Assessing the performance of urban public transport using population grid data

Hugo Poelman and Linde Ackermans

European Commission, DG Regional and Urban Policy

Introduction

A working paper published in 2020[1] presented several indicators describing the availability, proximity and performance of public transport in European cities. In that paper, performance (at the level of any populated place in an urban centre) is defined as the number of people in the urban centre that can be reached within x minutes by means of public transport combined with walking (= accessibility), divided by the number of people in the urban centre living within a radius of y km around the place of departure (= proximity).

This definition is fully compatible with the EC-OECD framework for transport performance indicators but applying it within urban centres poses a few challenges.

1) The urban centre is considered as a closed system: both proximity and accessibility are constrained by the boundaries of the urban centre. If we want to assess accessibility, proximity and performance for all populated places in an urban centre in the same way, we should also be able to take into account destinations beyond the boundaries of the urban centres.
2) Accessibility (using public transport combined with walking) is measured without any straight-line distance limit (apart from the distance to the outer boundaries of the urban centre). Extending this concept beyond the boundaries of the urban centre leads to an enormous computational burden because of the huge amount of potential destinations to which travel times need to be calculated.

To answer this twofold challenge we tested an adapted version of the performance indicator framework and applied it to a wide selection of urban centres using grid-based population data.

Scope of the analysis

The aim is to assess the performance of (urban) public transport by comparing accessibility and proximity of residential population. To ensure comparability of the results among cities the analysis focuses on urban centres (high-density clusters)[2]. Accessibility and proximity are measured without constraining them by the boundaries of the urban centre. Hence, the close surroundings of the urban centre are also taken into account as possible destinations.

For every populated place inside an urban centre, we want to answer the question to what extent people that are living within a certain radius can be reached within a certain travel time, by means of public transport and/or by means of a short walk. To grasp the variability of accessibility and proximity within the urban centre a high-resolution representation of the spatial distribution of residential population is required. As the urban centres are defined as clusters of 1-km² population grid cells, it seems a logical

[1] Poelman, H., Dijkstra, L. and Ackermans, L., 2020.

[2] For a detailed description of the definition of grid-based urban centres, see: Eurostat, 2019.

choice to use (smaller) grid cells to measure the spatial distribution of population in the urban centre and its surroundings.

This leads to the use of the following concepts and definitions:

Urban centres: clusters of high-density grid cells defined by means of the GEOSTAT 2011 1-km² grid[3].

Places of departure (origin): all centroids of populated grid cells of 200 x 200 m that are located within the urban centre[4].

Potential places of destination: centroids of populated grid cells of 200 x 200 m that are within a 12-km (or 8-km) radius around the place of departure. Consequently, places of destination can also fall outside the urban centre boundaries.

Proximity: population within a 12-km radius around the place of departure. This radius is used as benchmark for trips up to 45 minutes (i.e. the radius corresponding to a straight-line trip at a speed of 16 km/h). For trips up to 30 minutes proximity is defined as the population living within an 8-km radius around the place of departure.

Accessibility: population accessible within 45 minutes by public transport (combined with walking), insofar the destination falls within the 12-km radius. For shorter trips accessibility is defined as the population accessible within 30 minutes insofar the destination falls within an 8-km radius. The maximum allowed walking time (or distance) is set at 15 minutes (or 1.2 km). This walking time can be combined with public transport (access, transfer and/or egress time) or can be used for walking directly to a destination nearby without the need to use public transport.

Performance: accessibility / proximity x 100. As accessibility is constrained by the 12-km (8-km) radius, performance can be described as the share of population in a 12-km (8-km) radius that can be reached within 45 (30) minutes by public transport (and/or by walking, insofar the maximum walking time of 15 minutes is respected).

Public transport performance in EU urban centres

We analysed performance in 197 urban centres in the EU[5]. Due to limitations in data availability the analysis cannot yet be carried out for all urban centres in Europe. Progress in the availability of open and integrated timetable data is fastest in many northern and western parts of the EU. This explains the unequal geographical distribution of the analysed urban centres. This skewed distribution should be kept in mind when interpreting any result of this analysis.

Data are not consistently available for the same timeframe neither. Most often the timetables we used refer to some period during the years 2019 to 2021. All results refer to services operating during morning peak hours of a typical weekday outside main holiday periods. For these morning peak periods the best

---

[3] The GEOSTAT 2011 grid is published at https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat

[4] Population with reference year 2018 has been estimated by 100 x 100 m grid cell by JRC (see Batista e Silva e.a., 2021). For the purpose of the transport performance analysis the values of the 100 m grid have been aggregated to 200 x 200 m grid cells. Sources and downscaling methods of the JRC-GEOSTAT grid are described in a technical note available at https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat (see JRC GRID 2018).

[5] Based on the GEOSTAT 2011 grid the EU counts 683 urban centres. Using integrated timetable data it would also be possible to extend the analysis to urban centres in the EFTA countries.

available trip between a place of departure and of arrival is taken into account. This is the trip that provides the fastest connection of both places.

