Geographic location of events

Martijn Tennekes

April 10, 2018

1 Introduction

For statistical inference on mobile phone data, geographic location is one of the most important variables of the data. However, in many applications, the exact geographic location is either not measured or not stored. Data collected by mobile network antennae are primarily logged for billing customers and for network analysis. For these tasks, only the identification number of the serving antenna is logged rather than the approximated geographic location of the events. There exist advance geographic pinpointing techniques such as triangulation and Timing Advance (Calabrese, Ferrari, and Blondel 2014). However, they are often unavailable since it demands some special infrastructure and data storage and analysis not often found in practice.

We shall avoid the terms antenna and mast as much as possible, since they may cause confusion. Rather, we will use the term *cell*, which refers to both the antenna and the geographic area that is be served by this antenna. Note that this term is also used in cell phones and cellular networks. A *cell site* or shortly a *site* is the location of one of more cells. When we refer to *antenna*, we mean the physical object that receives and transmits signals.

Table 1 lists the major site types. The most commonly known site type is the cell tower, which usually contains three cells which have coverage within approximately 120 degrees radius. Cells in other site types are omnidirectional, i.e. the cell operates evenly in all directions.

[1] Usual number of cells per unique location
[2] Distributed Antenna System

Type	Description	Number of cells[1]	Range
Cell tower	Tower constructed to support cells	3	500 meters to 40 km
Rooftop site	Cell located on rooftops	3	2 to 40 km
Small cell	Small sized cell	1	500 meters or less
Outdoor DAS[2]	Set of small outdoor cells	1	500 meters or less
Indoor DAS[2]	Set of small indoor cells	1	500 meters or less

Table 1: Types of cell sites and their characteristics.

The vast majority of studies on mobile network data use Voronoi tessellation (Okabe et al. 2000) to distribute the geographic location of logged events. The geographic area is divided into Voronoi regions such that each Voronoi region corresponds to the geographic location of a cell and each point in that region is closer to that cell than to any other cell.

There are a couple of downsides to use Voronoi tessellation to estimate the geographic location of devices. First of all, it assumes that all cells are onmidirectional. As described above, most cells are placed in cell towers or on rooftops and are directional. The second downside of Voronoi tessellation is that the coverage range of cells vary across cell types. Table 1 shows that the range depends on the cell type and moreover, on the configuration of the cells. Third, cells have overlap, especially in urban areas. This is because of load balancing; if a cell has reached full capacity, neighbouring cells that also have coverage are able to take over communication with mobile devices. This means that a mobile phone is not always connected to the nearest cell with the best signal. In urban areas, a mobile phone switches almost continuously between cells[3].

We present a Bayesian model to estimate densities of mobile phone devices. The likelihood function takes the estimated signal strength of nearly cells into account. Optionally, prior information about where devices are to be expected can be used. This information can be extracted from land use registers, building registers, or OpenStreetMap (*Open Street Map Foundation* 2018).

[3]There are several smart phone apps that show where the connected cell is located, e.g. Network Cell Info Lite (*Network Cell Info Lite app* 2018).

2 Method

In the proposed method, we will use a raster of the geographic area of interest. As an illustration, we use 100×100 raster cells using the Dutch National Grid projection (*EPSG* 2018). The main advantage to use raster cells is that different geospatial vector datasets can be combined without the need to calculate spatial intersections, which is a time consuming operation. Besides, the mathematics described below is easier since all raster cells have the constant area size.

The key of the proposed localization method is Bayes' formula, which is used in the following way:

$$\mathbb{P}(i|j) \propto \mathbb{P}(i)\mathbb{P}(j|i) \tag{1}$$

where i represents a raster cell and j the polygon of a cell. $\mathbb{P}(i)$ represents prior information about the relative frequency of events at raster cell i. The likelihood term $\mathbb{P}(j|i)$ is the probability that a device is connected to cell j given that it is actually located in raster cell i.

2.1 Prior information

The prior function can be used to specify where devices are expected to be. For instance, you would expect more devices on a road than on a grass field next to it. Also, more devices are expected to be inside buildings than outside when normalized per squared kilometer.

Geographic auxiliary information such as land use registers, building registers, and geographic data of roads and railways can be translated into a prior probability of presence of a device per raster cell i.

In the absence of prior information, $\mathbb{P}(i)$ can set to 1. In that case, it is assumed that devices are uniformly distributed across the geographic areas in which they are logged.

