

Analyzing air pollutants and routing cyclists via official and Citizen Science data in Stuttgart, Germany

1. Introduction

In recent years, the city of Stuttgart has been in the focus of media interest for excessive levels of air pollutants. Exposure to high concentrations of air pollutants like particulate matter (PM), tropospheric ozone (O₃) and nitrogen dioxide (NO₂) can cause a broad spectrum of health issues – particularly of the respiratory system - both in the long and short term. At the same time, while cycling has no direct adverse effect on the air quality - cyclists themselves are directly subjected to air pollution within a city, as cycle paths are usually in immediate proximity to roads. Analyzing and mapping this concentration and distribution of particulate matter - and to map a 'least-polluted' route through particularly affected areas - is a step towards informing about - and protecting users from - air pollution in the city of Stuttgart.

The study will thus concentrate on mapping air pollution within Stuttgart's central districts Nord, Süd, West, Ost, Feuerbach and Bad Cannstatt and offering a basic routing service for cyclists.

2. Literature research

The following papers were considered for an introductory overview of methods for the project:

Using PM monitoring site data, LI ET AL. (2017) offer a literature review of interpolation methods (LI ET AL. 2017, pp. 9103) and identify the Inverse Distance Weight (IDW) method as most widely used in epidemiological studies on air pollution (LI ET AL. 2017, p. 9104).

In an early approach of mapping air pollution using data from low-cost sensors, BRIGGS ET AL. mention Kriging as standard method of geographic interpolation; but developed a multiple regression model fitted for different cities individually (BRIGGS ET AL. 1997, p. 701). The use using weighted land cover instead of simple distance weight (BRIGGS ET AL. 1997, p. 707) appears useful for inner-city mapping. For modelling air pollution, Matějček et al. (2006) used a simple Inverse Distance Weighting method and ordinary Kriging with estimates of variability (MATĚJČEK ET AL. 2006, p. 266).

The Office for the Environment, Measurements and Nature Conservation of the Federal State of Baden-Württemberg (LUBW) has conducted research into the comparability of the Nova Fitness SDS011 sensor used by Luftdaten.info (see 3.1), and a calibrated sensor by Grimm GmbH. The SDS011 sensor performs comparable to sensors used by the LUBW in conditions of 50-70% air humidity and below 20µg/m³ PM concentration. Above these thresholds, or under changing atmospheric conditions, the SDS011 data significantly higher values (LUBW 2017, pp.17). The tests were conducted under 'laboratory' conditions and direct comparison between one LUBW measuring station and one nearby SDS011 (LUBW 2017, pp. 17).

3. Data research

3.1. Air pollutants

The LUBW operates 44 air quality measuring stations within the State of Baden-Württemberg, with 5 stations located within Stuttgart itself. The site provides an hourly average of Ozone (O₃) and Nitrogen dioxide (NO₂); and a moving 24-hour average of particulate matter (PM_{2.5}, PM₁₀).

Upon request, direct access to the measuring station data was not provided by the LUBW. Thus, one goal of the project will be to access and download the output of the LUBW's publicly available data at <https://www.lubw.baden-wuerttemberg.de/luft/messwerte-immissionswerte>.

To supplement these measurements, the Stuttgart-based project Luftdaten.info was brought into life. As a 'Citizen Science'-project, it uses inexpensive sensors and offers multiple APIs at <https://github.com/opendata-stuttgart/meta/wiki/APIs> to access the recorded data.

3.2. Road network and building geometries

The road geometries for the bicycle paths in the area of interest will be downloaded from the OpenStreetMap (OSM) project. There exist several interfaces to access OSM-data, such as the OHSOME API (<https://api.ohsome.org/>) or the Overpass API (<https://z.overpass-api.de/api/interpreter>). Structures will be downloaded to evaluate methods of refining simple interpolation methods.

4. Software tools

4.1. Data download

All of the data sources Offer API

For the download of the air quality and geometries data, both Windows or Linux operating systems provide adequate tools like wget or curl, which can be controlled via simple scripting e.g. in bash. For a scheduled download, tools like cron or Windows Task Scheduler can be used to initiate the script.

4.2. Data storage and ingestion

The public API of the LUBW measuring stations are provided as GeoJSON. The data provided by the Luftdaten.info API is in JavaScript Object Notation Format (JSON). Both datasets will be downloaded perpetually (see 4.1) in regular intervals, creating a time-series dataset. A database with a fitting data model will be implemented for further analysis. PostGIS, an extension to the Free and Open Source relational Database Management System PostgreSQL, offers support of geographic data types and simple geographical analyses.

To ingest the data into the database, the raw downloaded files data need to be parsed and inserted. Both the Python and PostgreSQL's PL/pgSQL scripting language offer support for the (Geo)JSON datatype.

4.3. Comparing official and Citizen Science data

The implications of the LUBW's comparisons between sensors by Grimm and Nova Fitness are twofold:

First, to further assess the comparability of official air quality measurements with Citizen Science data under ‘field conditions’, a simple comparison between either projects in close proximity of each other will be conducted using simple statistical methods, which can be implemented in the database or via scripting languages such as Python. Since the smallest temporal resolution offered by the LUBW is one hour, and the accuracy of the SDS011 data seems to benefit from using aggregate data, the project will use hourly data from the Luftdaten.info project.

4.4. Combining air pollution and road network

The analysis needs to interpolate the measured concentration of particulate matter at one station with its surrounding environment. This analysis aims for a two-dimensional modelling of the atmospheric concentration of PM via Free and Open Source GIS. PostGIS offers a limited range of interpolation methods, such as simple Buffering and Inverted-Distance Weighting. These methods can be expanded using additional FOSSGIS-software such as GRASS GIS. GRASS offers an interface to import and export from PostGIS databases and additional interpolation methods such as Kriging, different methods of Splining (B-Splines, Regularized Spline with Tension) and Inverse Cost Weighting. In light of the studied literature, the latter, as a modified IDW method, appears to be most promising for this project.

4.5. Displaying the results

Geoserver is a FOSSGIS project to publish data from a PostGIS database via Open Web Services (OWS). A Geoserver instance will be established on top of the database providing at least a WebMapService (WMS) of the resulting data.

4.6. Routing

Based on the OpenRouteService project (<https://openrouteservice.org>), a simple routing service is to be offered based on the data analysis in order for users to be able to avoid areas with the highest air pollution within Stuttgart.

5. Sources

- BRIGGS, D. ET AL. (1997) Mapping urban air pollution using GIS: a regression-based approach, *International Journal of Geographical Information Science*, 11 (7), 699-718
- LI, L. ET AL. (2014): Fast Inverse Distance Weighting-Based Spatiotemporal Interpolation: A Web-Based Application of Interpolating Daily Fine Particulate Matter PM_{2.5} in the Contiguous U.S. Using Parallel Programming and k-d Tree. In: *International Journal of Environmental Research and Public Health*, 11 (9), 9101-9141
- LUBW (2017): Messungen mit dem Feinstaubsensor SDS01. Ein Vergleich mit einem eignungsgeprüften Feinstaubanalysator. Online source: https://pudi.lubw.de/detailseite/-/publication/90536-Ein_Vergleich_mit_einem_eignungsgepr%C3%BCften_Feinstaubanalysator.pdf
- MATĚJČEK ET AL. (2006): A GIS-based approach to spatio-temporal analysis of environmental pollution in urban areas: A case study of Prague's environment extended by LIDAR data. In: *Ecological Modelling*. 199 (3), 261-277