Classification: Social sciences

Spatially Designed Measures of Human Mobility Using Mobile Phone Records: Developing Big Data for Social Science

Nathalie E. Williams [a,b,g]
Timothy Thomas[a, g]
Matthew Dunbar[c, g]
Nathan Eagle[d]
Adrian Dobra[e,f,g]

February 2014

a Department of Sociology
[b] Jackson School of International Studies
[c] Center for Studies in Demography and Ecology
[d] Harvard University
[e] Department of Statistics
[f] Department of Biobehavioral Nursing and Health Systems
[g] University of Washington, Seattle, WA 98195

Corresponding author:
Nathalie Williams, Department of Sociology, Box 353340, University of Washington, Seattle, WA 98195, USA; phone: 206-616-6270; email: natw@uw.edu.

Keywords:
Big data
Call data record (CDR)
Mobility
Migration

Abstract

In the past decade, large scale mobile phone data has become available for the study of human movement patterns and holds immense promise for studying human behavior on a vast scale and with a precision and accuracy never before possible with censuses, surveys or other existing data collection techniques. There is already a significant body of literature that has made key inroads into understanding mobility using this exciting new data source, and there have been several different measures of mobility used. However, there has been little discussion and analysis of these measures. It is unclear what exactly these measures tell us and we argue that existing measures are contaminated by infrastructure and demographic and social characteristics of a population. These issues would be best addressed immediately as they will influence future studies of mobility using mobile phone data. In this article, we discuss problems with existing mobile phone based measures of mobility and describe new methods for measuring mobility that address these concerns. Our measures are designed to address the spatial and social nature of human mobility, to remain independent of and unadulterated by social, economic, political, or demographic characteristics of context, and to be comparable across geographic regions and time. We also contribute a discussion of the variety of uses for these new measures in developing a better understanding of how human mobility influences micro-level human behaviors and well-being and macro-level social organization and change.

Significance statement

Human mobility likely has significant influences on health outcomes and behaviors, demographic behaviors, social, economic and political indicators of well-being, and macro-level social organization. However, due to constraints in measuring mobility, social science has been able to develop only limited understanding of this critical behavior. Mobile phone call records have recently become available and show immense promise for recording human mobility on a large scale and with detail and accuracy never before possible. We contribute to the study of mobility through developing and assessing new measures of mobility from mobile phone data. We discuss how these measures can be used to significantly advance the social science of human mobility and the numerous health and behavioral outcomes it likely influences.

Introduction

Human mobility, or movement over short or long spaces for short or long periods of time, is an important yet under-studied phenomenon in the social and demographic sciences. While there have been consistent advances in understanding migration (more permanent movement patterns) and its impact on human well-being, macro-social, political, and economic organization (Donato 1993; Durand et al. 1996; Harris and Todaro 1970; Massey 1990; Massey et al. 1993; Massey and Espinosa 1997; Massey et al. 2010; Stark and Bloom 1985; Stark and Taylor 1991; Taylor 1987; Todaro 1969; Todaro and Maruszko 1987; VanWey 2005; Williams 2009), advances in studies of mobility have been stymied by difficulty in recording and measuring how human move on a minute and detailed scale. This gap is particularly glaring given that mobility is likely a key factor in social behaviors and macro-level social change, with likely associations with key issues that face human societies today, including spread of infectious diseases, health behaviors and outcomes, economic, social, and political well-being, and migration. In this context, new methods for measuring human mobility could lead to significant advances in the social and demographic sciences.

Mobile phone data has recently become available for the study of human mobility. Call data records (CDRs), which are records of each call and the location at which the mobile phone call was placed, have proven particularly interesting, by providing the possibility of recording movements over time of individual people and aggregate movements of whole populations. This exciting new type of data holds immense promise for studying human behavior on a vast scale never before possible and with a precision and accuracy never before possible with surveys or other data collection techniques. Indeed, as mobile phone penetration increases dramatically worldwide, to an estimated 89.4 mobile phones per 100 people even in poorer countries (ITU

2013), selection in who uses mobile phones is decreasing, making CDRs ever more appropriate for studying human mobility.

Consequently, there is a significant body of literature that has already made key inroads into understanding mobility using this exciting new data source, and there have been several different measures of mobility used (Blumenstock 2012; Gonzales, Hidalgo, and Barabási 2008; Lu, Bengtsson and Holme 2012; Phithakkitnukoon, Smoreda, and Olivier 2012; Song et al. 2010; Weslowski et at. 2012, 2013a, 2013b). However, there has been little discussion and assessment of these measures. Consequently, we understand little about what they actually measure and how they perform. Indeed, we argue in this paper that existing measures of mobility from CDRs inherently measure more than just mobility and are therefore not ideal to advance mobility studies with CDR data. Further, we argue that design of improved measures of mobility would be best addressed immediately as this will influence the conclusions of future studies of mobility using mobile phone data.

In this paper, we develop and assess five measures of mobility derived from CDRs. We define key dimensions of mobility, propose and analyze several measures that directly address each dimension as well as two comprehensive measures of mobility, and carefully assess our measures using CDR data from Rwanda between 2005-2009. One key difference in our proposed mobility measures from those used previously is that they are fundamentally based on existing spatial analytical methods, reflecting the spatial nature of mobility. A second key difference in our proposed mobility measures is that they account for how humans actually move, which is most often via road networks and through many places, instead of by apparition from one place to another. An important consequence of our spatial and movement perspectives is that they produce pure measures that address only movement of humans and are unadulterated by endogenous social, political, or economic factors. A second consequence of these spatial methods is that they are designed to be broadly applicable to different geographic settings, from Kansas City to Kigali, regardless of human behavioral patterns or variation in context. This article ends with a discussion of the new ways in which these measures can be used to advance the scientific study of human mobility.

