New challenges in the definition of functional geographies

Mike Coombes (CURDS, Newcastle University),

Lucas Martínez-Bernabeu &

José Manuel Casado-Díaz (IEI, Alicante University)

mike.coombes@ncl.ac.uk

The definition of Functional Geographies (FGs)

FGs long-established in many countries as official geographies (sets of areas), that a National Statistical Institute (NSI) uses for data on labour markets

FG boundary definitions:

Eurostat & OECD c.2020 research studies identified one standard method (based on the UK's method to define "TTWAs") for all to use in future

***continuing need* is small zone data to capture locally-specific linkage patterns**

***pre2020*: Census journey-to-work (J2W) data the basis of most definitions**

! *Covid 19* radically affected J2W behaviour during the 2020/2021 Censuses

***so* data from then unsuitable for FG definitions for mid-2020s and beyond**

New challenge for FG definitions is finding data on 'new normal' labour markets

Leading candidate source is data from mobile phone use

Dataset* here is from Spain 2021 but are the <u>issues generic & persistent?</u>

Key process applied to each anonymised mobile phone's call records is the <u>anchor point</u> algorithm which uses call time/day to identify the most likely…

home location

and workplace location …and thus inherently…

the estimated 'J2W' flow for that mobile

[Annex A specifies the algorithm used for the data here]

zones = polygons round masts frovide the geolocations in the dataset

! zones will have to be grouped to approximate official data areas

Limitations of this (and any such?) dataset include:

! data sensitivity to algorithm

! 6% of mobiles had no identifiable 'workplace'

? mostly working people with irregular work times/locations?

**** The raw data measures MOBILITY in general, it's tested here as proxy J2W***

Overview comparison of Census & mobile phone data

Potential issue	Census J2W data	Mobile phone data
mid-2020s	**unavailable in most countries**	available
labour supply measure	does not include all jobs and thus not all commuting trips	
no. of days per week	unknown so estimating the number of commuting trips not possible	
dependence	no problem because the data is in-house for NSIs	*potential problems in negotiations to get appropriate data at low cost, and uncheckable data quality*
grossing-up	NSIs have established methods to deal with low non-response	*potential problem of correcting for limited coverage of provider(s), and any geographical bias in market share*
selective under-coverage	*some countries have declining Census response rates in certain sub-groups*	probable that not using a mobile phone is now rare
locally distinct linkages	very high coverage should allow local patterns to be identifiable	
granularity	very high level of spatial precision of data output should be possible	high spatial precision of data output *(reduced by 'best-fit' to NSI's units and any problem of 'flicker')*
home location	*missing question of where commute started creates some 'implausible' flows*	*minor issue of anchoring algorithm assumptions reverses commuting flows of night workers*
work location	unproblematic	*intrazonal flows not identified in large antenna zones*
different workplace types	*not available in all countries*	*anchoring algorithm assumptions in effect prevent their identification*
only working people included	yes	**major issue of non-workers false commuting trips, with spatial bias to this over-estimation in the data**

normal text = no substantial problem of similarity and/or validity of the 2 types of dataset

plain italic = *minor issue for that dataset, minimal impact on similarity of datasets*

bold = **major issue for that dataset, impacting on the similarity of datasets**

Key issue with the data from mobile phones

Mobile datasets are anonymized: unknowable which mobiles owned by workers

J2W flows from anchoring algorithm include mobility of people not in work

regular movements by people not in work probably shorter distance

SO their inclusion in 'J2W' data is not random but introduces bias

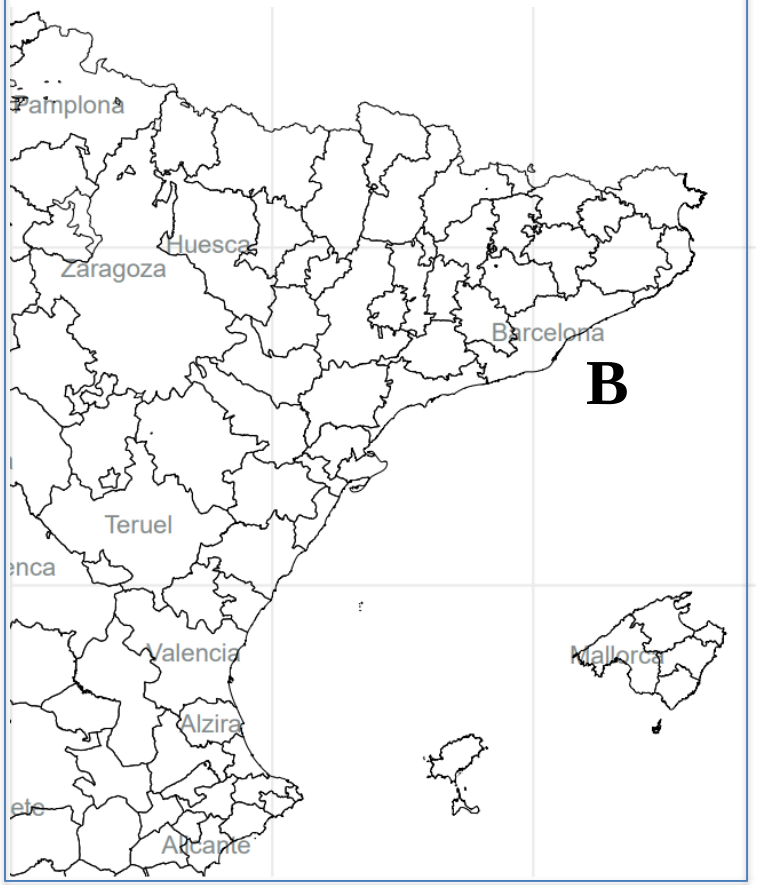
FG definitions need accurate values of local flows, including intrazonal flows

