Air Quality Adjusted Routing for Cyclists and Pedestrians

Sebastian Müller
Freie Universität Berlin
Takustr. 9
Berlin, Germany
sebastian.mueller@fu-berlin.de

Agnès Voisard
Freie Universität Berlin and Fraunhofer FOKUS
Takustr. 9
Berlin, Germany
agnes.voisard@fu-berlin.de

ABSTRACT
Air quality adjusted routing can provide health benefits for users and encourage users to use more environmental friendly vehicles. This is particularly true for urban areas with varying zones for different air quality levels. We use PM10 total emission data from the state of Berlin, Germany. This data provides the amount of PM10 pollution for a 1x1km grid for Berlin as it represents a major indicator for air pollution. The data is available in Soldner Berlin format (EPSG:3068). As routing software we use the Open Source Routing Machine (OSRM) with map data for Berlin from OpenStreetMaps (OSM) available at geofabrik.de. Our approach uses the possibility of OSRM to access a PostGIS database from a transportation mode profile. The evaluation compares our approach with a baseline approach which assumes the same speed for all roads. Compared to the baseline approach, the routes calculated with our approach tend to avoid the city center as much as possible, but are still acceptably short (for long distances about 22% longer than the shortest path).

Categories and Subject Descriptors
H.2.8 [Database Management]: Database Applications—Spatial Databases and GIS

General Terms
Algorithms, Performance, Experimentation

1. INTRODUCTION
Air quality adjusted routing can provide health benefits for users and encourage users to use more environmental friendly vehicles. Air pollution has strong relations to various kinds of diseases [4]. However, being outside and walking or cycling can have health benefits that in total outweigh the health risks [2]. The best health benefit can therefore be achieved by cycling with the avoidance of air pollution. Navigation systems for cyclists and other transportation modes where one breathes unfiltered air can help the user to avoid air pollution and weigh time and distance against health benefits. The importance of navigation systems is already in the focus of the GIS research community [1].

In this work, we test the feasibility of currently available open-source software and publicly accessible data to provide health-aware navigation options for the public. We see this work as a start to work in a promising advancing field of research. The aim of this work is the integration of the public data into the open-source software to provide a health-aware routing service. We evaluate our work with a scenario test and statistics which show the increase in distance compared to the shortest-path algorithm.

Related work can be categorized into work that has already been done in the same or similar field of research and work about methods and technology which we use in our approach.

Sharker et al.[15] describe a health-aware routing for pedestrians. They describe a weighting model which takes into account static and dynamic parameters of the environment as well as individual user attributes. However, the evaluation is not based on a real-world scenario, but on simulation data which does not prove the practical feasibility of the approach. Nevertheless, walkit.com[19] offers a pedestrian navigation which can be set to consider air-pollution for the routing calculation. The integration of background data is also used in scenic-routing approaches[20]. The example of the integration of background data to the Open Source Routing Machine (OSRM) [5] comes closest to our approach. There, the routing of cyclists includes data from OpenStreetMaps (OSM) and avoids industrial zones. Our approach includes external data which is more directly related to air pollution than the avoidance of industrial zones. However, it is technically similar.

The routing software used by our approach is OSRM[18] which was already introduced to the research community[6]. Furthermore, we also use the possibility to include background data from PostGIS[10] via a LUA[17] transportation mode profile. We use the pollution data from the “Senatsverwaltung für Stadtentwicklung und Umwelt Berlin”[13]. The geographical conversion between the Soldner format (EPSG:3068) and the pseudo-mercator projection from OSM (EPSG:3857) is done in PostGIS[10].
The remainder of this paper is organized as follows: Section 2 describes our approach for background data integration into OSRM with PM10 emission data in Soldner Berlin format. The evaluation is presented in Section 3 and includes test scenarios and statistical information. Section 4 concludes this paper and gives an outlook on future work.