Existing CDR-Based Measures of Mobility

Existing measures of mobility from mobile phones, as well as our new proposed measures, use CDRs. CDRs, which are permanently recorded by mobile phone service providers all over the world for billing purposes, are generated every time a user makes or receives a call or text message. Each CDR contains a unique identifier of the caller and the callee, identifiers for the initial and final cellular antennae (towers) that handled the call, the date and time the call was placed, and the duration of the call. Coupled with a dataset describing the locations (latitude and longitude) of cellular towers, these massive datasets provide the approximate location of the caller when placing the call. Because mobile phone owners generally make multiple calls during the course of even a single day and evening, CDRs can be used to study the movements, or mobility patterns, of individuals. Of course, the more calls a person makes, the more accurate will be our understanding of their actual mobility.

Existing measures that are commonly used in the literature include the radius of gyration (RoG), number of towers used (NTU), and maximum distance traveled (MD) with the RoG being the most common (Blumenstock 2012; Gonzalez et al. 2008; Lu et al. 2012; Weslowski et al. 2012; Weslowski et al. 2013a, 2013b). Each of these measures can refer to any specified period of time, such as the number of towers used in one month or in one week. The NTU measure is simply that, it counts the number of mobile towers from which a person called in the requisite period of time. Ostensibly, the higher the NTU, the higher the mobility of a person. The MD measure calculates the distance between towers that a person used. For example, if one made a call from tower 1, 5, and 9, the MD would be the sum of the distance between towers one and five and five and nine. The RoG is calculated by first finding the center of gravity or all mobile towers that a person used. Then the distance from the center of gravity to each tower is calculated and used to determine the overall RoG. For each of these measures, there is a clear logic for how the apply to measuring mobility.

There are several problems with these measures. The first is that the mobility rating of any person is affected by the density of mobile towers. As a simple example, consider a person who lives in an urban area with 50 towers within a five mile radius. This individual could regularly move within only this five mile radius, but CDR records would document them as using 50 towers and their mobility could be then calculated as high. Alternately, consider a person living in a rural area with only one tower in a five mile radius of their home. Even if they move about this five mile radius zone as often as our urban individual, the rural individual would only ever use this one tower and thus be classified as not moving anywhere and attain the lowest mobility rating. Thus, if not taken into account, variations in tower density create variations in mobility that do not actually exist. This problem is exacerbated by the fact that mobile towers are not placed randomly or evenly spaced. Instead, they are placed more often in urban areas with high population density, politically important areas, such as capital cities, or wealthy areas with higher mobile phone penetration. This problem, whereby mobility measures are confounded with social context, affects existing CDR-based measures of mobility including the NTU, MD, and RoG.

A second, and related, concern with CDR-based measurement of mobility is that the placement of mobile towers varies with time. In many countries, where the mobile infrastructure has not yet reached saturation, new mobile towers are built every year. For example, as shown in Figure 1, the number of towers in Rwanda increased from 72 in January 2005 to 240 by December 2009. In addition, in any country, mobile towers are dismantled or taken off the grid for various reasons. This creates a situation where the density of mobile towers, which is already a problem for existing CDR-based mobility measures, is time-varying. In other words, there is temporal variance in the spatial variance in inflation of mobility measures such as NTU, MD, and RoG.

[Figure 1 about here.]

In addition to the problem that existing measures are confounded with tower density, they are also inherently confounded with call frequency. The more often a person calls, the more towers at which they will be registered. In this situation, a person who uses their phone frequently will have an inflated mobility rating, compared to a person with the same mobility but lower calling frequency. This problem is particularly acute given that call frequency is selective of men and wealthier people.

A fourth problem with existing measures of mobility is that it is not entirely clear what aspects of mobility they are measuring. In addition to an overall concept of mobility, which has not been clearly defined in the literature, several dimensions of mobility have been suggested, including: distance traveled, number of places visited, and number of trips taken. Although these dimensions are generally accepted as relevant in multiple disciplines, they yet suffer from definitional difficulties and variance. There is no clear definition of what is a ‘place’ or a ‘trip’. This is further complicated when we consider different contexts and that a ‘place’ or ‘trip’ is likely very different from Kansas City compared to Kigali Rwanda. In any case, there is no clear understanding of what aspects of mobility are addressed by existing measures of mobility NTU, MD, and RoG.

A fifth problem with existing measures is that there are implicit, yet unrealistic assumptions about the nature of human movement. They measure distances in straight lines between mobile towers. Most often, humans do not travel in straight lines or “as the crow flies”, but longer distances on roads. In addition, when thinking about places that people visit, it is important to note that outside of air travel (which we discuss in Appendix A) humans do not usually apparate from one place to another. Instead they travel via road on the ground. As such, they exist, for some short period of time in many places between the places that they make calls.

Proposed CDR-Based Measures of Mobility

Given these concerns about existing measures of mobility, our intent is to design new measures that 1) are pure and unadulterated and independent of demographic, social, political, or economic characteristics of the context or mobile infrastructure; 2) independent of call frequency; 3) measure clearly defined aspects of mobility; and 4) are relevant and comparable across contexts, countries, and time.

The first foundation of our measures is a system of grid cells of even size placed across a country or area of study. As shown in Figure 2, some grid cells have a mobile tower in them, some do not, and some have multiple mobile towers. With the grid system, if an example person calls from a mobile tower, we register them as having been at the centroid of the corresponding grid cell. Movement is then calculated only when a person moves from one grid cell to another. Thus if our example person calls from another tower in the same grid cell, then they are registered in the same grid cell, and thus have not moved. If the person calls from a tower in a separate cell, then they have moved. Consequently, this system entirely disposes of mobile towers and instead replaces them with grid cells. In addition to other purposes described below, this system entirely negates the problem of varying tower density and creates measures that are pure and unadulterated by the demographic, social, economic, or political characteristics that regularly influence mobile tower density. The grid system is subject to grid cell density, but since grid cells are of even size and non-overlapping, there is no variance in grid cell density. All of our measures described below are entirely based on this grid system.

