International Cartographic Association
Association Cartographique Internationale

Leveraging spatial big data technology for country-wide analysis of mobility using Floating Car Data

Kamil Choromański [a,*]

[a] *Warsaw University of Technology, Faculty of Geodesy and Cartography, kamil.choromanski@pw.edu.pl*

* Corresponding author

Keywords: Spatial Big Data, Floating Car Data, Mobility Data, Big Data Cartography

Abstract:

Conventional spatial data processing tools such as GIS software (e.g. QGIS) or programming libraries (e.g. geopandas) running on a single computer are very often not sufficient to extract information efficiently from currently available datasets. In order to solve this problem, big data technology (represented, for example, by the Apache Spark framework) is used. Over time, libraries for analyzing big data with a spatial component have also emerged, such as GeoMesa or RasterFrames.

However, using these tools is more complex than GIS software. Therefore, before performing specific analyzes, it is necessary to analyze which technological path will be optimal in a given case. Not every big dataset requires the tools outlined above. Numerous sources (e.g. Schroeck et al. (2012), Khan et al. (2018)) characterize data as big data based on meeting requirements such as Volume, Velocity, Variety, Veracity, and Value. Examples of spatial data that have these characteristics are Floating Car Data (FCD). These are spatial data extracted from GNSS receivers of vehicles (or devices inside them). Trajectory points of the individual vehicles are then assigned to the corresponding section of the road network. In this way, it is possible to calculate data on the number of vehicles and their speed on road sections with a very high spatial and temporal resolution. The results of FCD data analysis have many applications, such as traffic monitoring, dynamic routing, or detecting traffic anomalies (Ajmar et al. (2019)).

This paper presents the results of the analysis of anonymized FCD data acquired using the Yanosik mobile app. The Yanosik application is a very popular vehicle navigation and monitoring application in Poland that includes an effective tool for sharing information between drivers. In record periods, it is used by nearly 300,000 drivers from whom location and other type of data is collected at once. Such a large number of active users can generate extensive data sets. The dataset analyzed in the research covers the entire area of Poland from a two-year period. Each section of the OpenStreetMaps road network was assigned information on the average speed of vehicles traveling along it at an interval of 5 minutes. The raw data in the form of CSV files takes up 14.8TB of disk space. In terms of size, temporal and spatial resolution, as well as high information value, this dataset exhibits the characteristics of Spatial Big Data. In order to process it efficiently, the CENAGIS computing infrastructure (Gotlib et al. (2022)) was used.

In the first step, the data was converted from CSV files to a parquet file in GeoMesa-FS format. The columnar structure of the parquet files allowed optimization of disk space and acceleration of the Apache Spark distributed environment. A traffic anomaly analysis based on FCD data was performed on this set, based on the work of Altintasi et al. (2017).

The detection of anomalies such as traffic jams is based on the analysis of adjacent segments of the road network. For this purpose, it was necessary to design a distributed spatial join method that allows efficient assignment of neighboring segments in the whole set. Then, depending on the traffic level of the adjacent segments, a class representing the traffic state (e.g. Free Flow, Transition Flow, or Congested Flow) was assigned to each segment. The dataset prepared in this way allowed convenient data aggregation for visualizations. For the purposes of this study, a set was prepared presenting data only from Fridays, from 4 p.m. to 6 p.m. In this scope, the number of occurrences of the Congested Flow class was counted. Results are presented in Figure 1 below.

This study demonstrates how spatial big data technology can extract valuable insights from large spatial datasets.. While these tools can be challenging to implement, they are often essential for analyzing datasets that exceed the capabilities of conventional methods. Without the ability to process such extensive data, creating various types of maps can become highly difficult or even unfeasible.

