Supplementary Appendix for

Urban Air Pollution May Enhance COVID-19 Case-Fatality and Mortality Rates in the United States

Donghai Liang[1]*, PhD, Liuhua Shi[1]*, ScD, Jingxuan Zhao[2], MPH, Pengfei Liu[3], PhD, Joel Schwartz[4], PhD, Song Gao[5], PhD, Jeremy Sarnat[1], ScD, Yang Liu[1], PhD, Stefanie Ebelt[1], ScD, Noah Scovronick[1], PhD, Howard H Chang[6], PhD

1 Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

2 Surveillance and Health Services Research Program, American Cancer Society, Atlanta, Georgia, USA

3 School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

4 Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

5 GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI, USA

6 Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

Correspondence to
Donghai Liang, PhD
Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University
1518 Clifton Rd NE, Atlanta, GA 30322
Telephone: (404)712-9583
E-mail: donghai.liang@emory.edu

* These authors contributed equally to this work

Supplementary Tables and Figures

Table S1 Model Effect Estimates on Zero-inflated Negative Binomial Mixed Models to Examine the Associations between Long-term Exposure to Air Pollution and COVID-19 case-fatality rate or mortality

	COVID-19 Case-Fatality			COVID-19 Mortality		
Pollutant	**Main Effect Estimate***	**95% Confidence Interval**	**p-values**	**Main Effect Estimate**	**95% Confidence Interval**	**p-values**
Single Pollutant Model—Non-zero Components						
NO_2	1.015	1.003 to 1.028	0.02	1.023	1.007 to 1.039	<0.001
$PM_{2.5}$	0.996	0.953 to 1.041	0.87	1.049	0.995 to 1.107	0.08
O_3	0.993	0.975 to 1.010	0.42	0.986	0.964 to 1.008	0.22
3- Pollutant Model—Non-zero Components						
NO_2	1.016	1.003 to 1.029	0.02	1.022	1.005 to 1.038	<0.001
$PM_{2.5}$	0.991	0.947 to 1.037	0.70	1.054	0.996 to 1.115	0.07
O_3	0.992	0.974 to 1.010	0.36	0.979	0.957 to 1.002	0.08
Single Pollutant Model—Zero Components						
NO_2	0.963	0.938 to 0.988	<0.001	0.943	0.917 to 0.969	<0.001
$PM_{2.5}$	0.843	0.779 to 0.912	<0.001	0.689	0.630 to 0.754	<0.001
O_3	0.860	0.828 to 0.892	<0.001	0.792	0.760 to 0.825	<0.001
3-Pollutant Model—Zero Components						
NO_2	0.969	0.945 to 0.994	0.02	0.949	0.922 to 0.978	<0.001
$PM_{2.5}$	0.902	0.831 to 0.980	0.01	0.772	0.703 to 0.848	<0.001
O_3	0.869	0.838 to 0.902	<0.001	0.804	0.772 to 0.838	<0.001

*****Effect estimate based on per unit increase in air pollutants

Table S2 Moran's I test for spatial autocorrelation in residuals from tri-pollutant models for COVID-19 Case-fatality Rate and Mortality Rate for each US state