There are several key details to the use of this grid system that influence mobility measurement, including the arbitrary nature of how a grid system is placed on a map and the size of grid cells. These issues and solutions are discussed in Appendix B of the supplementary material.

[Figure 2 about here.]

The second foundation of our measures is a set of assumptions about how humans travel: they most often travel via roads, will take the fastest, easiest road route from one place to another, and the speed of travel is affected by speed limits and quality of road surfaces. With these assumptions, we can create routes of travel from one place to another that are not straight lines, but are based on publically available information of road systems. An example is shown in Figure 3. With this information, it is possible to calculate an assumed route of travel between any two points in a country, where the assumed route has the shortest possible travel time compared to all other routes. Because all our measures are based on a grid system, we create assumed routes of travel that begin at the centroid of the grid cell from which a person placed a call, take the shortest distance route to the nearest road from the cell centroid, travel the shortest route of travel to the grid cell in which their next call was placed, and end at the centroid of that grid cell.

[Figure 3 about here.]

The third foundation of our measures is that humans most often travel on the ground. Even if they do not make calls at every place they visit, we can assume they existed for some amount of time in every place along a road route, between two subsequent calls. This assumption, which is further discussed under Measure 2 below, at least partially addresses the influence of call frequency on mobility measurement. In existing measures, only places where a person made calls are included in mobility measures, thus higher call frequency inflates mobility ratings. Here, because we account for places where people made calls and places where they did not, then call frequency is less confounding. In the case, where an individual makes a call at the origin of a trip and another call at the destination of trip, then call frequency in between has no effect on our mobility measures. While this case comprises another assumption, it is not entirely unrealistic.

Using these foundations, and assumptions, we create five different mobility measures, three of which directly measure three dimensions of mobility (distance traveled, places visited, and trips taken and two of which are measures of overall mobility.

Measure 1: Distance traveled

This measures the distance between grid cells where consecutive calls were placed. The distance is via road travel and the fastest of all possible routes between the two points. An example of this calculation is graphically presented in Figure 4a for a person who placed a call from rural Rwanda and a second call from the capital Kigali. Note that the traditional measure of a straight line between these two points would yield a distance of 30.06 kilometers, while our grid cell and road-based measure yields a distance nearly twice as long, at 58.4 kilometers.

[Figure 4 about here.]

Measure 2: Places visited

As described above, in order to determine the number of places visited, it is necessary to define what a place is. This has historically proven difficult with great variation in definitions depending on context and research questions. In our case, we propose a spatially determined and pure definition of place that can be compared across countries, contexts, and time: a grid cell in which a mobile tower is placed. This assumes that there is something important about where a mobile tower is placed, either high population density, high through traffic, near an important area, at a cross-roads, etc. The reason that mobile towers are located in certain areas might differ between contexts and across time, but what does not differ is that there is likely a reason for mobile tower placement. We use this particular assumption because it assumes the least possible in order to define a place and is therefore the most comparable across contexts and time.

The second assumption that is required for this measure is that all places in which a person exists for any amount of time could be important. Some of these places are marked by a person making a call. However, there are other places that a person travels through on a road route in which they did not make a call. The logic behind this assumption is fundamentally that of a missing data problem: we do not know how long a person stayed in each place, how important was each place to a particular person, or if places where they made calls were more or less important than other places they traveled through. Thus, this measure assumes all places along a person's road route are of equal importance and counts all places (grid cells with mobile towers) through which a person between two subsequent calls. The origin grid cell and destination grid cell are counted and one grid cell must be deleted from this measure in order to avoid overcounting places visited. Figure 4b presents how this measure would be calculated for an example person. As you can see, from the starting to the ending points (where the person made phone calls), the shortest road route would take them through the origin, destination, and three enroute grid cells in which there were towers. Thus, this person is recorded as visiting five places.

Measure 3: Trips taken

Similar to places, in order to calculate a measure of trips taken, it is necessary to define what a trip is. Here again, in order for our measures to be independent of and comparable across context and time, we use the only information we have from CDRs and no information about context to define a trip as: movement between two subsequent phone calls that were placed in different grid cells. Thus, as shown in Figure 4c, if an individual makes a call from one grid cell and their next call is from a different grid cell, then this is one trip. If an individual makes a call from one grid cell and their next call is from the same grid cell (regardless if it is from a different tower) then this is not a trip. In Figure 4c, our example person made five phone calls from five different grid cells. Thus we calculate that this person made four trips, between points A and B, between points B and C, and between C and D, and D and E.

Measure 4: Comprehensive grid cells measure

The final two measures we propose are intended to characterize overall mobility. If a researcher does not need to isolate a specific aspect of mobility, such as distances, trips, or places, then these measures will allow them to use a specific and parsimonious record of mobility. In this regard, the fourth proposed measure is a comprehensive measure that includes to some extent two of the three dimensions of mobility in a single measure—distance traveled and places visited. Again, this is inherently based on the grid cell system and road routes. As shown in Figure 4c, this measure is calculated by counting the number of grid cells that a person travelled through from one call to the next and subtracting one. Here we again assume that people travel along roads, so grid cells along roads are used. Within four trips, our example person travelled through forty grid cells. Subtracting one grid cell, this person's comprehensive grid cell measure is thirty-nine.

Clearly this measure is related to distance traveled, but instead of counting kilometers, we count grid cells traveled through. It also incorporates the idea of 'place' but with a slightly different logic than Measure 2. In this case, all grid cells along a road route are assumed to be of equal importance and therefore worth counting. Thus, this measure, which counts grid cells that a person existed in between subsequent calls is more comprehensive by incorporating both 'places visited' as well as 'distance travelled'.

Measure 5: Combined index measure

The final measure we propose uses basic statistical procedures to combine measures of the three dimensions of mobility into one general mobility measure. This measure simply standardizes and then adds all three measures to create a single index of mobility. For example, if a particular person was 0.5 standard deviations above the mean for distance traveled, 0.3 for places visited, and 0.4 for trips taken, then their combined index measure would be 1.2.

