Table 1: Pearson correlation coefficients between various forms of pollution, including light at different spatial scales. This table shows the correlations calculated using the simulated residences, not the entire map. Numeric Headers refer to resolution (m) of Light at Night measurement. TN: traffic noise, AN: All Noise, BC: Black Carbon, UFP: Ultrafine particles

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **5** | **50** | **100** | **200** | **500** | **1000** | **2000** | **TN** | **AN** | **BC** | **NO2** | **NO** | **PM2.5** | **UFP** |
| **5** | 1.00 | 0.93 | 0.86 | 0.69 | 0.55 | 0.41 | 0.29 | 0.04 | 0.05 | 0.14 | 0.21 | 0.21 | 0.06 | 0.31 |
| **50** | 0.93 | 1.00 | 0.94 | 0.76 | 0.61 | 0.46 | 0.33 | 0.02 | 0.03 | 0.14 | 0.22 | 0.22 | 0.07 | 0.32 |
| **100** | 0.86 | 0.94 | 1.00 | 0.83 | 0.66 | 0.51 | 0.36 | -0.01 | 0.01 | 0.14 | 0.24 | 0.23 | 0.09 | 0.32 |
| **100** | 0.69 | 0.76 | 0.83 | 1.00 | 0.72 | 0.59 | 0.43 | -0.02 | 0.00 | 0.13 | 0.29 | 0.26 | 0.13 | 0.33 |
| **500** | 0.55 | 0.61 | 0.66 | 0.72 | 1.00 | 0.78 | 0.62 | -0.05 | -0.01 | 0.13 | 0.36 | 0.31 | 0.19 | 0.35 |
| **1000** | 0.41 | 0.46 | 0.51 | 0.59 | 0.78 | 1.00 | 0.77 | -0.02 | 0.05 | 0.18 | 0.44 | 0.39 | 0.25 | 0.35 |
| **2000** | 0.29 | 0.33 | 0.36 | 0.43 | 0.62 | 0.77 | 1.00 | 0.01 | 0.08 | 0.19 | 0.44 | 0.35 | 0.18 | 0.35 |
| **TN** | 0.04 | 0.02 | -0.01 | -0.02 | -0.05 | -0.02 | 0.01 | 1.00 | 0.77 | 0.48 | 0.31 | 0.47 | 0.15 | 0.33 |
| **AN** | 0.05 | 0.03 | 0.01 | 0.00 | -0.01 | 0.05 | 0.08 | 0.77 | 1.00 | 0.48 | 0.34 | 0.48 | 0.17 | 0.29 |
| **BC** | 0.14 | 0.14 | 0.14 | 0.13 | 0.13 | 0.18 | 0.19 | 0.48 | 0.48 | 1.00 | 0.52 | 0.52 | 0.40 | 0.54 |
| **NO2** | 0.21 | 0.22 | 0.24 | 0.29 | 0.36 | 0.44 | 0.44 | 0.31 | 0.34 | 0.52 | 1.00 | 0.60 | 0.33 | 0.44 |
| **NO** | 0.21 | 0.22 | 0.23 | 0.26 | 0.31 | 0.39 | 0.35 | 0.47 | 0.47 | 0.52 | 0.60 | 1.00 | 0.23 | 0.51 |
| **PM2.5** | 0.06 | 0.07 | 0.09 | 0.13 | 0.19 | 0.25 | 0.18 | 0.15 | 0.17 | 0.40 | 0.33 | 0.23 | 1.00 | 0.08 |
| **UFP** | 0.31 | 0.32 | 0.32 | 0.33 | 0.35 | 0.35 | 0.35 | 0.33 | 0.29 | 0.54 | 0.44 | 0.51 | 0.08 | 1.00 |

