Melanie Imfeld Word count: 1180

Geographic Information Systems and Science
Hidden Traces – Tracking individual air pollution exposure

Introduction

Air pollution remains a major health hazard in urban environments all over the globe. The four most common air pollutants are Particulate matter (PM), Nitrogen Oxide (NO_2), Sulfur Dioxide (SO_2) and ozone (O_3). Air pollution emerges from a variety of sources such as vehicle exhaust, industry, power generation, domestic fuel burning and construction activities.[1] Research suggests that both PM and NO_2 can cause or aggravate asthma and impair lung development.[2] In response to the adverse effects of air pollution on humans, the European Union has specified standards for the amount of pollutants in the air.[3] Despite congestion charging and a citywide low-emission zone, the administration of London has failed to lower its air pollution levels to the EU limit values.[4]

At the same time, the emerging trend of the 'Quantified Self' has a big impact on health consciousness and self-optimization. Self-tracking data streams such as wearables, mobile phones and cloud-based services are booming.[5]

By using self-tracking, this tool aims to overlay data about pollutants in London's air with the positions from daily commutes and to visualize air pollution exposure at a given place and time within one day.

Data

The following datasets are used to model air pollution exposure along different paths.

London Air Quality Network (LAQN): Air pollution measurements

- Source: The Files are imported from a remote server operated by King's College London through the package Openair in R.[6]
- Chosen variables: Date / time of measurement, coordinates and hourly mean NO_2 measurements for 73 measurement stations.

LAQN Measurement stations: Metadata

- Source: From King's College London through the Openair package in R.
- Chosen variables: Codes of 55 measurement stations in the LAQN.

Google location history

- Source: If enabled, the Google Maps location history records the position of a cell phone. It can be downloaded as JSON through Google Takeouts.[7]
- Chosen variables: Date, time and coordinates for all positions in one day.

---

[1] World Health Organization. (2005). WHO Air Quality Guidelines for particulate matter, ozone, nitrogen sulfur dioxide. Geneva dioxide and, page 5.

[2] Royal College of Physicians. (2016). Every breath we take: The lifelong impact of air pollution.

[3] See footnote 1

[4] Kelly, F.J., Zhu, T., 2016. Transport solutions for cleaner air. Science 352, 934–936.

[5] Swan, M., 2013. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. Big Data 1, 85–99.

[6] The openair project is led by the Environmental Research Group at King's College London and is available under http://www.openair-project.org/

[7] Google takeouts is available under https://takeout.google.com/settings/takeout