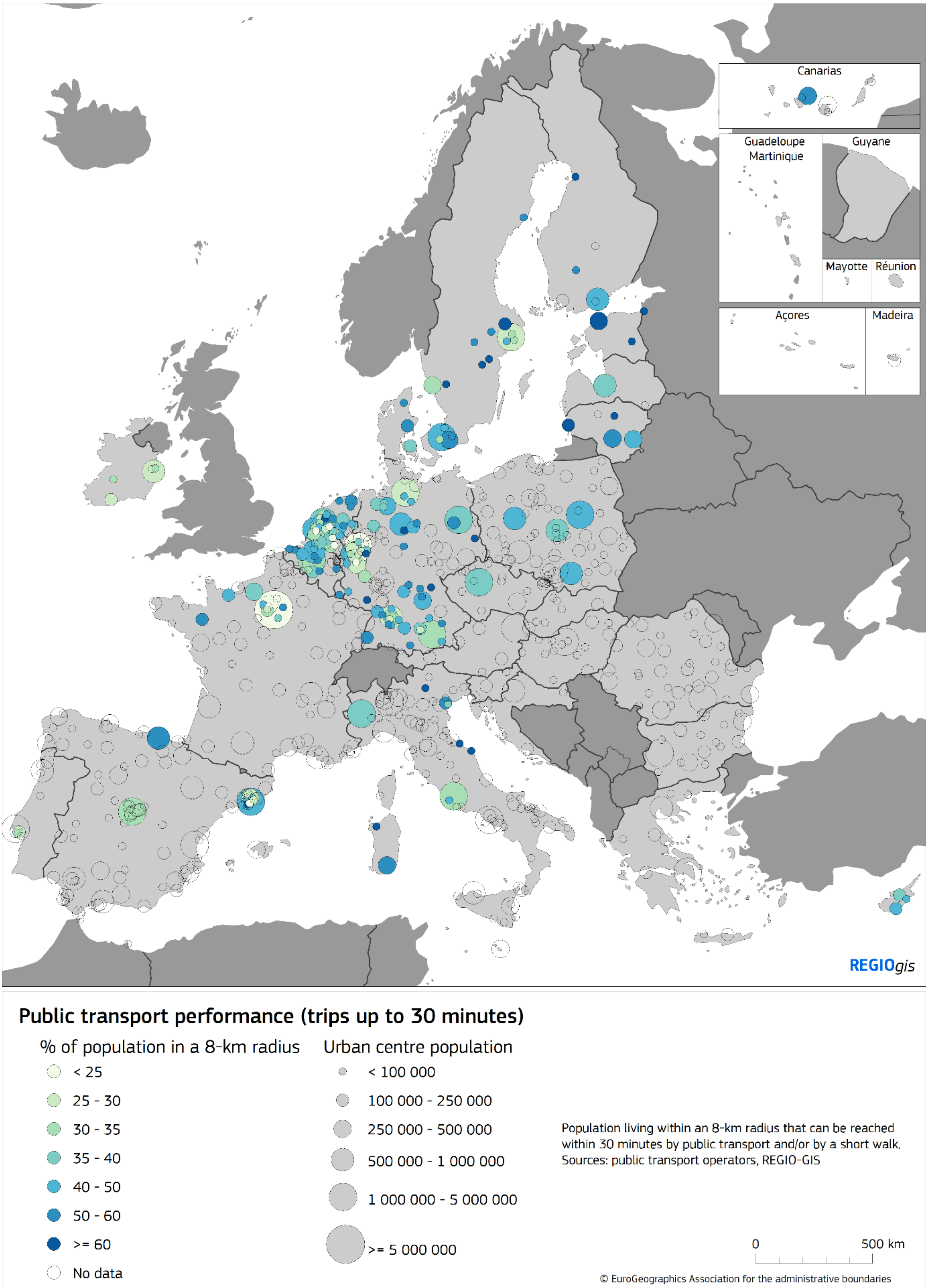
The results by urban centre show a wide variety in transport performance, both for trips up to 30 minutes and up to 45 minutes. On average, in the analysed urban centres 36 % of people living in an 8-km radius can be reached within 30 minutes. Trips up to 45 minutes allow to reach almost 54 % of people living in a 12-km radius. For short trips (up to 30 minutes) public transport tends to perform better in smaller cities than in large ones while the relationship between city size and performance is less pronounced for trips up to 45 minutes. Table 1 provides an overview of the average, minimum and maximum of performance by population size class of urban centres.

Population size class	Number of urban centres	Trips up to 30 minutes, within 8 km radius			Trips up to 45 minutes, within 12 km radius		
		minimum	average	maximum	minimum	average	maximum
50k-100k	100	15.4	45.9	90.1	21.1	53.1	89.6
100k-250k	51	26.6	43.1	83.1	16.7	56.7	89.3
250k-500k	14	23.8	41.8	63.7	31.7	63.3	83.9
500k-1m	17	24.8	39.3	56.8	31.3	59.8	79.1
1m - 2.5m	10	26.0	35.8	47.1	37.4	59.5	71.1
2.5m +	5	20.4	28.1	38.7	32.4	44.5	61.7
all analysed urban centres	**197**	**15.4**	**36.2**	**90.1**	**16.7**	**53.9**	**89.6**
Averages are weighted by the population of the urban centres.							

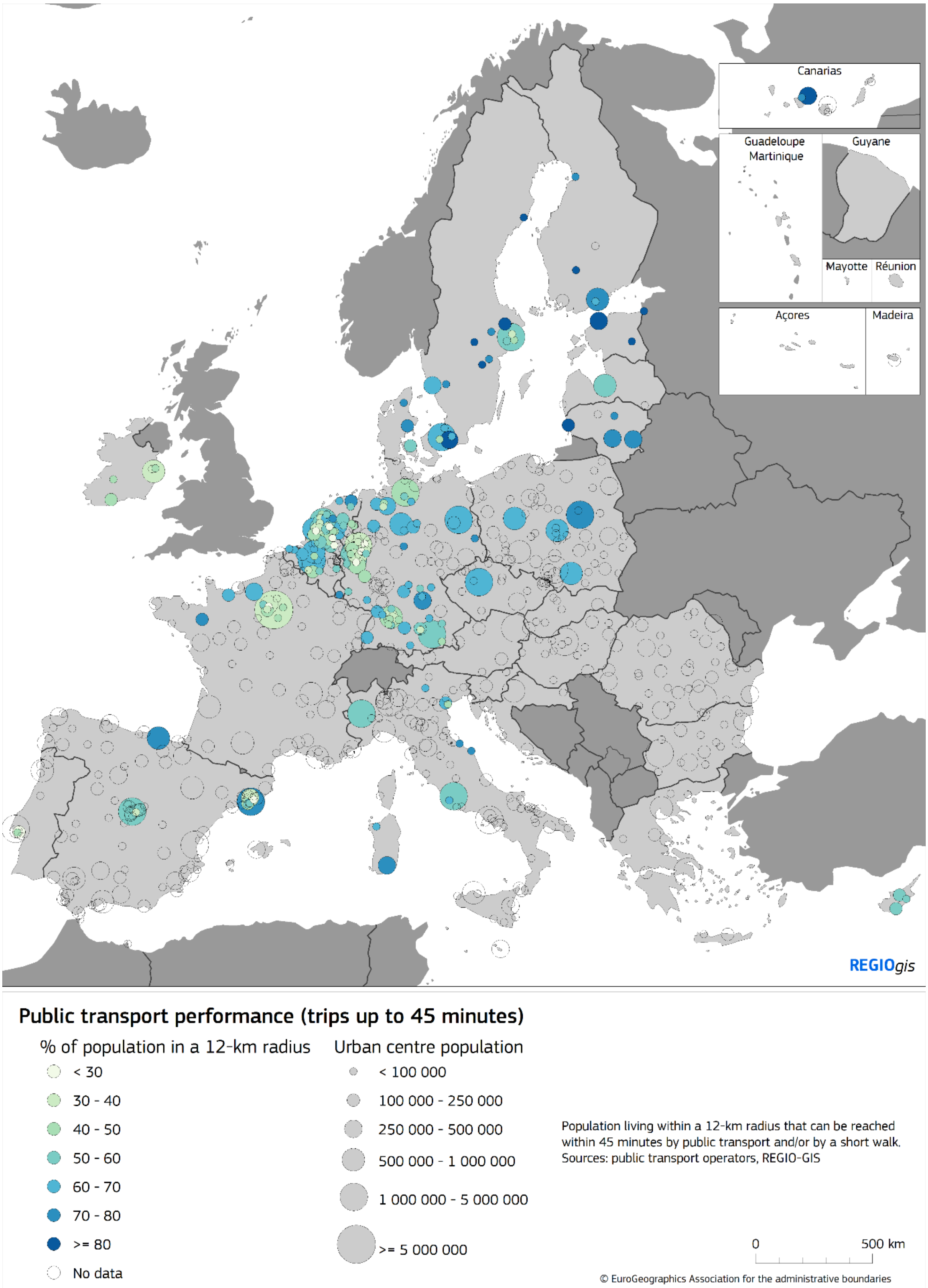
Table 1: Performance by population size class of urban centres

