

Estimating Population Distribution with Landsat Imagery and Volunteered Geographic Information

Gabor Bakos¹ and Andrea Ballatore²

Department of Geography, Environment, and Development Studies

Birkbeck, University of London

January 13, 2017

Summary

Knowing where people live is crucial to provide timely and effective relief in humanitarian crises. To date, the population distribution in vulnerable parts of the developing world is insufficiently mapped, particularly in low population density areas. Based on work in the Missing Maps project, we propose a method to obtain cheap estimates of the population that combines automated processing of satellite imagery and a crowdsourced, manual classification collected through the MapSwipe app. The method produces granular estimates of rural population and is illustrated on a case study in a vulnerable border region in Sierra Leone.

KEYWORDS: Population distribution; Rural population; Volunteered Geographic Information (VGI); Missing Maps; MapSwipe; Humanitarian mapping; Landsat 8

1. Introduction

In humanitarian catastrophes, data about the affected area is needed to increase the effectiveness of responses. Notably, spatial information about the distribution of population and the road network is critical for emergency management, and yet is missing for large portions of the world. The absence of information greatly limits the abilities of organisations to plan and deliver an adequate response in case of emergency. Across many parts of Africa, this essential spatial information is often of limited quality, outdated, or even completely missing (Linard et al., 2012). Moreover, the available data often has low spatial resolution and does not provide sufficient detail about rural and urban settlement patterns, and is therefore of limited value to determine how to respond to natural and man-made disasters. Since the population in Africa is primarily rural (UN, 2015), the availability of data about rural areas is highly important. To date, estimating the population distribution in areas of low density with existing databases is challenging (Linard et al., 2013).

Existing geodemographic databases use a combination of population counts and detailed satellite-derived settlement extents to map population distributions across the planet (Linard et al., 2012). However, these data sources have some disadvantages. Although many land cover and population density datasets can be used free of charge, they are often of variable quality and timeliness (Friedl et al., 2002), limiting their usefulness for disaster management. A notable example is that of China's GlobeLand30, completed in 2014, with high spatial resolution and advanced cloud-removal algorithms, which reflected the status of the Earth's land cover from 2010. Moreover, the spatial resolution of the population distribution databases is often 500–1,000 meters only, leaving small towns and farms in the vicinity of large settlements de facto invisible. Such data can lead to poor planning choices for relief organisations.

¹ gabor.bakos@aol.co.uk

² a.ballatore@bbk.ac.uk

2. Humanitarian mapping and volunteered geographic information (VGI)

The discovery and mapping of “invisible” rural areas represent a complex challenge. When facing natural disasters, wars, and outbreaks, international organisations such as the Red Cross, Médecins Sans Frontières/Doctors Without Borders (MSF) and UN agencies need target areas to be mapped rapidly and accurately. Since 2010, volunteered geographic information (VGI) and collaborative mapping have been deployed in these scenarios with notable success (Zook et al., 2010). Local community volunteers map the main characteristics of settlements, including evacuation centres, potential sites for temporary hospitals, and so forth. Typically, field mapping is then complemented by remote mapping based on satellite imagery, which can be performed by volunteers on the Web.

Systematic efforts in this direction are made by the Humanitarian OpenStreetMap Team (HOT), which coordinates volunteers to perform mapping tasks in emergency crises (Goodchild & Glennon, 2010). In this arena, the Missing Maps project emerged in November 2014 as a collaboration between MSF, HOT, the American Red Cross, and the British Red Cross, to map vulnerable places in the world (Michael, 2014). In many instances, the learning curve of the digital mapping tools is steep, and only computer-literate and motivated volunteers tend to remain involved. As broadening participation is one of the fundamental challenges to increase the quality of crowdsourced data (Ballatore and Zipf, 2015), efforts are being made to reduce the entry barrier to the remote mapping process by developing more usable and accessible tools and techniques for remote, crowdsourced mapping.