Assessment of Proposed Mobility Measures

The assessment of CDR mobility measures is limited by the reality that there currently exists no standard measure of mobility, or no gold standard to which we can compare new measures. In this regard, the most important assessment tool available is face validity. In other words, the best assessment tool is a careful discussion of which measures make sense and if they actually measure what we think they should be measuring. Part of this face validity discussion is above in the description of the measures and sources of error. We aim for this article to stimulate further discussion in the literature on validity of mobility measures.

Other options to assess measure validity are to compare them with each other and across time. With this strategy, if two measures produce different results, we have no ability to decide which is more accurate or appropriate. We can simply understand that they are different. Our first such comparison, shown in Figure 5, shows how our measures and the existing NTU, MD, and RoG measures track across time in Rwanda. As you can see, _______________________.

[Figure 5 about here.]

A third validity assessment tool is to compare how similar the measures are for an example person. We use the most mobile person in Rwanda, in terms of distance traveled, and then examine their mobility score for all other measures. We do this for all measures and compare the results in Table 1. As you can see, _______________________.

[Table 1 about here.]

Discussion

The series of measures of mobility that we propose in this article constitute an important advance in the use of big data sources in social science. Being almost entirely spatially-derived, these new measures circumvent many of the problems inherent in existing mobility measures and are independent of social, political, economic, demographic, or infrastructure characteristics in the setting they are used. They are thus relevant and comparable regardless of geographic area, from Kansas City to Kigali. Our primary goal with this paper is to propose these pure measures of mobility, to stimulate discussion on mobility measures using big data, and promote to social science research on the causes and consequences of human mobility.

In this regard, we herewith discuss the many ways in which these new CDR-based measures of mobility can be used to enhance and expand our understanding of human well-being and social organization. First, these new measures can simply replace older measures, often based on sample surveys, to improve understanding of existing mobility related questions. The benefit here is clear, given that CDR measures can significantly increase the accuracy, detail, and time period over which mobility can be recorded. CDR-based measures can be collected and calculated for respondents in sample surveys, giving the researcher not only immense detail about respondent mobility, but the opportunity to compare it with immense surveyed detail about other characteristics and behaviors.

Second, these new measures open up entirely new avenues of research. For example, because CDRs can cover millions of people, it is possible to calculate population-level mobility measures. For example, one can calculate a measure of general mobility for a city, state, province, or region. This could then be compared to individual level behaviors and outcomes (for example, how does population mobility influence individual migration, tuberculosis infection, or women's labor force participation). Population level mobility can also be related to population-level characteristics (such as HIV prevalence rates, birth rates, social norms, economic well-being, or political participation). With sample surveys, it has never before been possible to calculate population level characteristics, thus CDR-based measures, if appropriately calculated to be independent of social, economic, political, or demographic contextual characteristics, create new and possibly groundbreaking opportunities for social science.

While CDR-based measures can create immense new opportunities for understanding human mobility, there are several limitations of researchers must be aware. As with all organic big data (Groves 2011), selection is a major concern. For mobile phone data, mobile phone users are included in the data set and non-users are excluded. Research suggests that users are more likely to be male, educated, and live in urban areas (Blumenstock and Eagle 2012; Weslowski 2013b). Alternately, research has also shown that there are an estimated 128.2 mobile phones per 100 people in wealthier countries and 89.4 per 100 in poorer countries (ITU 2013). Considering that mobile phone penetration statistics are largely analogous to response rates in surveys, we can say that CDR based data essentially have 89.4% response rates in poorer countries, which is generally considered good if not excellent, regardless of selection.

Another key limitation to the use of CDR-based mobility measures is inherent error. The primary problem is that although mobile calls are recorded as occurring at a tower, the person making the call is rarely at the tower. Instead they are likely to be within five or ten kilometers of the tower, depending on the type of antenna used in the tower and topography. Further, when we replace towers with grid cells, there is the possibility of increasing the error in a person's location. In most cases, it is likely that the combined error (uncertainty of a person's location in relation to a tower combined with additional grid system locational error) is negligible. In the most extreme circumstance, with a five by five kilometer grid system, towers that broadcast to a 10 kilometer radius, and a tower that is in a far corner of a grid cell, a person's location could be calculated as being up to 13.5 kilometers[1] from their actual location. Note that the majority of the error here (10 out of 13.5 kilometers) is due to tower-location uncertainty and the minority of error (3.5 kilometers) is due to the imposition of the grid system. This maximum possible error of 13.5 kilometers is likely not as problematic when measuring mobility on a national scale compared to a smaller local scale. When measuring mobility on a smaller scale in areas with higher tower density, or towers that are closer to each other than 10 kilometers, the maximum

[1] This maximum error is calculated by adding the 10 kilometer radius of the tower antenna to the distance from the corner to the centroid of a five by five grid cell (10 + sqrt(12.5)).

possible locational error will be less. It can be further reduced by decreasing the size of grid cells. Thus, locational error must be carefully considered when CDR-based data is used to measure mobility and further work should be done to assess the effects of selection and locational error. However, we argue that the benefits of CDR-based mobility measures vastly outweigh the detractions, especially when compared to the alternative of survey-based measures with inherent error due to human difficulties in recalling location, time, and movement accurately.

Materials and Methods

Our mobility measures are designed to be applicable to any research setting, from wealthy countries with well-developed mobile phone and transportation infrastructure, to poorer countries that are yet developing transportation and communication networks. However, it is necessary to use a setting and specific data to demonstrate and assess our measurements, and for this we use data from one mobile phone company in Rwanda covering the period 2005-2009. Our data include about 1.5 million, or about 3% to 24% of the entire population of the country in 2005 and 2009 respectively.