2. BACKGROUND DATA INTEGRATION

We use PM10 total emission data from the state of Berlin, Germany. This data provides the amount of PM10 pollution for a 1x1km grid for Berlin. The data is available in Soldner Berlin format (EPSG:3068). Figure 1 shows the data. The data has high peaks so that we decided to color the grids in groups with equal distribution. Conclusively, the color is not linearly dependent on the emission of PM10. We decided to use the data for the total emission of PM10 as it represents a major indicator for air pollution. Other available data is restricted to traffic, heating with wood or coal, and industry. We decided for the total amount because we cannot really tell the effect of the emission on the immission of the different sources. Some peaks in Berlin are created by coal power plants. These power plants have no direct effect on the neighboring area because they emit, e.g., in a height of 122 meters. However, it cannot be said that every industrial emission has no effect on the neighboring area. Conclusively, we see this as best indicator and are well aware that it is only an indicator, but not a fact. We decided for PM10 because worse pollutants have been more and more eliminated from the city areas by new technology and PM10 is commonly used as air-quality indicator by numerous institutes. Another practical choice would have been PM2.5. However, these data are more or less similarly distributed.

The Soldner format (EPSG:3068) has its reference point at the trigonometric point 1 (TP1) at Müggelberg with the WGS84[7] coordinates 13° 37’ 37”, 9332 (longitude) and 52° 25’ 07”, 1338 (latitude). In order to avoid negative numbers for Berlin, this point is shifted to an artificial point 40000m to the west and 1000m to the south.[12] The reference shape for the Soldner system is the Bessel ellipsoid. The Soldner Berlin system will be replaced in Berlin by the ETRS89/UTM[16] reference system.[14]

As routing software we use the Open Source Routing Machine (OSRM) with map data for Berlin from OpenStreetMaps (OSM) available at geofabrik.de[3]. Our approach uses the possibility of OSRM to access a PostGIS database from a transportation mode profile. Algorithm 1 shows the SQL statement which is used in the LUA transportation mode profile to collect the emission data of PM10 depending on the location of the road for which the weight is being calculated. The table pm_berlin contains the emission data and an attribute geom which represents the bottom left point of the grid. Therefore, another point is created which represents the top right point and then both points are aggregated as an envelope. A ST_DWithin function determines if the road lies within one or more grids.

Algorithm 1 SQL statement in LUA profile

```sql
SELECT AVG(area."PM10_Gesamtemissionen_2008/2009"in_t) AS val
FROM osm_roads way
LEFT JOIN pm_berlin area ON
ST_DWithin(ST_SetSRID(way.geometry,3857),
ST_ENVELOPE(ST_UNION(ST_Transform(
ST_SetSRID(ST_Point(ST_X(area.geom)+1000, ST_Y(area.geom)+1000),3068),
3857),ST_Transform(area.geom,3857)))
),0)
WHERE way.osm_id="..way:id().."
GROUP BY way.id
```

We use the input of the emission to reduce the speed with which the road can be driven on. We classify the emission in 4 groups (the choice of 4 will be explained thereafter), where t represents tons:

1. more than 8t
2. between 5t and 8t
3. between 2t and 5t
4. less than 2t

Within each class we interpolate the speed with the emission. We give the variation within the groups of less emission a higher variability than within the groups of higher emission to be able to still differentiate within the “better” grids. A similar behavior can be described by a root function. However, this function would have to be adjusted to fit the values in the data set best and the illustrative utility would be lost. Figure 2 shows a histogram of the distribution of the PM10 emission values. The y-axis shows the number of entries (#) and the x-axis shows the amount of emitted PM10 (t/km²). The values are sorted into classes with a distance of 0.25 such that each ton of emission has 4 bars. Most of all emission values are below 5 tons per km². The median value is 0.65 and the average value is 1.13. This means that most of the values fall into the last group. This classification into 4 groups provides feasible results for Berlin. However, for different cities the thresholds for the speed reduction should
be evaluated again because other cities have other normal levels of PM10 emission. It would also be possible to adjust these thresholds to the distribution of the data.

Figure 3 shows the architecture of our routing setup. The processes are numbered. The first process is the database import. The OpenStreetMap data from geofabrik.de and the PM10 emission data are imported into the PostGIS database. For the OpenStreetMap data there is a tool (imposm[8]) available which does an automated import of the data. The PM10 emission data has to be downloaded as .xsl file first, then saved as CSV file. Next, the file can be imported into he PostGIS database. Therefore, we made use of QGIS[11] which additionally creates a point geometry (step 2) from the two columns which provide the x and y coordinates. The third step is the OSRM extraction process. For this process, the LUA profile is used to create a connection and perform queries for each road on the PostGIS database with the PM10 emission data. The fourth step (preparation) needs the same input for calculation and additionally the created extraction files from step 3. After this processing, all files are ready to start the web service (5) which offers an API for the OSRM frontend (6).