	Case-fatality Rate		Mortality Rate	
State	**Moran's I**	***p*-value**	**Moran's I**	***p*-value**
Alabama	0.023005	0.666791	-0.016454	0.997143
Arizona	-0.069966	0.971086	0.159086	0.183565
Arkansas	-0.045066	0.776283	0.045421	0.548356
California	0.024098	0.498309	0.058227	0.219995
Colorado	-0.019308	0.900768	-0.069735	0.60085
Connecticut	-0.249311	0.733068	-0.213497	0.835009
Delaware	-0.338612	1	NA	NA
District of Columbia	NA	NA	NA	NA
Florida	-0.033006	0.771992	-0.046262	0.617719
Georgia	0.056318	0.172308	0.051514	0.200098
Idaho	-0.053368	0.95988	-0.003237	0.717008
Illinois	0.056139	0.412049	0.096366	0.177248
Indiana	-0.116814	0.169918	0.172651	0.02579
Iowa	-0.054103	0.568083	-0.027161	0.8795
Kansas	0.07953	0.061874	-0.03842	0.901546
Kentucky	-0.047995	0.623543	0.062124	0.36778
Louisiana	-0.029227	0.902967	0.035875	0.539011
Maine	-0.135679	0.44648	-0.141616	0.384822
Maryland	0.028251	0.611963	0.236883	0.041622
Massachusetts	-0.041547	0.826158	-0.096154	0.947948
Michigan	-0.03011	0.730351	0.038998	0.252498
Minnesota	0.019843	0.417719	-0.01777	0.986115
Mississippi	0.096789	0.043456	0.044718	0.271883
Missouri	0.030381	0.500362	0.073753	0.192159
Montana	0.052994	0.060993	0.076018	0.031501
Nebraska	-0.040916	0.86608	-0.059581	0.722833
Nevada	0.337241	0.038866	0.38327	0.035115
New Hampshire	0.131492	0.042746	0.103783	0.081307
New Jersey	0.175217	0.087009	0.126441	0.178484
New Mexico	-0.045416	0.979142	0.153729	0.258981
New York	0.013377	0.579773	-0.033898	0.807794
North Carolina	-0.050907	0.34774	-0.032161	0.647075
North Dakota	0.109892	0.49598	0.354751	0.100111
Ohio	-0.041737	0.707563	0.107715	0.139459
Oklahoma	0.06193	0.010202	0.102011	0.000142
Oregon	0.007053	0.350889	0.019009	0.24168
Pennsylvania	0.108983	0.133666	0.348912	0.000011
Rhode Island	-0.256215	0.973717	-0.225587	0.897816
South Carolina	0.062342	0.49757	0.102716	0.23941

South Dakota	-0.034629	0.740424	-0.059895	0.913713
Tennessee	-0.007701	0.958532	0.110452	0.166716
Texas	0.00567	0.215483	-0.00599	0.965107
Utah	0.445684	0.000382	0.152707	0.073403
Vermont	0.019688	0.629256	0.269936	0.117271
Virginia	0.410102	0.000001	0.298699	0.000003
Washington	0.005076	0.781592	0.05457	0.507951
West Virginia	0.112795	0.000103	0.187581	0.06824
Wisconsin	-0.039313	0.67382	0.095091	0.018566
Wyoming	-0.215725	0.005326	-0.148779	0.047632

