Deriving Hospital Catchment Areas from Mobile Phone Data

Bernd Resch[1,2], Azmat Arif[1], Gautier Krings[3], Guillaume Vankeerberghen[3], Marc Buekenhout[4]

[1]University of Salzburg, Department of Geoinformatics – Z_GIS, Schillerstrasse 30, 5020 Salzburg, Austria
Email: bernd.resch@sbg.ac.at, azmat.arif@stud.sbg.ac.at

[2]Harvard University, Center for Geographical Analysis, 1737 Cambridge Street, Cambridge, MA 02138, USA

[3]Real Impact Analytics, 5 Place du Champ de Mars, 1050 Brussels, Belgium
Email: {gautier.krings; guillaume.vankeerberghen}@realimpactanalytics.com

[4]Email: mbuekenhout@msn.com

Abstract

Delineating catchment areas of medical facilities is essential for estimating the quality of a health-care system and to maximise the efficiency of health service provision. One critical shortcoming of previous approaches are manifested in their comprehensive assumptions about a hospital's patients by using census data or gravity models. In contrast, our approach uses anonymised mobile and landline phone data to derive hospital catchment areas. Our goal is not to assess the quality of the health care system, but to identify the geographic areas, in which people actually use a hospital. Thus, our results reveal new insights into the catchment areas of hospitals by minimising assumptions about demographic factors.

1. Introduction and Related Work

Adequate provision of health services is a central priority of health professionals and policy makers worldwide. More, the efficiency of health care systems, i.e., the provision of best possible service with minimum resources, is critical to public providers (Fransen et al. 2015). These requirements have led to a number of studies to analyse catchment areas and service quality of medical facilities. Even though geospatial analysis methods have existed for decades, there is still a general lack of studies that have mapped and examined health service catchments in practice, not only from a theoretical viewpoint (Schuurman et al. 2006).

Previous approaches for delineating medical catchment areas comprise statistical population-to-provider ratios, gravitational models, travel cost estimation, analysis of the physical distance between hospitals, census-based patient-origin analysis, commuter-based approaches of modelling spatial accessibility (Fransen et al. 2015; Wang and Wheeler 2015), or the two-step floating catchment area method based on the physician-to-population ratio (Luo and Wang 2003). The major drawback of these approaches is that they make far-reaching assumptions about a hospital's patients by applying census data, travel times or gravity models. Moreover, they do not take heterogeneous activity and mobility patterns into account or only derive them from static census data.

Thus, the approach proposed in this paper uses anonymised mobile and landline phone data to delineate hospital catchment areas. Like this, we aim to identify the geographic areas, in which people use a hospital instead of assessing the quality of the health care system per se. Therefore, we analyse calls to and from hospitals in Trinidad and Tobago. This goes beyond just using patient records in that we are able to draw conclusions from a wider range of communication with a hospital (enquiries, arrangement of appointments, follow-up care, visitors, etc.), beyond patients' hospital stays.

2. Data Sources

For retrieving the **hospital polygons**, we manually digitised 117 hospitals in Trinidad and Tobago using Open Street Map (OSM).

The **cell phone network data**, which was provided by a national operator with a market share of more than 50%, contain locations of the cells, including the antennas' mounting angles. The data also contain the properties of each **call** (originating cell, destination cell, duration, etc.). Overall, we used more than 4,000 antennas and about 2.5 billion calls over a period of six months. The call data are fully anonymised, i.e., no conclusions can be drawn to individual subscribers.

In addition to the operator's data, we used network data from **OpenCellID** (http://opencellid.org), a collaborative open-source project collecting data about mobile cells around the world. OpenCellID enrich our provider's data by additional cell locations and the estimated range of each antenna.

3. Methodology and Results

3.1 Determining Service Areas for Mobile Cells

To determine the service areas for each antenna, we use the following parameters: location and range (in meters) from OpenCellID, direction (mounting angle) from the cell phone network operator, and azimuth (beam width).

