Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2017/05, 46 pàg.
Working Paper 2017/05, 46 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2017/03, 46 pàg.
Working Paper 2017/03, 46 pag.

"Computing Functional Urban Areas Using a Hierarchical Travel Time Approach: An Applied Case in Ecuador"

Moisés Obaco A, Vicente Royuela and Xavier Vítores

Research Institute of Applied Economics

Regional Quantitative Analysis Research Group

I|R|E|A|
Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics
WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

AQR
Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group
WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies,
(ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

Abstract

Identifying integrated urban areas is an important issue for urban analysis and policy evaluation. In this paper, we extend the OECD's methodology to identify Functional Urban Areas to countries where there is not commuting data. We do so substituting such socioeconomic flows by available information on road structure, which allow us to work with accessibility based on travel time. The main advantage of our procedure is its applicability to most countries in the world, as it only uses GIS data. In this paper we apply the procedure two border countries: Colombia, which has a recent census with commuting data, to calibrate our approach, and Ecuador, where there is not commuting census. We perform several sensitivity analysis and robustness checks to Ecuador with alternative sources of socioeconomic flows.

JEL Classification: R12, R14, R52.

Keywords: Functional Urban Areas. GIS data. Ecuador. Colombia. Travel time.

Acknowledgements

We thank Landscan project for the density database on Colombia and Ecuador. M. Obaco thanks AGAUR for founding part of this paper. V. Royuela thanks the support of ECO2016-75805-R. Authors are especially gratful to Daniel Sanchez Serra, Javier Romaní, José Manuel Casado Díaz, and Kyle Thomson for providing valuable feedback..

Moises Obaco: AQR Research Group-IREA. University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain. Email: mobacoal7@ub.edu.
Vicente Royuela: AQR Research Group-IREA. University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain. Email: vroyuela@ub.edu.
Xavier Vítores, Independent Statistical Researcher, Email: xavier.vs78@gmail.com

1. Introduction

Integrated cities are generated by the urban and functional expansion of cities beyond their administrative boundaries. Identifying the right dimension of urban zones is important, as urban areas concentrate most of the population and economic activity, and are the engines of their respective regions and countries.

There is a large diversity of names for such urban areas (Metropolitan areas, funtional regions, urban zones, conurbations, urban regions, large urban areas, metropolis, etc.) what illustrates the complexity of the phenomenon. Despite many country-specific definitions of urban areas, until recently there was no harmonised economic definition of a city.

One of the more ambitious developments in this regard has been the OECD cross-country analysis of cities. Together with the European Commission, they developed a new definition of a city and its commuting zone in 2011 under the label of Functional Urban Areas (FUAs). This initiative increases international comparability and helps collecting statistical data. The methodology identified 1,179 urban areas of different size in 29 OECD countries, which gave as a further result the OECD metropolitan dataset, which considers 275 cities with 500,000 population and more.[1]

The method applied by OECD is grounded on the use of population density to identify urban cores and of commuting flows to identify policentricity and urban hinterlands. The latter data is available in most (if not all) developed countries, but this is usually not the case in developing countries. Consequently, some additional work is needed in order to generalise such methodology to the rest of the world. Our paper takes the witness and approaches the OECD definition of integrated cities in a country where there is a lack or that has poor administrative information on the socio-economic links to be used to connect spatial units.

We do this by using GIS data such as LandScan, Google maps and Open Street Maps. LandScan stores information about the density of a country in grid cells of one squared kilometre that allows to identify urban areas. Google and Open Street Maps give information on the road network system to connect urban areas. We show that travel time proxies accessibility, what can be a reasonable substitute of commuting information. Our approach is a feasible and robust solution that can be applied to most developing countries. Consequently, it can be very important pivot to measure and implement the Sutainable Development Goals.[2]

[1] The data base is publicly available at http://measuringurban.oecd.org/
[2] See http://www.un.org/sustainabledevelopment/cities/. Goal 11 aims at making cities inclusive, safe, resilient and sustainable.

The rest of the paper is structured as follows. Section 2 presents the background of the study. Section 3 shows the methodology while section 4 introduces the case of study and the used data. Results are displayed in section 5. Section 6 presents a robutness checks and Section 7 concludes summarizing the main outcomes of our work.

2. Functional Urban Areas

Administrative regions are "the expression of a political will: their limits are fixed according to the tasks allocated to the territorial communities, according to the sizes of population necessary to carry out these tasks efficiently and economically, and according to historical, cultural and other factors" (Eurostat, 1999, p.7). Even though they are not spatially random units, they are not the best spatial units to perform socio-economic analysis. One way to overcome the problems associated with administrative units is the identification and modification of political divisions in order to shape them in an existing social-economic relationship (Cörvers et al., 2009; Frey & Speare, 1992; Karlsson & Olsson, 2006). In this line, an FUA can be understood as the harmonized economic definition of "city": a functional economic unit (OECD, 2013). It has preference over the political definitions when we aim at analysing, designing or considering urban policies, although this creates tensions and causes planning problems, since several local governments are responsible for planning, which calls for cooperation between agents within an integrated city.

The process of clustering spatial units according similar characteristics or attributes is generally considered as regionalization procedure (Duque et al., 2007; Kim et al., 2013; Kim et al., 2015). Kim et al., (2016) identify three main shapes of regionalization: districts, coverage and incomplete coverage. Metropolitan areas usually involve the third definition, as they are based on centers of spatial concentration that are not exhaustive in space. The final coverage of every area is defined in terms of socio economic flows among spatial units. Such functional clusters provide a better way to define integrated urban systems. In this line, we find different approaches to define in the better way integrated areas as spatial clusters. See Davoudi, (2008) for a critical review and Adams et al., (1999); Tong & Plane, (2014) for particular applications.

We have to understand that urban agglomerations are the result of urbanisation processes, including the transformation of land cover and land use to categorize an area from non-developed to being developed (Pham et al., 2011; Weber, 2000). An urbanised space is characterised by its population density and population size. Nevertheless, an urbanised area is not only dense, but also integrated. Connected urban zones define the new boundaries (also known as hinterland or fringe). The connection can be defined by considering many alternatives, the most common being daily interactions in the labour market (Casado-Díaz & Coombes, 2011; Feria et al., 2015; Flórez-Revuelta et al., 2008;

Klapka & Tonev, 2013; Smart, 1974).[3]

The process of delimiting of FUAs followed by the OECD (2012, 2013) is applied in three identification steps. Firstly, urban cores are identified, according to some density measure. All areas above some minimum threshold of population density are then characterised as potential urban cores. Such threshold may vary for every country. The OECD applied a threshold of 1,500 inhabitants per km^2, a sill that was lowered to 1,000 inhabitants per km^2 for US and Canada.[4] Land cover using satellite imagery has been widely used in this identification step. Nowadays this information is available and easy to gather for most countries in the world (some recent examples of its use are Ferreira et al., 2010; Gisbert & Marti, 2014; Herold et al., 2003; OECD, 2013; Weng, 2012). The quality of such data will depend of the quality of the satellite images and the further recognition of density.

In this first step, a second condition must be fulfilled: areas need to contain a minimum of population size to be considered as an urban core. These minimum thresholds are established by the OECD at 50,000 inhabitants for Europe, US, Chile and Canada and 100,000 for Japan, Korea and Mexico, where cities are, on average, larger. In addition, as geographic areas usually do not coincide with administrative areas, the method assumes that a municipality is part of an urban core if the majority (at least 50%) of its population lives within the urban cluster.

The second identification step connects urban areas resulting from the first step that may not be contiguous but that belong to the same integrated space. This way FUAs account for polycentric urban structures. Two non-contiguous areas are associated if they show some amount of accessibility. The OECD uses labour commuting data and poses that two urban cores are integrated and belong to the same FUA, if at least 15% of the residence population of any of the cores commutes to work in the other core.

The third and final step of the methodology defines the hinterland or worker catchment area, this is, the area of influence of the urban cores, again considering such influence in terms of accessibility, materialised in labour commuting. The OECD defines this hinterland as all municipalities with at least 15% of their employed residents working in a certain urban core.

In the developing world, the scarcity of data is a huge problem for developing a suitable identification of these existing relationships in space. In turn, it becomes a very difficult task

[3] There are alternative approaches in the literature such as: services (Green, 1950), land prices (Bode, 2008), person-environment interactions (Murray et al., 2005; Van de Voorde 2011), quality of life aspects (Royuela et al., 2009).
[4] Recently (OECD, 2015) applied such methodology for identifying Chinese cities and lowered the threshold to 550 inhabitants per km^2.

to carry out any kind of analysis related to urban policies, planning or socio-economic analysis. Hence, this part of the world is hidden in most applied socio-economic analysis. Coombes (2004) proposed some alternative approaches to the use of commuting data to integrate urban system, such as internal migration flows, concentration indexes or cluster analysis. Internal migration requires again a good range of data and it presents some problems, being the biggest one the fact that migration does not only take place within urban areas, what can be interpreted as a substitute of commuting, but also between them.[5] Concentration indexes require again so much detailed information that in general is not available. Finally, cluster analyses do not consider integration links, which makes it a poor proxy.

In order to overcome the lack of commuting data, the gravity approach is a common option in territorial studies, including migration and trade (Ahlfeldt & Wendland, 2016; Cohen et al., 2008; Wang & Guldmann, 1996). The simplest expression derives flows as a result of a limited amount of data, including masses of population and distance between units (Goh et al., 2012). The gravity approach has been used to study commuting patterns: Vries et al., (2009) analyze the distance-decay function for commuting, while Persyn & Torfs (2015) derive a theoretical model for commuting and use count data models to overcome the problems of zeros in a large commuting matrix.

Recently, the radiation model has been used to estimate flows such as commuting or migration. Such models appeared first in physics to study the travel process of energetic particles or waves through vacuum. The model is parameter free, which makes it suitable for predicting flows when there is no data for setting parameters in gravitational models (Masucci et al., 2013; Simini et al., 2012).

Some authors have performed the task of identifying FUAs in developing countries. Commuting data is available in few (and recent) cases. Duranton (2015) uses commuting census of 2005 to define local labour markets in Colombia, while (Sanchez-Serra (2016) uses the OECD methodology to identify FUAs for this country. Several other works apply the concept of accessibility: the OECD used road network availability and gradient density in China (OECD, 2015) to identify FUAs. Rodrigues da Silva et al., (2014) use cluster analysis and road supply index in the Brazilian region of Bahia to identify functional regions. Gajovic (2013) uses artificial neural networks, isochrones and cluster analysis in Serbia. Arsanjani et al., (2014) propose that new techniques for FUA identification should be: easy to apply, requiring few data, and able to predict urban boundaries precisely.

[5] Jones (2010) and Royuela & Vargas (2009) use migration flows to define Housing Market Areas. The level of self-containment is substantially lowered compared to the one used with commuting algorithms. According to Royuela & Vargas (2009) commuting data is preferred over migration data to define Housing Market Areas.

In our work, we propose using the concept of accessibility expressed in terms of travel time on the road network system. It allows to measure and define proximity between urban cores and the extension of the worker catchment areas. This alternative has been already considered in other multinational experiences, such as in the ESPON project "Study on Urban Functions" (ESPON, 2005), where isochrones were fixed at 45 minutes to determine the boundaries.[6] Thus, our work connects previous experiences and links them to the standardized procedure based on the OECD definition of FUA. We present a thecnique that can be easily calibrated, and for which the data is available for most regions in the world.

3. Methodology

We follow OECD's methodology based on three steps. Next we describe those phases.

1. *Identifying urban cores*: Our first step is identical to the OECD's procedure. We identify high population density areas by using satellite data reporting grid cells, which are classified in terms of inhabitants per km^2. An area is categorised as high density if it is beyond a minimum threshold. We identify clusters of contiguous grid cells of high population density according to the majority rule.[7] The resulting high-density area is required to have a minimum population size to be considered an urban core. Finally, an administrative unit, e.g. a municipality, is part of an urban core if at least 50% of its population lives within the urban cluster

2. *Connecting non-contiguous urban cores that belong to the same functional area*: As described above, two non-contiguous urban cores belong to the same FUA if they are connected, what allows for poly-centricity in FUAs. This step requires the estimation of travel time between urban cores to infer if they are close enough to have social-economic interactions. Next, we introduce the assumption that urban cores follow a hierarchical pattern in space, having some areas a superior role than others. Then, a clustering algorithm sorts urban cores using the hierarchical variable, population size. Next, we test iteratively if any urban core is within a time threshold t, defined as the travel time from centroid to centroid of each urban core. The traveltime can be fixed for all urban cores or vary as a function of the area of every urban core. For the latter, we propose using a generic expression such as $Tc_i = \alpha_1 * A_i^{\beta_1}$, where Tc_i is the time in minutes from the urban core

[6] Travel time have also been considered in coverage analysis, where the main porpuse is to identify the spatial extent of the functional form. It usually involves covering the total demand for private or public services such as: emerging systems, fire stations, police stations, markets areas, etc (Togeras et al., 1971).

[7] To fill gaps in a high-density cluster we use iteratively the majority rule. Following the OECD (2013) procedure, the majority rule means that if at least five out to the eight cells surrounding a cell belong to the same high-density cluster, the lower-density cell will be added. This procedure is repeated until no more cells are merged.

and A_i is the geographical area of the urban core. Parameters α_1 and β_1 will vary according to every analysed case (country), what calls for some calibration.[8] If two urban cores are within such threshold of time, they are clustered, being the one with the lower hierarchy assigned to the one with the higher hierarchy. This procedure is repeated until there are no possible additional merges.

3. *Identifying the hinterlands or fringe*: The worker catchment area uses a new threshold, defined as travel time from the centroid of each urban core to surrounded political divisions that are not covered by urban cores. We can consider a fixed travel time for every urban core (e.g. 60 minutes). Alternatively we can think the maximun commuting time of a location is proportional to its size. Consequently, we derive a city-specific hinterland,[9] related to the dimension of each urban core by means of the following formula: $Th_i = \alpha_2 * A_i^{\beta_2}$, where Th_i is the time in minutes for the hinterland, A_i is the geographical area of the urban core and parameters α_2 and β_2, again, will need some calibration. Finally, if one area is linked to two urban cores, it will be associated to the largest FUA, as it represents the highest position in the urban hierarchy.

4. The Case study: FUAs in Ecuador and Colombia

We use Ecuador and Colombia as case study. Both are South American countries, with a total population at around 16 (Ecuador) and 47 (Colombia) million of people in 2014. Ecuador has a total territorial extension of 283,560 Km^2, close to Great Britain or Italy, although each of these two countries has about 60 million inhabitants. Colombia, on its side, is much a bigger country with 1,141,748 Km^2, doubling the size of Spain. The urbanization rate is around 65 percent for Ecuador and 75 for Colombia, being the average of Latin America around 70 percent.