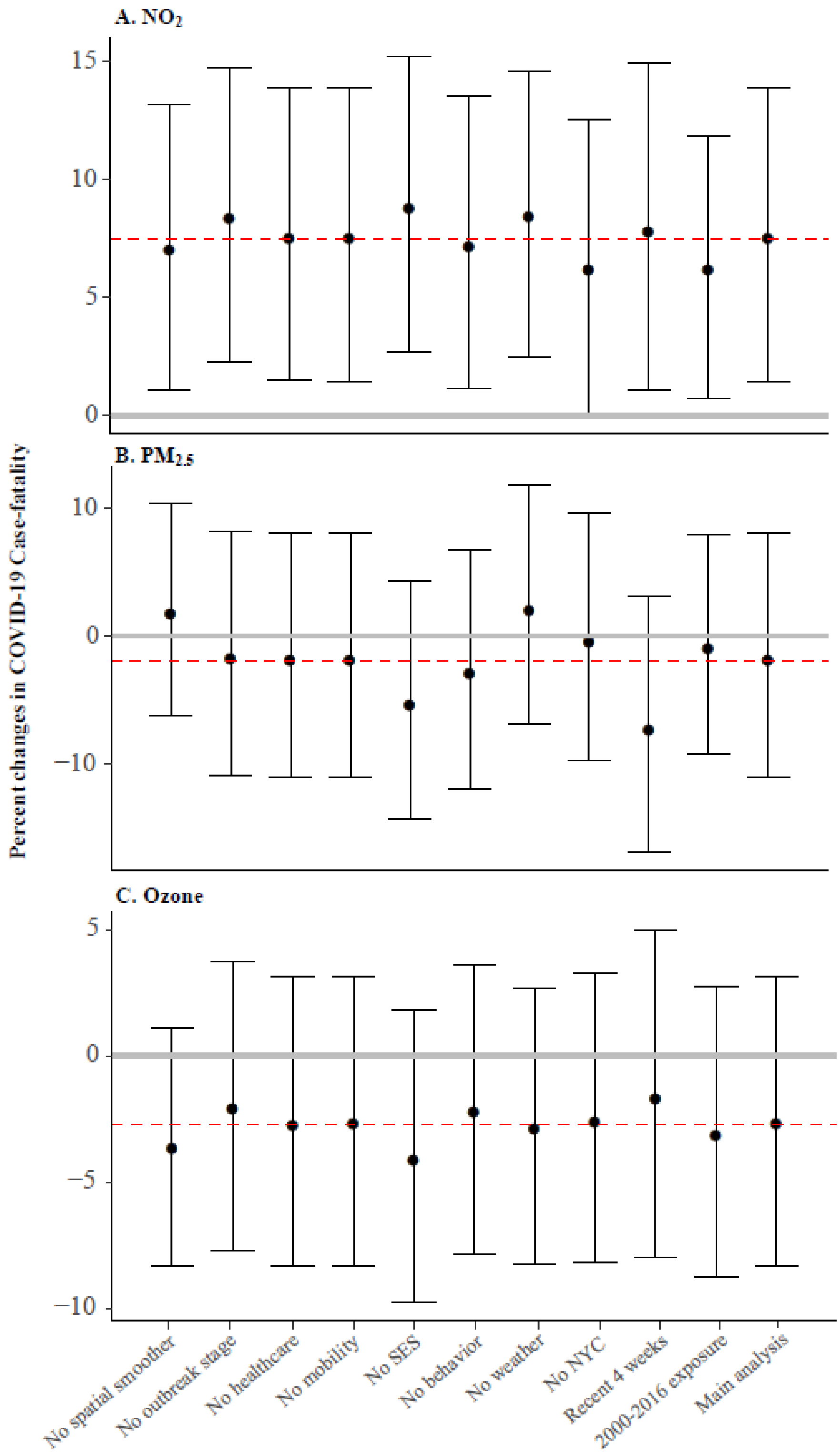

Figure S1 Percent Change in COVID-19 Case-fatality Rate Per Inter Quartile Range (IQR) increase in (A) NO_2, (B) $PM_{2.5}$, and (C) Ozone Concentrations in the Sensitivity Analysis. The red line represents the estimated effects in the main analysis. All results were derived from the tri-pollutant models.