3. MapSwipe

In July 2016, MSF released the MapSwipe app (<http://mapswipe.org>), which aims at reaching users to support the Missing Maps missions, adopting the approach of gamification. Through the smartphone app, users are shown satellite imagery about small portions of a target area, and are asked to indicate whether roads and settlements can be seen. MapSwipe divides satellite imagery into tiles of 256 x 256 pixels (about 24,000 m²), with a spatial resolution of 150 meters. Users are asked to tap directly on the tiles (Figure 1). To date, 13,000 users collectively scanned millions of square kilometres of imagery, making more than 8.8M contributions to record signs of human presence such as buildings or roads. As the number of individuals involved in the process tends to increase the mapping quality (Haklay et al., 2010), several volunteers are asked to classify the same tiles, resulting in a raster dataset representing the likelihood of presence of human settlements and/or roads.

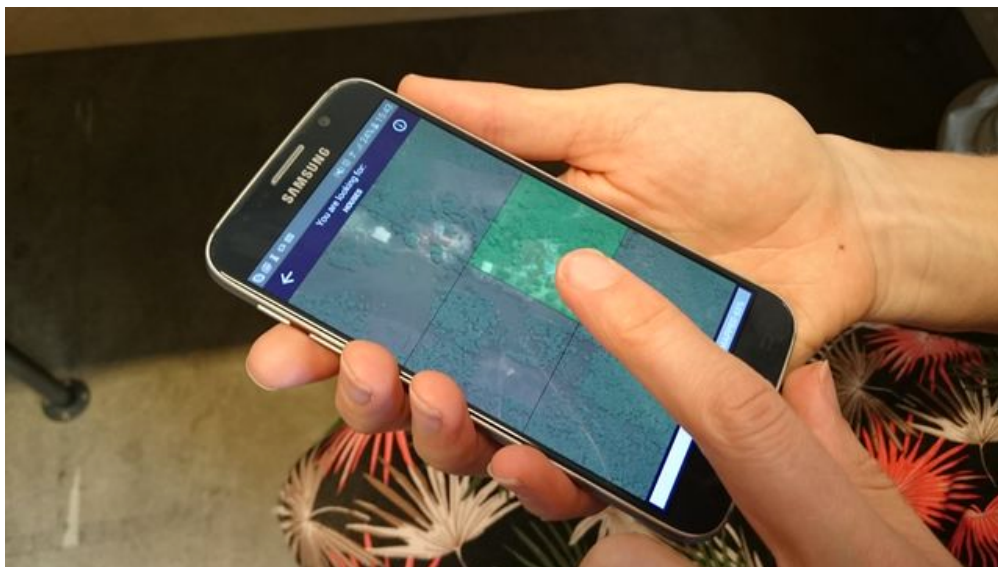


Figure 1: Classification task in the MapSwipe app (Source: <https://wiki.openstreetmap.org>)

4. Method for the estimation of population in high- and low-density areas

Thanks to collaborative mapping efforts, it is now possible to obtain information about the distribution of the population in vulnerable areas, providing a new, high-quality data source to governments and relief organisations. To overcome the challenge of mapping low-population areas, we outline a new information fusion approach, selecting a border region in West Africa as a case study. The proposed method is able to determine an examined area's settlement structure in a fast and cost-effective way, combining existing satellite imagery and data generated by volunteer mappers. In this case study, Landsat imagery provides the high-density settlement structure, while MapSwipe data enables the identification of the extents of low-density areas, such as villages and farms.

4.1 High population density areas

In our approach, high population density areas are identified by settlement extents provided by processing Landsat imagery.³ Well-known methods were selected to pre-process the data, starting with dark-object subtraction (DOS), a correlation method used because of the atmospheric scattering reduces the inaccuracy caused by scattering. Then, we used the pan-sharpening method to create a high-resolution colour image, combining lower resolution (30 m) multispectral “colour” bands (1, 2, and 3) with the higher resolution (15 m) band 8. This image helps recognise objects when selecting regions of Interest (ROI). The converted land cover map was filtered to obtain the settlement pattern, i.e., high population density areas, as a raster dataset. The result was then compared with OpenStreetMap building data, showing that it covered ~70% of the OSM buildings. As this approach is insufficient by itself, it was extended to be more sensitive to low-density areas.