3. EVALUATION

The evaluation compares our approach with a baseline approach which assumes the same speed for all roads. We show test scenarios as figures and as statistical data with length and length deviation. The tests are performed on a Dell Latitude E7440 notebook with an Intel Core i7-4600U CPU @ 2.10GHz (×4) and 16 GB memory. While the extraction and data preparation are influenced by the integration of background data, the actual route calculation is not. For the tests we used the Berlin-wide OpenStreetMaps data set from geofabrik.de which uncompressed size is 927 MB. The addition of background data increases the time needed for the extraction process from about 4 minutes to about 22 minutes and the preparation process from 5 minutes to 8 minutes. The amount of data used for the routing with OSRM and the routing calculation time did not increase.

Figure 4 shows four different long distance (about 30km) routes calculated by the algorithm with and without using background PM10 emission data. The long distance routes are the most illustrative because they present the largest deviation between the air quality aware routing and the shortest-path routing. Figures 4a and 4c show the routes calculated with and without air quality adjusted routing from south west to north east in Berlin. One can observe the strong avoidance of the city center. The route calculated with air quality adjusted routing makes use of many ways which go through forests and less-polluted, but also less populated, parts of the city. The route is about 22% longer than the reference route. Taking into account that PM10 emission avoidance tends to go along with traffic avoidance, the time difference should be even less. More of less the same accounts for the comparison between the routes on display on the Figures 4b and 4d which show different routes from the north west to the south east in Berlin. In Figure 4d one can see two routes from different starting points which are very close to each other. Both routes show how significantly the algorithm switches from one way around the city center to another way around the center.

Table 1 puts different start and end locations on display. Each entry lists the length of the route calculated by the algorithm using background PM10 emission data (t_{PM10}), the length of the route calculated by the simple shortest-path algorithm (t_{fix}), and the deviation between t_{fix} and t_{PM10} (dev) in percent. The deviation of both lengths is dependent on how long the length itself and on the fact if the route avoids the city center. The first three entries start at Wakenbergstraße and are comparable in terms of length. However, the very first route avoids the center (air quality aware) and therefore the deviation is higher. Both other routes go along areas which are farer to the center. Conclusively, the deviation is smaller, because the benefit of avoiding areas shrinks. The following routes are every time a little smaller and the deviation also shrinks. There are probably two reasons behind this: First, the data is not fine-grained enough to be able to detect exceptions in the city center where the pollution is low, e.g., in Tiergarten. Therefore, possible alternatives through the center are omitted because of a bad
(a) Without air quality adjusted routing from south west to north east in Berlin

(b) Without air quality adjusted routing from north west to south east in Berlin

(c) With air quality adjusted routing from south west to north east in Berlin

(d) With air quality adjusted routing from north west to south east in Berlin

Figure 4: Results with and without air quality adjusted routing
Table 1: statistical data from test scenarios

<table>
<thead>
<tr>
<th>Route</th>
<th>$t_{PM}$</th>
<th>$t_{fix}$</th>
<th>dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wackenbergstraße - Basaltweg</td>
<td>24.6km</td>
<td>20.9km</td>
<td>17.7%</td>
</tr>
<tr>
<td>Wackenbergstraße - Kyllmannstraße</td>
<td>20.9km</td>
<td>19.5km</td>
<td>7.18%</td>
</tr>
<tr>
<td>Wackenbergstraße - Ernst-Haeckel-Straße</td>
<td>18.9km</td>
<td>18.3km</td>
<td>3.28%</td>
</tr>
<tr>
<td>Petersburger Straße - Mecklenburgische Straße</td>
<td>13.6km</td>
<td>11.3km</td>
<td>20.4%</td>
</tr>
<tr>
<td>Heidelberger Straße - Lübecker Straße</td>
<td>10.4km</td>
<td>9.9km</td>
<td>5.6%</td>
</tr>
<tr>
<td>Jungstraße - Auguststraße</td>
<td>6.0km</td>
<td>5.5km</td>
<td>10.9%</td>
</tr>
<tr>
<td>Bissingzeile - Meierottostraße</td>
<td>3.63km</td>
<td>3.45km</td>
<td>5.2%</td>
</tr>
<tr>
<td>Hohenzollerndstraße - Palmzeile</td>
<td>3.53km</td>
<td>3.46km</td>
<td>2.02%</td>
</tr>
<tr>
<td>Hinter dem Zeughaus - Hafenplatz</td>
<td>2.93km</td>
<td>2.89km</td>
<td>1.38%</td>
</tr>
<tr>
<td>Steinstraße - Sophiestraße</td>
<td>0.66km</td>
<td>0.66km</td>
<td>0%</td>
</tr>
</tbody>
</table>