For information on road networks with which we create our routes, we find that OpenStreetMap currently provides the most detailed and up-to-date road network information available to the public. In addition to road maps, they provide assessment of road quality that can be categorized as primary, secondary, tertiary, or unpaved. With this information, we set probable speeds for each road type, allowing us to calculate probable travel times between any two points in the country.

As described above, our method of overlaying a grid on a map of Rwanda creates a systematic method of circumventing the major problem of spatial variance in mobile tower density. However, there are several special cases where the broad system did not apply well. In all such cases, our primary aim was to remain systematic so that all special situations are addressed at least similarly to the general method. Appendix C in the supplementary material describes these situations and how we addressed each.

References

Blumenstock, Joshua E. 2012. "Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda." *Information Technology for Development* 18(2):107-125.

Blumenstock, Joshua and Nathan Eagle. 2012. "Divided we call: disparities in access and use of mobile phones in Rwanda." *Information Technology and International Development* 8(2): 1-16.

Donato, Katharine. 1993. Current trends and patterns of female migration: Evidence from Mexico, *International Migration Review* 27(4): 748-771.

Durand, Jorge, William Kandel, Emilio A. Parrado, and Douglas S. Massey. 1996. "International migration and development in Mexican communities." *Demography* 33(2):249-264.

Gonzales, Marta C., César A. Hidalgo, and Albert-László Barabási. 2008. "Understanding individual human mobility patterns." *Nature* 453: 779-782.

Groves, R. (2011). "Designed data" and "organic data." Director's Blog, U.S. Census Bureau. Retrieved from http://directorsblog.blogs.census.gov/2011/05/31/designed-data-and-organic-data/ on 2/27/2014.

Harris, John R. and Michael P. Todaro. 1970. Migration, unemployment and development: A two-sector analysis, *American Economic Review* 60(1): 126-142.

International Telecommunication Union. 2013. The World in 2013: ICT Facts and Figures. Report accessed online 2/26/20414 at: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx.

Lu, Xin, Bengtsson, and Petter Holme. 2012. "Predictability of population displacement after the 2010 Haiti earthquake." *PNAS* 109(29): 11576-11581.

Massey, Douglas S. 1990. Social structure, household strategies, and the cumulative causation of migration, *Population Index* 56(1): 3-26.

Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J.Edward Taylor. 1993. Theories of international migration: A review and appraisal, *Population and Development Review* 19(3): 431-66.

Massey, Douglas S. and Kristin E. Espinosa. 1997. What's driving Mexico-U.S. migration? A theoretical, empirical, and policy analysis, *American Journal of Sociology* 102(4): 939-999.

Massey, Douglas S., Nathalie Williams, William G. Axinn, and Dirgha Ghimire. 2010. Community services and out-migration, *International Migration* 48(1): 1-41.

Phithakkitnukoon, Santi, Zbigniew Smoreda, and Patrick Olivier. 2012. "Socio-geography of human mobility: a study using longitidunal mobile phone data." *PLoS ONE* 7 e39253.

Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. "Limits of predictability in human mobility." *Science* 327(5968):1018-1021.

Stark, Oded, and David E. Bloom. 1985. The new economics of labor migration, *American Economic Review* 75(2): 173-178.

Stark, Oded, and J. Edward Taylor. 1991. Migration incentives, migration types: The role of relative deprivation, *The Economic Journal* 101(408): 1163-1178.

Taylor, J. Edward. 1987. Undocumented Mexico-U.S. migration and the returns to households in rural Mexico, *American Journal of Agricultural Economics* 69(3): 616-38.

Todaro, Michael P. 1969. A model of labor migration and urban unemployment in less developed countries, *The American Economic Review* 59(1): 138-148.

Todaro, Michael P., and Lydia Maruszko. 1987. Illegal immigration and US immigration reform: A conceptual framework, *Population and Development Review* 13(1): 101-114.

VanWey, Leah K. 2005. Land ownership as a determinant of international and internal migration in Mexico and internal migration in Thailand, *International Migration Review* 39(1): 141-172.

Weslowski, Amy, Nathan Eagle, Andrew J. Tatem, David L. Smith, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2012. "Quantifying the impact of human mobility on malaria." *Science* 338(6104):267-270.

Weslowski, Amy, Caroline O. Buckee, Deepa K. Pindolia, Nathan Eagle, David L. Smith, Andres J. Garcia, and Andrew J. Tatem. 2013a. "The use of census migration data to approximate human movement patterns across temporal scales." *PLoS ONE* 8(1): e52971.

Weslowski, Amy, Nathan Eagle, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2013b. "The impact of biases in mobile phone ownership on estimates of mobility." *Journal of the Royal Society Interface* 10(81):

Williams, Nathalie. 2009. Education, gender, and migration in the context of social change, *Social Science Research* 38(4): 883-896.

Figure Legends

Figure 1. Change in number of mobile phone towers in Rwanda 2005-2009.