2.2 Likelihood function

The main advantage of using this Bayesian model compared to the Voronoi tessellation is that it takes the overlap of cells into account. This information is contained in the likelihood, which is defined as

$$\mathbb{P}(j|i) = \begin{cases} 0 & \text{if raster cell } i \text{ is not in cell} j, \\ \frac{s(i,j)}{\sum_k s(i,k)} & \text{if raster cell } i \text{ is in cell} j, \end{cases} \tag{2}$$

where $s(i,j)$ represents the signal strength that device i receives from cell j. Before defining it, let us define $S(i,j)$ which is an approximation of the actual signal strength denoted in dBm, which stands for decibels relative to one milliwatt. For onmidirectional cells, it is defined as

$$S(i,j) = S_r(r_{ij}) \tag{3}$$

where r_{ij} is the distance between the middle point of raster cell i and the antenna of cell j in meters. The function $S_r(r)$ returns the signal strength as a function of distance r:

$$S_r(r) = S_0 - 10\log_{10}(r^2/r_0^2) = S_0 - 20\log_{10}(r) \tag{4}$$

where S_0 is the signal strength at $r_0 = 1$ meter distance from the antenna.

A directional cell has an antenna which is directed at a specific angle. Along this angle, the signal strength is received at its best. However, the signal can also be good in other directions. It is comparable to a speaker which produces sound in a specific direction. The sound will be audible in many directions, but at the sides and the back of the speaker, the sound will much weaker. The directional beam of antenna j can be specified with four parameters:

- The horizontal/azimuth angle α_j is the angle from the top view between the north direction and the direction in which the antenna is pointed. Therefore, in reality this angle can be anywhere between 0 and 360 degrees. Note that cell towers and rooftop cells often contain three antennas with 120 degrees in between.
- The vertical/elevation angle β_j is the angle between the horizon plane and the tilt of the antenna. Note that this angle is often very small, typically only four degrees. The plane that is tilt along this angle is called the elevation plane.

- The horizontal beam width γ_j specifies in which angular difference from the azimuth angle in the elevation plane the signal loss is $3dB$ or less. In other words, the angles in the elevation plane for which the signal loss is $3dB$ correspond to $\alpha_j \pm \gamma_j/2$. In reality, these angles are around 65 degrees.

- The vertical beam width θ_j specifies the angular difference from β_j in the vertical plane orthogonal to α_j in which the signal loss is $3dB$. The angles in which the signal loss is $3dB$ loss correspond to $\beta_j \pm \theta_j/2$. In reality, these angles are around 9 degrees.

Let ϵ_{ij} be the angle from the side view between the line along the elevation angle β_j and the line between the center of antenna j and the center of grid cell i. Let δ_{ij} be the angle in the elevation plane between the azimuth angle α_j and the line between the center of antenna j and the center of grid cell i orthogonally projected in the elevation plane. Then, the signal strength for directional cells is defined by

$$S(i,j) = S_r(r_{ij}) - S_{e\ell}(\epsilon_{ij}, \theta_j) - S_{az}(\delta_{ij}, \gamma_j) \tag{5}$$

where $S_{e\ell}$ and S_{az} specify the signal loss based on the angular difference with the elevation and azimuth, respectively.

Each antenna type has its own radiation pattern for both the azimuth and elevation angles. These patterns define the relation between signal loss and the offset angles, i.e., δ_{ij} for the azimuth and ϵ_{ij} for the elevation angles. We used a Gaussian distribution to model the radiation pattern. The result is shown in Figure 1. The black line shows the relation between signal loss and angle in the azimuth plane (left) and elevation plane (right). The grey circles correspond to the signal loss; the outer circle means $0dB$ loss (which is only achieved in the main direction), the next circle corresponds to $5dB$ loss, etcetera. The red lines correspond to the angles corresponding to $3dB$ loss. So the difference between the red lines is γ_j in the Azimuth plane and θ_j in the Elevation plane.

Although these models approximate the general curve of real radiation patterns, the radiation pattern are more complex in reality, e.g. they often contain local spikes caused by so-called side and back lobes.