average value for the emission of PM10. Second, having more alternative routes with varying PM10 emission rating increases the search space in terms of area to be compared and therefore more alternative ways away from the shortest path are taken into account.

4. CONCLUSION

An air pollution aware routing with open source software and open data is feasible and straightforward to implement. The results show routes which are practically meaningful and promise a real improvement according to the air quality a bicycle driver exposed to. Compared to the baseline approach, the routes calculated with our approach tend to avoid the city center as much as possible, but are still acceptably short (for long distances at most 22% longer than the shortest path based on the test scenarios).

The PM10 emission data is not a direct measurement of the air pollution, a cyclist is exposed to. Furthermore, the 1x1km grid is not precise enough to find routes which might have less pollution within an area which has on average a high pollution. For example, in Berlin many main roads cross in Tiergarten which is also a park with bicycle ways. Supposedly, the air pollution within the park is very low, because of the filtering effect of the trees. However, on average the PM10 emission is high because of the main roads. Figure 5 shows a 1x1km area in Berlin Tiergarten with a PM10 emission of 3.6881 tons. The emission is one of the higher emission values for Berlin in total and comparable to other areas in the city center. However, most of the area is forest, park, and lake which does not emit any PM10 at all. Supposedly, the main roads around the “Siegessäule” account for most of the PM10 emission due to the high traffic. In order to avoid PM10 pollution one would not use the main road, but cycle through the forest and park which is very suitable. Conclusively, there are ways within this area, but they are not evaluated by the algorithm because the algorithm only uses the average value. In order to be able to use the ways in the park as attractive alternative to other ways one could add the landuse information in OpenStreetMaps to the background information and calculate how far away is this way from the main road and how many trees are between such that this much of pollution is left as exposure on this way. One could even use the information from the “Senatsverwaltung für Stadtentwicklung und Umwelt Berlin” for the different emission sources and distribute traffic emission to “osm_mainroads”, heating with wood or coal emission to “osm_buildings”, and industry emission to “osm_landusages.type=industrial”. While the buildings and industries are represented as area, the area for the main roads has to be estimated by road type, number of lanes, speed, etc. Next, the diffusion of PM10 to other areas would have to be estimated, e.g., a pathway in a park next to the highway is stronger polluted by the neighboring main road than a pathway which is at least 5 km away from any main road. This should be possible the same way as heat maps are generated.

For future work we plan to expand the routing prototype to multiple dimensions. These dimensions are:

1. Transportation Modes: Air quality related background information is not only interesting for bicycles, but for all other transportation modes. The challenge is to create profiles which fit the transportation mode to the effect which can be calculated out of the air quality data. Here, also user interaction can be used to find out the individual effect. Furthermore, the cyclist profile can be improved to the current standard of OSRM with inclusion of surface type and more.

2. Area: We want to provide this routing prototype for multiple cities and cover in total more area. The challenge is the integration of different data sources.
3. Performance: Whilst adding more data the reduction of calculation time is important. A possible first approach to counter the increasing calculation costs would be an in-memory database.

4. Health Data: This prototype is restricted to a PM10 indicator. Possible extensions are ozone, pollen, and excessive temperature. Using the same format, the "Senatsverwaltung für Stadtentwicklung und Umwelt Berlin"[13] provides data for population density, noise, heat load, and other pollutants (nitrogen oxides, sulfur dioxide, PM2.5) which can easily be integrated.

5. Dynamic Routing: With more and more comprehensive dimensions the data preparation costs in a hierarchy-based routing algorithm increase. A dynamic routing approach should be evaluated.

To provide the best utility for the end user, the systems has to include a utility function for an end user. For one user the pollen information does not matter while for another one very much. Furthermore, not everybody is sensible for noise as everybody else. How this is implemented is also dependent on whether a static or dynamic routing approach will be used.

5. REFERENCES