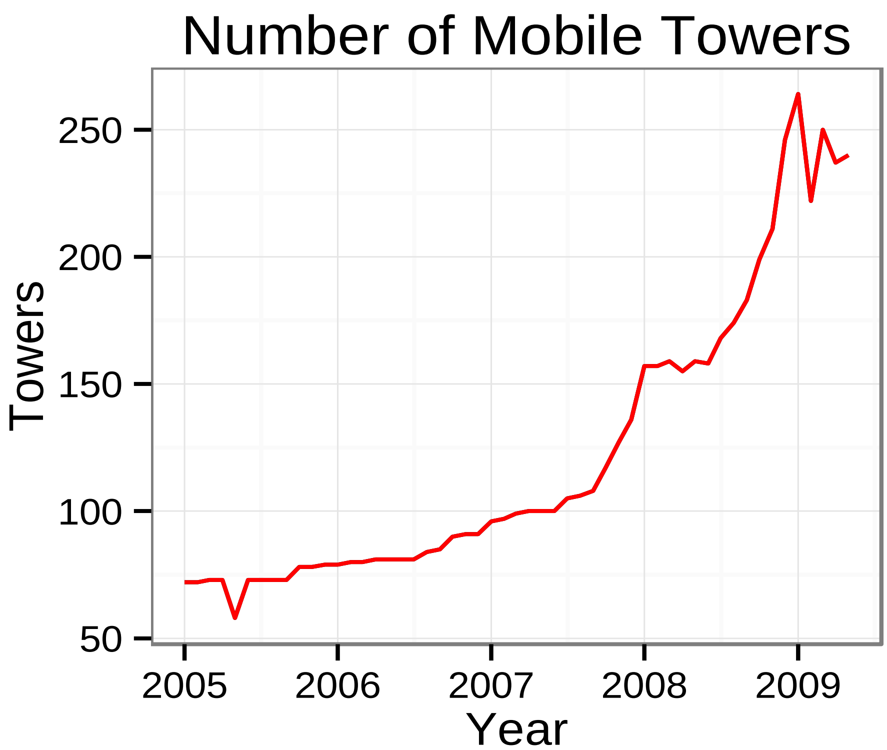

Figure 2. Grid system for mobility measurement placed over map of Rwanda. Grid cells are five by five kilometers in size.

Figure 3. Grid system for mobility measurement in Rwanda.

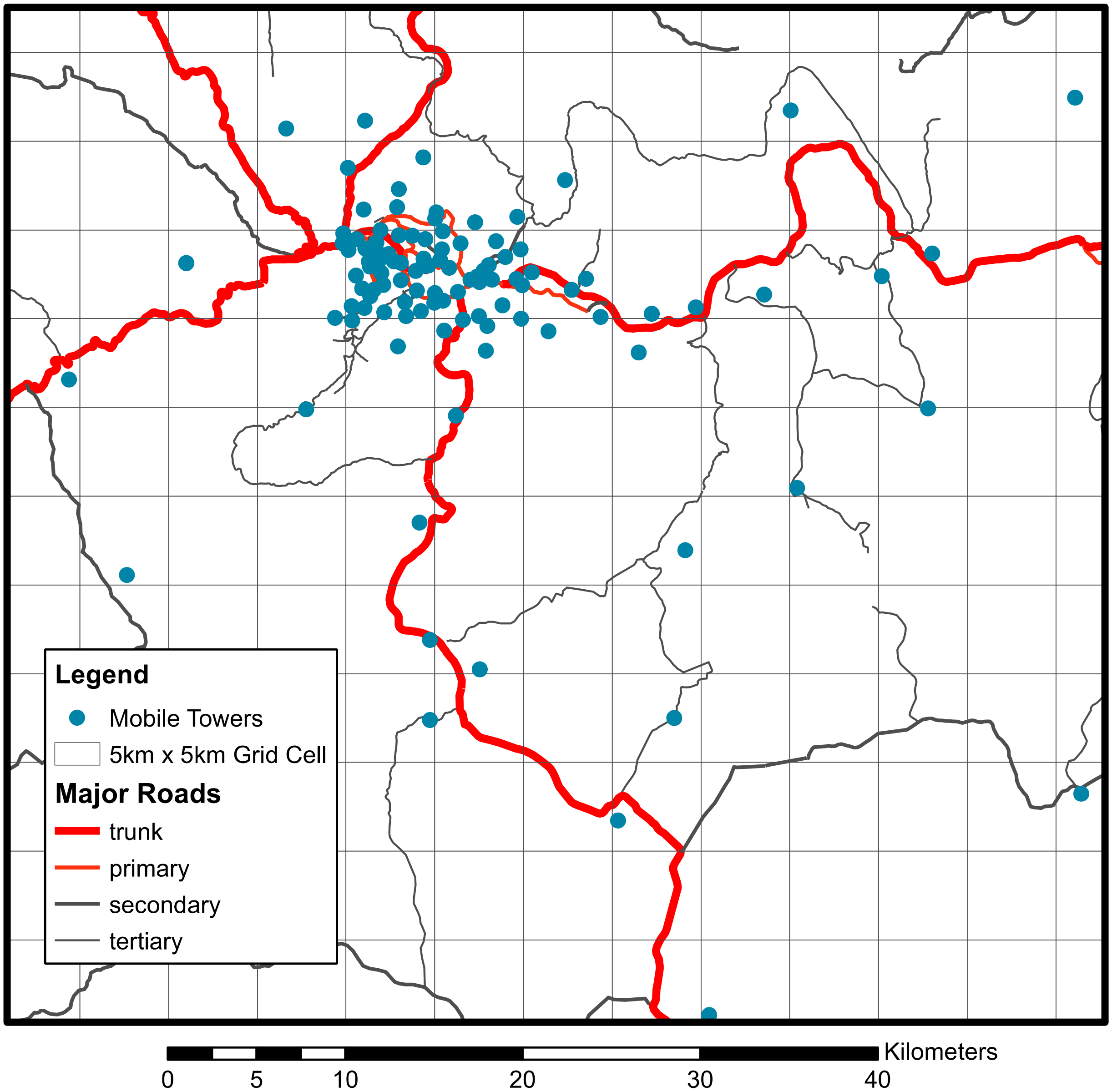

Figure 3. Example of road route compared to straight line route between two grid cells where a person made calls.