4.2 Low population density areas

Data generated by volunteers through the MapSwipe app was used to determine the spatial extent of the low population density areas. The case study area was a border region in West Africa, between Sierra Leone and Guinea, which was badly hit by the 2014 Ebola epidemic and remains particularly vulnerable. The MapSwipe task for the case study partially covered 8 wards in Guinea and 12 in Sierra Leone, but only the Toli ward (Kono District, Sierra Leone) was completely covered, and was therefore selected to illustrate our method. The MapSwipe data was then converted into 3 different values for each tile (green, yellow, and red). Figure 2 shows the MapSwipe data for the ward generated by 1,031 volunteers in 2016, including pre-existing OSM data. The MapSwipe data clearly identifies more settlements and roads than those defined in OSM.

³ For satellite image processing, we used QGIS and the Semi-Automatic Classification Plugin (SCP), followed by GeoDa for the settlement pattern validation (Dormann et al., 2007).

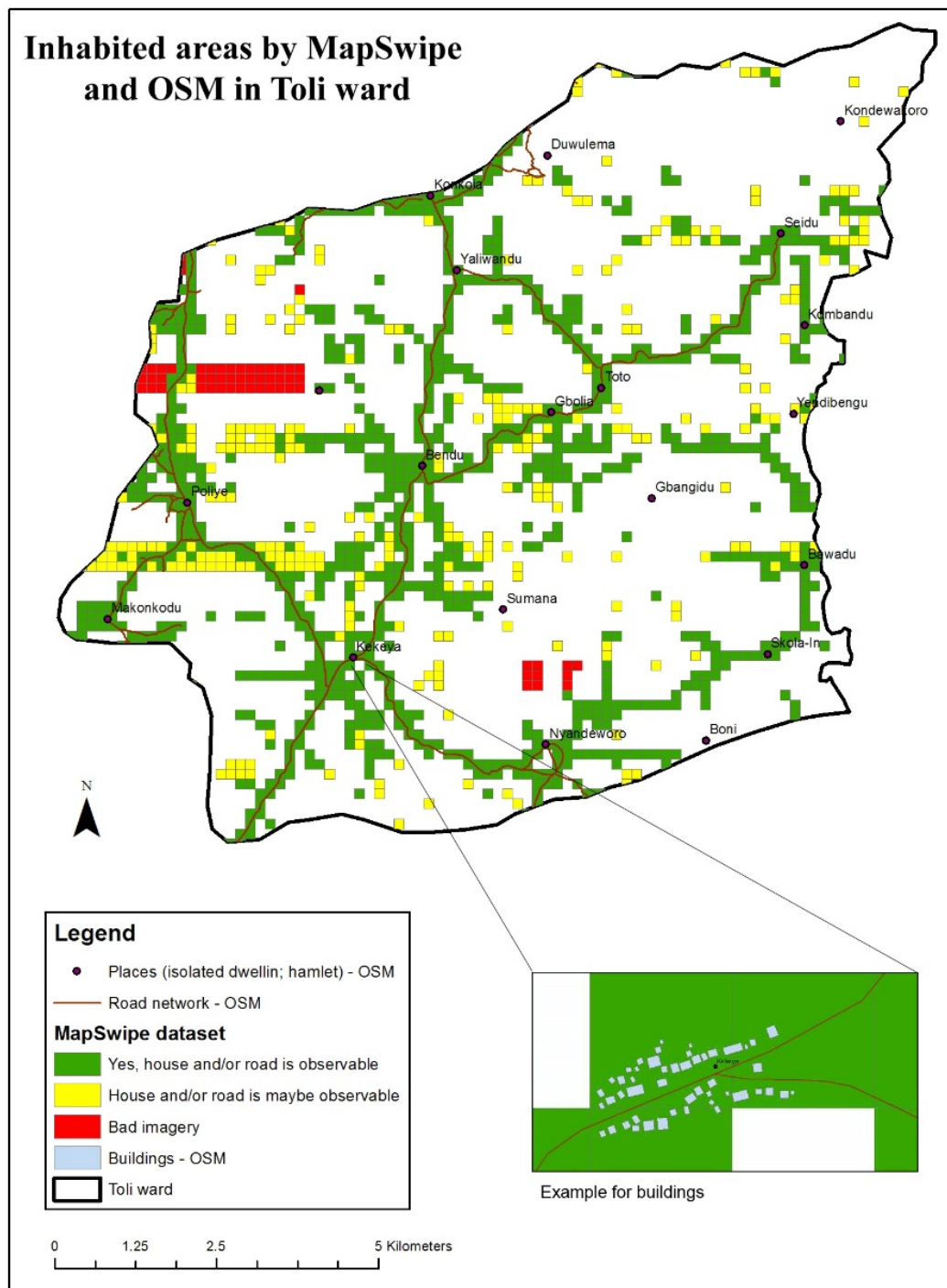


Figure 2: MapSwipe and OSM data for Toli ward (Sierra Leone)

Finally, results of the settlement structure from Land Cover map created from Landsat imagery were combined with data obtained by MapSwipe. As an input, each pixel of the settlement structure map can have two possible values (0 or 1), while the MapSwipe map has four possible values (“no”, “yes”, “maybe”, “bad imagery”). Through a dedicated algorithm, the two raster datasets were combined into 8 possible values, as shown in Table 1. The map in Figure 3 shows the results of this method spatially. It is possible to note that the resulting map provides much richer and granular estimation of population, than those obtained solely from satellite imagery and OSM.

Pixel value	Description	Estimate
0	Both input zero	No population
1	LandCover Yes, MapSwipe No	Classification error
2	LandCover No, MapSwipe Yes	Low population
3	LandCover Yes, MapSwipe Yes	High population
4	LandCover No, MapSwipe Maybe	Low population
5	LandCover Yes, MapSwipe Maybe	Low population
6	LandCover No, MapSwipe Bad image	No data
7	LandCover Yes, MapSwipe Bad image	Unknown

Table 1: Raster values of combination of high and low population density estimates

