Implemented as part of the project „Improvement in Conditions for Decentralization and Accessibility of State Administration in the Territory“ financed from EEA and Norway Funds 2014 – 2021.

ISSN: 2510-2591 DOI: 10.48340/ecscw2023_p02

Václav Jaroš, Jakub Jaňura and Petr Svoboda (2023): Geolocation data as a research tool for the organization of the settlement system Case study of the spatial mobility model in Czechia. In: Proceedings of the 21st European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies - Posters, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2592), DOI: 10.48340/ecscw2023_p02

Geolocation Data as a Research Tool for the Organization of the Settlement System

Case Study of the Spatial Mobility Model in Czechia

Václav Jaroš, Jakub Jaňura, Petr Svoboda
(Ministry of the Interior, Czech Republic)

Implemented as part of the project „Improvement in Conditions for Decentralization and Accessibility of State Administration in the Territory“ financed from EEA and Norway Funds 2014 – 2021.

Abstract

Geolocation data is a widely used source of the spatial information. Their great potential might be also used for population mobility research to identify spatial interactions forming the hierarchical structure of the settlement system. For this purpose, a model of data acquisition and their preliminary analysis was developed. This model represents an effective tool for mapping the mobility behaviour of the population. Using the example of Czechia, significant commuting links are identified, which are subsequently analysed in detail using GIS tools. Therefore, important commuting centres of different hierarchical levels are defined by the volume and nature of spatial interactions. This approach is used as a source of important expertise for the proposals on subsequent Czech public administration reform. Nevertheless, the entire model is generally transferable, and the entire method of using the geolocation data for mapping the hierarchy within the settlement system can be replicated in other countries as well.

Research explanation

The research is based on assumption that the settlement system represents a large set of complex processes between particular components of the society and the landscape variable in time and space. This process results in socio-spatial differentiation, which manifests itself the most as a spatial concentration of activities within society. The spatial concentration of activities is a natural process of development of social systems. A certain form of concentration is necessary, as it is not possible to ensure the availability of all activities, which have different degrees of rarity, in all locations equally. This is the very essence of the formation of settlement systems, which the concentration of activities allows to arise.

The interconnectedness of individual processes within the settlement system is so complex that it cannot be easily identified or even measured in any way. However, the external manifestations of these processes are measurable. These take the form of spatial interactions and manifest themselves as commuting relationships different at various hierarchical levels[1]. They are thus realized through transport links which have been measured for long period by transport geographers[2]. In general, the described spatial interactions can be called population mobility. It contains not only the actual journeys but also a reflection of the overall spatial pattern of each individual's behaviour. Hence, the mobility/commuting behaviour of an individual takes into account the repeatability of certain elements of spatial behaviour, which also determine the hierarchical position of the commuting destination and its relationship to the place of origin. In conclusion, the spatial behaviour of the population completely reflects the relationships and processes within the settlement system and therefore, it is a suitable object of measurement for their explication.

A wide range of tools can be used in both local and large-scale statistical surveys focused on the traffic behaviour/mobility of residents, such as questionnaire surveys, traffic diaries, GPS loggers, measuring passengers transported by individual modes of transport, or measuring traffic intensity. In the Czech Republic, queries about commuting to work and schools are even part of the census, however, these available statistics have a low return in recent censuses, and it is assumed that up to 40 % of commuting flows are missing from the census statistics. With this in mind, a significant potential for mobility measurement can be seen in the use of the **geolocation data of mobile operators**. Due to the high penetration of the population by mobile devices, and the possibility of tracking movement in unlimited random periods, this approach combines both the advantages of population-wide data collection and detailed (movement tracking) studies as well.

[1] See e.g., Hampl 2007; Hampl, Marada 2015

[2] See e.g., Marada 2010; Marada et al. 2016, Jaroš 2017, El-Geneidy, Levinson 2007.