3.2 Identifying Calls to Hospital Landline Numbers

Next, we identify the calls that were made from and to hospital landline numbers to and from the cell phone network through a simple attribute join on the hospitals' phone numbers. The geometric structure of the resulting graph resembles a star-like pattern with a hospital in its centre, as expected. Through mapping the calls, we can distinguish between the different levels of spatial influence of the hospitals.

3.3 Identifying Calls to and from Cell Phones Located within a Hospital

Then, we determine calls that are made from and to cell phones which are potentially located within a hospital. Therefore, we model each antenna's radiation pattern as a 3-dimensional distribution curve, which reflects the fact that the probability of a cell phone actually being connected to an antenna decreases with distance to the antenna. We approximate the model from Buvaneswari et al. (2007) by using ellipses, as shown in Figure 1.

We developed the following six-step algorithm. 1.) for each cell, we calculate the centre of the ellipse, which represents the **antenna's coverage** given its location and range. 2.) for each cell, we create **ellipse segments** representing a cell's decreasing coverage over distance, according to the coverage model by Buvaneswari et al. (2007). 3.) we orient the ellipses using the antenna's azimuth value, and intersect them with the opening angle of the cell (90°). 4.) we assign a **proportional percentage of the cell's calls to each segment**, which represents the probability that a call has been made in a segment. Therefore, we use the well-known path loss formula $L=10*n*log(d)$, which characterises the decrease of the receivable power radiated by an antenna with increasing distance (d). In the formula, n is the free space component, which is five in our setting – a mixture between urban, indoor and rural places. 5.) we intersect the ellipse segments with the **hospital polygons**. 6.) we calculate the **number of calls in a cell**, which are made to and from cell phones located within the hospital using a spatial join and subsequent summation.

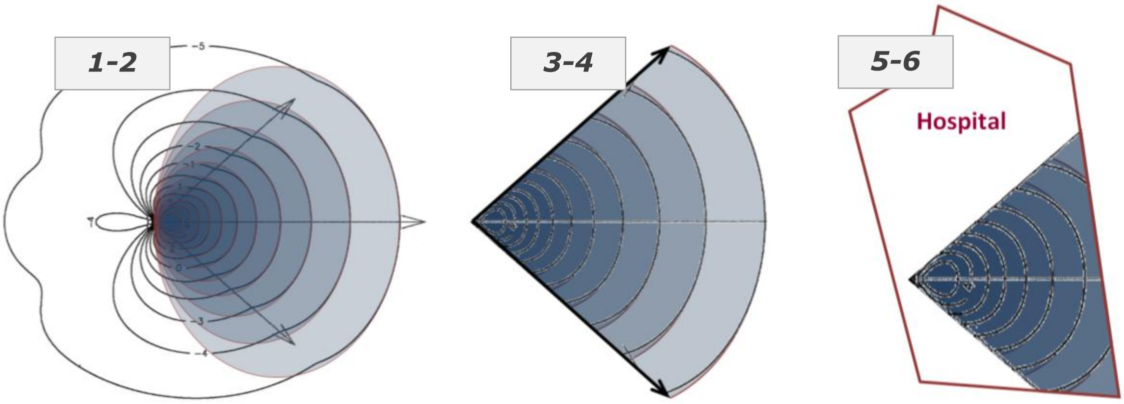

Figure 1: Identifying Calls from and to Cell Phones in a Hospital.

3.4 Computing the Catchment Areas for each Hospital

In the next step, we determine the catchment area of each hospital, where a hospital is identified using landline calls and cell phone calls (s. sub-sections 3.2 and 3.3). First, we compute the probability that a cell phone call has been made from or to a hospital to or from a mobile phone in a remote cell. This is done by identifying distinct origin-destination pairs for each call – a hospital and a remote cell. Then, we use a regular grid (cell size 1x1 km) to compute the number of calls made to and from mobile phones located within each grid cell, analogously to step 6 described above. Finally, we compute the catchment area for each hospital by summing up for each grid cell the number of calls to and from a hospital. Figure 2 shows an exemplary catchment area for one hospital.