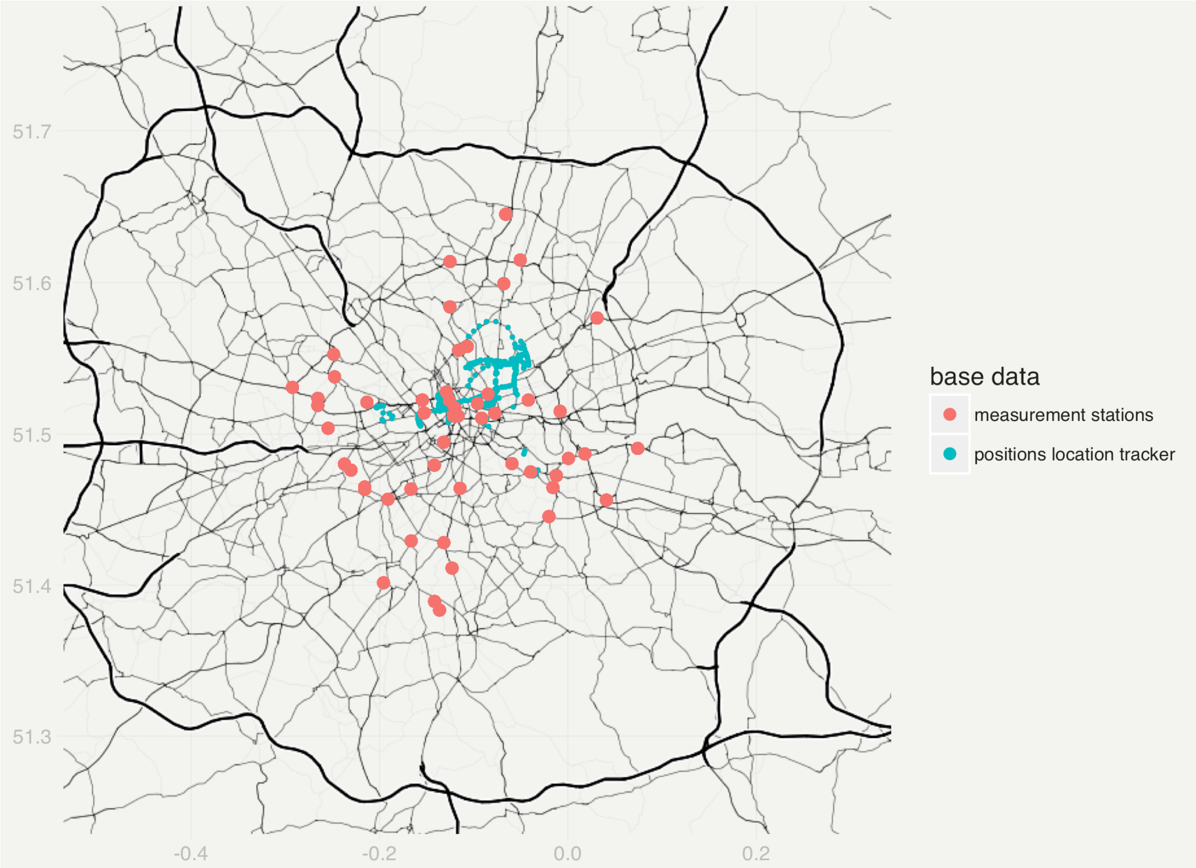


Figure 1: LAQN stations (red) and location history positions (blue)

Analysis

The tool is written in R and the open source environment R Studio. R is chosen because it provides access to the library 'Openair', by which air pollution data can be sourced directly from a remote server operated by King's College London.

The main challenge is that the measurement stations from the LAQN only provide information on air pollution for their specific location (Figure 1). Based on these measurements, the information for other locations can be estimated through a variety of methods. In this paper, spatial interpolation and more specifically Inverse Distance Weighting (IDW) is used to assign an hourly average amount of NO2 in µg/m3 to each recorded position from the Google location history. IDW is chosen because it is a popular nonlinear method with a fairly simple application. The method of IDW assigns a weight to each observed measurement (in this case NO2 measurements), where the sum of all weights is equal to 1.[8] The weight is given by the following expression: [9]

$$w_{ij} = {d_{ij}}^{-b}$$

Where w is the weight, and b a parameter that affects the decay of the weight with distance d between two points i and j. For this tool, several IDW powers *b* were tested. From a visual perspective, a power of $b = 2$ generated the most continuous

[8] Wong, D.W., Yuan, L., Perlin, S.A., 2004. Comparison of spatial interpolation methods for the estimation of air quality data. Journal of Exposure Science and Environmental Epidemiology 14, 404.

[9] Longley, P., Goodchild, M.F.,2005. Geographic information systems. John Wiley & Sons, Ltd, p.94

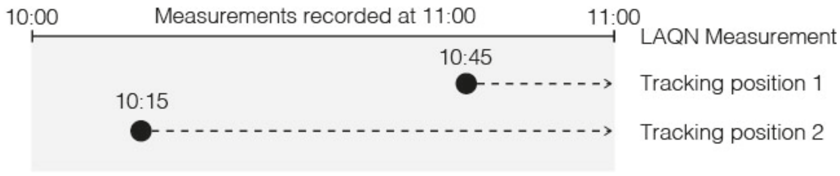


Figure 2: Time approximation for the location history data

results. In addition, the NO2 measurements from all stations are included in the interpolation. The interpolated air pollution values for the location history positions are not only influenced by spatial but also by temporal aspects. The measured hourly average amount of NO2 for one measurement station is not the same throughout the day. Thus, the observed measurements that coincide with a specific position in time of the Google location history have to be selected before interpolation. For that reason, both air pollution measurements and the location history data are binned by hour. The time of each position from the location history is rounded to the next full hour to match the LAQN measurements of hour *t* that are assigned to the next hour *t+1* (Figure 2). The IDW is then performed on each hourly interval.

Figure 3: Timeseries for four different days

Results

The plotted predicted values for the location history dataset suggest different amounts of air pollution exposure at different positions and different times of the day (Figure 3 and 4). A higher amount of NO2 can be observed in the city center compared to the outskirts of the city, where the sample routes from the location history start and end. Tests with several days have not reached a NO2 exposure that exceeds the EU limit of 200 in µg/m3 per hour.[10]

Limitations and Potentials

Currently, the limitations of the tool mostly lie within the method of interpolation. The prediction could be optimized by testing different methods of interpolation, as for example Kriging, Nearest Neighbor methods or Spatial Averaging. To improve the results of the IDW, the mean squared error between measured and predicted value could be calculated to find the ideal IDW power *b* and the optimal number of LAQN measurement stations to be included in the interpolation.[11] In addition, the interpolation lacks precision due to the incomplete recording of the Google location history. Most of the commutes in London are spent underground and therefore not recorded. Generally, no difference in exposure is made between indoors and outdoors.