In Ecuador, there is not commuting data, and consequently can be labelled as the focus of our work. Analysing Colombia allows for working with a developing country with available commuting data on a recent census (2005).[10] In addition, Colombia case allows to calibrate the parameters for Ecuador, because they share common characteristics; both Ecuador and Colombia are countries of regions, with large disparities and idiosyncratic characteristics in geographical, economic and socio-cultural terms; and road is the main network connection system. Large cities are found both in the mountainious areas (Bogotá, Medellín and Cali for Colombia and Quito for Ecuador) and in coastal areas (Barranquilla and Cartagena for Colombia and Guayaquil for Ecuador). In addition, both countries have an Amazon region,

[8] In a perfect circle $r_i = \sqrt{1/2\pi} * A_i^{-1/2}$, where r_i represents the radius. Parameters α and β will capture aspects such as average speed, geography, etc.
[9] This expression follows Ahlfeldt & Wendland (2016).
[10] It was gathered from the Departamento Administrativo Nacional de Estadística (DANE).

which, in fact, are not very populated compared with the others two regions. These similarities made them a good couple for comparison purposes. In particular, Ecuador is our best option for several other reasons. a) it is representative of many developing countries in the world in terms of population size;[11] b) its urbanization rate and population size characteristics allow for analysing changes in minimum thresholds; and c) it has not been previously analysed, and consequently it expands present knowledge in the applied literature.

We use land cover information, transport network, and demographic information at the lowest political division, municipalities for Colombia and parishes for Ecuador. The LandScan (2005)™ and LandScan (2013)™ dataset, developed by Oak Ridge National Laboratory, provide the land cover information based on Satellite Imagery[12]. It uses approximately 1 Km^2 resolution (30" x 30") and represents an ambient population (average over 24 hours). It is practically Raster information vectorized into SHP format. The roadways information comes from Google maps and Open Street databases[13] 2013. Political division at the local level comes from INEC (Instituto Nacional de Estadística y Censo)[14] for Ecuador and El Departamento Administrativo Nacional de Estadística (DANE) for Colombia.

Colombia has five natural regions: two on the coast (Pacific and Caribe), one on the Andean central highlands (Andes) and two on the planes (Amazonia and Orinoquia). The Landscan data sets report 334,215 grid cells of population density[15] (see figure A1.1 in the Appendix). Ecuador has four natural regions: the coastal plain (Costa), inter-Andean central highlands (Sierra), Eastern jungle (Oriente), and the Galapagos Islands (Insular). The final Landscan dataset considers 122,544 valid grid cells of 1 km^2 of population density. These are mainly concentrated at Coastal plain and inter-Andean central highlands regions (see figure A1.2 in the Appendix) in two specific urban poles, one located at the Coastal plain region (Guayaquil) and the other at the inter-Andean central highlands region (Quito).

In 2013 there are 1,046 parishes in Ecuador and 1,120 municipalities in Colombia for 2005. The mean of population density is around 120 inhabitants per km^2 in Ecuador and 128 inhabitants per km^2 in Colombia, and the median is around 35 inhabitants per km^2 for the former and 10 inhabitants per km^2 for the latter. In line with other countries, the distribution of population over municipalities follows a very lumpy and concentrated

[11] This is close to the average size of a country in the world once we exclude the 10 largest and 10 smallest countries.
[12] In this regard, we follow OECD, as they have also used the LandScan database.
[13] OpenStreetMap can be accessed at http://download.geofabrik.de
[14] We have considered the use of official data. However, there are other international databases with the same available information e.g. http://www.gadm.org/country, http://www.diva-gis.org/gdata and http://www.statsilk.com/maps/download-free-shapefile-maps .
[15] The Amazon region, which is a low populated region and with accecissibilty problems, the quality of the satallite imagery does not have a very good quality as the other regions of Colombia.

distribution. In addition, they are largely spatially heterogeneous.

In order to perform further robustness analysis in Ecuador, where there is not commuting data, we consider the Survey of Households' Living Conditions (SHLC) of 2014. Even though this survey is not designed to map the commuting pattern of the whole country, it reports information of this variable for a large sample of individuals. We use this source to report the average commuting time in Ecuador. Finally, we use the Ecuadorean National Census of Population 2010 to perform additional robustness checks based of the analysis of internal migration flows patterns, and the computation of commuting flows based on the gravity and radiation models.

5. Results

The first analysed country is Colombia. In this case we can use both the OECD methodology using commuting data and our approach considering road accessibility. Having data for the OECD approach allows for calibrating several parameters for the second procedure. We finally use such parameters for the Ecuadorean case.

A first decision has to be made on minimum thresholds for population density and urban size. As usual, chosen thresholds will depend on the type of considered policy. In our case, our decision must allow for capturing the maximum presence of urban settlements in the whole country, including the less populated regions, but may have representative urban settlements. For Colombia, previous examples are Metropolitan Areas with more than 100,000 inhabitants (Duranton, 2015) and FUAs with minimum population in clusters of 50,000 inhabitants and minimum density of 1,500 inhabitants per km^2 (Sanchez-Serra, 2016). Duranton (2015) considers a preferred threshold for commuting flows of 10%, while Sanchez-Serra (2016) follows the standard OECD criterium of 15% for commuting, although he also plays with lower figures, such as 10%, which is the threshold set in the national methodology to delimitate FUAs in Colombia (DNP, 2012). In order to work with less developed countries, usually less urbanised, we go below those minimun thresholds. Consequently, we set the minimum threshold for density in 500 inhabitants per km^2, which represent 2.5% of total grid cells for Colombia. As for the minimum threshold of population size of the urban core, we propose the use of 25,000 inhabitants. Finally, we set the minimum threshold for commuting flows at 10% for obtaining results for Colombia that will be used for calibrating our method. Consequently all thresholds are lower that the ones used in most developed world, what allows us to identify urban settlements in most parts of the country, where otherwise, small urban settlements would be invisible.

In line with several authors (Adams et al., 1999; Puderer, 2008), we assume that any technique and any threshold are somehow arbitrary. Nevertheless, our decisions are not far from other experiences. ESPON (2005) uses 650 inh./km² at the NUTS-5 level (municipalities) to identify level urban areas in Europe. OECD (2015) applies a minimum threshold of 550 inh./km² in China. Even, there has been also considered by authorities an urban density of 400 inh./km² (Demographia, 2015). In the same vein, the minimum size threshold is somehow flexible: Toribio (2008) argues that the typical population size to delimit a municipality as central core inside of a Metropolitan Area is 50,000 inhabitants. However, he used a minimum of population size of 100,000 inhabitants because he considered that Spain is a big country in demographic terms. The OECD used 50,000 for Europe and Gisbert & Marti (2014) used for Spain the minimum threshold of 1,500 inh./km² and 50,000 inhabitants to consider urban centers. Our decisions are consistent with the objective of maximising the number of FUAs in developing countries, where small and median cities are expected to grow in the near future (a process that is taking place in Ecuador, as explained in Royuela and Ordóñez, 2017). Later, we will analyse the sensitivity of our procedure to alternative thresholds.

Table 1 shows the results of the OECD methodology using commuting flows on the number of FUAs in Colombia based on 500 inh./km² as a minimum threshold for population density. We present the number of FUAs identified at three different minimum sizes for urban cores: 25,000, 50,000 and 100,000 inhabitants. The results are also presented for two alternative tresholds for commuting flows: 10% and 15%. Sánchez-Serra (2016) identifies 53 FUAs for a minimum population size of 50,000, 15% of commuting links and 1,500 inh. per km², while we identify 58 FUAs with a lower density threshold (500 inh. per km²). Our results show how increasing the minimum population size of urban cores reduces significantly the number of FUAs, and that increasing the threshold of commuting for merging urban cores drives to a larger amount of isolated FUAs.

With our preferred thresholds, we obtain 76 FUAs in Colombia. We use these units to calibrate the parameters of connectivity of step 2 of our methodology. Urban cores resulting from the 1st step can be linked by a fixed traveltime or vary as a function of the area of every urban core. We compute the average travel time of connected urban cores using the OECD methodology that considers commuting data.[16] This average figure is about 40 minutes, this is: on average, urban cores within 40 minutes of travel time belong to the same FUA. A second alternative is to allow that such time threshold vary with city size ($Tc_i = \alpha_1 * A_i^{\beta_1}$). By using again the information of connected urban cores we estimate this expression and get $Tc_i = 13 * A_i^{1/4}$.[17]

Step 3 is used to compute the hinterland of the FUAs. As result of the administrative division of the country, only 19 FUAs report hinterlands with additional municipalities to the

[16] Appendix 2 displays the considered options for getting these distances.
[17] Appendix 3 reports the basic computations for getting this expression.

original urban cores. We observe that larger urban cores are the ones with hinterlands, as they usually have better road connectivity. Again, we use this outcome, resulting from the OECD methodology that considers commuting data, to calibrate travel time as an expression of accessibility. As in the previous step, we can use a fixed travel time or a threshold that depends of the area of the urban core ($Th_i = \alpha_2 * A_i^{\beta_2}$). In the Colombian case this formula is $Th_i = 4.5 * A_i^{1/3}$.[18]

Bottom panel of table 1 displays the results based on road accessibility. We obtain the same amount of FUAs than using the commuting-based connectivity approach (76), being the descriptive statistics reasonably close. Such similarities hold while increasing the minimum threshold for population size. We obtain better aggregate summary statistics using a varying travel time approach than considering fixed thresholds.

Table 1: FUAs in Colombia based on commuting flows and travel time approach in Colombia (population in miles)

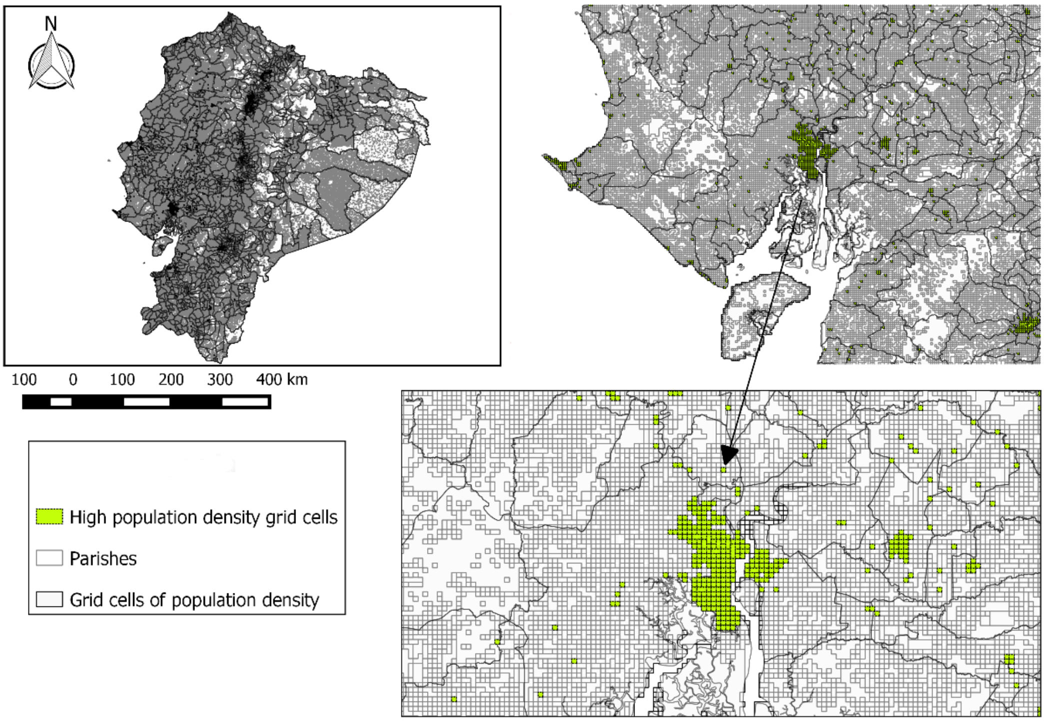
Min Pop	Urban cores	used Link	Total FUAs	Total Pop	Mean	Median	Min	Max	St. Desv.
Commuting based approach									
25	88	Commuting at least 10%	76	27,493	361	83	25	7,606	995
50	64		57	26,791	470	121	50	7,606	1,131
100	35		34	25,237	742	322	101	7,606	1,407
25	88	Commuting at least 15%	80	27,195	339	82	25	7,539	954
50	64		58	26,374	454	116	50	7,539	1,099
100	35		34	24,741	721	328	100	7,539	1,372
Accessibility based approach									
25	88	Fixed	69	27,214	494	149	25	7,654	1,156
50	64	travel	54	26,211	569	190	50	7,608	1,237
100	35	time	32	24,642	794	354	100	7,597	1,449
25	88	Varying	76	27,253	363	90	25	7,703	1,008
50	64	Travel	56	26,390	471	121	50	8,674	1,229
100	35	time	34	24,709	726	298	100	7,636	1,410

Once we have calibrated the parameters of our procedure for the Colombian case, we use them for computing Ecuadorean FUAs. For Ecuador, we have 3% of total grid cells as grid cell of high population density (above 500 inh. per km^2). Figure 1 displays the map for Ecuador with such cells together with a higher detail for the example of the largest city in the country, Guayaquil, which is composed by three administrative boundaries; Guayaquil, Durán and Samborondón.

[18] Appendix 4 explains the hinterlad computations and displays several maps for a group of FUAs including hinterlands.

Using our preferred thresholds, we identify 34 urban cores in Ecuador, which cover about 50% of total population and 80% of total urban population in the country in the considered year. In Appendix 5, table A5.1 displays some descriptive statistics of those urban cores, while Figure A5.1 maps the urban cores and the network system. Given its specific characteristics, we treated the Galapagos Island as a special case. If we aim at identify a city in such territory, the density threshold has to be set at 200 inhabitants for km^2 and a minimum population size of 10,000 inhabitants (see Appendix 5 for further details).

Figure 1. Grid cells of high population density. Detail for Guayaquil.