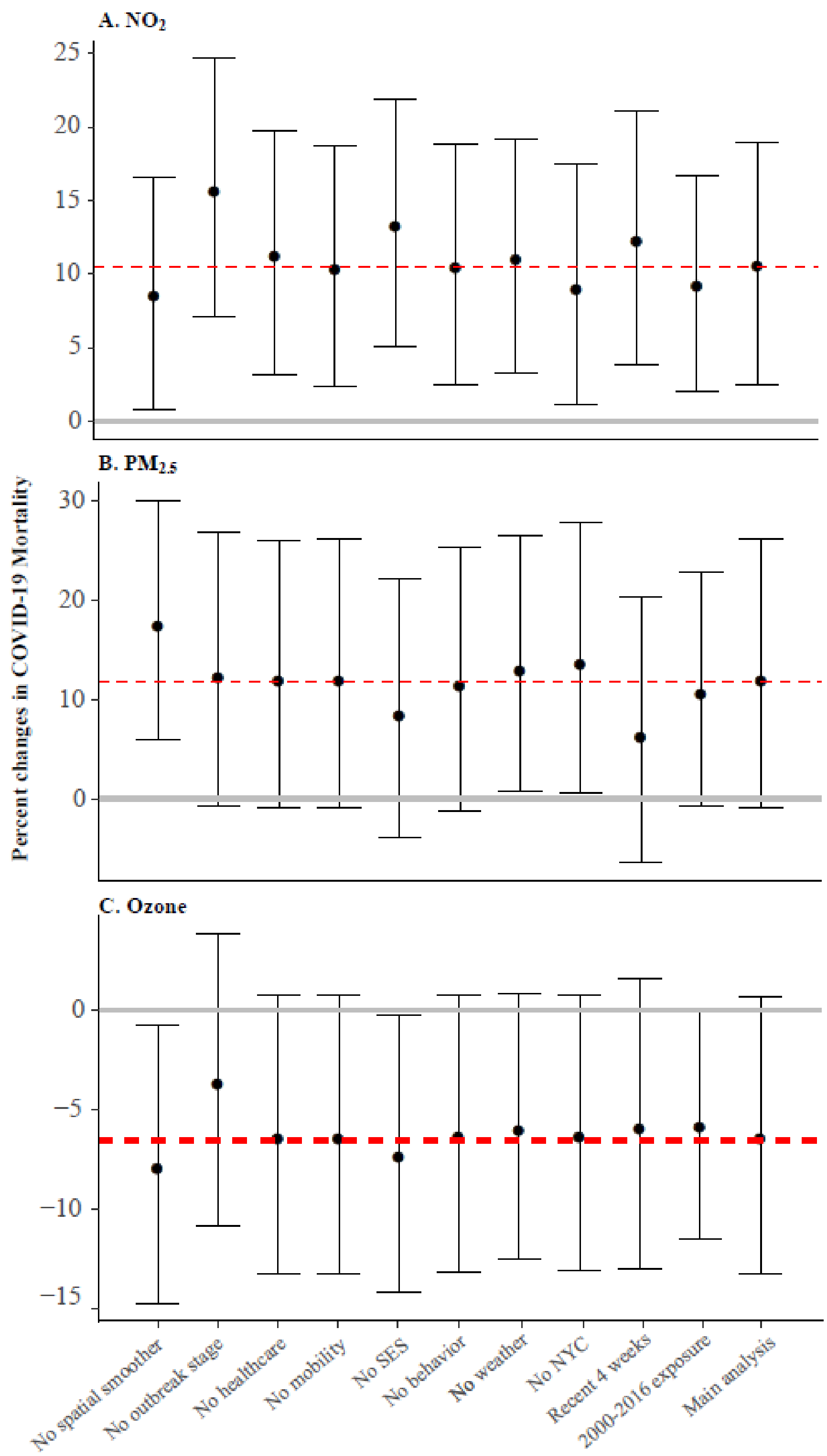

Figure S2 Percent Change in COVID-19 Mortality Rate Per Inter Quartile Range (IQR) Increase in (A) NO_2, (B) $PM_{2.5}$, and (C) Ozone Concentrations in the Sensitivity Analysis. The red line represents the estimated effects in the main analysis. All results were derived from the tri-pollutant models.

Appendix-Technical Appendix

COVID-19 case-fatality rate: We obtained the number of daily county-level COVID-19 confirmed cases and deaths that occurred from January 22, 2020, the day of first confirmed case in the US, through April 29, 2020 in the US from three databases: the New York Times, the USAFACTS, and 1Point3Acres.com. Each of these databases provide real-time data by retrieving information on official reports from state and local health agencies. After data acquisition from these sources, we compared the number of confirmed COVID-19 cases and deaths in each US county (identified by the Federal Information Processing Standards, FIPS code) across all databases for accuracy and consistency. In case of discrepancy, county-level case and death number were corrected by manually checking the data reported from the corresponding state and local health department websites. We calculated county-level COVID-19 case-fatality rate by dividing the number of deaths over the number of people diagnosed with COVID-19 for each US county with at least 1 or more confirmed case, as reported by April 29, 2020. Of all the data reported as of April 29, 2020, confirmed cases and deaths with unassigned counties were excluded in the analysis.