*Table 2: Characteristics of the pollution maps and underlying models.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pollutant | Method | Time period | Notes | Min, Q1, Q2, Q3, Max |
| Light | ISS photograph | March 31, 2013, 01:57 local time | Radiance based mainly on the green band | 0.00, 0.017, 0.036, 0.066, 0.223 |
| All noise | Model based on data about road traffic, aircraft, and railway noise | The datasets that build up the model are from different years. Validation data measured at 103 sites during 8:00-18:00 in February, March, and September of 2003. | Data is averaged over an entire day, although noise during evening and night was increased to account for increased sensitivity of residents. Model reports dB(A). | 3.01, 59.20, 65.00, 71.61, 14.82 |
| Street noise | Model based on road traffic data alone | 6.02, 55.80, 62.38, 69.89, 113.67 |
| Black carbon | Land use regression model | Validation data acquired during July 14-August 16, 2005, during peak afternoon traffic (~16:00-19:00). | Validation based on mobile monitoring of 39 locations. Model reports particle light absorption coefficient in 10-5 m-1. | 0.09, 1.08, 2.31, 4.27, 5.00 |
| NO | Land use regression model | Data acquired during October and November 2009, and April and May 2010. | Measurement data acquired at 116 sites. Model reports parts per billion. | 2.31, 10.93, 16.53, 31.56, 100.00 |
| NO2 | 0.00, 10.84, 12.35, 14.28, 29.52 |
| PM 2.5 | Land use regression model | Data acquired during 8:00-18:00 in February, March, and September of 2003. | Data acquired at 103 sites. Model reports µg/m3. | 0.00, 2.62, 4.24, 5.58, 10.00 |
| Ultrafine particles | Land use regression model | Data acquired during April and May 2010. | Data acquired at 80 sites. Model reports number/cm3. | 8876, 11350, 16377, 28507, 84871 |



Figure : Estimated sample size for 80% power vs. OR per 1 s.d. increase in full resolution light at night. Light resolution of the regression predictor varies by line type. Sample size is on a logarithmic scale for plotting. Results are presented from simulations with outcome rate of 0.05 (a), 0.2 (b), and 0.5 (c) at mean exposure.

**Appendix**

Below we present a glossary of terms. Since our group is interdisciplinary and we hope that the article has a broad readership, we wanted to ensure readers have a place to quickly look up terms that come from research fields outside of their area of experience.

|  |  |  |
| --- | --- | --- |
|  | |  |
| Adverse outcome | | A negative health outcome for a study participant, for example developing breast cancer. |
| Bias | | The expected difference between an estimator and the true value. Here, empirical bias is defined as the mean coefficient measured over 2000 simulations minus the simulation input coefficient (). |
| Bias curve | | A graph showing how bias changes in response to another variable (in our case how strongly an environmental factor is related to an adverse outcome). |
| Berkson error | | An error type arising from using aggregate measures across more data points than just the desired target. This type of error reduces the power of a study, but does not necessarily result in a bias. |
| Black carbon | | Fine particles of carbon that are produced through incomplete combustion (e.g. from biomass burning or diesel engines). A concern for both public health and climate. |
| Confidence interval | | A confidence interval is a prediction of a range in which the true value of a parameter is estimated to lie a given percentage of the time. For example, a 95% confidence interval from [0.5-1.5] means the researchers estimate there is only a 5% chance that the true value is below 0.5 or above 1.5. Keep in mind that at least 1 in 20 experiments the true value lies outside of the 95% confidence interval. |
| Correlated / correlation / Pearson correlation | | If two variables are correlated, then they have a predictable relationship to each other. For example, for a linear correlation, if one value is larger one expects the second will also be larger (e.g. tall parents are more likely to have tall children). The Pearson correlation coefficient specifies the strength of a linear relationship, with 0 being no correlation and 1 being perfect correlation. |
| DMSP-OLS | | The Defense Meteorological Program (DMSP) Operational Line-Scan System (OLS) was a sensor on a US defense satellite able to observe artificial light from space. It was used to produce the first global maps of artificial light pollution, with an intrinsic resolution in the range 2.5km to 5km. |
| DNB | | The Day-Night Band (DNB) is a part of the Visible Infrared Imaging Radiometer Suite (VIIRS) satellite radiometer. It measures light emissions worldwide at a resolution of ~750 meters each night. |
| Ground sample distance | | The distance between the centers of grid cells in a raster map. The ground sample distance could be larger or smaller than the resolution of the data itself. For example, the true resolution of individual DNB images is ~750m, but researchers often use reprojected data that has a smaller ground sample distance (~350 m in Vancouver). |
| LAN | | (Artificial) Light at Night |
| Logistic regression | | A modelling technique in which parameters are estimated so that the probability of a binary (yes-no) outcome can be expressed as a function of selected variables. In this study, we create data sets and perform logistic regression on them to simulate what a researcher may deduce if they had used logistic regression to predict the probability of an adverse outcome using a measure of LAN as a predictor.  If using one predictor variable X to predict the probability of outcome Y=1 (with alternative that Y=0), logistic regression estimates the values and to fit the following function: |
| NO, NO2 | | Nitric oxide and nitrous oxide, two common components of air pollution that cause adverse health effects |
| Noise | All noise:  Street noise: | All noise refers to noise generated from all transit sources (e.g. including rail noise).  Street noise refers to noise generated from activities on city streets (e.g. cars) |
| Odds ratio | | If an event has probability P, the odds of this event are defined as P/(1-P). In some contexts, such as for a logistic regression model, it is more convenient to discuss odds than probabilities. An odds ratio (OR) is the ratio of the odds for two different events. It is equal to  Odds(Event 1)/Odds(Event 2)  For logistic regression, it is common to discuss the odds ratio when the predictor variable X increases by some amount ‘k’. This means we take the odds of the outcome at some X value, and divide by the odds of the outcome at that X value minus k. This odds ratio tells us the strength of the relationship between X and the outcome.  In our study, determines the odds ratio of an increase in the exposure variable. The higher is, the higher the odds ratio will be for an increase in X. We choose to discuss increases in X by one standard deviation of X. |
| Outcome status | | A variable that says whether a person had an adverse outcome (1) or no effect (0). |
| Outcome rate at mean exposure | | For an exposure variable X, when selecting parameter values and to simulate data with, a particular and combination will result in a specific observed rate of the outcome at the mean value of the exposure variable. We select values for our simulations to achieve specific outcome rates at mean exposure. |
| P(Y|X) | | The probability of outcome Y happening, given that we know that X is true. For example, on a weekday I am in the office for 8 hours, so the probability that I am in the office at any given time is 1/3. However, the probability of my head being in my office given that you know that my feet are in the office is 1. |
| PM 2.5 | | A measure of air pollution. Fine particulate matter (PM) with a diameter of 2.5 μm or smaller. |
| Power | | Power is the probability that a study would observe a statistically significant effect when the exposure really does cause adverse outcomes. |
| Power curve | | The power curve shows how power varies depending on how strong the impact of exposure is. |
| PSF | | A point spread function specifies how an imaging system responds to a light source consisting of a single point. For example, when taking a photo of a star, the star will appear as a point if the PSF is narrow, but a blur if the PSF is broad. |
| Raster data | | Geographic data that consists of individual pixels with a single defined value within each pixel. |
| Resolution | | A value related to the size of the smallest object that can be observed in an image. For example, a house may be identified in a high resolution (1 m), but cannot be identified in a low resolution (1 km) image. (C.f. “ground sample distance”.) |
| Sample size | | The number of individual objects examined. Here it is the number of people in a simulated study. |
| Standard deviation | | A measure of the spread that a set of values have. Higher standard deviation means that a variable tends to have a greater spread away from the mean. For this study, because different exposures may take values in very different ranges, we discuss when exposures increase by one standard deviation, instead of when exposures increase by a specific number like 1, because LAN increasing by 1 might be very different than PM2.5 increasing by 1. |
| Statistical significance / significant | | In this case, a significant finding means concluding that there is a relationship between a pollutant and an adverse outcome. Different fields use different standards for statistical significance, in this study we use a significance level of 5%. Note that with this standard, one expects that 5% of studies in which no relationship exists will mistakenly observe a “statistically significant” relationship (a type 1 error). |
| Type 1 error / False positive | | This is when a factor is not related to an adverse outcome, but a study (wrongly) concludes that the factor is associated with the adverse outcome. |
| Type 2 error / False negative | | This is when a factor is related to an adverse outcome, but a study (wrongly) concludes that the factor is not associated with the adverse outcome. |
| UltraFine particles | | A measure of air pollution. Ultrafine particles have a diameter smaller than 0.1 μm. |
| UTM | | The Universal Transverse Mercator coordinate system is used to assign coordinates to locations on Earth, while minimizing spatial distortions due to the curvature of the Earth. |
| Xi | | The true LAN exposure of an individual (as measured using the highest spatial resolution map). |
| X\* | | The LAN exposure of an individual, as measured using a given spatial resolution map. |
| Yi | | Outcome status for an individual. Yi=0 means no adverse outcome, Yi=1 means adverse outcome. |
| λ1 | | A parameter that determines the strength of relationship between the exposure and the outcome. Higher λ1 values result in a stronger relationship between exposure and outcome. |
| λ0 | | A parameter that determines the base outcome rate in the population. |