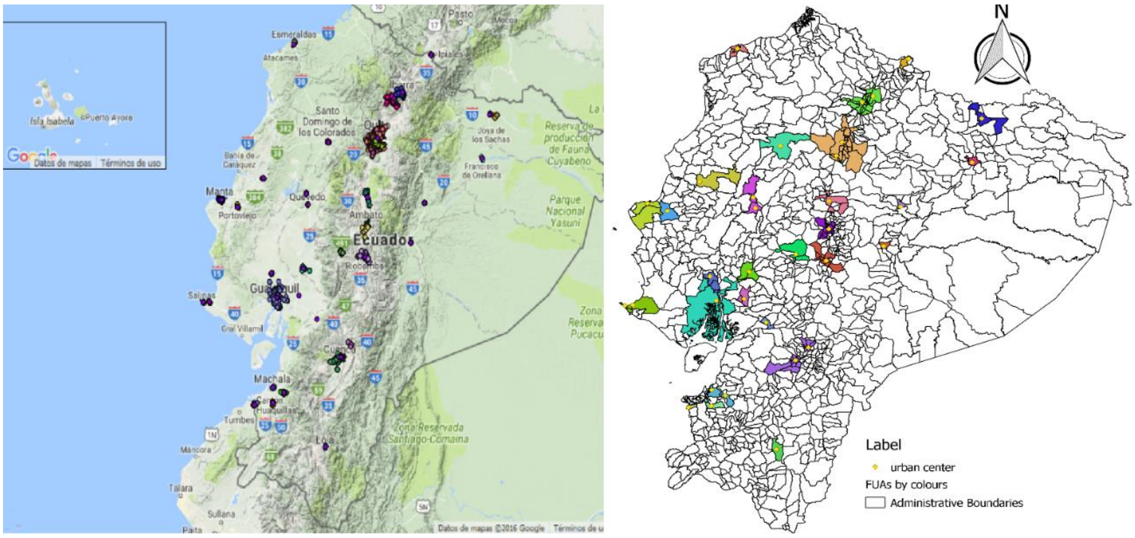
The second step connects non-contiguous urban cores that belong to the same functional area. Every urban core identified above is shaped into a polygon, for which we identify the centroid.[19] We then define the distance matrix by computing the time distance by road from centroid to centroid. In order to verify the travel time threshold for connecting urban cores, we have analysed the 2014 SHLC.[20] The survey contains information about 110,000 individuals, and around 50,000 are workers. We do not consider commuters within a city, we discard workers younger than 15 years old and workers than do not come back home in the same day. Finally 6,763 workers commute to another city per day, and 3,917 do it by bus the more popular transportation mode. The time of the median commuter using the bus

[19] One alternative is the use of the coordinates of the highest populated grid cell as the center of an urban core. We did not find significant changes by using either option.
[20] The SHLC 2014 is not representative at local level, only at the national or regional level. Similarly, this survey is not designed to capture commuting paterns.

is 60 minutes.[21] The median of people who commute using private car is about 30 minutes.

Like in Colombia, Google maps does not report actual travel time by public transport in Ecuador, but the one by private car, assuming roads in good conditions and fluent traffic. Developing countries usually have poor quality roads, congested traffic and buses networks with improvable efficiency. Consequently, we need to translate the 60 minutes by bus inferred from the SHLC into time distance by road reported by Google maps. We do so by comparing commutes reported at the SHLC with the time resulting from Google maps. We verify that 30 minutes by private car mode reported by Google maps is equivalent to 1 hour by bus.[22] Once we set such threshold, we apply our algorithm based on a hierarchical travel time approach[23]. By applying the clustering algorithm with such thresholds, we merge 4 high-density urban cores and we finally identify 30 FUAs, some of them being polycentric. If we allow varying the travel time, using the same equation for Colombia,[24] we identified 28 FUAs.

Figure 2: Results of FUAs in Ecuador



The final step delimitates the hinterland of every FUA, using the equation calibrated above.

[21] The mean of commuting by bus is 83 minutes. The median and mean of all commuters are 46 and 68 minutes respectively. The global average, then, is close to one hour of travel time and it is supported with the Marchetti's constants that fixes the average amount of commuting time in approximately one hour (Marchetti, 1994).
[22] Appendix 6 shows how we fit Google maps road distance with survey time distance.
[23] The first and third steps have been done using QGIS software. The second step has been programmed in Stata. The scritps are available upon request.
[24] The SHLC only identifies 326 people commuting between urban cores and only three urban cores can be connected using this information. In this case, is preferred applying the accessibility approach over using incomplete survey data.

Any municipality at a lower distance of the threshold is set to be part of the FUA.[25] The final list of FUAs are shown in the appendice. Figure 2 shows the hinterland analysis on the left side and the result in terms of administrative boundaries on the right side (different FUAs by colour). Appendix 7 reports the detailed list of FUAs.

5.1. Sensitivity analysis

This section explores the changes in our results for alternative minimum thresholds in Ecuador. Table 2 reports the number of urban cores when increasing the minimum threshold for density, minimum population size and travel time. As expected, increasing such thresholds imply a reduction in the total number of urban cores. No definition should be preferred a prior, although, in our view, in a country where urbanization is taking place, the identification of the maximum number of FUAs is preferred.

Table 2 reports an interesting situation for several threshold combinations. The highest minimum threshold of population density (1,500 inh./km^2) with a minimum population size of 25,000 inhabitants results in the fragmentation of large urban cores, and the creation of new and independent urban cores when compared to a lower threshold for density (1,000). Consequently, we believe that in the Ecuadorean case the chosen lowest minimum threshold of population density was more representative of urban cores across the country.

We also checked the influence of fixed versus size varying thresholds for connecting urban cores. Using varying thresholds we connect two urban cores that are also reported as having significant flows by the SHLC (San Jacinto de Buena Fe with Quevedo, and La Libertad with Santa Elena). Consequently, as happened in the Colombian case, size varying thresholds are preferred over fixed time thresholds.

[25] For instance, under a velocity of 60 km/h, the threshold for Quito, the urban core with the largest area, above 474 km^2 is set at 35 minutes by car and for the smallest FUA, San Jacinto de Buena Fe, with just 10 km^2, the threshold is set at 10 minutes by car.

Table 2: Sensitivity test of urban cores based on travel time

Density threshold	Grid cells	Minimum Size Threshold	Initial Number of Urban Cores	Results: # FUAs			
				Varying travel time (in minutes)	Fixed travel time (in minutes) 30m	60m	90m
500 inh./km2	3,699 (3%)	25,000	34	28	30	23	16
		50,000	21	20	20	16	14
		100,000	16	15	15	13	12
1,000 inh./km2	2,114 (1.75%)	25,000	29	27	28	22	15
		50,000	20	20	20	16	14
		100,000	16	15	15	13	12
1,500 inh./km2	1,532 (1.25%)	25,000	33	29	31	22	15
		50,000	21	20	20	16	14
		100,000	16	14	15	13	12

6. Robustness checks

In this section, we compare the FUAs obtained for Ecuador using our accessibility approach against urban clusters derived from actual and generated socioeconomic flows, as there is not commuting data. Next we describe all considered alternatives to use or generate such flows:[26]

- **Survey of Household's Living Conditions 2014 (SHLC 2014):** as reported above, this survey has information of commuters, although it is not designed for having a representative picture at the local level, as we need. There is information of 6,763 commuters from around 50,000 workers. It is a matrix of 641 parishes of origin by 540 of destination, but only 2,800 pairs of origin-destinations have non-zero values. The percentage of commuting flow is obtained from the total outflow of commuters from origin *i* to destination *j*, divided by total interviewed in *i*.

- **Gravitational Approach:** We use a gravitational approach to estimate the full matrix of commuting for the whole contry at the local level. The parameters of the gravitational function are obtained by using the commuting information of the SHLC 2014 and the National Census of Population 2010. The specification is a Zero Inflated Negative Binomial model of the between-urban mobility. The considered variables in this model are the rescaled commuting flow, total population, and geographical distance.

[26] Additional details for every method are reported in Appendix 8.

- Radiation model: We consider the radiation model (Simini et al., 2012) reports flows between municipalities without any parameterization. This method requires the total outflow of commuters from the origin municipality, and population in origin and destination, that we obtain from the National Census of Population of 2010.

- Internal Migration: We use matrix of internal migration among parishes between 2005 and 2010, which was gathered from the National Census of Population 2010. We can expect migration flows to be strong within FUAs, proxying commuting flows. Nevertheless, these movements are mixed together with migration between cities, what is far more complicated for commuting. Consequently, we have to differenciate between “movers” and migrats (Zax, 1994). The number of parishes at the migration matrix considers 1,149 origins and 1,211 destinations. We impose a geographical distance restriction between urban cores so that any move beyond such threshold will be a migration between FUAs rather than within them. The restriction of distance was 30 minutes by car, what according with Google maps is, on average, 35 km. In this case we use a flows threshold at 15%, in line with other works comparing these methodologies (Royuela & Vargas, 2009).

Table 3 presents some descriptive statistics of the flows resulting from the reported alternatives. As can be expected they are are relatively similar. As can be expected, the rescaled number of total commuters from the SHLC 2014 reports several outliers. Similarly, migration flows are quite heterogeneous compared with what we find in gravity and radiation models.

Table 3. Descriptive statistics of commuters

	Obs.	Min	Max	Mean	Median	St. Dev.
Rescaled SHLC	558,902	0	91,403	2.99	0	161.88
Gravity equation	1,024,140	0	4,537	1.54	0	28.71
Radiation model	1,024,140	0	7,563	0.94	0	29.91
Migration flows	1,024,140	1	13,453	12.03	2	98.55

Every described alternative report different flows between municipalities. We use as starting point the 34 urban cores resulting from the first step of the procedure, which were identified using the minimum density of 500 inh./km^2 and minimum population size of 25,000 inhabitants. Then we incorporate the computed flows into the OECD procedure to create alternative FUAs, which we compare with the ones obtained using our accessibility approach. The OECD procedure using commuting flows assumes a minimum threshold of at least 10%, while it is set at least 15% for internal migration.

Table 4: Comparative analysis of results among all applied methodologies in terms of population contained in each FUA

	FUAs (1)	Min (2)	Max (3)	Mean (4)	Median (5)	St. Dev. (6)	Population in FUAs (% of Total) (7)
Accessibility (varying Travel time)	28	37,663	2,812,609	357,320	172,578	663,008	10,004,967 (63.80%)
Commuting SHLC (10%)	31	53,237	2,930,848	340,763	150,258	658,285	10,222,899 (65.15%)
Commuting Gravitational (10%)	33	37,663	2,769,539	295,143	107,129	618,271	9,739,748 (62.07%)
Commuting Radiation (10%)	32	33,186	2,492,869	296,305	161,022	572,811	9,481,786 (60.05%)
Migration flows (15%)	29	59,312	2,558,798	417,070	280,325	634,405	11,260,940 (71.77%)

Table 4 displays the comparison table of FUAs in Ecuador. Column (1) shows the number of identified FUAs. Columns (2) to (6) present some descriptive statistics of population contained in those FUAs. Column (7) is the total population contained by those FUAs, and the percentage of population with respect of the country. Differences arise when using computed commuting flows, usually connecting less FUAs than our accessibility procedure. On the contrary, internal migration flows are the method connecting more urban cores, as expected, due to the presence of longer distance migration moves. Similarly, the migration option is the one capturing more population living in FUAs (over 11 million), while the other methods report about 10 million inhabitants. The hiternlands resulting from every method may differ in spatial terms, although the differences in population terms will be small, as every additional spatial unit can be expected to be small.

7. Conclusions

This paper identifies Functional Urban Areas when the researcher has no data on commuting flows. Here, we proxy the OECD methodology to identify FUAs by using accessibility and proximity expressed in travel time rather than actual flow data. Our starting point is the use of satellite imagery to identify urban cores. Next, we use travel time in a hierarchical approach to define potential interaction between urban cores and their hinterlands.

We apply our approach to Colombia, and then we extend it to Ecuador, a small country that we believe that can be representative of other developing countries. We test different minimum thresholds to identify cities and we calibrate our procedure with Colombian records, for which labour commuting data is avaliable. Low thresholds seem to better identify the largest number of cities in a country where urbanisation is taking place. We identify 34 urban cores that result in 30 FUAs using a fixed travel time and 28 FUAs using a size-varying travel time, two of them (Quito and Guayaquil) significantly large (2.5 and 2.8 million inhabitants respectively) the remaining of smaller size. Such areas account for more than 60% of total Ecuadorean population.

We perform some robustness checks based on survey and census data that are available for Ecuador. We consider commuting patterns directly derived from the SHLC 2014, and commuting flows resulting from a gravitational approach and a radiation model. We also compare our results with algorithms using internal migration flows. All methodologies report similar results, highlighting an important concentration of urban population in those identified urban cores.

Our approach, then, allows researchers, policy makers and planners to have a better perspective of the integrated cities in the developing world and introduce a methodology that can be applied minimizing the need of administrative information. Still, several drawbacks are present. First of all, any approach based on accessibility is actually mixing labour market outcomes with other socio economic flows such as leisure or study commuting. A detailed calibration with labour data of a close country is advisable to overcome this potential problem. In addition, we admit that our approach is based on GIS Google and Open sterrt maps assumptions for speed. For example, we do not model explicitly for congestion in larger cities, even though we try calibrate our approach comparing survey and maps travel time in order to partially overcome this problem. Clearly, both aspects could be tailored with improved data, which is usually the lacking dimension that motivates our work.

References

Adams, J. S., VanDrasek, B. J., & Phillips, E. G. (1999). Metropolitan Area Definition in the United States. *Urban Geography*, *20*(8), pp. 695–726.

Ahlfeldt, G. M., & Wendland, N. (2016). The spatial decay in commuting probabilities: Employment potential vs. commuting gravity. *Economics Letters*, *143*, pp. 125–129.

Arsanjani, J. J., Helbich, M., & Mousivand, A. J. (2014). A Morphological Approach to Predicting Urban Expansion. *Transactions in GIS*, *18*(2), pp. 219–233.

Bode, E. (2008). Delineating Metropolitan Areas Using Land Prices. *Journal of Regional Science*, *48*(1), pp. 131–163.

Casado-Díaz, J. M., & Coombes, M. (2011). The delination of 21st century local labour market areas: a critical review and a research agenda. *Boletín de la Asociación de Geógrafos Españoles*, (N.57).

Cohen, J. E., Roig, M., Reuman, D. C., & GoGwilt, C. (2008). International migration beyond gravity: a statistical model for use in population projections. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(40), pp. 15269–15274.

Coombes, M. (2004). Multiple Dimensions of Settlement Systems: Coping With Complexity, in: *New forms of urbanization: beyond the urban-rural*, p. Chapter 16.

Cörvers, F., Hensen, M., & Bongaerts, D. (2009). Delimitation and Coherence of Functional and Administrative Regions. *Regional Studies*, *43*(1), pp. 19–31.

Demographia. (2015). Demographia World Urban Areas & Population Projections. *Demographia*, (January), p. 132.

DNP (2012), *Algunos aspectos del análisis del sistema de ciudades colombiano*, Bogotá, Colombia.

Duranton, G. (2015). A proposal to delineate metropolitan areas in Colombia. *Desarrollo y Sociedad*, (75), pp. 223–264.

Duque, J. C., Ramos, R. & Suriñach, J (2007). Supervised Regionalization Methods: A Survey. *International Regional Science Review,* 30(3), pp. 165-220.

ESPON. (2005). *Potentials for polycentric development in Europe*.

EUROSTAT. (1999). Regio Database, User's Guide, Methods and Nomenclatures. Official Publication Office, Luxemburg.