Maps 1 and 2 illustrate the diversity in performance among urban centres and document the spatial distribution of these cities. Data are available for almost[6] all cities in the Nordic and the Baltic countries, Ireland, Netherlands, Belgium, Luxembourg and Cyprus. Furthermore, data are available for many German cities, as well as for some major cities in Spain, France, Italy, Czechia and Poland.

[6] In several countries nationwide timetable data are available. Nevertheless, the analysis was not carried out for some urban centres in those countries, in particular in cross-border situations. This is due to difficulties in combining timetable data from different countries and/or from different sources.



Map 1: Public transport performance for trips up to 30 minutes.



Map 2: Public transport performance for trips up to 45 minutes.

When the travel time between populated places is calculated the algorithm also returns information if a destination has been reached by means of a combination of public transport plus walking, or exclusively by means of a (short) walk. Consequently, we can calculate the share of the performance that is due to short walks (of maximum 15 minutes) without the use of public transport. On average, for destinations within an 8-km radius, 9.1 % of the accessible people are reached by means of a short walk. For destinations within a 12-km radius, the average share of accessibility by means of a short walk drops to 3.9 %.

Figures 1 and 2 illustrate the relationship between performance and the share of accessibility that is due to walking without the use of public transport. Not surprisingly, walking accounts for higher shares in accessibility in smaller urban centres. In urban centres of less than 100 000 inhabitants, the average shares are 8.2 % for destinations up to 12 km and 14.8 % for destinations up to 8 km.

A combination of a high share of accessibility by means of walking and a relatively low level of overall performance indicates situations where the role of public transport in reaching nearby destinations is limited.

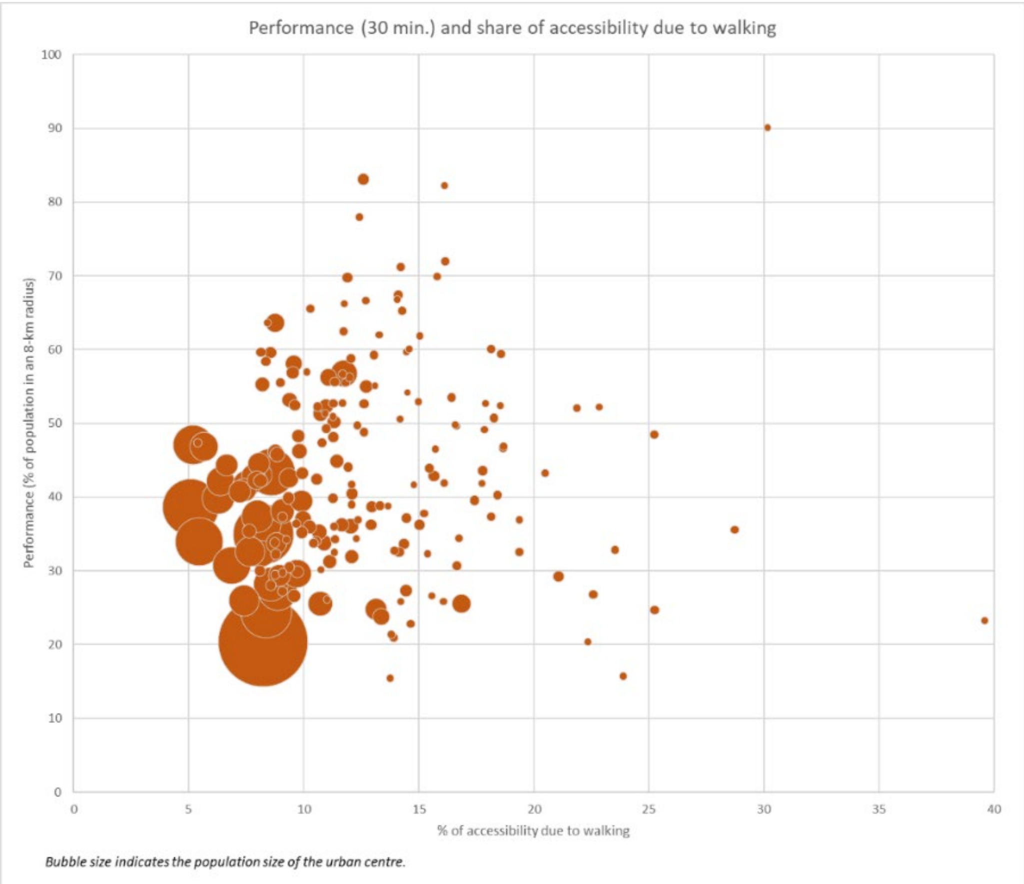


Figure 1: Performance (for trips up to 30 min.) and share of accessibility due to walking.