The essence of the method are the records in the geolocation network, which are created every few minutes by every device joined to the GSM network via SIM cards. Determining the location is approximate by this technique, as only the transmitter that registered the recording is precisely located. From the signal coverage map of individual transmitters, the approximate location of the SIM card can be deduced with an accuracy of hundreds of meters in urbanized areas and up to a few kilometres in rural areas.

In order to obtain this type of data, it is necessary to set up a complex mechanism of tools analysing more than 10 million SIM cards (the case of the Czech Republic), each of which produces thousands of records within the measured periods. In addition to the technical solution and considerable computing capacity required for Big Data processing itself. Besides, it is also essential to consistently establish methodological procedures for the preliminary processing of primary records for the creation of databases of citizens' mobility/travel behaviour.

In the past, it was the method of data claiming that was the main obstacle in the use of geolocation data concerning their low validity[3]. Research carried out in the past in the field of data analysis of mobile operators had to solve problems of representativeness of data and their evidential value when generalizing to the population[4]. Although this shortcoming is not an obstacle for use in research from the technical fields aimed at measuring the volume of journeys made or data transmitted, in the field of social geography the question of the generalizability of data to the population and the projection of spatial patterns of behaviours onto entire society in space and time is absolutely essential. For this purpose, a unique model was created, including a complete range of interconnected processes, which captures the mobility of the population and projects it on the social and settlement networks.

The whole model is based on the presumption that mobile phones move together with their users for most of the day[5]. Based on this assumption, the model eliminates records created by other devices than mobile phones, thereby largely eliminating the problem of duplicate records of a single user of multiple devices. Similarly, rarely used SIM cards that do not make enough records in the network are neglected. Furthermore, the assumption of high penetration of the population by mobile phones is also crucial. In general, it can be concluded that in contemporary societies of developed countries, both assumptions are fulfilled.

When detecting the movement or stay of the SIM card, the proposed model must solve the problem of the inconsistency of the administrative boundaries of the municipalities with the boundaries of the signal

[3] Mentioned in more detail e.g., by Mazouch et al. 2017; Novák 2010; MV 2020.
[4] See e.g., Šveda, Barlík 2018; Halás et. al. 2021. [5]
Mentioned in e.g., Novak 2010, MV 2020.

transmitter´s service area (cells). In reality, it is common for one transmitter to serve several municipalities or their parts at the same time. In addition, overlapping of service areas (cells) of different transmitters is common as well. The detection of stay/movement itself and its assignment to particular territorial units (municipalities) occurs via the cell-mapping process. This tool distributes the measured records between specific settlements (municipalities) according to the amount of intravillan (build-up area) of each settlement extending into a specific cell (service area of the signal transmitter) and also reflecting the population density of each settlement. In conjunction with the clustering algorithm, cell mapping can eliminate the unwanted effects of so-called cell jitter ("random" switching between neighbouring transmitters)[5]. Databases obtained through this model also removes other undesirable elements that worsen the evidential value of the data, such as the share of virtual operators using the network, ownership of multiple SIM cards by one user, or on the contrary, not owning a mobile phone and thus no SIM cards. All these aspects are taken into account by the model. In addition, the model also contains relocation mechanisms that are capable to correct retroactively any model errors in assigning the records to individual territorial units. At the same time, sufficient anonymization of the final data is ensured.

This model is flexible in terms of the output databases produced. According to the primary setting, it produces a total of 15 attributes on the territorial detail of individual municipalities, structured into 3 basic interconnected datasets: a) statistical data for individual municipalities and characteristics of their residents, b) OD matrix showing commuting directions taking into account a total of 6 types of commuting intensity, and c) the average number of currently present population in every hour of the week (24/7) in each municipality with a breakdown by particular attributes.

The method of assigning attributes to individual users in the network is also unique. Basically, the method does not monitor the actual volumes of the trips made but analyses the commuting rhythms and the overall spatial commuting behaviour of each SIM card user. During the monitored period (28 days), "labels" of attributes are assigned to each SIM card according to its unique pattern of spatial behaviour. Each individual (SIM card user) can only have one label for each municipality, but he can have several labels for several municipalities - can be a resident in one municipality, commute to another for work or school, commute for services, or be an occasional visitor to another etc. The output databases themselves do not indicate specific measured values for a certain day or period, but each attribute represents basically the number of people who reports a given type of behaviour. This is no longer the geolocation data itself, but a summary of time-spatially aggregated statistics about geolocation data.