Air pollution: Three major criteria ambient air pollutants were included in the analysis, including NO_2, a traffic-related air pollutant and a major component of urban smog, $PM_{2.5}$, and O_3. We recently estimated daily ambient $PM_{2.5}$, NO_2, and O_3 levels at 1 km^2 spatial resolution across the contiguous US an ensemble machine learning model with ground measurements, satellite-data products, chemical transport model output, meteorological and land-use information as predictors[22,23]. We calculated the daily average for each county based on all covered 1 km^2 grid cells, and then further calculated the annual mean (2010-2016) for $PM_{2.5}$ and NO_2 and the warm-season mean (2010-2016) for O_3, defined as May 1 to October 31, as surrogates for long-term $PM_{2.5}$, NO_2, and O_3 exposures, respectively. More recent exposure data were not available at the time of this analysis. However, county-specific mean values of an air pollutant among different years are highly correlated.

Covariates: We compiled county-level information for several covariates that could also explain heterogeneity in the observed COVID-19 rates and may confound associations with long-term air pollution exposure. Healthcare capacity was measured by the number of intensive care unit (ICU) beds, hospital bed, and active medical doctors per 1000 people. Number of ICU beds were based on Kaiser Health News analysis of 2018 and 2019 hospital cost reports filed to the Centers for Medicare & Medicaid Services. Numbers of active medical doctors and hospital bed of 2017 were obtained from the Area Health Resources Files. Based on the number of COVID-19 tests performed in each state, we calculated a positive rate (i.e., the percentage of specimens tested that are positive for COVID-19). Travel mobility index, based on anonymized location data from smartphones, was used to account for changes in travel distance in reaction to the COVID-19 pandemic. Socioeconomic status (SES) was measured by social deprivation index, a composite measure of area-level deprivation that takes into account income, education, employment, housing, household characteristics, transportation, and demographics. Sociodemographic covariates included population density, percentage of elderly (age $\geq$ 60), and percentage of male. We also obtained behavioral risk factors including population mean BMI and smoking rate, and meteorological variables including air temperature and relative humidity (converted from specific humidity). All covariates were linked to the COVID-19 data using FIPS code and additional details on data source and process are given in the Supplementary Appendix.

Statistical methods: We fit zero-inflated negative binomial mixed models (ZINB) to examine the associations between long-term exposure to $PM_{2.5}$, NO_2, and O_3 and COVID-19 case-fatality rate or mortality. The ZINB model comprises a negative binomial log-linear count model and a logit model for predicting excess zeros[30,31]. The former was used to describe the associations between air pollutants and COVID-19 case-fatality rate among counties with at least one reported COVID-19 case. The latter can account for excess zeros in counties that have not observed a COVID-19 death as of April 29, 2020. We fit single-pollutant, bi-pollutant, and tri-pollutant models, with all analyses conducted at the county level. For the negative binomial count component, results are presented as percent change in case-fatality rate or mortality rate per interquartile range (IQR) increase in each air pollutant concentration. IQR was calculated on national levels. Similar results are presented as odds ratio for the excess zero component. We included a random intercept for each state because observations within the same state tend to be correlated due to similar COVID-19 responses, quarantine and testing policies, healthcare capacity, sociodemographic, and meteorological conditions.

As different testing practices may bias outcome ascertainment, we adjusted for state-level COVID-19 test positive rate (i.e. high positive rate might imply that the confirmed case numbers were limited by the ability of testing, and the case-fatality can be biased high). To model how different counties may be at different time points of the epidemic curve (i.e., phase-of-epidemic), we adjusted for days both since the first case and since the 100^{th} case (i.e., case counts reaching 100) within a county through April 29 as a measure of epidemic timing. To account for how people may have reacted to the social distancing guidelines imposed during the COVID-19 outbreak, we adjusted for county-level travel mobility index. In addition, we considered potential confounding by county-level healthcare capacity, sociodemographic, SES, behavior risk factors, and meteorological factors. Because county-specific population densities span 5 orders of magnitude, we adjusted for density using a logarithmic transformation. To control for potential residual spatial trends and confounding, we included spatial smoothers within the model using natural cubic splines with 5 degrees freedom for both county centroid latitude and longitude. We further calculated Moran's I of the standardized residuals of tri-pollutant main models for each state, to examine the presence of spatial autocorrelation in the residuals.