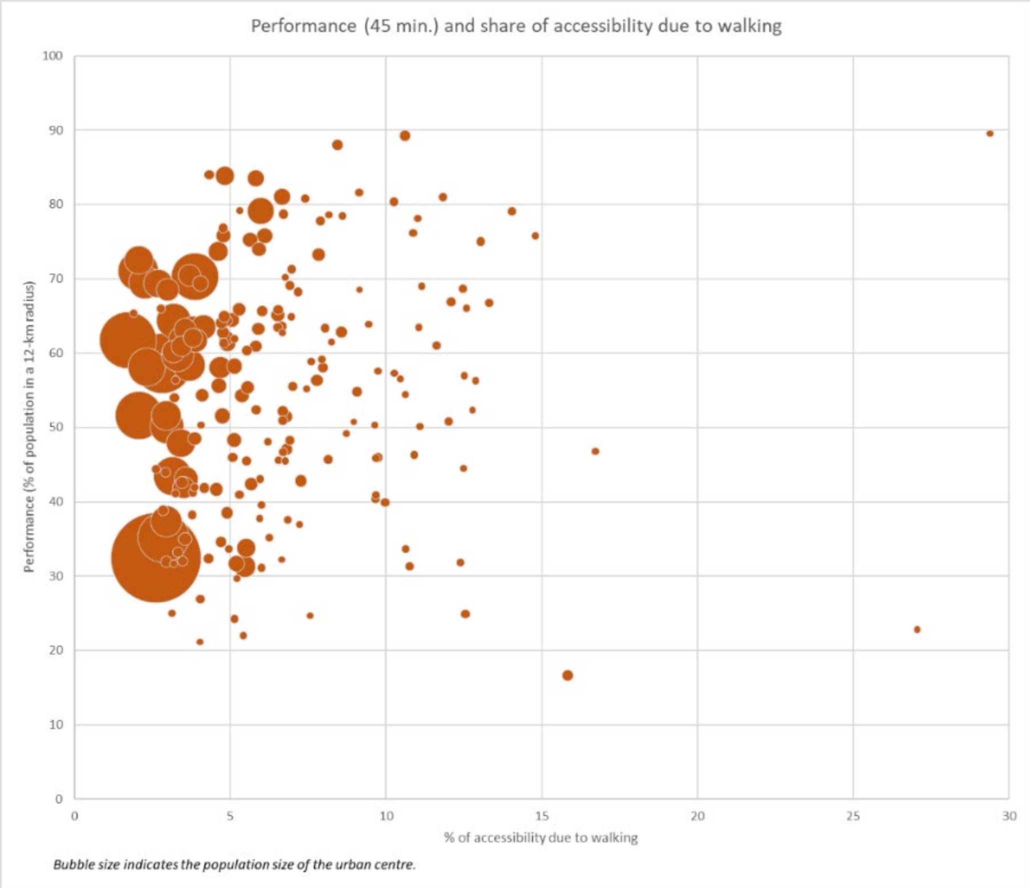


Figure 2: Performance (for trips up to 45 min.) and share of accessibility due to walking.

Reaching high levels of performance is easier in places with a higher population concentration in the neighbourhood. To assess this relationship we calculated, for each populated place in the urban centres, the population-weighted average distance to grid cells within a 12-km and an 8-km radius. A low value means that – on average – population is closely concentrated around the places of departure.

Figures 3 and 4 illustrate the relationship between performance and the average distance to population within the proximity radius and help to highlight where public transport in urban centres performs better or worse than expected on the basis of population concentration.

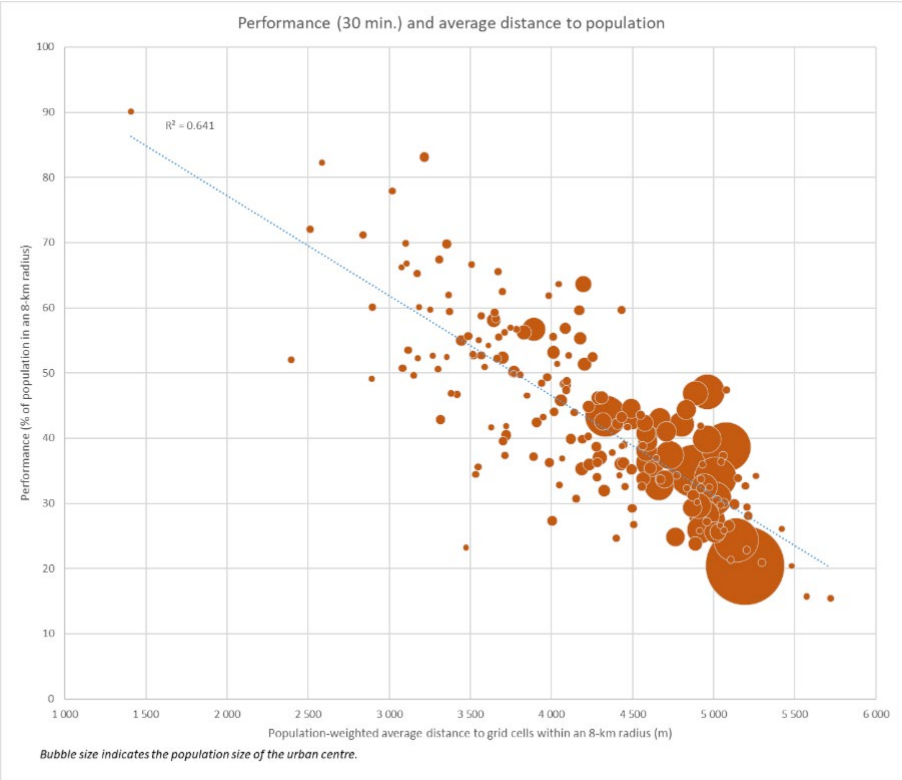


Figure 3: Performance (for trips up to 30 min.) and average distance to nearby population.