[5] For more details see MV 2020; Mazouch 2017.

The output databases enable subsequent applications of geographic analyses identifying functionally integrated regions and their central areas at different hierarchical levels. Based on the principle of commuting to certain centres, the intensity and volume of these interactions, relatively closed (in terms of functional closeness of the interactions) and internally integrated regions are formed.

Primarily, the method is set to identify functional micro-regional commuting links. Microregions are territories in which a resident should be able to secure all his daily activities necessary and important for his everyday life. Their centres are primary commuting destinations for their surroundings and provide a sufficient range of job opportunities, primary and secondary education, health services, shops, etc. Visiting centres of a higher hierarchical levels providing services of a higher grade, however, is not needed daily. Nevertheless, thanks to the robustness of the analysed data, it is also possible to define centres of higher (mezoregional) or lower (submicroregional) levels. Therefore, this method enables the implementation of a complete socio-economic regionalization of the state at individual hierarchical size-levels, including their hierarchical relationships.

This approach was used in the Czech Republic for a comprehensive revision of the spatial units of the public administration structure. The purpose of this activity was to harmonize the administrative units with natural commuting regions. Particularly, the aim was to ensure that public administration offices were located where people naturally concentrated. This leads to streamlining and deconcentration of the public administration and its adaptation to the needs of citizens. Based on this application example, it is also possible to conclude about the transferability of this approach and its applicability both in other territories (states) or in other scientific fields.

References:

EL-GENEIDY A., LEVINSON M., (2006). Access to Destinations: Development of ccessibility Measures. Access to destinations study series, Report No. 1. Minnesota Department of Transportation.

HALÁS, M., BLAŽEK, V., KLAPKA, P., KRAFT, S. (2021). Population movements based on mobile phone location data: the Czech Republic. Journal of Maps 17 (1), 116–122. ISSN 1744-5647, DOI: 10.1080/17445647.2021.1937730.

HAMPL, M. (2005): Geografická organizace společnosti v České republice: transformační procesy a jejich obecný kontext. Univerzita Karlova, Praha, 147 s.

HAMPL M., MARADA M. (2015): Sociogeografická regionalizace Česka. Geografie, vol. 120, no. 3, pp. 397–421. DOI:10.37040/geografie2015120030397

JAROŠ V. (2017). Social and transport exclusion. Geographia Polonica, vol. 90, no. 3 pp 247-263. DOI: 10.7163/GPol.0099.

MARADA M., FRÁNĚ L., JANOŠ V., JAROŠ V., KRAFT S., KŘÍŽ M., KOWALSKI M. (2016). Rychlá spojení metropolitních oblastí: Dopady (nové) dostupnosti na pracovní trh. Project TAČR TB0500MD005: Fast connections among metropolitan areas: impact of (new) accessibility on labour market.

MAZOUCH, P. a kol. (2017): Limity využití mobilních sítí ve statistických šetřeních ČSÚ. Project TAČR TD03000452. Limits of data from mobile sites in statistical surveys of Czech statistical office.

MV ČR (2020): Využití geolokačních dat mobilních operátorů pro potřeby veřejné správy. EAA grants, GG-PDP1-001, Improvement of preconditions for decentralisation and availability of public administration in the territory.

NOVÁK, J. (2010): Lokalizační data mobilních telefonů: Možnosti využití v geografickém výzkumu. Disertační práce. Katedra sociální geografie a regionálního rozvoje PřF UK, Praha, 175 s.

ŠVEDA, M. - BARLÍK, P. (2018). Daily commuting in the Bratislava metropolitan area: case study with mobile positioning data. In Papers in Applied Geography, 4(4): 409-423.