Abstracts of the International Cartographic Association, 10, 40, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-40-2025

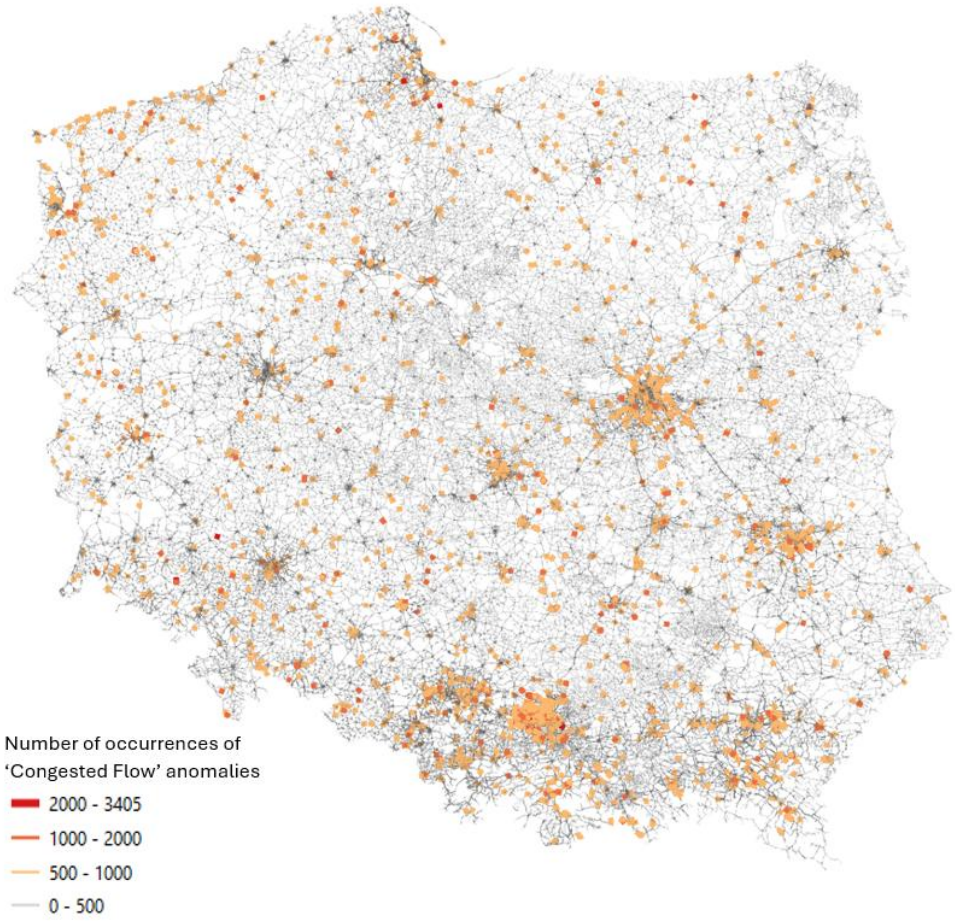

Figure 1. Map visualization of FCD data anomalies detection results for class "Congested Flow" occurring between 4 p.m. and 6 p.m. on Fridays in years 2019-2020.

Acknowledgements

The author is thankful for the opportunity to use the CENAGIS infrastructure, developed under the project entitled "CENAGIS - The Centre for Scientific Geospatial Analyses and Satellite Computations including Laboratories for testing and authorising geomatic products," which is co-financed from the European Fund for Regional Development within Priority Axis I "Utilisation of research and development activities in economy" (RPMA.01.01.00-14- 9871/17 from 11.12.2018).

References

Schroeck M, Shockley R, Smart J, Romero Morales D and Tufano P Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report IBM Inst. Bus. Value Said Bus. Sch. Univ. Oxford

Khan N, Alsaqer M, Shah H, Badsha G, Abbasi A and Salehian S 2018 The 10 Vs, issues and challenges of big data Proceedings of the 2018 International Conference on Big Data and Education (New York, NY, USA: Association for Computing Machinery) pp 52–6

Ajmar, A., Arco, E., Boccardo, P., & Perez, F. (2019). Floating car data (fcd) for mobility applications. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 1517-1523.

Gotlib, D., Choromański, K., & Kaczałek, B. (2022, September). CENAGIS Geo-Cyberinfrastructure as a R&D Environment for a Surveyor 4.0. In Volunteering for the future-Geospatial excellence for a better living, FIG Congress, Warsaw, Poland (pp. 11-15).

Altintasi, O., Tuydes-Yaman, H., & Tuncay, K. (2017). Detection of urban traffic patterns from Floating Car Data (FCD). Transportation research procedia, 22, 382-391.