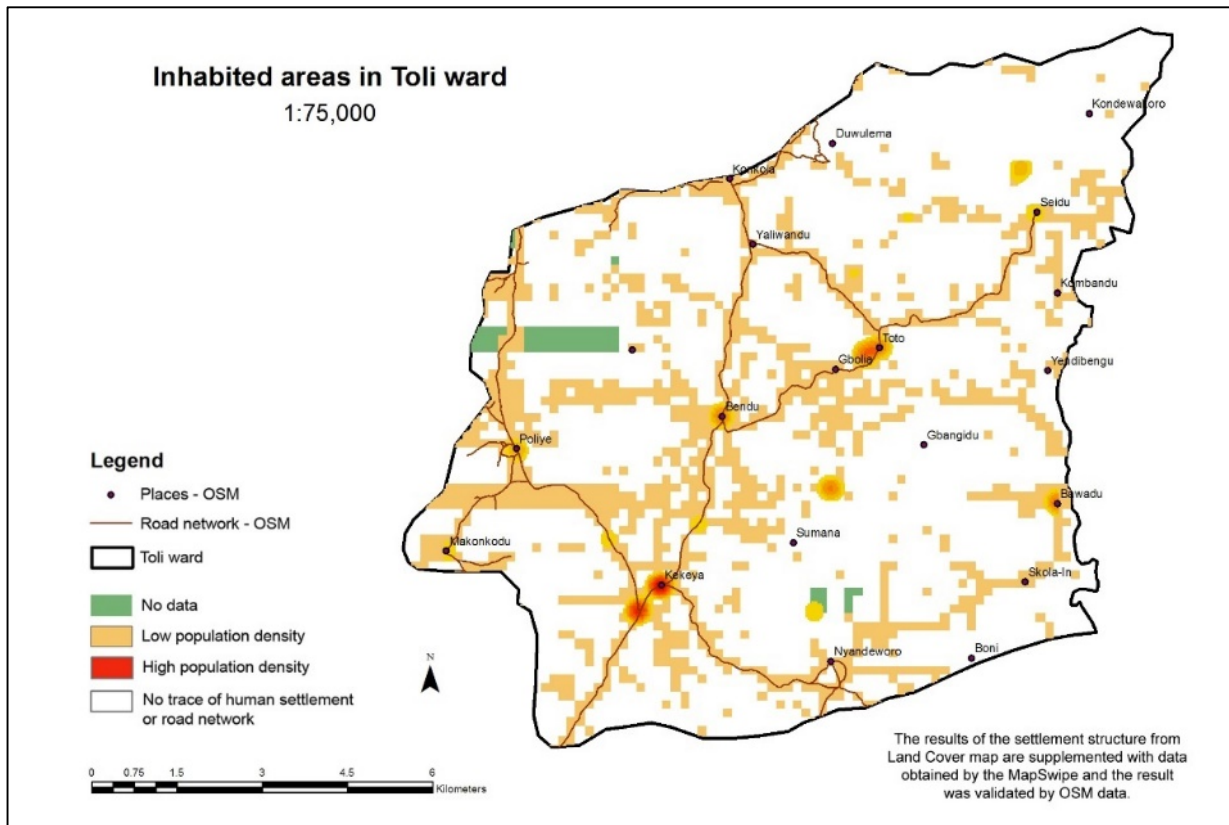


Figure 3: Population distribution in the Toli ward (Sierra Leone), based on the combination of Landsat and MapSwipe data

The combined use of MapSwipe and Satellite images can provide a fast, cheap and efficient solution, and has several practical applications. For example, thanks to these estimates, an organisation like MSF can determine suitable locations for hospitals and/or supply bases, reaching the previously under-represented rural population in the area (see Figure 4).