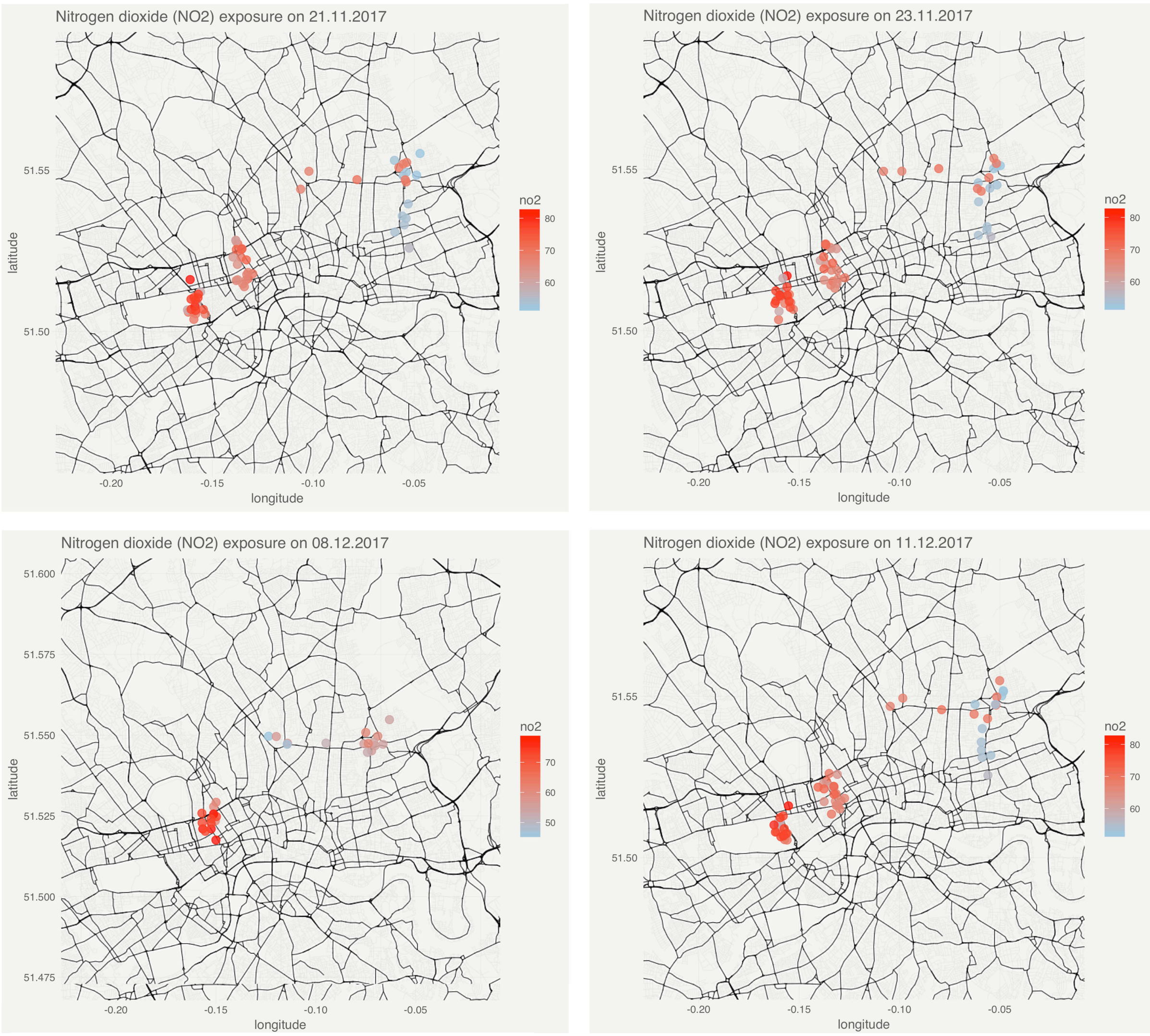


Figure 4: Maps with NO2 exposure for four days

[10] See footnote 1

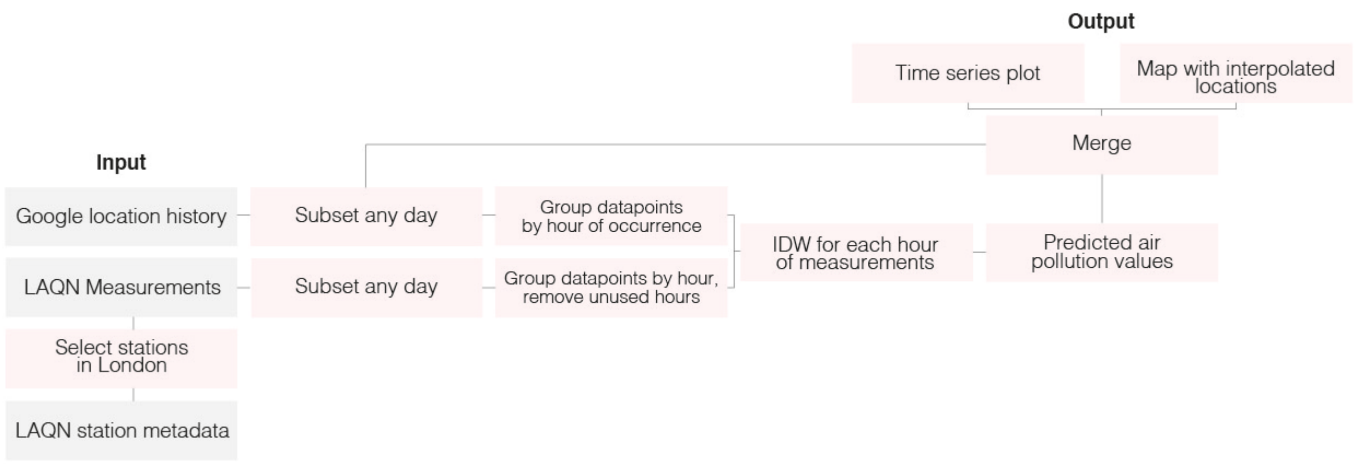
[11] Chen, C., Zhao, N., Yue, T., Guo, J., 2015. A generalization of inverse distance weighting method via kernel regression and its application to surface modeling. Arab Journal of Geosciences 8, 6623–6633.

If applied to other cities around the world, the tool could aid in decision-making and enable its users to avoid areas with high air pollution levels. This is particularly relevant for developing countries. According to the WHO, more than more than 2 million premature deaths each year can be attributed to the effects of air pollution, most of them in developing countries. [12] In addition, the tool could be integrated in commonly used mapping services such as Google Maps or Openstreetmap and could influence decisions about travelling before a trip is taken.

User documentation

1) Install package 'Openair'.
2) Download Google location history as JSON and save file under 'Location History.json' in the project directory. Currently, analysis can only be performed on data from the City of London. There is already a file containing location data from 19.11.2017 to 13.12.2017 in the project folder that can be used.
3) Specify desired date and desired type of air pollution.
4) Run script and save output if desired.

Flow chart



[12] World Health Organization, 2005. WHO Air Quality Guidelines for particulate matter, ozone, nitrogen sulfur dioxide. WHO Press. p. 5.

References