Feria, J. M., Casado-Díaz, J. M., & Martínez-Bernabéu, L. (2015). Inside the metropolis: the articulation of Spanish metropolitan areas into local labor markets. *Urban Geography*, *3638*(October), pp. 1–24.

Ferreira, J. A., Condessa, B., Castro e Almeida, J., & Pinto, P. (2010). Urban settlements delimitation in low-density areas-An application to the municipality of Tomar (Portugal). *Landscape and Urban Planning*, *97*(3), pp. 156–167.

Flórez-Revuelta, F., Casado-Díaz, J. M., & Martínez-Bernabeu, L. (2008). An evolutionary approach to the delineation of functional areas based on travel-to-work flows. *International Journal of Automation and Computing*, *5*(1), pp. 10–21.

Frey, W. H., & Speare, A. (1992). *Metropolitan areas as functional communities a proposal for a new definition. Research Report No. 92-245.*

Gajovic, V. (2013). Comparative analysis of different methods and obtained results for delineation of functional urban areas. *Spatium*, (29), pp. 8–15.

Gisbert, F. J. G., & Marti, I. C. (2014). El concepto europeo de ciudad : una aplicación para España. *, 30*, pp. 145–156.

Goh, S., Lee, K., Park, J. S., & Choi, M. Y. (2012). Modification of the gravity model and application to the metropolitan Seoul subway system. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *86*(2), pp. 1–6.

Green, F. H. W. (1950). Urban Hinterlands in England and Wales : An Analysis of Bus

Services. *The Geographical Journal*, *116*(1), pp. 64–81.
Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: Measurement, analysis and modeling. *Remote Sensing of Environment*, *86*(3), pp. 286–302.
Jones, A. M., & Ones, a N. M. J. (2010). Models For Health Care M ODELS FOR H EALTH C ARE. , (January).
Karlsson, C., & Olsson, M. (2006). The identification of functional regions: theory, methods, and applications. *The Annals of Regional Science*, *40*(1), pp. 1–18.
Kim, H., Chun, Y., & Kim, K. (2013). Delimitation of Functional Regions Using a p-Regions Problem Approach. *International Regional Science Review*, *38*(3), pp. 235–263. Retrieved from http://irx.sagepub.com/cgi/doi/10.1177/0160017613484929
Kim, K., Chun, Y., & Kim, H. (2016). p-Functional Clusters Location Problem for Detecting Spatial Clusters with Covering Approach. *Geographical Analysis*, pp. 101–121.
Kim, K., Dean, D. J., Kim, H., & Chun, Y. (2015). Spatial optimization for regionalization problems with spatial interaction: a heuristic approach. *International Journal of Geographical Information Science*, *30*(3), pp. 451–473.
Klapka, P., & Tonev, P. (2013). FUNCTIONAL REGIONS OF THE CZECH REPUBLIC : COMPARISON OF SIMPLER AND MORE ADVANCED METHODS OF REGIONAL TAXONOMY. , *44*(1), pp. 45–57.
Marchetti. (1994). Anthropological Invariants in Travel Behavior. *Technological Forecasting and Social Change*, *47*, pp. 75–88.
Masucci, a. P., Serras, J., Johansson, A., & Batty, M. (2013). Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *88*(2), pp. 1–9.
Murray, G., Judd, F., Jackson, H., Fraser, C., Komiti, A., Hodgins, G., et al. (2005). The Five Factor Model and Accessibility/Remoteness: Novel evidence for person–environment interaction. *Personality and Individual Differences*, *39*(4), pp. 715–725.
OECD. (2012). *List of urban areas by country*.
OECD. (2013). *Definition of Functional Urban Areas (FUA) for the OECD metropolitan database*.
OECD. (2015). *OECD Urban Policy Reviews: China 2015*.
Persyn, D., & Torfs, W. (2015). A gravity equation for commuting with an application to estimating regional border effects in Belgium. *Journal of Economic Geography*, *16*(March 2015), pp. 155–175.
Pham, H. M., Yamaguchi, Y., & Bui, T. Q. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. *Landscape and Urban Planning*, *100*(3), pp. 223–230.
Puderer, H. (2008). *Defining and Measuring Metropolitan Areas : A Comparison Between Canada and the United States Geography*. Canada.
Rodrigues da Silva, A. N., Manzato, G. G., & Pereira, H. T. S. (2014). Defining functional urban regions in Bahia, Brazil, using roadway coverage and population density variables. *Journal of Transport Geography*, *36*(2014), pp. 79–88.
Royuela, V. & Ordoñez, J (2017) "Internal migration in a developing country: A panel data analysis of Ecuador (1982-2010)", *Papers in Regional Science*, forthcoming
Royuela, V., Romaní, J., & Artís, M. (2009). Using quality of life criteria to define urban areas in Catalonia. *Social Indicators Research*, *90*(3), pp. 419–440.
Royuela, V., & Vargas, M. A. (2009). Defining Housing Market Areas Using Commuting and Migration Algorithms: Catalonia (Spain) as a Case Study. *Urban Studies*, *46*(11), pp. 2381–2398.
Sanchez-Serra, D. (2016), "Functional Urban Areas in Colombia", OECD Regional

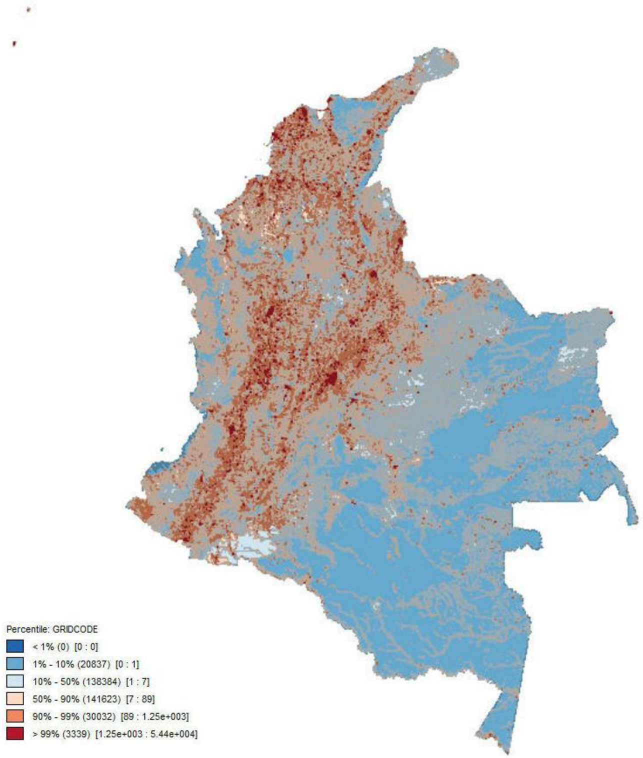
Development Working Papers, 2016/08, OECD Publishing, Paris.
Simini, F., González, M. C., Maritan, A., & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, pp. 8–12.
Smart, M. . (1974). Labour market areas: Uses and definition. *Progress in Planning*, *2*(July), pp. 239–353.
Tong, D., & Plane, D. A. (2014). A New Spatial Optimization Perspective on the Delineation of Metropolitan and Micropolitan Statistical Areas. *Geographical Analysis*, *46*(3), pp. 230–249.
Toregas, C., Swain, R., ReVelle, C., & Bergman, L. (1971). The Location of Emergency Service Facilities. *Operations Research, 19*(6), 1363-1373.
Toribio, J. F. (2008). Un ensayo metodológico de definición de las áreas metropolitanas en España a partir de la variable residencia-trabajo. *Investigaciones Geográficas*, *46*, pp. 49–68.
Van de Voorde, T., Jacquet, W., & Canters, F. (2011). Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data. *Landscape and Urban Planning*, *102*(3), pp. 143–155.
de Vries, J. J., Nijkamp, P., & Rietveld, P. (2009). Exponential or power distance-decay for commuting? An alternative specification. *Environment and Planning A*, *41*(2), pp. 461–480.
Wang, F., & Guldmann, J. M. (1996). Simulating urban population density with a gravity-based model. *Socio-Economic Planning Sciences*, *30*(4), pp. 245–256.
Weber, C. (2000). Urban agglomeration delimitation using remote sensing data, in: *Remote sensing and urban analysis: GISDATA 9*, p. 304.
Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, *117*(August), pp. 34–49.
Zax, J. S. (1994). When is a move a migration? *Regional science and urban economics*, *24*(3), pp. 341–360.

Appendices for "Building Functional Urban Areas based on road accessibility: an applied case in the developing world"

Appendix 1. Colombia and Ecuador description

Colombia: the data came from the Departamento Administrativo Nacional de Estadística (DANE). We use LandScan 2005 to identify urban cores municipalities. See figure A1.1.

Figure A1.1: Colombia: Population Distribution of High Population Density: cells with at least 500 inhabitants.



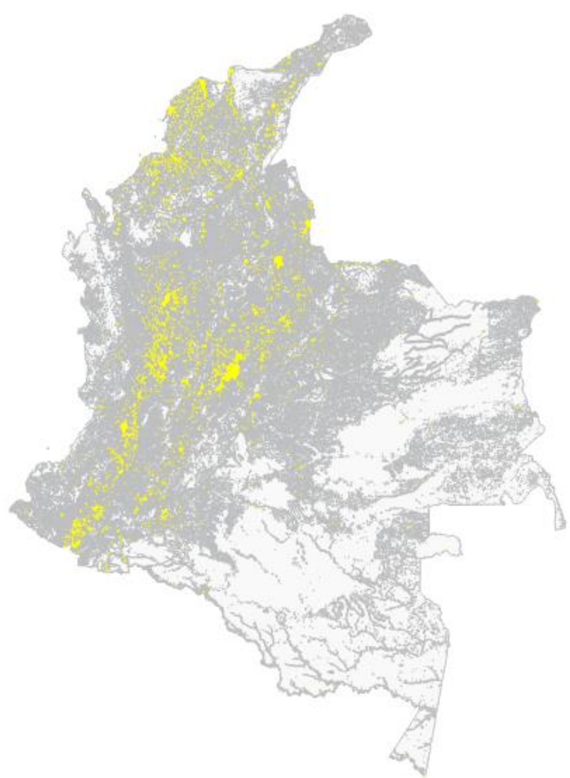
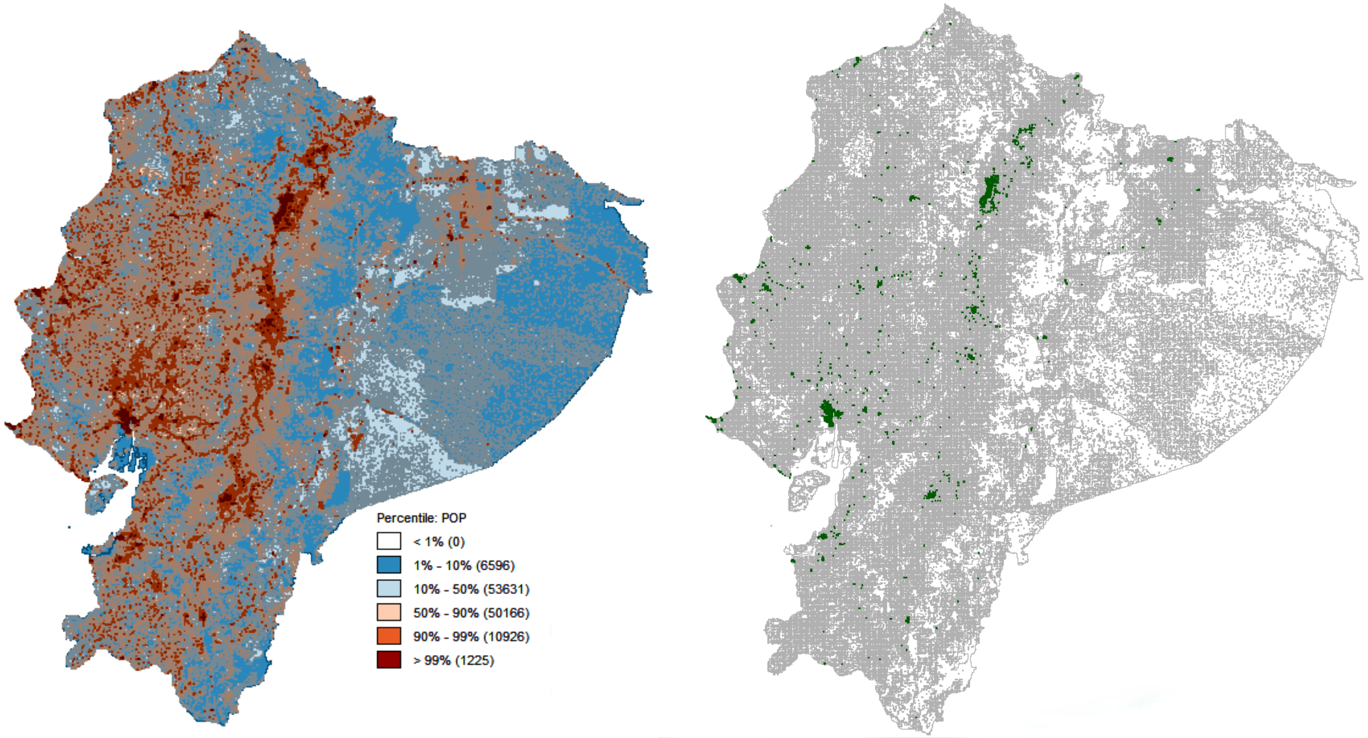


Figure A1.2: Ecuador: Population Distribution of High Population Density: cells with at least 500 inhabitants.



Appendix 2. Colombia and Ecuador description

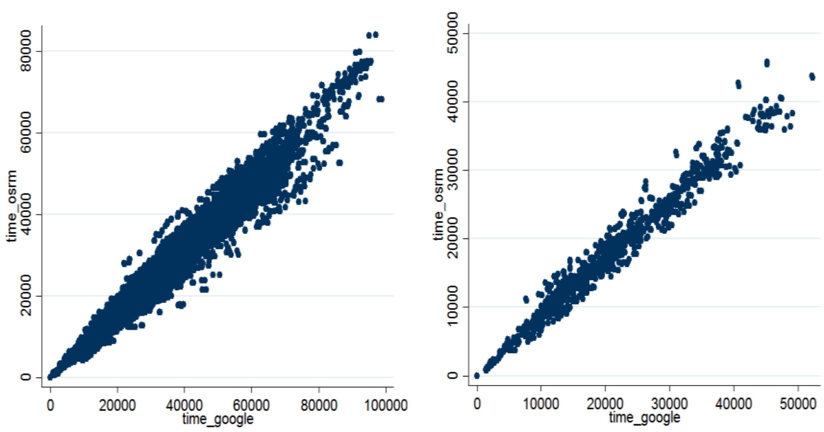
We consider up to three possibilities to compute geographic distances:

1. API Google maps: It is useful when the distance between urban cores there is not computed yet, so it computes at that moment using the Google maps service. We compute these distances by means of the traveltime3 Stata command. We notice that Google maps service has a limitation in the computation of distance per day, around 5,000 distances. See http://jearl.faculty.arizona.edu/sites/jearl.faculty.arizona.edu/files/traveltime3%20geocode3_b.pdf
2. Open Street Maps: It works in a similar way, but using the OSRM database. We use *osrmtime* Stata command. While there is not a limitation in the computation of time per day, the database needs to be downloaded, and installed previously (also updated). Consequently, it needs more minimum hardware requirements for working. See https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2691551
3. Origin-destination matrix: we leave open the possibility to upload a self-computed distances matrix, for instance coming from surveys or alternative data sources.