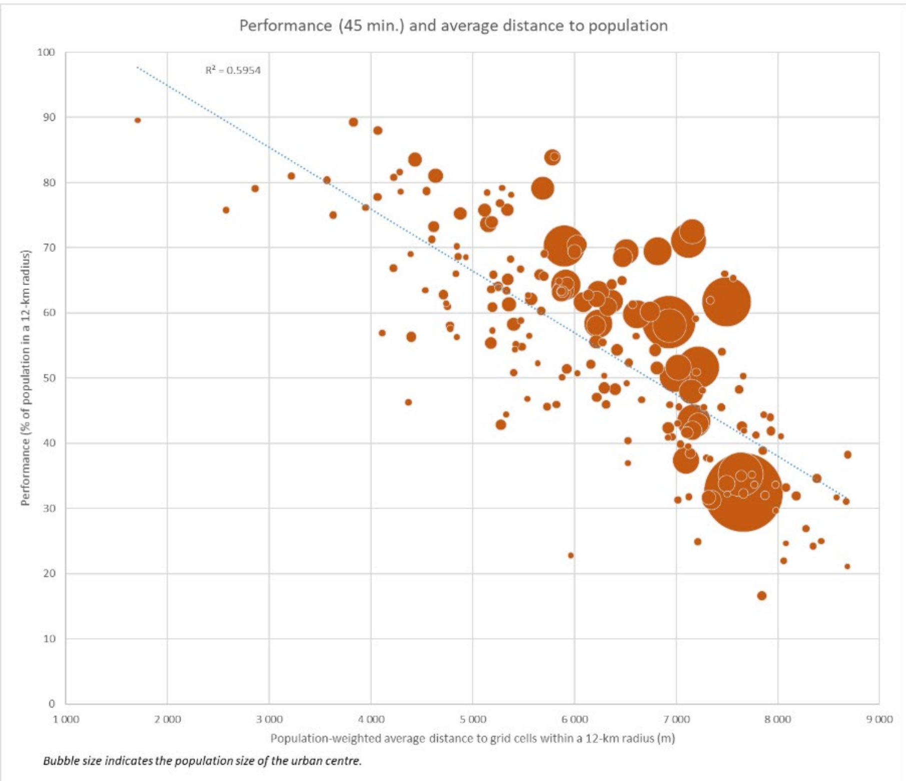


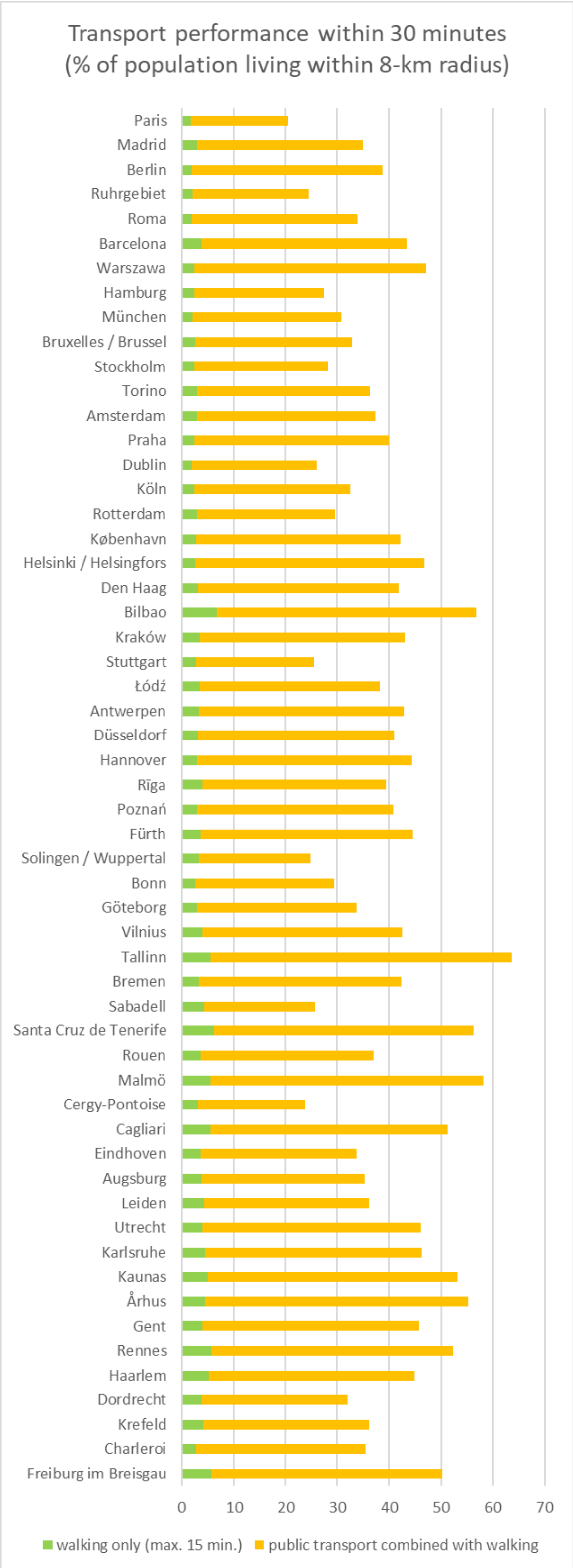
Figure 4: Performance (for trips up to 45 min.) and average distance to nearby population.

Detailed data at the level of individual grid cells inside each of the analysed urban centres can be explored by means of an interactive map viewer[7]. The viewer includes grid data on performance, accessibility, proximity, the share of accessibility that is due to walking, and the average distance to nearby population.

Figures 5 and 6 illustrate the diversity in public transport performance in urban centres of more than 200 000 inhabitants. The urban centres are ranked in descending order of population size. The transport performance shown on these graphs contains two components: a (small) part of the performance is due to destinations that can be reached easiest by means of a short walk, i.e. a walk of maximum 15 minutes. The other part of the performance relates to destinations that are reached by means of a combination of public transport and some walking. In all urban centres shown on these graphs performance for trips up to 45 minutes is higher than performance for trips up to 30 minutes. Only in a minority of smaller urban centres (not shown on Figures 5 and 6) the performance for shorter trips is higher than the performance for longer trips. Among the very large urban centres quite low performance values for Paris and Ruhrgebiet are found. Given the large area size of both centres, performance for trips within a radius of 12 km does not systematically cover trips to or from the centre of the city.

Using population data at grid cell level in combination with the grid-based performance metrics we can draw up simple graphical profiles of the distribution of population by performance level within each city. Figures 7-10 provide such graphs for urban centres of more than 200 000 inhabitants and for some additional capital cities of less than 200 000 inhabitants. The graphs are ranked by descending order of the urban centres' population size.

[7] https://ec.europa.eu/regional_policy/mapapps/public_transport/city_acc_grid.html