World Health Organization, 2005. WHO Air Quality Guidelines for particulate matter, ozone, nitrogen sulfur dioxide. WHO Press, page 5.

Royal College of Physicians, 2016. Every breath we take: The lifelong impact of air pollution. The Lavenham Press, Suffolk. www.rcplondon.ac.uk/projects/outputs/every-breath-we-take-lifelong-impact-air-pollution

Kelly, F.J., Zhu, T., 2016. Transport solutions for cleaner air. Science 352, 934–936. https://doi.org/10.1126/science.aaf3420

Swan, M., 2013. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. Big Data 1, 85–99. https://doi.org/10.1089/big.2012.0002

Carslaw, D.C., Ropkins, K., 2012. openair — An R package for air quality data analysis. Environmental Modelling & Software 27–28, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

Wong, D.W., Yuan, L., Perlin, S.A., 2004. Comparison of spatial interpolation methods for the estimation of air quality data. Journal of Exposure Science and Environmental Epidemiology 14, 404. https://doi.org/10.1038/sj.jea.7500338

Longley, P., Goodchild, M.F.,2005. Geographic Information Systems. John Wiley & Sons, Ltd, page 94

Chen, C., Zhao, N., Yue, T., Guo, J., 2015. A generalization of inverse distance weighting method via kernel regression and its application to surface modeling. Arab Journal of Geosciences 8, 6623–6633. https://doi.org/10.1007/s12517-014-1717-z

A Brief Introduction to Spatial Interpolation. http://www.bisolutions.us/A-Brief-Introduction-to-Spatial-Interpolation.php (accessed 12.28.17).

Google takeout. https://takeout.google.com/settings/takeout