We compare the differences in travel time between the Open Street Maps and Google time. On average Open Street Maps distances travel time are faster. Our preferred option is the use of Google maps. However, its limitation in use per day and the unavailability to download the roads makes OSRM the best complementary data base. Consequently, we suggest using Google time in the second step and OSRM time in the third step.

Figure A2.1: Google Maps vs Open Street Maps travel time between urban cores

(a) Colombia **(b) Ecuador**



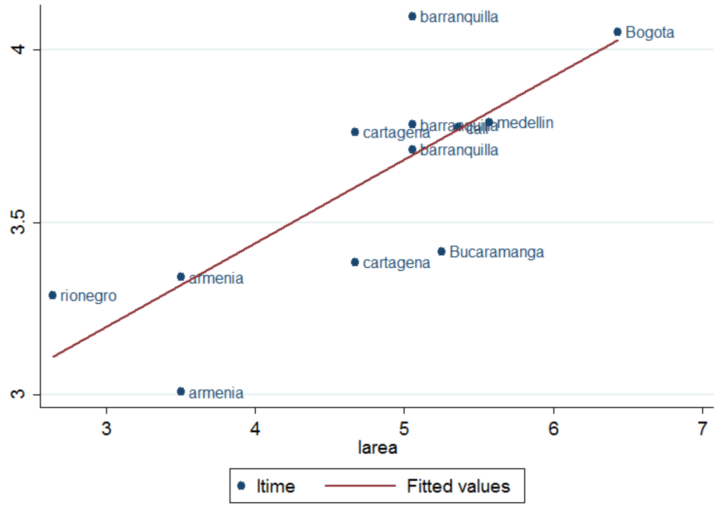
Appendix 3. Calibration of parameters for connecting urban cores (step 2)

Table A3.1 reports the 12 urban cores (origin) that are connected with other urban cores of higher hierarchical level (destination). This information allows us to display the average travel time of connected urban cores, that we set at 40 minutes. A fixed travel time can be a good proposal, but it may be not the optimal one. We explore the relationship between commuting patterns and urban size. Figure A3.1 shows the scatterplot between the log of the area of the destination urban core and the log of time between connected urban cores.

Table A3.1. Connected urban cores at 10% commuting flow (identified at 500 inh., 25,000 inh.)

Origin ID	Dest ID	Urban Core Origin Name	Population Size Origin	Urban Core Destination Name	Population Size Destination	Origin-Destination Commuting Flow	Origin-Destination Time	Area (size) Destination
5308	5001	Girardota	42566	Medellin	2214494	0.1891	44	263.22
5148	5615	El Carmen	41012	Rionegro	100502	0.1331	27	14.03
8638	8001	Sabanalarga	86631	Barranquilla	1146359	0.1285	60	156.87
8078	8001	Baranoa	51571	Barranquilla	1146359	0.2665	41	156.87
8634	8001	Sabanagrande	25399	Barranquilla	1146359	0.2921	44	156.87
25175	11001	Chia	97896	Bogota	6840116	0.2301	57	620.78
13052	13001	Arjona	60407	Cartagena	892545	0.1831	43	106.56
13836	13001	Turbaco	63046	Cartagena	892545	0.3362	30	106.56
63401	63001	La Tebaida	33504	Armenia	280930	0.1241	28	33.15
63130	63001	Calarca	73741	Armenia	280930	0.1818	20	33.15
68547	68001	Piedecuesta	117364	Bucaramanga	516512	0.2411	30	190.48
19573	76001	Puerto Tejada	44324	Cali	2119908	0.1137	44	213.54

Figure A3.1. Log(time) vs log(area) between connected urban cores



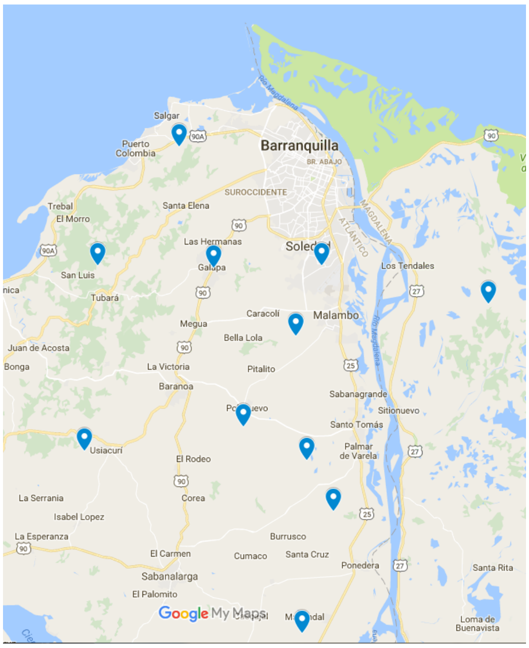
We finally regress log of time and the log of the area of the head of the FUA. We have a reasonable adjustment (R^2 about 60%). The constant is 2.473152 and the parameter 0.2417572, both significant at 1%. The final expression is: $Tc_i = 13 * A_i{}^{1/4}$.

Appendix 4. Calibration of parameters for computing hinterlands (step 3)

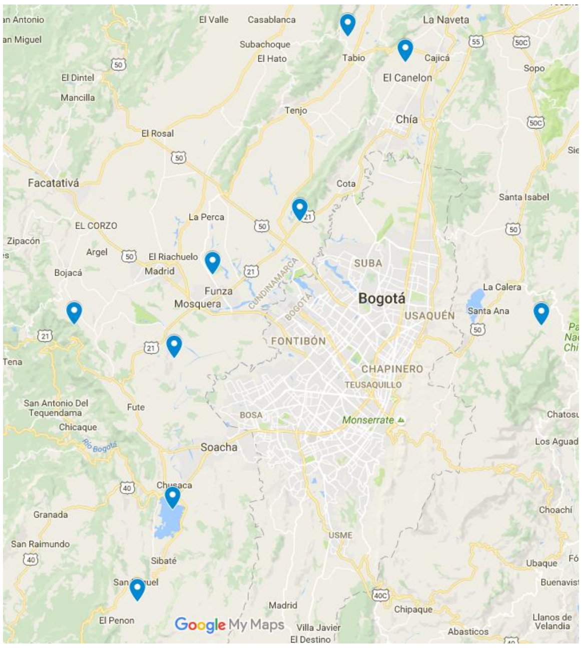
Figures A4.1 display the hinterlands for five Colombian cities: Barranquilla, Bogotá, Cartagena, Medellín and Cali. Every blue-point reports a municipality that belongs to the hinterland of every FUA. We consider as the hinterland distance the one of the farthest municipality of every FUA.

Figure A4.1. Hinterland zones in Colombia

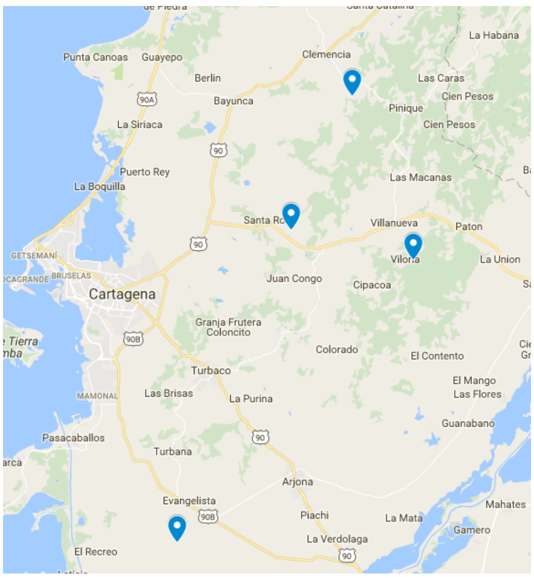
Hinterland of Barranquilla



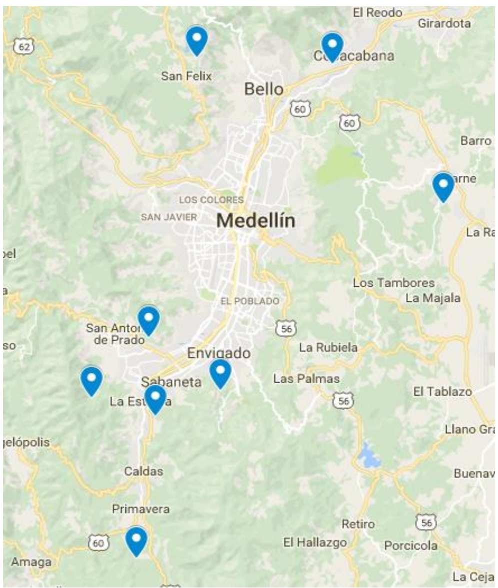
Hinterland of Bogota



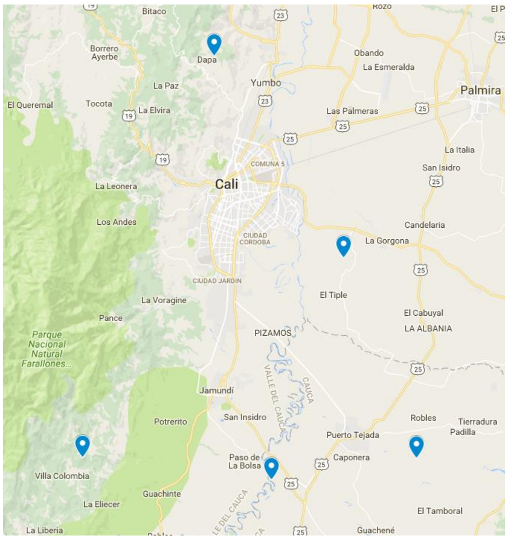
Hinterland of Cartagena



Hinterland of Medellin

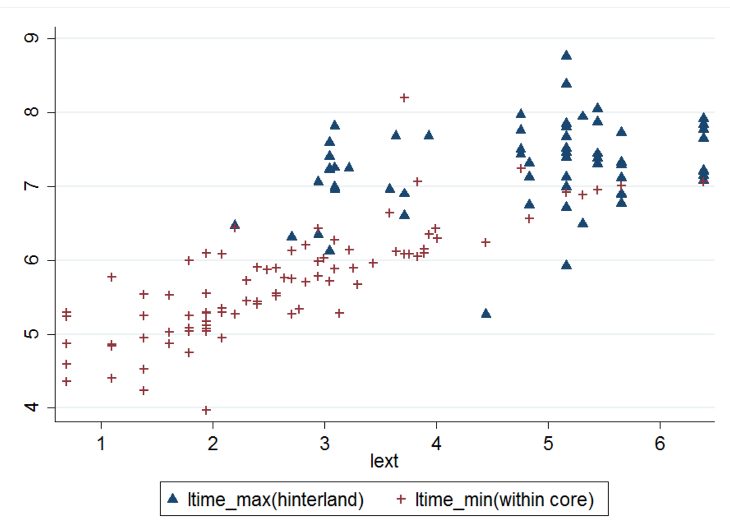


Hinterland of Cali



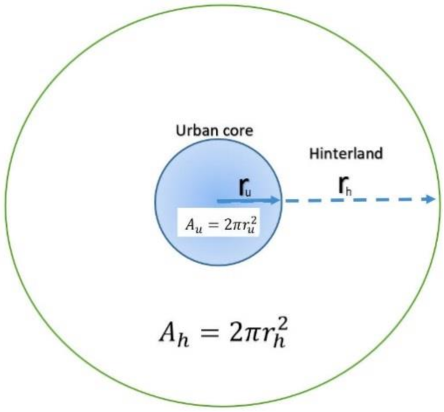
We now consider the relationship between the area of the *urban core*, and the distance of the farthest municipality in the *hinterland*. We assume that every FUA will have a hinterland (low density area) out of the urban core (characterised by high density). The administrative division of space, i.e. municipalities or parishes, makes that these hinterlands are usually within municipalities. We have computed the distance between the centroid of every urban core and the farthest coordinate of the FUA. Figure A4.2 plots the linear relationship between the size of the urban core of *all* FUAs and the maximum distance to every hinterland. Blue triangles represent those FUAs capturing alternative municipalities in the hinterland, while red crosses characterize FUAs where the hinterland is included in a single municipality. Consequently, in practical terms we only have to capture the hinterland of those FUAs adding new municipalities.

Figure A4.2. Hinterland zones in Colombia



Then, we look for a relationship where the area of the urban core is used to find the size of the hinterland (see figure A4.3.)

Figure A4.3. Hinterland approach



We can say that the hinterland area, A_h, is a function the urban core area, A_i.

$$A_h = \alpha_2 {A_i}^{\beta_2} \tag{A4.1a}$$

or

$$\log(A_h) = \ln(\alpha_2) + \beta_2 \ln(A_i) \tag{A4.1b}$$

Where, α_2 is an expansion factor and $\beta_2$is an adjustment factor. We may obtain the radius of the hinterland area as a function of the urban core, where the radius measured in distance is equal to time multiplied by a given velocity.

$$r_i = Dist_i = Th_i * speed = \sqrt{\frac{A_i}{2\pi}} = \sqrt{\frac{\alpha_2 {A_i}^{\beta_2}}{2\pi}} \tag{A4.2}$$

Considering a speed of 60km/h, we get an expression that allows estimating the maximum of travel time as a function of the area of the urban core. The empirical model becomes as:

$$\log(Th_i) = \frac{1}{2}\ln\left(\frac{\alpha_2}{2\pi}\right) + \frac{\beta_2}{2}\ln(A_i) \tag{A4.3}$$

$$\log(Th_i) = \alpha_2' + \beta_2' \ln(A_i) \tag{A4.4}$$

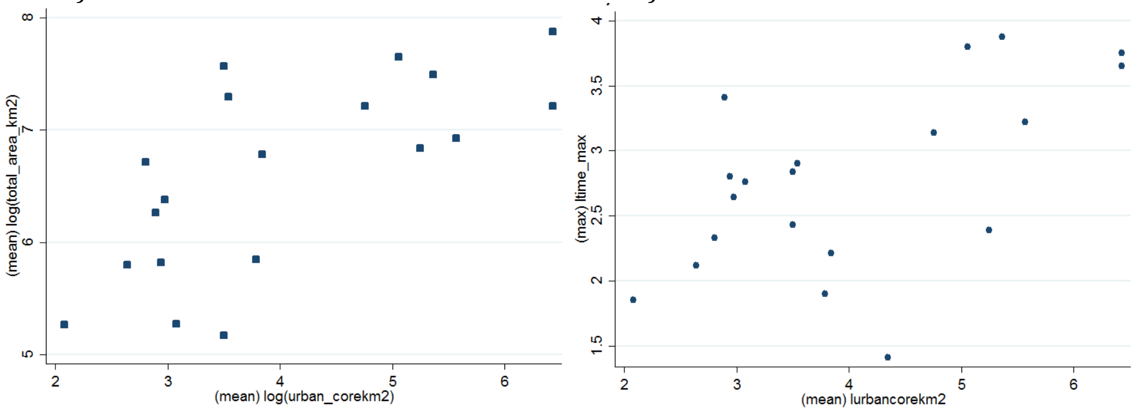
Equation (A4.4) is a simple linear equation that allows computing the size of the hinterland as a function of the size (area) of the urban core, what is particularly useful when there is not commuting data of the hinterland, as happens in the Ecuadorean case. To estimate equation (A4.4), we need the hinterland generated by urban cores and, for those hinterlands, we need the maximum travel time by urban core.[27] As can be expected, the areas of both urban cores and hinterlands, are not even close to a circle. In addition, administrative boundaries are relatively large compared to real settlements of those municipalities that belong to the hinterland. These characteristics are very close to the Ecuadorean case, where the administrative boundaries are relatively large compared with the municipalities extension as well. Finally, the radius using travel time, generated by using road

[27] We use maximum of travel time because to the mean or the minimum of the hinterland time do not have a significant slope with the size of urban core.

network measured in extension of Km, tend to be larger than the geographical radius. Figure A4.4.a) shows the relationship between the areas of urban core and urban hinterland, while A4.4.b) shows the relationship between maximum of hinterland time and the area of urban core.