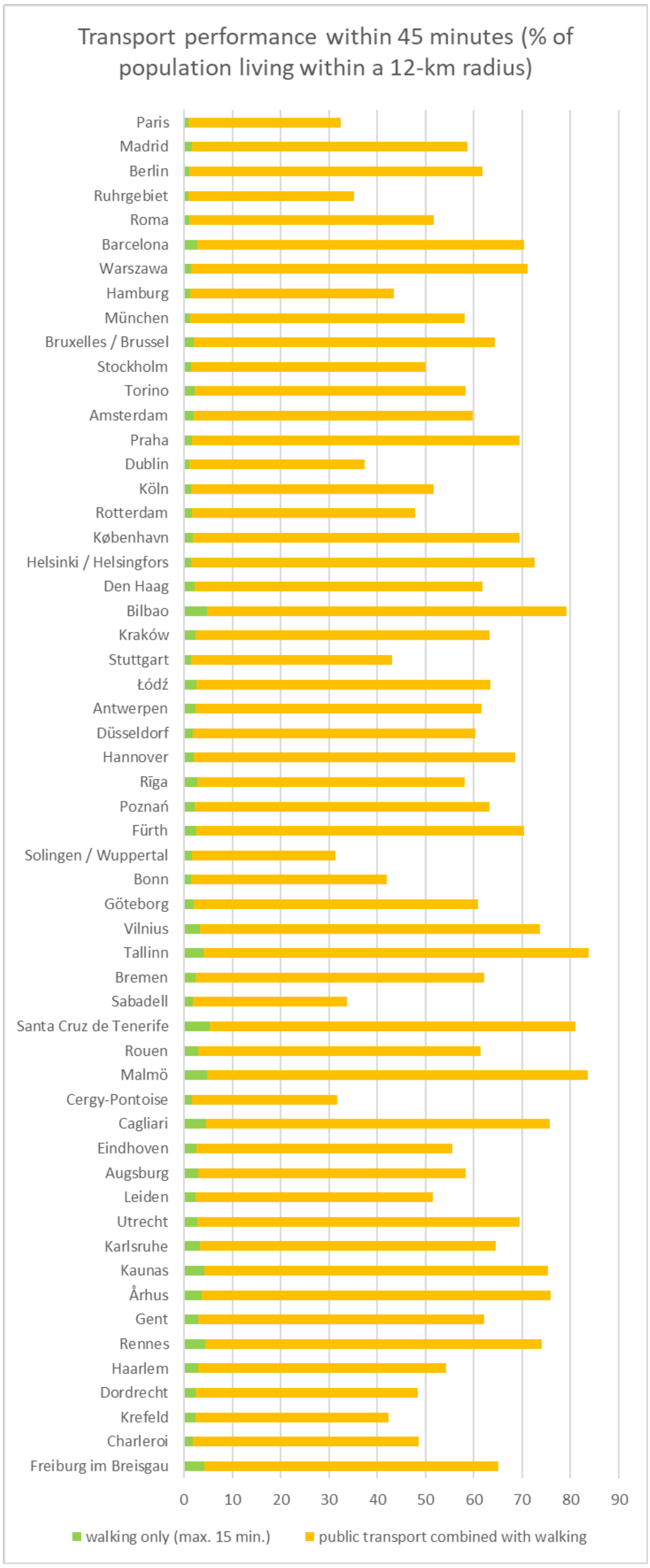


Figure 5 and 6: Transport performance in urban centres of more than 200 000 inhabitants, ranked by total population.

Figure 7: Share of population by level of performance (trips up to 30 min.) in urban centres with more than 500 000 inhabitants.

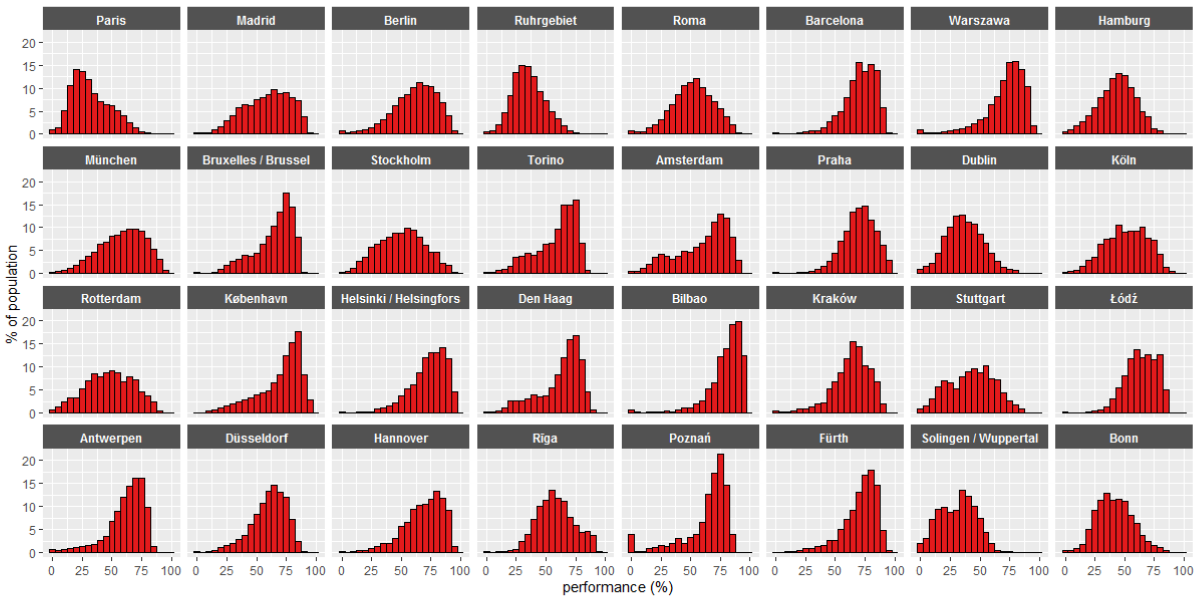


Figure 8: Share of population by level of performance (trips up to 45 min.) in urban centres with more than 500 000 inhabitants.

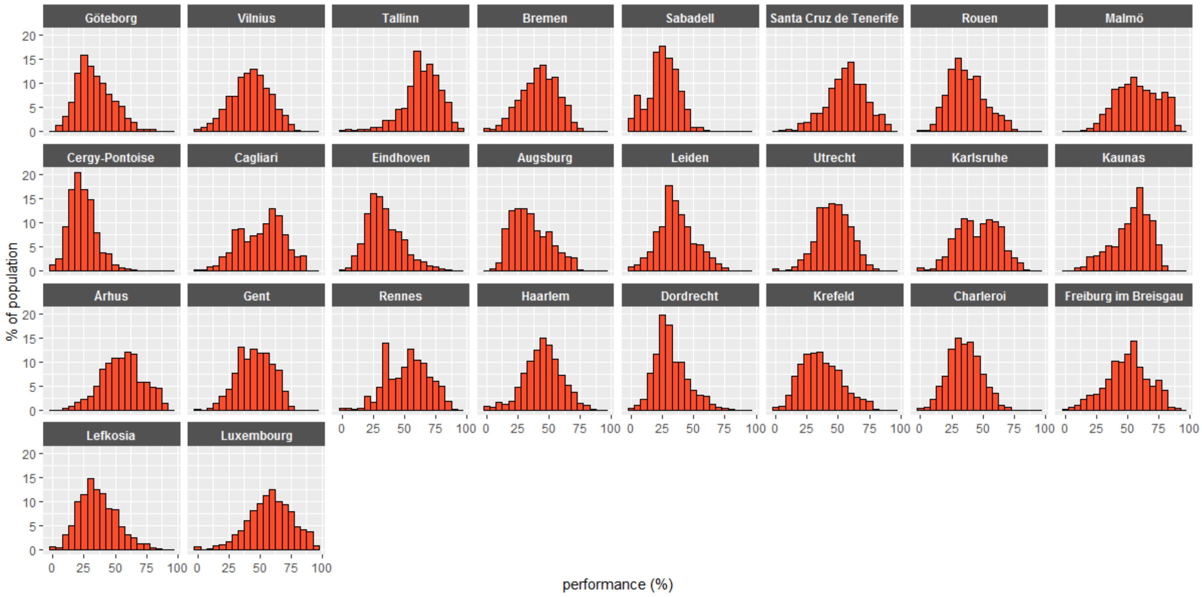


Figure 9: Share of population by level of performance (trips up to 30 min.) in urban centres with a population between 200 000 and 500 000, including smaller capital cities.