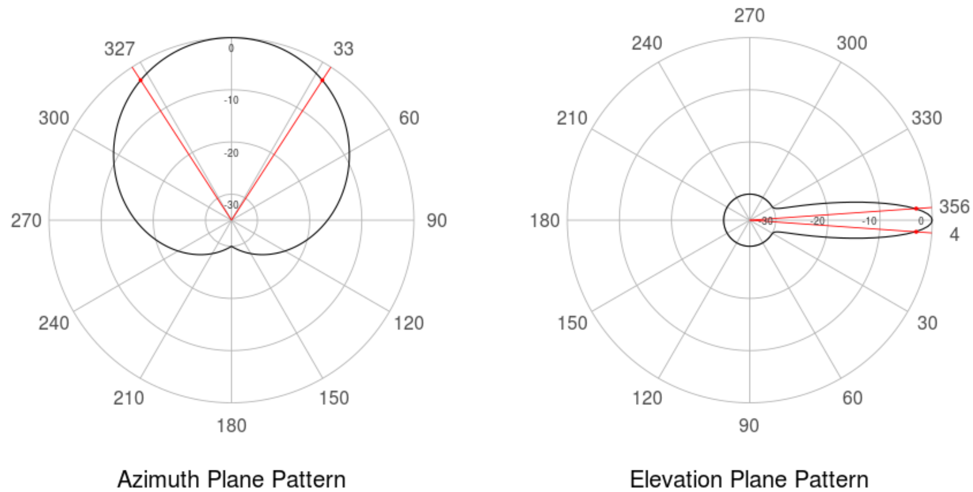

Figure 1: Radiation patterns for the azimuth and elevation planes

Signal strength (dBm)	Indication
-70 or higher	excellent
-90 to -70	good
-100 to -90	fair
-110 to -100	poor
-110 or less	bad or no signal

Table 2: Signal strength indication.

Figure 2 illustrates the signal strength at the ground level from above for a specific cell. In this case, the cell is placed at $x = 0$, $y = 0$ at 55 meters above ground level. The cell is directed eastwards with an elevation angle (tilt) of 5 degrees, a horizontal beam width of 65 degrees and a vertical beam width of 9 degrees. Table 2 describes how the signal strength values are interpreted by the network. Notice that the signal strength close to the cell, which means almost under the cell, is lower than at a couple of hundred meters distance. This is caused by a relatively large ϵ angles at raster cells nearby the cell.

It is often unclear how to load balancing mechanism works in practice. In the connection process between a device and the cell network, it can be

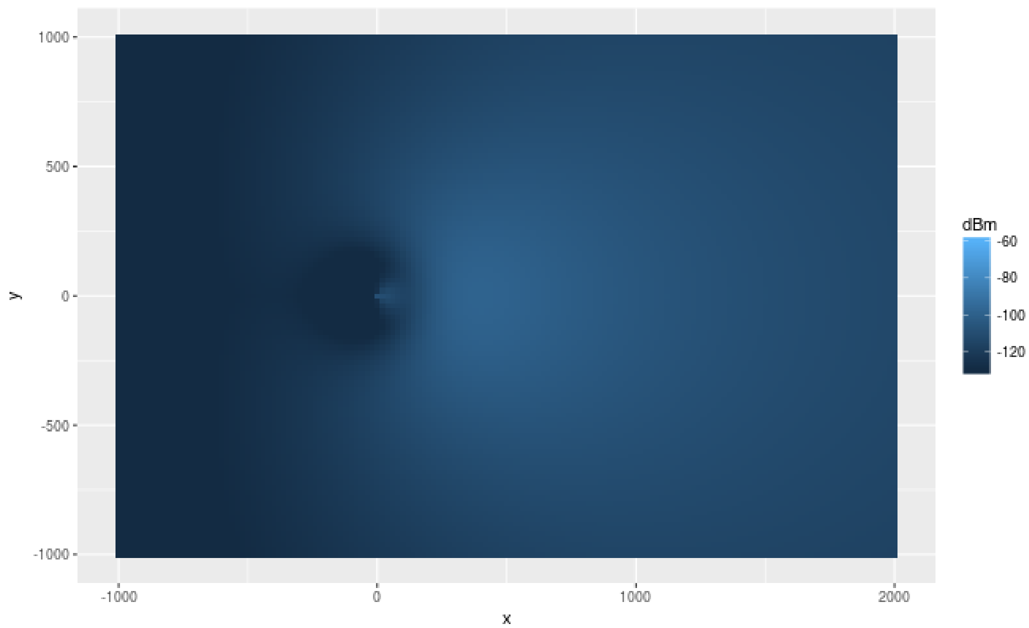

Figure 2: Signal strength at ground level

assumed that when there are a couple of cells with a signal strength above a certain threshold, say $-100dBm$, the cell is selected that has the highest capacity available. Therefore it is less important if the signal strength is -70 or -90 than -90 or -110. To model this load balancing mechanism, we have used a logistic function that translates the signal strength $S(i,j)$ to a relative signal strength measure $s(i,j)$ which we used to define the likelihood function (2).

$$s(i,j) = \frac{1}{1 + e^{-T(i,j)}} \tag{6}$$

where

$$T(i,j) = \frac{S(i,j) - S^{mid}}{S^{width}} \tag{7}$$

where S^{mid} and S^{width} are parameters that define the mid point and width of the curve respectively. Figure 3 shows the relation between the signal strength S_{ij} on the x-axis and the relative signal strength s_{ij} on the y-axis.

