NA
Mean_tas_natar, mean_tmax_natar, mean_tmin_natar	Average mean, minimum and maximum temperature of the natural cells aggregated by city, month, and year, respectively.	K		NA
Mean_aer_natar	Average aerosol optical depth of the natural cells aggregated by city, month, and year.	µm of particulates scaled from 0 to 1		NA
Mean_sulf_natar	Average sulfate dioxide of the natural cells aggregated by city, month, and year.	µg m <sup>-3</sup>		NA
Mean_nitr_natar	Average nitrate dioxide of the natural cells aggregated by city, month, and year.	billion molecules mm <sup>-2</sup>		NA
Glass_delta	Effect size of the CG lightning between urban and natural areas to define unambiguous lightning enhancement.			NA
CG_ratio	Ratio between CG lightning in urban and natural areas.			NA
Lat_group	Location of the city based on their latitude. Nothern (>23.5°), tropical (23.5° < city < -23.5°) and southern (<-23.5°).			NA
Total_ratio	Ratio between the average total lightning in urban and natural areas.			NA



Elev_ratio	Ratio between the average elevation between urban and natural areas.			NA
Prec_ratio	Ratio between the average precipitation between urban and natural areas.			NA
Tas_ratio, tmax_ratio, tmin_ratio	Ratio between the average mean, minimum and maximum temperature between urban and natural areas, respectively.			NA
Aer_ratio	Ratio between the average aerosol optical depth between urban and natural areas.			NA
Sulf_ratio	Ratio between the average sulfate dioxide between urban and natural areas.			NA
Nitr_ratio	Ratio between the average nitrate dioxide between urban and natural areas.			NA
Longitude_km	Longitude in kilometers.	km		NA
Latitude_km	Latitude in kilometers.	km		NA
Dist_to_water	Shortest distance to large water bodies among the urban cells per city.	km		NA

### Ancillary files: software, code, protocols

- Analyzing\_the\_dataset\_08142024.R
- creating\_urban\_lightning\_dataset\_08142024.R

## Data provenance

Group	Variable	Data type	Dataset DOI or URL	Creator (name & email)	Contact (name & email)
Climate	Regional lightning frequency	Electrical ground sensor network	<a href="https://ams.confex.com/ams/91Annual/webprogram/Paper183895.html">https://ams.confex.com/ams/91Annual/webprogram/Paper183895.html</a>	Liu and Heckman (sheckman@earthnetworks.com)	Stan Heckman (sheckman@earthnetworks.com)
	Average air temperature	Reanalysis of weather station data	<a href="https://doi.org/10.1038/sdata.2017.122">https://doi.org/10.1038/sdata.2017.122</a>	Karger, D. N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (dirk.karger@wsl.ch)	Dirk Nikolaus Karger (dirk.karger@wsl.ch)
	Maximum air temperature				
	Minimum air temperature				
	Local precipitation		<a href="https://doi.org/10.1038/sdata.2017.122">https://doi.org/10.1038/sdata.2017.122</a>		
Pollution	Total aerosols	Satellite sensors	<a href="https://doi.org/10.1038/nature01091">https://doi.org/10.1038/nature01091</a>	Kaufman, Y. J., Tanré, D., & Boucher, O. (kaufman@climate.gsfc.nasa.gov)	Yoram J. Kaufman (kaufman@climate.gsfc.nasa.gov)
	NO <sub>2</sub>		<a href="https://doi.org/10.5067/Aura/OMI/DATA2017">https://doi.org/10.5067/Aura/OMI/DATA2017</a>	Krotkov, N. A., & Veefkind, P. (nickolay.a.krotkov@nasa.gov)	Nickolay A. Krotkov (nickolay.a.krotkov@nasa.gov)
	SO <sub>2</sub>	Reanalysis of satellite data	<a href="https://doi.org/10.1175/JCLI-D-16-0758.1">https://doi.org/10.1175/JCLI-D-16-0758.1</a>	Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., ... Zhao, B. (ron.gelaro@nasa.gov)	Ronald Gelaro (ron.gelaro@nasa.gov)
Topography & Geography	Elevation	Satellite sensors	<a href="https://globalsolaratlas.info">https://globalsolaratlas.info</a>	Solargis (contact@solargis.com)	Solargis (contact@solargis.com)
	Distance to water bodies	Satellite sensors	<a href="https://www.marineregions.org">https://www.marineregions.org</a>	Flanders Marine Institute (info@marineregions.org)	Salvador Fernández Bejarano (info@marineregions.org)

Urban chars.	Urban area	Satellite sensors	<a href="https://lpdaac.usgs.gov/products/mcd12q1v006/">https://lpdaac.usgs.gov/products/mcd12q1v006/</a>	Friedl, M., & Sulla-Menashe, D. (friedl@bu.edu)	Mark Friedl (friedl@bu.edu)
	Population	Population census	<a href="https://population.un.org/wup/Download/">https://population.un.org/wup/Download/</a>	UN (population@un.org)	Patrick Gerland (population@un.org)
	Greenspace	Satellite sensors	<a href="https://lpdaac.usgs.gov/products/mcd12q1v006/">https://lpdaac.usgs.gov/products/mcd12q1v006/</a>	Friedl, M., & Sulla-Menashe, D. (friedl@bu.edu)	Mark Friedl (friedl@bu.edu)