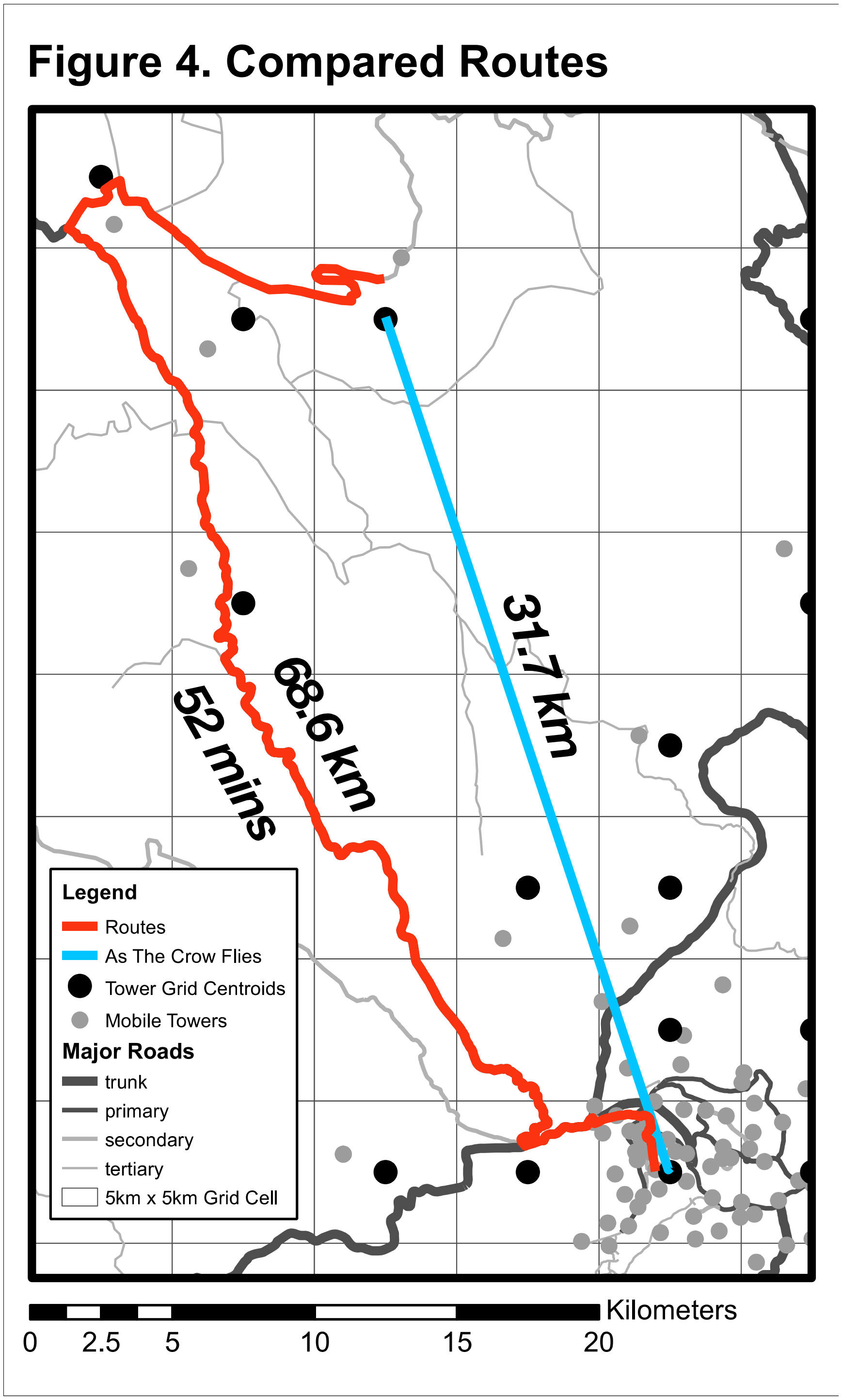

Figure 4. Calculation of proposed mobility measures for example person in Rwanda.