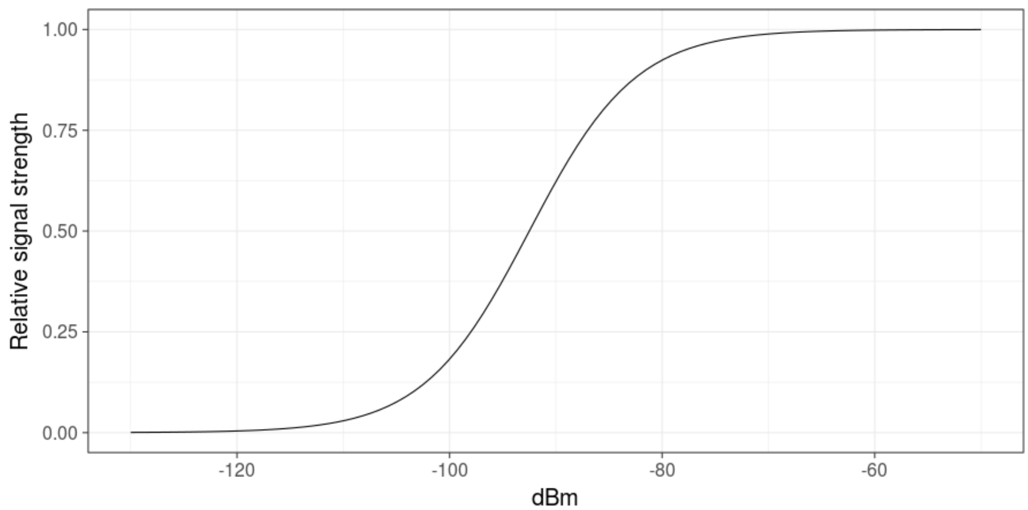

Figure 3: Signal strength at ground level

The relative signal strength at the ground level is shown in Figure 4. The probability values that are shown are normalized such that they sum up to one. Compared to the absolute signal strength shown in Figure 2, this distribution puts more emphasis on the geographic area that is in the spotlight of the cell.

References

Calabrese, Francesco, Laura Ferrari, and Vincent Blondel (2014). "Urban sensing using mobile phone network data: A survey of research". In: *ACM Computing Surveys* 47.2, pp. 1–20.

EPSG (2018). *Coordinate Systems Worldwide*. URL: `http://www.epsg.org`.

Network Cell Info Lite app (2018). URL: `https://play.google.com/store/apps/details?id=com.wilysis.cellinfolite&hl=en_419`.

Okabe, A. et al. (2000). *Spatial tessellations: concepts and applications of Voronoi diagrams (2nd ed)*. Ed. by Wiley.

Open Street Map Foundation (2018). URL: `https://www.openstreetmap.org`.

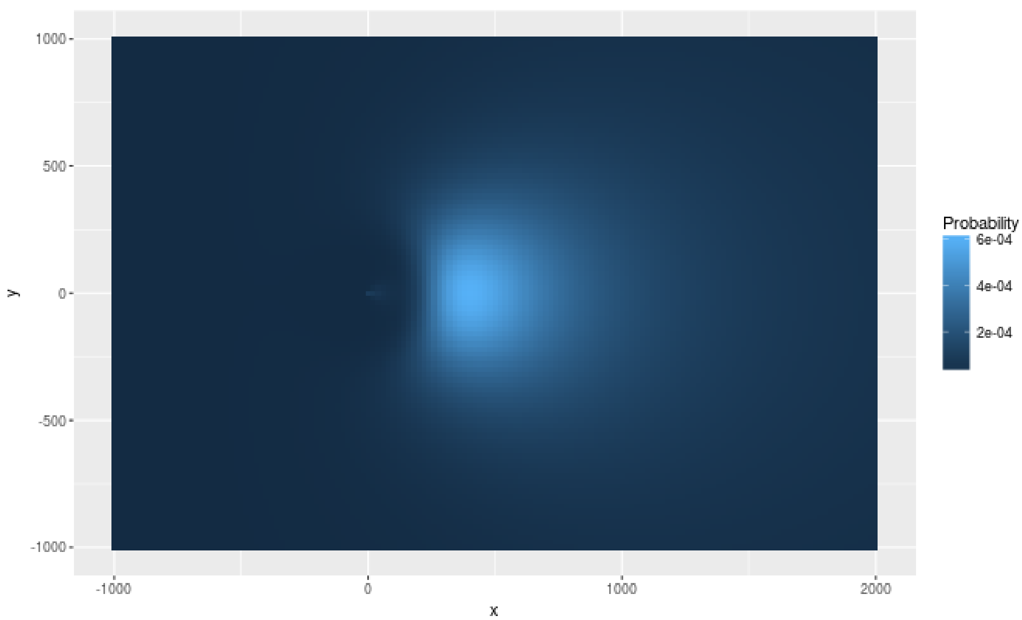

Figure 4: Relative signal strength at ground level