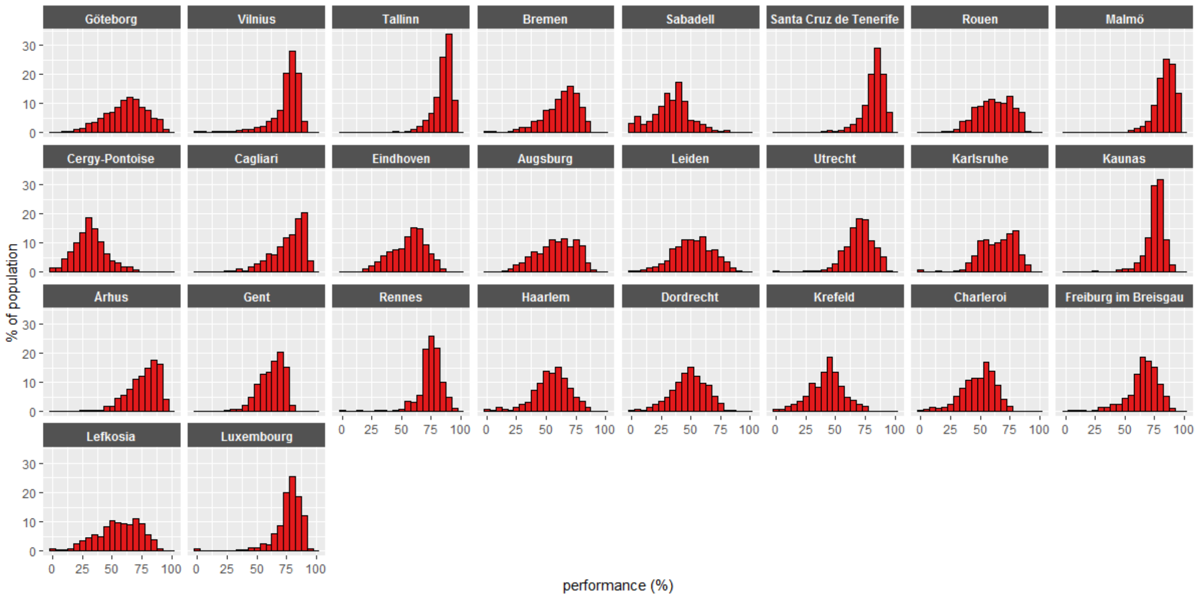


Figure 10: Share of population by level of performance (trips up to 45 min.) in urban centres with a population between 200 000 and 500 000, including smaller capital cities.

Concluding remarks

Using an extensive collection of public transport timetables in combination with high-resolution data on the location of population, this analysis enabled the production of a set of transport performance indicators that highlight the diversity in the way public transport serves urban population to reach nearby destinations in the urban centres and their close surroundings. Additional indicators that can be computed at grid cell level help to interpret the differences in performance, such as the destinations that can be reached by means of a short walk, or the average distance to population living in the proximity of the place of departure.

While the analysis already covers a wide range of urban centres, this collection cannot be considered representative for the situation in all European cities, due to the persistent limitations in the availability of machine-readable and open public transport timetable data in large parts of the EU. Further progress in timetable data availability, harmonisation and openness is needed to be able to extend and update the analysis. The parameters used for the calculation of the performance indicators can be considered realistic and appropriate for the majority of urban centres. For very large cities if might be interesting to investigate how accessibility, proximity and performance change if somewhat higher values for the parameters would be used. Furthermore, a regular production of authoritative population grid data is an important prerequisite for future analysis.

Workflow

Data preparation

The JRC-GEOSTAT 2018 grid comes with a 100 m version, derived from the publicly available 1 km² version[8]. At European level, this 100 m grid is currently the most recent and highest-resolution grid available, based on relatively high-resolution authoritative sources of population figures. Using a 100-m grid for origin-destination computations would require massive amounts of calculations. To somewhat reduce this computational burden while keeping a relatively high resolution of population distribution we used a grid of 200 x 200 metres[9].

Several steps of data preparation have been performed on the entire territory:

1) Aggregation of the 100 m grid to 200 m grid cells by summing the population of the 100 m cells. The 1-km² population grid is used as snap grid, ensuring that the 200 m cells neatly follow the boundaries of the urban centres (resulting grid = POPL_GR_200_2018)
2) Calculation of two proximity grids at 200 m resolution: focal sum of the population in a circular neighbourhood of 12 km and of 8 km radius.
3) Creation of a point layer representing the grid cell centroids of the 200 m population grid. Each of the points has a unique integer identifier (pointid).
4) Creation of a point layer by selecting only the points with population > 0 from the entire 200 meter point layer.
5) Spatial join of the point layer with the urban centre polygons to allocate the urban centre codes to the cell centroid points.
6) Extract the values of the proximity grids to the 200 m point layer. These steps result in a 200 m point layer containing a unique integer identifier, as well as the following items:
    - Population of the grid cell
    - Population within a 12 km radius
    - Population within an 8 km radius
    - Urban centre code

    (resulting point layer = POPL_GR_200_2018_NZ_PT)
7) Creation of a 12 km wide buffer around all urban centres. To determine which urban centres are fit for an analysis of public transport performance, the buffered urban centres are compared with the spatial extent of the sources of public transport timetable data. As buffered urban centres

[8] Batista e Silva e.a., 2021.

[9] A preliminary test has compared the accessibility values by urban centre when using a 100 m grid or a 200 m grid. This test found the differences between the two versions to be very small.

may spill over regional or national borders, several sources of timetables may need to be combined for certain urban centres.

The population-weighted average distance within a 12-km and an 8-km radius is also calculated for all populated 200 m grid cells:

8) For each 200 m grid cell, calculation of the focal sum of population within a 12-km (8-km) radius, whereby each cell is weighted by the straight-line distance from the origin. This operation uses the "kernel file" neighbourhood option of the ESRI Spatial Analyst focal statistics tool. The kernel file contains a matrix of straight-line distances between 0 and 12000 m (8000 m).
9) Division of each of the grids obtained in step 8 by the corresponding grids of the simple (non-weighted) focal sum of population in the circular neighbourhood (i.e. the gids created in step 2). This results in two grids representing the population-weighted average distance. In a hypothetical situation where population would be evenly spread over the territory, this average would be 8 km within a 12-km radius neighbourhood, and 5.3 km within an 8-km radius neighbourhood.