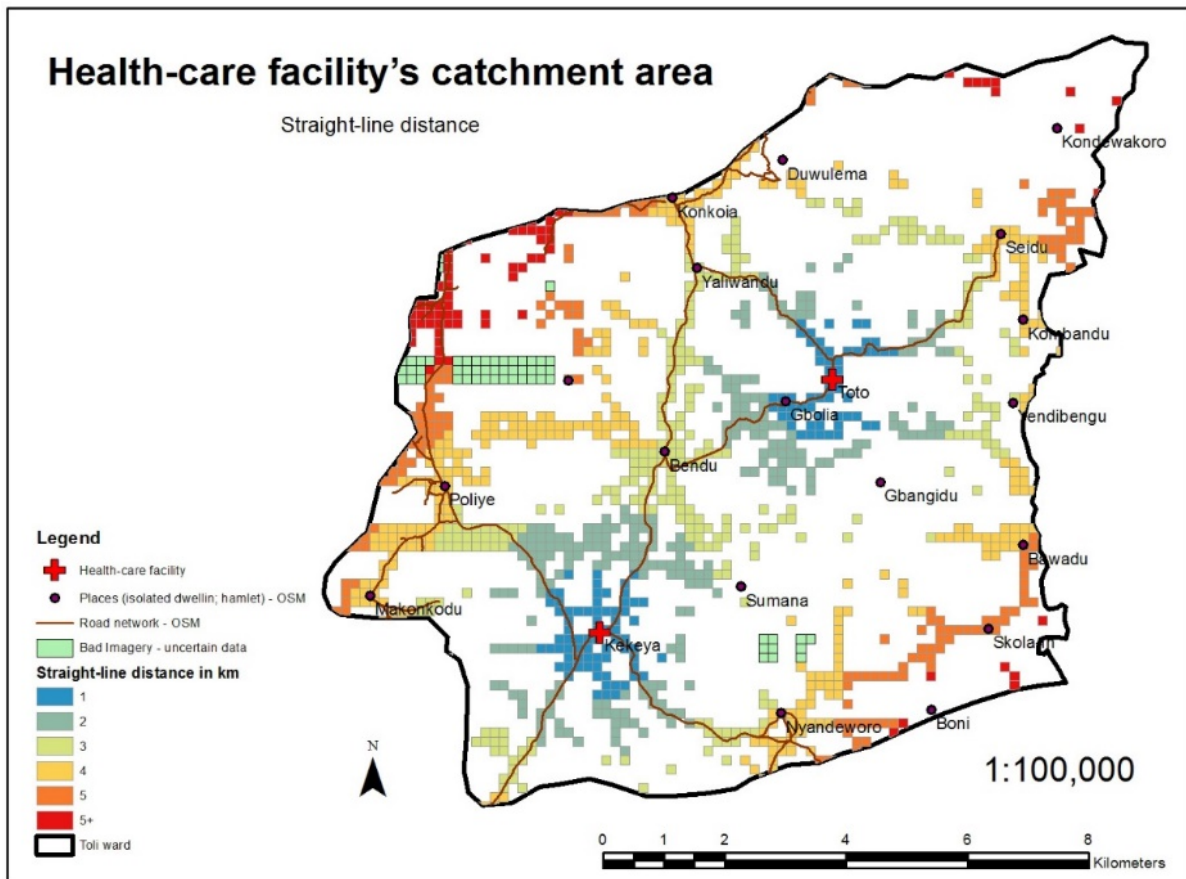


Figure 4: Health-care facility's catchment area in Toli ward (Sierra Leone), based on Landsat and MapSwipe population estimates

5. Conclusion

The proposed method of population estimation can identify inhabited areas by combining traditional satellite imagery and spatial crowdsourcing, as shown in a case study in Sierra Leone. As the necessary software and input datasets are open source and volunteered mapping practices are well-established, the method is cost-effective and scalable. More research is needed to evaluate the accuracy of the estimates, and other possible data sources to make the estimates more robust. Thanks to the usability and reach of MapSwipe, it will be possible to include increasingly more volunteers to estimate population in vulnerable parts of the world. The resulting spatial databases are useful not only to relief organisations, but also to local actors for planning and development, saving time, resources, and, most importantly, lives.

References

- Ballatore, A., and Zipf, A. (2015) "A Conceptual Quality Framework for Volunteered Geographic Information." In *Spatial Information Theory*, 9368:89–107. Berlin: Springer. doi:10.1007/978-3-319-23374-1.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. and Wilson, R. (2007) 'Methods to account for spatial autocorrelation in the analysis of species distributional data: a review.', *Ecography*, 30(5), pp. 609–628.
- Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C. (2002) 'Global

- land cover mapping from MODIS: Algorithms and early results', *Remote Sensing of Environment*, 83(s 1–2), pp. 287–302. doi: 10.1016/S0034-4257(02)00078-0.
- Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3), 231-241.
- Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, 47(4), 315-322.
- Linard, C., Gilbert, M., Snow, R.W., Noor, A.M. and Tatem, A.J. (2012) 'Population distribution, settlement patterns and accessibility across Africa in 2010', 7(2), p. e31743.
- Linard, C., Tatem, A.J. and Gilbert, M. (2013) 'Modelling spatial patterns of urban growth in Africa', *Applied Geography*, 44, pp. 23–32. doi: 10.1016/j.apgeog.2013.07.009.
- Michael, C. (2014) Missing maps: Nothing less than a human genome project for cities. Available at: <https://www.theguardian.com/cities/2014/oct/06/missing-maps-human-genome-project-unmapped-cities> (Accessed: 29 September 2016).
- United Nations, Department of Economic and Social Affairs, Population Division (2015). *World Urbanization Prospects: The 2014 Revision*, (ST/ESA/SER.A/366) (Accessed: 29 December 2016).
- Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*, 2(2), 7-33.