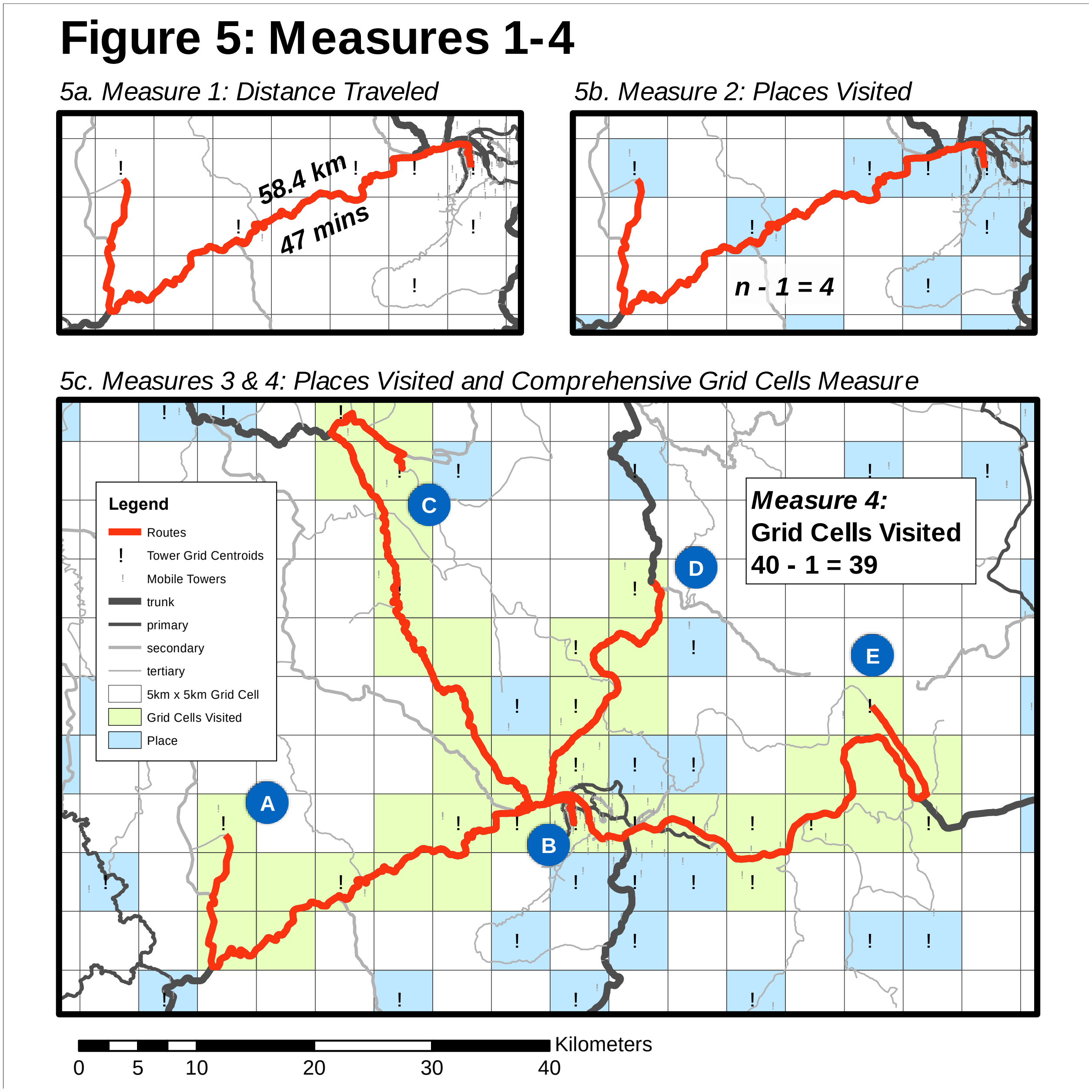

Figure 5. Average mobility, by existing and proposed measures, of Rwandan population from 2005-2009.

TO BE COMPLETED

Table 1. Comparison of measures, by most mobile person in Rwanda.

Highest mobile person by each measure:	**Mobility measurement** Number towers used	Max-imum distance	Radius of gyration	Distance traveled	Places visited	Trips taken	Compre-hensive	Com-bined
Number towers used								
Maximum distance								
Radius of gyration								
Distance traveled								
Places visited								
Trips taken								
Comprehensive								
Combined								

TO BE COMPLETED

Supplementary Material

Appendix A.

Addressing the possibility of air travel for proposed mobility measures.

Our proposed system of calculating mobility measures is entirely based on the assumption that most people travel via land and on roads. However, in any country there is the possibility of air travel and several adjustments must be made to our methods for this possibility.

The first step in adjustment is to identify when a person most probably traveled via air rather that road. For this, we use all information we have—time between calls. If the time period between two subsequent calls from different grid cells is shorter than the shortest possible time it would take to travel between those cells via road, then we can assume that a person traveled via air. As with any assumption, this inherently includes some error. This method will correctly identify all air travel for people who make calls soon before taking off and soon after landing. However, if a person does not make calls for some time before before and after flying, then their travel will not be identified as such. There will be more error in this manner for flights that cover short distances than for long distance flights.

Once air travel movement has been identified, then most mobility measures must be calculated differently. *Distance traveled* should be calculated as the straight line distance between the two points that represent air travel. *Places visited* should include only the grid cells in which the person made calls and no "places" (grid cells with mobile towers) on a route in between. No adjustments need to be made to the calculation of *trips taken*, *comprehensive mobility*, or *combined mobility*.

Appendix B.

Contingencies to using the grid cell system.

A possible detraction to using the grid cell method as a foundation for measuring mobility is that placement of the grid system creates arbitrary boundaries that could non-systematically influence mobility measures. For example, it is possible for a person to call from a mobile tower that is one meter from a grid cell boundary. They could then move two meters and call from another tower that is in a different grid cell. In this case, our example person would be registered as moving between grid cells, even with only two meters of actual movement. In another case, a person could call from a tower nearby a grid cell boundary, move almost five kilometers towards the other side of the grid cell and make another call. In this case, because our example person did not cross a grid cell boundary between calls, they would not be registered as moving.

There is a reasonably simple, though computationally intensive, solution to this problem. Consider a system of five kilometer by five kilometer grid cells. This system can be placed over a map of the study area and mobility measures calculated. Next, the entire grid system can be moved one kilometer east over the map and mobility measures recalculated. The grid system can be subsequently moved 25 times (by one kilometer east and one kilometer south each time) and 25 iterations of mobility measures calculated. The combination of these 25 iterations should then yield a single measure of mobility that is not subject to the arbitrary placement of grid cell boundaries.

Appendix C.

Special situations and systematic solutions for using the grid cell system.

While the grid system provides an overall standardized spatial relationship across an area, there are several minor caveats that needed manual adjustment to connect all the areas for analysis. The first adjustment is making sure that all the roads are connected to all areas. To maximize route efficiency, we chose major highways and roads as preferred transit options for our mobile users. The most complete road network we could find was the crowd sourced OpenStreetMap dataset. While this is a robust dataset, it is still volunteer created and holds some discrepancies in the interpretation of certain roads (i.e. trunk vs unclassified, etc.). After selecting all trunk, primary, secondary, and major tertiary roads for our routing scheme, several minor roads needed to be added to our network to make sure all areas were connected. This includes 67 lesser category roads, most of which were labeled as “unclassified” or “residential.” We went to great lengths to spot-check each of the areas and roads to make sure they were logical and relevant decisions.

The second caveat was making sure each place – a grid cell with a tower – was connected to all the roads (i.e. all places overlap the nearest road). We found 11 places off-route from the road network and moved their centroid to an adjacent grid cell centroid. In a few cases this led to an overlap between two places, such as the dense area of Kigali. We believe this is not a major problem because, as mentioned above, tower radio coverage is roughly 5 kilometers radius. The distance between two centroids is exactly 5 kilometers and each of the roads that were barely missed by the adjusted place was within 3 kilometers of its centroid.

The final caveat involves four places with similar problems to the second caveat. To calculate the start and finish of a route, the routing program finds the closest section of road to the centroid, or junction. In four cases, the closest junction was just outside of a place's grid cell. To fix this, the centroid of these four locations was moved to a section of the road inside the place grid cell so that we could manipulate the junction location to stay within the place.