For each urban centre to be analysed (i.e. where timetables of networks covering the entire buffered area are available), some specific geodata are prepared:

1) A point layer containing the 200 m cell centroids with population > 0, located within the buffered urban centre (spatial selection of the wall-to-wall grid of populated 200 m cells, falling within the selected buffered urban centre).
2) A shapefile of the OSM street network covering at least the urban centre plus a buffer of 13 km width. This network will be used to connect the points of origin and destination to the public transport stops. The buffer for the OSM selection is taken somewhat wider than the analysis area to ensure that destinations close to the outer limits of the buffered urban centre can still be reached if a road segment needed to connect to them falls outside the 12 km zone.
3) The point layer of point 1) is snapped to the edges of the street network, using a maximum snapping distance of 1000 m. This process is needed to ensure that all populated centroid points can actually serve as origin and/or destination of trips involving the use of the street network.
4) The XY coordinates of the points (in meters, under the Lambert Azimuthal Equal Aera EPSG:3035 coordinate reference system) are added to the snapped point layer.

An integrated and complete set of timetable data covering the entire urban centre and its surroundings is crucial for a successful analysis. When several complementary datasets are needed (for instance in cross-border situations or if several operators are active in the same area and they each provide a separate GTFS dataset) the validity dates of the timetables in each of these datasets need to be compared in order to find a common date selected to run the travel time calculations. Given the variety in GTFS implementations finding such dates is not always possible. Preparing the GTFS datasets for combined use often requires (manual) adaptations to the datasets.

Travel time calculation

Using OpenTripPlanner Analyst (OTPA 1.5), origin/destination travel time calculations are performed for requested departure times every five minutes during a 30 minutes' morning peak period. As part of the previous analysis project the appropriate timetable data sources for each urban centre were identified[10]. From the combination of these timetable datasets we computed the total number of departures per half an hour, for all stops located within each of the urban centres. This results in a time profile of departures for each of the urban centres. This time profile is slightly smoothened by calculating, for each half an hour

[10] Poelman, H., Dijkstra, L. and Ackermans, L., 2020.

starting at time t, the total number of departures in a full hour starting at time t. From this series we take the maximum value observed during morning hours. Time t corresponding to this maximum then becomes the start time of the 30 minutes' morning peak period used for the origin/destination calculations. Hence, for each origin/destination pair, the travel time calculations result in up to 6 different travel times. From this set, the shortest available total travel time is taken.

The total travel time between origin and destination is composed of up to three elements:

- Walking time from the point of origin to the public transport stop (or to the point of destination).
- Travel time by public transport from the departure stop to the arrival stop (transfers are allowed, i.e. the trip can be composed of more than one leg).
- Walking time from the arrival stop to the point of destination.

A destination is defined as accessible if all following conditions are met:

- The destination is located within the 12-km radius (alternatively within the 8-km radius).
- The shortest total travel time is maximum 45 minutes (alternatively 30 minutes).
- The total walking time does not exceed 15 minutes. This means that destinations that can be reached by walking only (thus without using public transport) should not require more than 1.2 km of walking. If public transport is used, the total complementary walking time should not exceed 15 minutes. This walking time can include access, transfer and/or egress time.

The criterion on maximum travel distance (i.e. 12 or 8 km) is evaluated using the straight-line distance between the coordinates of the points of origin and destination.

Accessibility, proximity and performance

For each populated grid cell (represented by its cell centroid point) located within the urban centre, we calculate the sum of population of all accessible destinations. To that sum the population the cell's own population is added. This means we include self-accessibility (at 200 m cell level). This is consistent with the definition of the proximity metrics, where the population of the departure cell is included as well. Moreover, self-accessibility within a 200 m grid cell will always be covered by a short walk of maximum 15 minutes.

Within the total number of accessible people we also calculate the number that is reached by means of a combination of public transport and walking, i.e. by excluding the destinations that are reached by walking only.

The point layer of the urban centre's grid cell centroids contains the proximity metrics calculated in the first preparatory phase.

Hence, at grid cell level, the transport performance can easily be calculated as:

accessibility / proximity * 100

At the level of the urban centre, population weighted averages of accessibility, proximity and performance are computed, i.e. the averages of the three metrics, each weighted by the population of the grid cell of departure.

The grid cell level data can also be used to compute a distribution of urban centre population by level of performance.

Assessment of the calculation results

The results - both at grid cell level and aggregated at urban centre level - should be carefully inspected. In particular, aggregated values that are unexpectedly high or low may indicate flaws in the input data. Before using the timetable data we applied validation tools to check the quality of the GTFS timetable data, but such validation tools cannot detect every possible anomaly in the data and can even less spot incomplete data. Visualising the grid results can then help to detect possible anomalies, as the spatial patterns of the indicators inside the urban centre normally reveal the presence of the main public transport lines. For instance, a test analysis for some Belgian cities revealed that some GTFS input datasets contained fictitious timetables pertaining to bus-on-demand services, which needed to be removed before running the travel time calculations. Inspecting the share of accessible people that are reached exclusively by walking is also useful to detect anomalies: if this share is excessively high it indicates that the timetable data are flawed, incomplete or not suitable for routing.

Data challenges and opportunities

This analysis faced multiple challenges in terms of data availability and quality. The following improvements would help to enable a better coverage and a more efficient analysis workflow:

- More open (timetable) data, accompanied by efficient tools to find and harvest them.
- More integration of timetable data in datasets covering the entire offer in a region or country, regardless of the operators of the services.
- Timetable datasets covering the same time span, allowing for an easier integration for routing purposes.
- A more harmonised implementation of data models (GTFS or Netex), avoiding variants or incomplete implementations that lead to validation issues.
- Efficient retrieval and integration of cross-border data (regarding timetables and/or street networks).

References

Batista a Silva, Filipe, Poelman, H. and Dijkstra, L., 2021, The JRC-GEOSTAT 2018: a novel small-area population estimation for Europe. European Forum for Geography and Statistics conference.
https://www.efgs.info/wp-content/uploads/2021/09/Session5b-4_Batista_Poelman_Dijstra.pdf

Eurostat, 2019, Methodological manual on territorial typologies – 2018 edition.
https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-008

Poelman, H., Dijkstra, L. and Ackermans, L., 2020, How many people can you reach by public transport, bicycle or on foot in European cities?
Measuring urban accessibility for low-carbon modes. Working paper, EC DG Regional and Urban Policy,
https://ec.europa.eu/regional_policy/information-sources/publications/working-papers/2020/low-carbon-urban-accessibility_en