Innovative method to adjust Spanish dataset to mitigate this key problem:

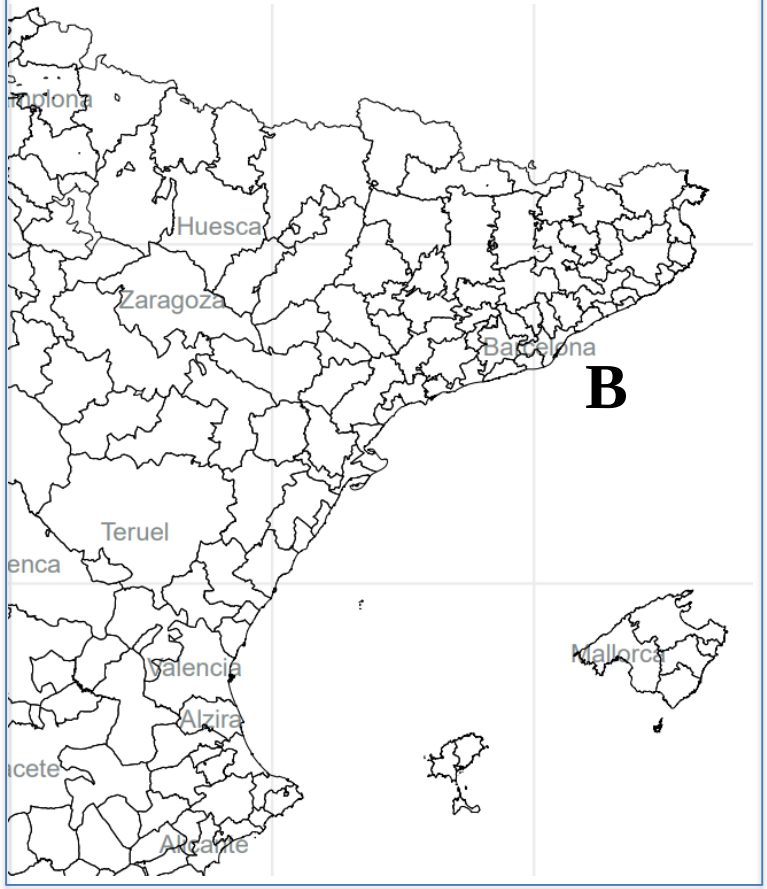
- **people not in work mostly move locally SO we assume local = intrazonal**
  - **access data on numbers of people (not) working in each zone**
    - **reduce the anchoring algorithm's count of intrazonal flows by the number of people in home location (zoneO)**
- **This creates a transformed J2W dataset possibly suited to defining FGs after additional zone grouping to create NSI-like zones for the analysis**

Spanish FG definitions based on…

2011 Census



2021 transformed MPF



B: Barcelona **example of highly fragmented metropolitan areas**

less urban areas show much more stability (eg. Mallorca)

FG ‘test case’ of mobile phone data: result?

Note that the FGs based on transformed mobile phone data here are from ongoing work SO are they sufficiently plausible to suggest that mobile phone data can be transformed to be taken as proxy J2W flows?

- **!answer may depend on how far the increased fragmentation of FGs vis-à-vis 2011 is due to the shift to home-working in the pandemic**

  (nb. there is evidence the greatest shift to home-working was in the most urban areas, which is where these FGs are most fragmented)

Full evaluation awaits the availability of J2W data from a Census (or some equivalent survey) for the same small zones covering the same period as robust analyzable mobile phone data

Acknowledgements

Research on this dataset undertaken in IEI Alicante University as part of 2 research projects:

PID2020-114896RB-I00 (funded by Agencia Estatal de Investigación, AEI / 10.13039/501100011033; Spanish State Plan for R&D&I 2017-2020)

AICO/2021/062 (funded by Generalitat Valenciana, Dept. of Innovation, Universities, Science and Digital Society, R&D&I Programme of the Comunidad Valenciana).

Annex A
mobile data & anchor point algorithm used here

***Background** 2021 Spanish Census did not collect commuting data, unlike in 1991-2011*

(some workplace data estimated using register data, but provincial capitals often used as zoneD due to high % of missing data, so data yields implausible 'TTWAs')

Spanish equivalent of ONS (INE) published a full population mobility dataset providing daily snapshots of mobile phone flows for all 2021 Sundays and Wednesdays

> dataset covers over 80% of mobile phones, and is anonymized

> each signal is timed and geo-located via the receiving mast

Anchor point algorithm directly applied by the network provider to identify…

home location (zoneO): where the mobile spent more time in the previous night

destination location (zoneD): where the mobile spent at least 4 hours between 10am and 4pm (if no such place, destination is unknown)

This algorithm finds for each mobile a main daytime mobility destination: it is probably a workplace of people in work, and so that mobile's potential 'J2W' (zoneO, zoneD)

!mobile phones can 'flicker' between masts, but this is only a low % and with only a localised impact (most of which is probably within cities or metropolitan areas)