Figure A4.4. Relationship between size of the urban core and size of the hinterland.

a) Total area vs urban core area b) Time hinterland vs urban core area



Distances were computed using the road network of Open Street Maps with a fixed speed of 60km/h in order to make the computations easier. In the same context, the area was expressed in km² and the travel time was recorded in minutes.

Table A4.1 introduces the results of estimate equation (A4.4) in column (1), equation (A4.1b) in column (2) and the radius of the hinterland against the total size of the hinterland (computed as the total area of all municipalities in the FUA) as robust check in column (3). All parameters are statistically significant and their values are the expected values within the confidence of interval. The adjustment of all regressions is quite similar, being larger

Table A4.1. Hinterland estimation

VARIABLES	(1) $\ln(time_h)$	(2) $\ln(area_h)$	(3) $\ln(ltime_h)$
$\ln(area_i)$	0.334***	0.459***	
	(0.0862)	(0.114)	
$\ln(area_h)$			0.501***
			(0.133)
Constant	1.498***	4.752***	-0.462
	(0.364)	(0.480)	(0.888)
Observations	19	19	19
R-squared	0.469	0.490	0.453

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

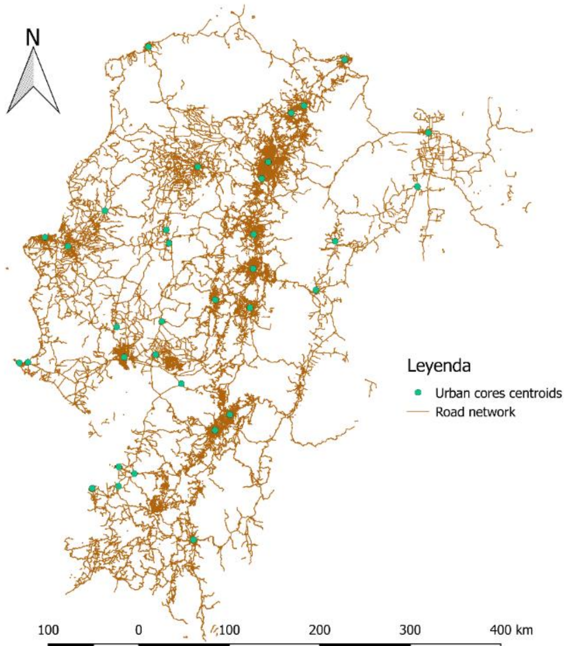
Using the parameters of column (1) find the final expression of the hinterland equation: $Th_i = 4.5 * {A_i}^{1/3}$. This time hinterland equation is an equivalent function of the maximum travel time, on average, that an urban core may have according its geographical extension.

Appendix 5. Ecuador: urban cores description

Table A5.1: Descriptive Statistics of Core Population (threshold of 500 inhabitants per grid cell)

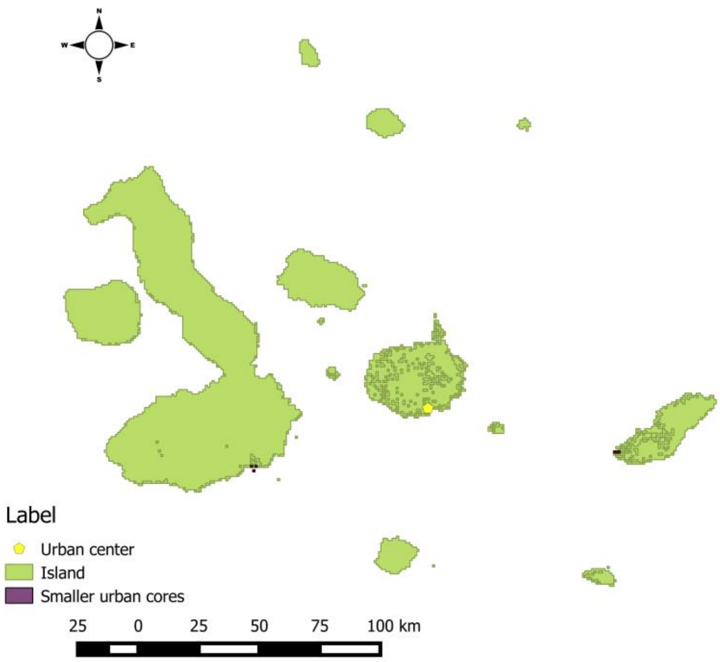
Reference Name	Pop. Size	Pop. Mean	Pop. Median	Pop. Max	Pop. Min	Pop. St.Dev.	Total cells	Fringe (minutes)	Area km2	Reference Region
Guayaquil	2553993	8238.69	5008.5	39800	0	9150.31	310	30	297	Coastal
Quito	2166700	4142.83	1753	41536	3	4950.62	523	35	474	Highland
Cuenca	347371	3581.14	1770	39473	92	4809.74	97	21	93	Highland
Manta	294618	3682.73	1910.5	21696	11	4337.59	80	19	70	Coastal
Santo Domingo	286186	8943.31	5531	31110	58	9217.87	32	14	29	Highland
Ambato	276507	2248.02	729	19390	7	3589.86	123	22	113	Highland
Machala	250088	6099.71	4272	43145	91	8935.10	41	15	36	Coastal
Portoviejo	212192	4330.45	1891	35823	112	7233.95	49	16	42	Coastal
Loja	180342	4293.86	1318	36652	392	7853.18	42	15	37	Highland
Esmeraldas	174433	4714.41	1849	19467	28	5388.00	37	15	32	Coastal
Riobamba	169165	4572.03	2008	24266	275	5950.39	37	15	33	Highland
Otavalo	167157	1168.93	893	5528	10	1229.94	143	23	127	Highland
Quevedo	158623	6100.88	2091	37498	563	1474.82	26	13	22	Highland
Libertad	157929	4644.97	2353	34035	0	6560.96	34	14	31	Coastal
Milagro	131806	5272.24	5213	12202	525	3317.68	25	13	22	Coastal
Ibarra	130131	3173.93	1755	19276	0	4062.01	41	15	37	Highland
Latacunga	79710	4195.26	1625	16304	535	4764.16	19	12	16	Highland
Babahoyo	71684	7964.89	2205	32503	819	1376.39	9	10	10	Coastal
Daule	69750	5812.5	1169.5	23606	511	7706.03	12	10	11	Coastal
Tulcan	55855	5585.5	4081.5	25846	599	7258.40	10	10	9	Highland
Nueva Loja	53787	2241.13	1778	5147	14	1536.48	24	13	21	Amazon
Huaquillas	49012	4455.64	3353	15801	1143	4119.98	11	10	9	Coastal
Chone	46159	3077.27	2250	7564	712	2498.53	15	11	13	Coastal
Pto.Orellana	45711	1987.43	1202	11981	3	2568.07	23	6	2	Amazon
Tena	39696	3308	1514.5	13105	223	3954.61	12	10	11	Amazon
Pasaje	39235	5605	3385	15888	892	5164.67	7	9	6	Coastal
Puyo	38318	3831.8	2035.5	11683	591	3962.50	10	10	9	Amazon
La Troncal	36678	4584.75	2986	19000	769	5959.36	8	9	7	Coastal
Santa Elena	35830	3981.11	2891	8839	81	3589.01	9	9	8	Coastal
Santa Rosa	32693	2335.21	1753.5	5987	256	1772.24	14	5	1	Coastal
Azogues	31361	2613.42	677	13855	398	4428.09	12	10	10	Highland
Cutuglahua	27797	1737.31	1241	6319	159	1508.30	16	11	14	Highland
Guaranda	27649	5529.8	5974	10648	1365	3626.97	5	8	5	Highland
S.J. de Buena Fe	25820	2347.27	1574	7580	732	1953.85	11	10	10	Coastal

Figure A5.1. Urban cores and road network system



For the Insular region (Galapagos Islands), in order to find an urban settlement we set the minimum density threshold at 200 inhabitants per km^2 and a minimum population size for the urban core at 10,000 inhabitants. As there is no road connection between cities in different islands, we applied a minimum distance between them is around 80 km from the largest urban core.

Figure A5.2. Galapagos' Islands



Appendix 6. Fitting Google maps road distance with survey time distance.

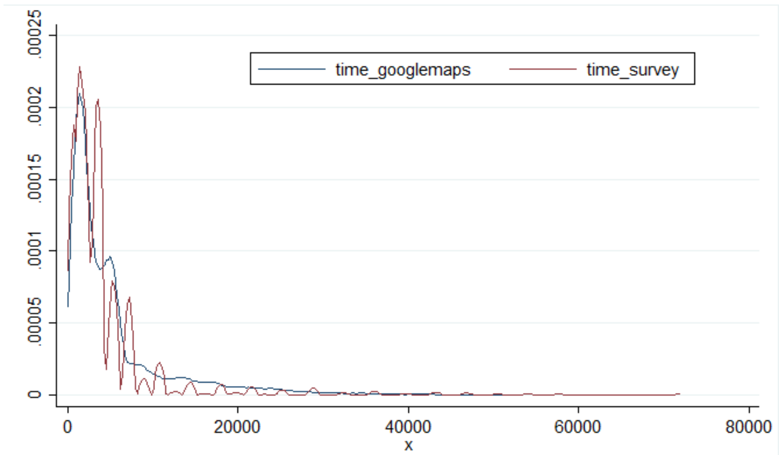
Here we fit Google maps road distance with survey time distance. We compare informed time of commuting at SHLS, from which we know origin and destination, against travel time by car computed using Google maps. The information at SHLS allows for considering the mode of transportation. We exclude marginal transportation modes, such as rides on animals, boats, airplanes, planes and those usual for short distances, such as walking and biking. Table A6.1 and figure A6.1 display some descriptive statistics.

Table A6.1. Travel time Survey vs Travel time google maps

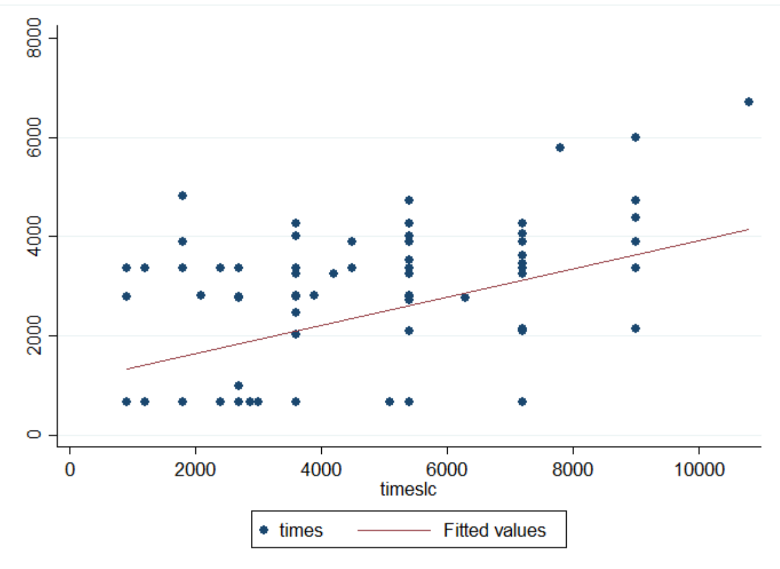
	Google time	Survey time
Mean	2161.907	3847.814
Std. Dev.	(1490.121)	(2065.578)