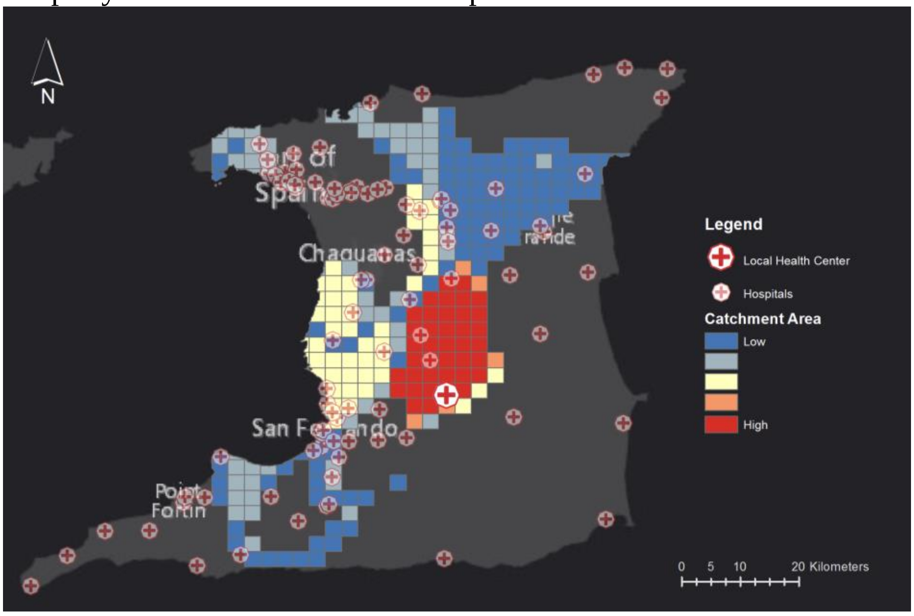

Figure 2: Catchment Area for a Hospital.

4. Discussion and Conclusion

This paper presents an approach to derive the catchment areas of hospitals from anonymised mobile and landline phone data. In contrast to previous approaches, we identify the geographic areas, in which people actually use a hospital. Therefore, we analyse calls to and from hospitals. Our results reveal new insights into the catchment areas of hospitals by minimising assumptions about demographic factors. Even though our case study uses data from Trinidad and Tobago, our approach is transferable to other regions worldwide.

In performing our research, we identified a number of limitations. First, not all calls that are made or received within a hospital are necessarily related to the hospital's "business", which induces an unknown bias. Moreover, we have not accounted for the mobile cells' spatial density in our analysis. Thus, our current method does not consider that the probability of a cell phone being connected to a particular antenna decreases if many cells' coverage areas overlap. Furthermore, the use of grid cells in the calculation of the catchment areas may lead to a modifiable areal unit problem (MAUP). This could potentially be mitigated by intersecting all ellipse segments, which was too computationally intensive given our extensive dataset. Finally, evaluation and validation of our results are difficult as no data for ground-truthing or comparable studies exist.

References

Buvaneswari A, James DA, Liu C, Graybeal JM, Lambert D and MacDonald WM, 2007, A Statistical View of the Transient Signals That Support a Wireless Call. *Technometrics*, *49*(3):305–317.

Fransen K, Neutens T, De Maeyer P and Deruyter G, 2015, A Commuter-based Two-step Floating Catchment Area Method for Measuring Spatial Accessibility of Daycare Centers. *Health and Place*, *32*:65–73.

Luo W and Wang F, 2003, Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region. *Environment and Planning B: Planning and Design*, *30*(6):865–884.

Schuurman N, Fiedler RS, Grzybowski SCW and Grund D, 2006, Defining Rational Hospital Catchments for Non-urban Areas Based on Travel-time. *International Journal of Health Geographics*, *5*(1):43.

Wang A and Wheeler DC, 2015, Catchment Area Analysis Using Bayesian Regression Modeling. *Cancer Informatics*, *14*(Suppl 2):71–79.