**PROJECT DESCRIPTION FOR THE GENERAL PULIC (by Andrzej Torój):** *Using geolocation data in the construction of interregional input-output tables with spatial econometric methods*

In this project, an interregional input-output matrix for Poland will be estimated for 77 sectors and 72 subregions (NACE-3 units). Currently, the Central Statistical Office is publishing the table for sectors only (i.e. of size 77×77), describing the transactions between individual sectors of the economy. The objective of the project is to construct a table sized $(77 \cdot 72) \times (77 \cdot 72)$, i.e. additionally measure the transactions between enterprises from individual 72 subregions (for each pair of sectors). The interregional table is useful to simulate the impact of an investment project, or the day-to-day activity of an enterprise, on the economy – not only on the national level, but also for individual regions.
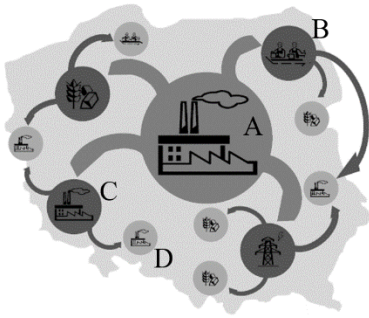
The regional breakdown of this impact is sometimes non-trivial, because some enterprises use specialized production inputs from remote locations and hence create some demand in another region of the country. For example, the interregional table should capture the fact that a power plant (point A in the figure beside) should generate some effects in nearby locations (i.e. engaging technical subcontractors, office cleaning services or water supplies; point B), but also in very distant ones (i.e. demanding coal or financial services; point C). In turn, the coal mine (C) will generate demand (via second-order effects) predominantly in its proximity (i.e. engaging further technical subcontractors; point D), etc.

In my previous research, I proposed to use publicly available data and spatial econometric methods to estimate such tables for Poland (and some other countries in Europe). The dataset consisted mainly of the data on value added (a measure close to GDP, up to intermediate taxes and product subsidies) in individual sectors and regions. However, this narrow dataset that could be considerably be extended to increase the accuracy of the resulting tables. The growing body of emerging big datasets, containing geographic information, can contribute to the previous knowledge about cross-regional transactions in the following ways:

- **Maps**, such as OpenStreetMap (open source) and ArcGIS LivingWorld repositories (commercial source) allow to flexibly construct a measure of supply-side potential (and diagnose supply-side limitations) to be used in the estimation process. For example, the satellite pictures allow to determine the coverage of a given region with forest, water, roads and buildings, traffic intensity and light at nighttime – all useful in determining the supply-side potential in a given subregion and sector.

- **Traffic information from Google Maps.** Using the data on the road distance between the main cities of the subregions, or the driving time, is economically more relevant than the straight-line distance between the generic centroids of the regions. Data from Google Maps can help to detect the pairs of regions that, in spite of geographic remoteness or proximity, are (respectively) well or badly connected and hence the likelihood of intense trade relationship is (respectively) higher or lower than implied by the pure distance visible in the map.

- **Reports prepared by the company Datarino sp. z o.o.** about the relocation of mobile devices in Poland during the day, allowing to capture the patterns of daily commuting. As a result, one can estimate how much demand in suburban areas is generated by people that earn their wages in big cities (that often constitute a separate subregion in Poland).

- Data collected by the Ministry of Finance from the **Single Audit File for Tax** system (pol. *Jednolity Plik Kontrolny*)[1] on the invoices issued in the recent years, including the net transaction values and the address of the contracting parties. Using this data, one could calculate the sum of trade flows between every pair of subregions, based on the invoicing addresses of the enterprises. Using this data table, one could either validate or update the results derived from other sources.

The inclusion of the above-mentioned datasets implies the use of spatial Bayesian econometric methods. They allow i.a. to combine different sources of data into a single statistical model (e.g. explicitly account for the economic intuition that advanced services are sold in spite of a high distance between buyer and seller, while heavy commodities are not).

After the successful completion of this project, to the best of my knowledge, Poland would be one of the first countries in the world, if not the first, where inter-regional tables would be available at NUTS-3 level and validated against an existing big data source on cross-regional trade flows. Future researchers could use the obtained results as a benchmark in future development of their own methods of regional input-output analysis. Also, the regional economic impact assessment toolbox (for the purpose of corporate social responsibility reports or evaluation of foreign investment plans) would be considerably enhanced.

---

[1] At the moment of the grant application, it is not yet clear whether the Ministry of Finance will officially agree to use the database for this purpose. The initial dialogue is promising.