Figure A6.1. Google_time vs Survey_time

(a) Histogram



(b) Scatter plot



Appendix 7. Ecuadorean FUAs

HEAD	NAME	CODE	NAME	POP	Total Pop	Area (km2)	Total Area
10150	CUENCA	10167	SININCAY	17507	544619	2.477.326	8.549.378
10150	CUENCA	30151	COJITAMBO	4070	544619	1.742.162	8.549.378
10150	CUENCA	30150	AZOGUES	41924	544619	5.909.766	8.549.378
10150	CUENCA	30153	GUAPAN	9768	544619	5.934.959	8.549.378
10150	CUENCA	10170	VALLE	26840	544619	4.320.871	8.549.378
10150	CUENCA	10151	BAÑOS	18602	544619	3.268.521	8.549.378
10150	CUENCA	30250	BIBLIAN	14812	544619	6.624.175	8.549.378
10150	CUENCA	30252	SAN FRANCISCO DE SAGEO	1870	544619	3.947.172	8.549.378
10150	CUENCA	10162	RICAURTE	21373	544619	1.380.344	8.549.378
10150	CUENCA	10168	TARQUI	11580	544619	1.373.907	8.549.378
10150	CUENCA	10150	CUENCA	366378	544619	7.617.434	8.549.378
10150	CUENCA	10169	TURI	9895	544619	2.667.744	8.549.378
20150	GUARANDA	20157	SAN SIMON (YACOTO)	4569	66680	9.666.064	6.426.204
20150	GUARANDA	20158	SANTAFE (SANTA FE)	1904	66680	2.647.649	6.426.204
20150	GUARANDA	20150	GUARANDA	60207	66680	5.194.833	6.426.204
30450	LA TRONCAL	30450	LA TRONCAL	48798	48798	120.773	120.773
40150	TULCAN	40150	TULCAN	65608	65608	1.364.669	1.364.669
50150	LATACUNGA	50550	SAN MIGUEL	33693	158706	1.803.323	5.348.432
50150	LATACUNGA	50158	POALO	6218	158706	58.052	5.348.432
50150	LATACUNGA	50150	LATACUNGA	107129	158706	2.644.992	5.348.432
50150	LATACUNGA	50153	GUAITACAMA (GUAYTACAMA)	10530	158706	2.833.077	5.348.432
50150	LATACUNGA	50652	CHANTILIN	1136	158706	3.628.956	5.348.432
60150	RIOBAMBA	60155	LICAN	8598	242563	2.279.192	5.832.877
60150	RIOBAMBA	60754	SAN ANDRES	14419	242563	1.602.319	5.832.877
60150	RIOBAMBA	60150	RIOBAMBA	169232	242563	6.237.932	5.832.877
60150	RIOBAMBA	60450	CHAMBO	12702	242563	164.182	5.832.877
60150	RIOBAMBA	60152	CALPI	6985	242563	5.392.896	5.832.877
60150	RIOBAMBA	60750	GUANO	17667	242563	9.049.141	5.832.877
60150	RIOBAMBA	60161	SAN LUIS	12960	242563	2.928.216	5.832.877
70150	MACHALA	70950	PASAJE	58366	324200	1.317.941	3.509.497
70150	MACHALA	70953	LA PEAÑA	3929	324200	1.681.887	3.509.497
70150	MACHALA	70150	MACHALA	261905	324200	2.023.368	3.509.497
70750	HUAQUILLAS	70750	HUAQUILLAS	53237	53237	6.352.836	6.352.836
71250	SANTA ROSA	71250	SANTA ROSA	57497	57497	1.823.571	1.823.571
80150	ESMERALDAS	80166	TACHINA	4285	181657	7.004.777	211.222
80150	ESMERALDAS	80168	VUELTA LARGA	3224	181657	7.367.439	211.222
80150	ESMERALDAS	80150	ESMERALDAS	174148	181657	6.749.983	211.222
90150	GUAYAQUIL	92550	NARCISA DE JESUS	21989	2812609	1.367.417	3.088.488
90150	GUAYAQUIL	90750	ELOY ALFARO (DURAN)	263970	2812609	3.004.528	3.088.488
90150	GUAYAQUIL	90150	GUAYAQUIL	2466882	2812609	2.428.395	3.088.488
90150	GUAYAQUIL	91650	SAMBORONDON	59768	2812609	2.228.984	3.088.488
90650	DAULE	90656	LOS LOJAS	9894	109872	1.184.553	3.318.351
90650	DAULE	90650	DAULE	99978	109872	2.133.797	3.318.351
91050	MILAGRO	91050	MILAGRO	157608	163499	2.205.837	262.863
91050	MILAGRO	91051	CHOBO	5891	163499	4.227.928	262.863
100450	OTAVALO	100455	SAN JOSE DE QUICHINCHE	9215	370244	8.548.788	9.418.413
100450	OTAVALO	100250	ATUNTAQUI	25603	370244	2.632.024	9.418.413
100450	OTAVALO	100650	URCUQUI	5554	370244	6.185.766	9.418.413
100450	OTAVALO	100453	GONZALEZ SUAREZ	6120	370244	4.912.401	9.418.413
100450	OTAVALO	100350	COTACACHI	18221	370244	7.101.264	9.418.413
100450	OTAVALO	100356	QUIROGA	6861	370244	6.833.472	9.418.413
100450	OTAVALO	100458	SAN RAFAEL	5893	370244	1.785.818	9.418.413
100450	OTAVALO	100157	SAN ANTONIO	19140	370244	2.726.367	9.418.413
100450	OTAVALO	100154	LA ESPERANZA	8042	370244	3.422.664	9.418.413
100450	OTAVALO	100457	SAN PABLO	10764	370244	6.521.755	9.418.413
100450	OTAVALO	100254	SAN ROQUE	11145	370244	1.662.668	9.418.413
100450	OTAVALO	100450	OTAVALO	57352	370244	8.503.825	9.418.413
100450	OTAVALO	100150	IBARRA	152624	370244	2.416.631	9.418.413
100450	OTAVALO	100251	IMBAYA	1405	370244	1.197.713	9.418.413
100450	OTAVALO	100456	SAN JUAN DE ILUMAN	9332	370244	2.091.834	9.418.413

100450	OTAVALO	100451	DR. MIGUEL EGAS CABEZAS	5308	370244	8.415.863	9.418.413
100450	OTAVALO	100452	EUGENIO ESPEJO (CALPAQUI)	7998	370244	2.336.258	9.418.413
100450	OTAVALO	100252	SAN FRANCISCO DE NATABUE	6209	370244	1.338.881	9.418.413
100450	OTAVALO	100253	SAN JOSE DE CHALTURA	3458	370244	1.374.736	9.418.413
110150	LOJA	110150	LOJA	200217	200217	2.858.597	2.858.597
120150	BABAHOYO	120150	BABAHOYO	103837	126355	1.736.947	4.423.187
120150	BABAHOYO	120154	PIMOCHA	22518	126355	268.624	4.423.187
120550	QUEVEDO	120550	QUEVEDO	173559	230294	1.908.779	6.059.271
120550	QUEVEDO	121050	SAN JACINTO DE BUENA FE	56735	230294	4.150.492	6.059.271
130150	PORTOVIEJO	130150	PORTOVIEJO	239695	239695	4.182.158	4.182.158
130350	CHONE	130350	CHONE	78255	78255	8.289.122	8.289.122
130850	MANTA	132150	JARAMIJO	21489	338852	9.722.836	942.956
130850	MANTA	130950	MONTECRISTI	78793	338852	6.532.543	942.956
130850	MANTA	130850	MANTA	238570	338852	1.924.733	942.956
150150	TENA	150150	TENA	37663	37663	2.624.857	2.624.857
160150	PUYO	160150	PUYO	41228	41228	8.776.846	8.776.846
170150	QUITO	170176	PINTAG	19689	2499616	489.603	2431.5
170150	QUITO	170163	GUAYLLABAMBA	17803	2499616	5.568.621	2431.5
170150	QUITO	170357	UYUMBICHO	5152	2499616	2.094.473	2431.5
170150	QUITO	170151	ALANGASI	26630	2499616	2.917.464	2431.5
170150	QUITO	170152	AMAGUAÑA	34158	2499616	5.649.767	2431.5
170150	QUITO	170180	SAN ANTONIO	35531	2499616	1.116.152	2431.5
170150	QUITO	170551	COTOGCHOA	4416	2499616	3.639.438	2431.5
170150	QUITO	170353	CUTUGLAHUA	18730	2499616	2.843.727	2431.5
170150	QUITO	170155	CALDERON (CARAPUNGO)	167179	2499616	7.869.295	2431.5
170150	QUITO	170177	POMASQUI	31746	2499616	2.360.987	2431.5
170150	QUITO	170356	TAMBILLO	9304	2499616	4.647.712	2431.5
170150	QUITO	170164	LA MERCED	9217	2499616	3.197.443	2431.5
170150	QUITO	170186	ZAMBIZA	4411	2499616	7.535.862	2431.5
170150	QUITO	170179	PUEMBO	14926	2499616	3.172.738	2431.5
170150	QUITO	170170	NAYON	17169	2499616	1.598.328	2431.5
170150	QUITO	170157	CUMBAYA	34550	2499616	2.100.438	2431.5
170150	QUITO	170162	GUANGOPOLO	3359	2499616	1.028.442	2431.5
170150	QUITO	170150	QUITO	1778016	2499616	3.720.005	2431.5
170150	QUITO	170166	LLOA	1640	2499616	5.402.823	2431.5
170150	QUITO	170156	CONOCOTO	90124	2499616	388.751	2431.5
170150	QUITO	170550	SANGOLQUI	91024	2499616	5.710.419	2431.5
170150	QUITO	170184	TUMBACO	54844	2499616	6.548.754	2431.5
170150	QUITO	170175	PIFO	18278	2499616	2.543.441	2431.5
170150	QUITO	170165	LLANO CHICO	11720	2499616	7.763.803	2431.5
180150	AMBATO	180758	SALASACA	6363	333601	1.275.586	4.326.403
180150	AMBATO	180156	IZAMBA	15717	333601	2.904.289	4.326.403
180150	AMBATO	180166	TOTORAS	7444	333601	802.138	4.326.403
180150	AMBATO	180160	PICAIGUA	8939	333601	1.592.994	4.326.403
180150	AMBATO	180157	JUAN BENIGNO VELA	8047	333601	3.956.536	4.326.403
180150	AMBATO	180158	MONTALVO	4222	333601	9.923.595	4.326.403
180150	AMBATO	180950	TISALEO	11704	333601	2.991.772	4.326.403
180150	AMBATO	180162	QUISAPINCHA (QUIZAPINCHA)	14031	333601	1.209.317	4.326.403
180150	AMBATO	180151	AMBATILLO	5658	333601	1.242.292	4.326.403
180150	AMBATO	180150	AMBATO	192693	333601	4.684.655	4.326.403
180150	AMBATO	180155	HUACHI GRANDE	11455	333601	1.439.753	4.326.403
180150	AMBATO	180751	BENITEZ (PACHANLICA)	2360	333601	4.975.559	4.326.403
180150	AMBATO	180951	QUINCHICOTO	1411	333601	2.921.289	4.326.403
180150	AMBATO	180165	SANTA ROSA	22668	333601	3.707.983	4.326.403
180150	AMBATO	180152	ATAHUALPA (CHISALATA)	11074	333601	9.512.729	4.326.403
180150	AMBATO	180163	SAN BARTOLOME DE PINLLOG	9815	333601	121.039	4.326.403
210150	NUEVA LOJA	210150	NUEVA LOJA	64041	67098	3.789.613	6.315.534
210150	NUEVA LOJA	210152	DURENO	3057	67098	2.525.921	6.315.534
220150	PUERTO FRANCISCO DE ORELLANA	220150	PUERTO FRANCISCO DE ORELLANA	49558	49558	1.460.697	1.460.697
230150	SANTO DOMINGO DE LOS COLORADOS	230150	SANTO DOMINGO DE LOS COLORADOS	334740	334740	1088.75	1088.75
240250	LA LIBERTAD	240150	SANTA ELENA	59125	228006	5.373.146	6.237.903
240250	LA LIBERTAD	240352	JOSE LUIS TAMAYO	24864	228006	3.395.671	6.237.903
240250	LA LIBERTAD	240350	SALINAS	39205	228006	2.736.405	6.237.903
240250	LA LIBERTAD	240250	LA LIBERTAD	104812	228006	2.515.493	6.237.903

Appendix 8. Robustness checks

Commuting patterns

Table A8.1 shows the results of applying the algorithm between urban cores using the SHLC 2014. Urban cores connected in commuting terms are exactly those that were relatively close in travel time terms. Therefore, it gives validation to our proposed based on proximity. A minimum threshold of at least 10% of commuting flow (the same as the preferred threshold for the Colombia case reported by Duranton, 2016) gives the closest approximation to our approach using travel time.

Table A8.1: Sensitivity test of urban cores based on rescaled commuting patterns from SHLC

		Initial	Results / FUAs *(% min. commuting flow)*			
	Size	Cores	8%	10%	15%	20%
500 inh./km2	***25,000***	***34***	30	31	32	32
	50,000	21	20	20	20	20
	100,000	16	16	16	16	16
1000 inh./km2	25,000	29	26	27	28	28
	50,000	20	19	19	19	19
	100,000	16	16	16	16	16
1500 inh./km2	25,000	33	27	28	29	29
	50,000	21	19	19	19	19
	100,000	16	16	16	16	16

Figure A8.1: Functional Urban Areas based on commuting patterns derived from the SHLC (A) 10% threshold of commuting (B) 15% threshold of commuting

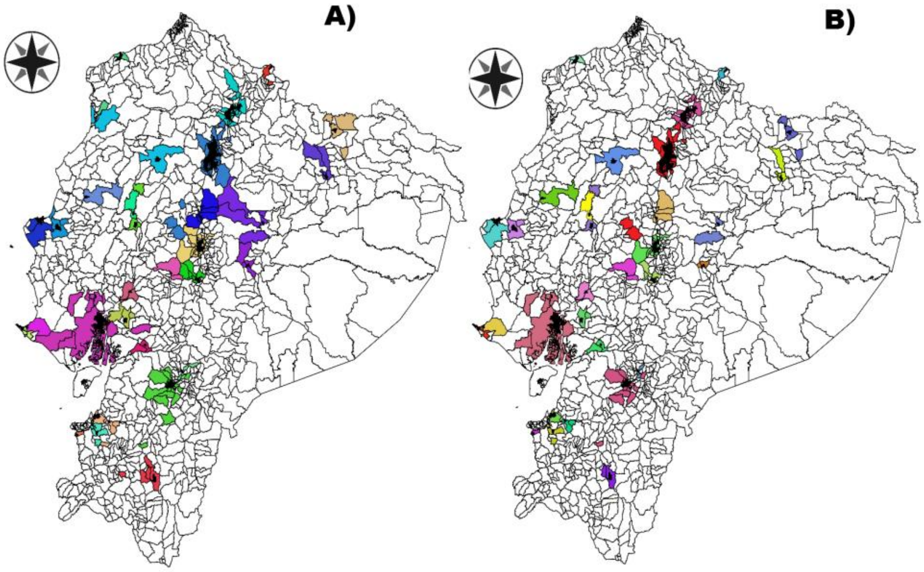


Figure A8.1 plots the FUAs with hinterlands computed using thresholds of commuting flow at 10% and 15%. In this case the hinterlands were very sensitive to the minimum threshold applied, what can be expected given the poor quality of the commuting data. Similar results of hinterlands, are obtained when we use a minimum threshold at 15% and 30 min of travel time using private car.

Gravitational approach

We use the gravity approach under the idea of extending the commuting flow to the whole population matrix of pairs of origin and destination. Using the SHLC 2014, we forecast the total expected number of commuting flows with respect to the total population in each area. In order to do that, we rescaled commuting flows resulting from the survey, multiplying the share of commuters by population size. We use a gravitational exponential decay function devoted to inter-urban mobility; where our dependent variable is the total rescaled commuting flow between origin and destination. This specification is preferred because it has a faster decay function with respect to distance, similar to commuting patterns. An alternative specification can be used to forecast migration patterns. The masses in origin and destination are total economically active population (pea) or whole population (pop). Distance is measured as straight geographical distance in meters (dist)[28]. The specification is the following:

$$Flow_{orig,dest} = f(Mass_{orig}; Mass_{dest}; Distance_{orig,dest}) \qquad \text{(A8.1)}$$

Flow is the rescaled commuting data from the survey. Mass represent the masses of origin and destination, D is the distance. We estimate a linear regression using a zero inflated negative binomial (ZINB) model as OLS overestimates commuters because we have a large amount of zeros in the matrix (Westerlund & Wilhelmsson, 2011). In the final estimation we include polynomial extension of origin and destination masses (see results at table A8.2). The flow of commuters was obtained from the ratio between the commuters from origin *i* to destination j, divided by population of origin i, $\sum F_{ij}/POP_i$.