Data sources on covariates: We adjusted for three county-level healthcare capacity covariates, including the number of intensive care unit (ICU) beds, hospital bed, and active medical doctor per 1000 people. Number of ICU beds were based on Kaiser Health News analysis of 2018 and 2019 hospital cost reports filed to the Centers for Medicare & Medicaid Services. Numbers of active medical doctors and hospital beds of 2017 were obtained from the Area Health Resources Files. State-level number of COVID-19 tests performed up to April 29, 2020 was derived from the Covid Tracking Project, based on which we calculated the positive rate in each state, i.e. the percentage of tests performed that are positive for COVID-19. The travel distance mobility data were released from the Descartes Labs and mapped by the GeoDS Lab using anonymized location data from smartphones (Warren and Skillman, 2020; Gao et al., 2020). The travel mobility index was a measure to compare the daily individual-level travel distance pattern to that in February. To enhance privacy, individual data are de-identified and aggregated to the county level. We calculated the county-level mean mobility index from March 1, 2020 to April 29, 2020 to represent the dramatic mean human mobility changes in reaction to the COVID-19. County-level socioeconomic status (SES) in 2015 was measured by social deprivation index, which is a composite measure of area-level deprivation based on seven characteristics, including income, education, employment, housing, household characteristics, transportation, and demographics. SDI has commonly served as an area-level composite measure of SES in other studies of health and health outcomes. County-level sociodemographic covariates in 2017 such as percentage of elderly (age$\geq$60) and percentage of male were derived from Area Health Resource Files, and population density was derived from the 2018 US Census. County-level behavioral risk factors, including population mean BMI (an indicator of obesity) and percentage of ever smokers, were derived from the 2011 US CDC Behavioral Risk Factor Surveillance System (BRFSS). From Phase 2 of the North American Land Data Assimilation System (NLDAS-2), we acquired hourly 1/8th degree gridded near-surface air temperature and specific humidity data from January 22, 2020 through April 29, 2020 (Xia et al., 2012), based on which we calculated the mean temperature and relative humidity for each 1/8th degree grid. We linked each county's centroid to the nearest 1/8th degree grid and assigned the mean temperature and relative humidity.

Sensitivity analyses: We also conducted a series of sensitivity analyses to test the robustness of our results to outliers, confounding adjustment, and epidemic timing (Supplementary Appendix Figures S1 and S2). Given that New York city has far higher COVID-19 cases and deaths than any other regions in the US, which can be a very influential observation, we excluded all five counties within New York city and repeated the analysis. In another set of sensitivity analyses, we restricted the study only to the most recent 4 weeks (April 1 to April 29), when the case count and death count may be more reliable and accurate than earlier periods and when COVID-19 tests were more available. We also conducted sensitivity analysis by using air pollution data averaged between 2000 to 2016. To assess the impact of potential bias of individual covariates, we fit models by omitting a different set of covariates for each model iteration while comparing effect estimates. Statistical tests were 2-sidedand statistical significance was determined with an alpha of 0.05. All statistical analyses were conducted used R version 3.4.

Technical Appendix Reference

The New York Times; 2020. https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html. Accessed 04/19/2020.

USAFACTS; 2020. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/. Accessed 04/19/2020.

1Point3Acres.com; 2020. https://coronavirus.1point3acres.com/. Accessed 04/19/2020.

Area Health Resources Files; 2019. https://data.hrsa.gov/topics/health-workforce/ahrf. Accessed 04/19/2020.

Social Deprivation Index. Robert Graham Center; 2019. https://www.graham-center.org/rgc/maps-data-tools/sdi/social-deprivation-index.html. Accessed 04/25/2019.

Warren, M.S. and Skillman, S.W., 2020. Mobility Changes in Response to COVID-19. *arXiv preprint arXiv:2003.14228*.

Gao, S., Rao, J., Kang, Y., Liang, Y. and Kruse, J., 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *Available at SSRN 3570145*.

Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko, (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, J. Geophys. Res., 117, D03109, doi:10.1029/2011JD016048.