Table A8.3 introduces the results of sensitivity test of urban cores. These results are similar to those presented using our travel time proposal and also close to the flows using rescaled commuting resulting from SHLC. Differences arise at lower thresholds, as the gravitational computed flows cannot connect very close urban cores, as other approaches do. Figure A8.2 displays the results considering hinterlands based thresholds 10% and 15% from commuting flows derived from the gravitational model. Again, hinterlands were very sensitive to those minimum thresholds.

[28] We preferred using travel time distance because parishes were too large compared with urban settlements, and consequently Google maps or Open Street Maps were reporting incorrect estimates in too many occasions.

Table A8.3. Gravity regression. Zero inflated binomial model estimation.

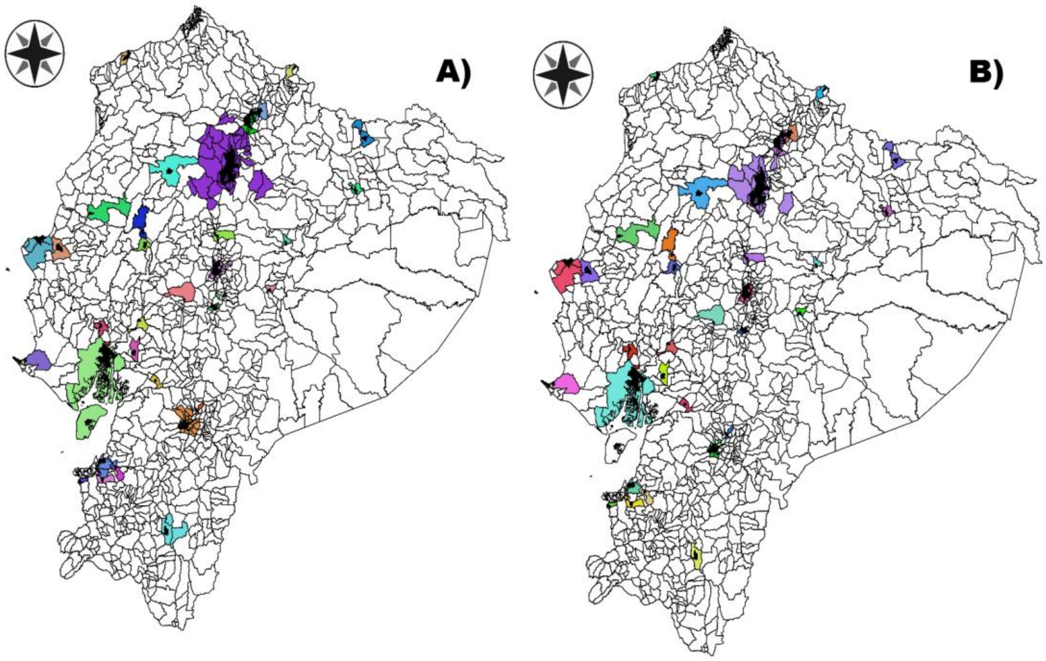
Variable	(1)	(2)	(3)
Count	Basic	squared population	cubic population
lpop_o	.4126328***	1.0552242***	1.0539545***
lpop_d	.26828608***	0.09047807	-1.5988409***
distance	-4.211e-06***	-.00001009***	-.00001607***
lpop2_o		-.0298708***	-.02942369***
lpop2_d		.00840842**	.16950979***
dist2		1.712e-11***	5.810e-11***
lpop3_d			-.00496326***
dist3			-6.471e-17***
Constant	-.53870537***	-2.7892201***	3.0278859
Inflate			
lpop_o	-.58417243***	.50902062***	.51579967***
lpop_d	-.81539292***	.17687747	6.313216***
distance	.00002385***	.00004249***	.00007095***
lpop2_o		-.05486598***	-.05582299***
lpop2_d		-.04945204***	-.64388137***
dist2		-6.179e-11***	-2.464e-10***
lpop3_d			.01862565***
dist3			2.746e-16***
Constant	15.248468***	4.2874004***	-16.970588***
lnalpha	-.42555607***	-.50448303***	-.52735042***
Statistics			
N	558,902	558,902	558,902
Lok Lik.	-31246.868	-30396.782	-30049.737
AIC	62511.736	60819.563	60129.474
BIC	62612.84	60965.602	60297.979

Note: Asteriscs account for significance * p<.05; **p<.01; *** p<.001

Table A8.2. Sensitivity test of urban cores based on gravitational approach

		Initial	Results / FUAs *(% min. commuting flow)*			
	Size	urban cores	5%	8%	10%	15%
500 inh./km2	*25,000*	*34*	33	33	33	34
	50,000	21	21	21	21	21
	100,000	16	16	16	16	16
1000 inh./km2	25,000	29	29	29	29	29
	50,000	20	20	20	20	20
	100,000	16	16	16	16	16
1500 inh./km2	25,000	33	33	33	33	33
	50,000	21	21	21	21	21
	100,000	16	16	16	16	16

Figure A8.2 Functional Urban Areas based on commuting patterns derived of the gravitational model (A) 10% threshold for commuting (B) 15% threshold for commuting



Radiation model

The radiation model for commuting is expressed in equation (A8.2).

$$F_{ij} = F_i * \frac{Pop_i * Pop_j}{(Pop_i + w_{i,j})\,(Pop_i + Pop_j + w_{i,j})} \qquad \text{(A8.2)}$$

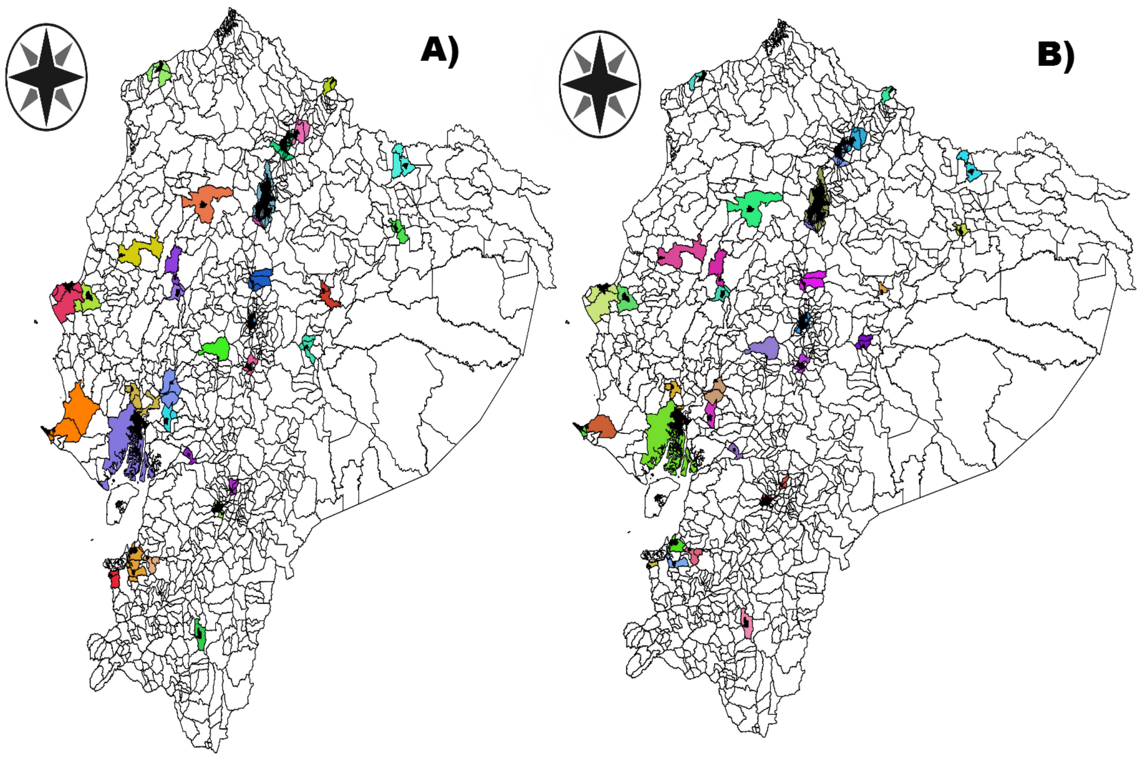
Where F_{ij} is the forecasted commuters from origin i to destination j; F_i is the total outflow of commuters from origin i; Pop_i and Pop_j are the total population in origin i and j destination respectively; and $w_{i,j}$ represents the population contained in a radius given by the distance between origin i and destination j, excluding both the population contained in origin i and destination j. One advantage of this approach is that is parameter free. We use the information at the National Census of Ecuador 2010; this census has a specific question that allows accounting for the proportion of workers commuting out of the parish.. Next, we programmed an algorithm in Stata to build the matrix W_{ij}.

We use the forecasted commuters as the source flow for OECD's algorithms. Table A8.4 reports the results and a sensitivity analysis for different thresholds. These outputs are pretty close to the ones derived from the travel time procedure, again at the 10% threshold of commuting. Figure A8.3 displays the FUAs including the hinterlands computed using 10% and 15% of commuting flows derived from radiation model. As before, the hinterland is the most sensitive part of the analysis.

Table A8.4. Sensitivity test of urban cores based on radiation model

		Initial	Results/FUAs (% min. commuting flow)			
	Size	urban cores	5%	8%	10%	15%
500 inh./km²	*25,000*	*34*	29	31	32	34
	50,000	21	20	21	21	21
	100,000	16	15	16	16	16
1000 inh./km²	25,000	29	24	26	27	29
	50,000	20	19	20	20	20
	100,000	16	15	16	16	16
1500 inh./km²	25,000	33	27	31	32	33
	50,000	21	20	21	21	21
	100,000	16	15	16	16	16

Figure A8.3: Functional Urban Areas based on commuting patterns derived of the radiation model (A) 10% threshold for commuting (B) 15% threshold for commuting



Internal migration

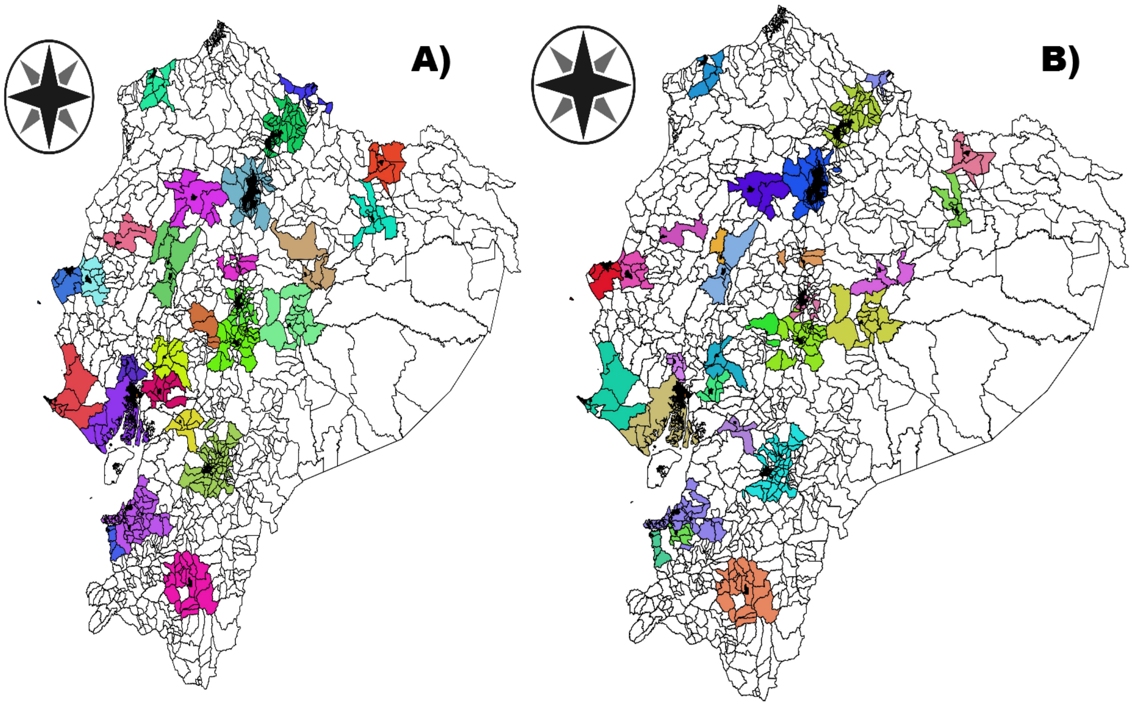
In this case we use internal migration patterns, gathered from the national census of population 2010 of Ecuador. There is information of internal migration between the years 2005 and 2010. The actual matrix is 1,149 parishes by 1,211parishes, as there were several changes in the boundaries of some parishes. We have identified large migration flows between the largest urban poles of the country. Consequently, we have opted for imposing a geographical distance restriction. This allows generating a correct identification of flows that can enter in the algorithm. We opt to use a hierarchical pattern and keep away those urban cores that are relatively far from each other. The restriction of distance was 34,765 meters, which according with Google maps is the distance by car with a half hour of travel time.

Table A8.5 shows the results of the algorithm for different thresholds. The algorithm was successful at connecting cities using a minimum threshold of internal migration, although the patterns are different to the results obtained from travel time and derived commuting flows. In this case, the closest approximation is obtained when using a threshold set at 15% of internal migration. As before, high minimum thresholds make the results more stable. Even if this is a good approach, the results seem very sensible and they were not very similar to commuting patterns. We also present in Figure A8.4, the hinterlands of each FUA at least 15% and at least 20% of internal migration. The results are relatively similar. However, the hinterlands are also too sensitive as the others approaches introduced previously. In this case, our best approximation of the hinterland was using the minimum threshold of at least 20% of internal migration.

Table A8.5. Sensitivity test of urban cores based on internal migration

		Initial	Results / FUAs *(% min. commuting flow)*			
	Size	urban cores	10%	15%	20%	25%
500 inh./km^2	25,000	34	27	29	*33*	33
	50,000	21	20	21	21	21
	100,000	16	15	16	16	16
1000 inh./km^2	25,000	29	26	27	29	29
	50,000	20	19	20	20	20
	100,000	16	15	16	16	16
1500 inh./km^2	25,000	33	27	29	32	32
	50,000	21	21	21	21	21
	100,000	16	15	16	16	16

Figure A8.4. Functional Urban Areas based on migration patterns (A) 10% threshold for migration (B) 15% threshold for migration



IREA

Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics

WEBSITE: www.ub-irea.com • **CONTACT:** irea@ub.edu

AQR

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona