

OD Stockholm

Data description

Summary

Description of the OD matrix for Stockholm estimated from cellular network data.

Contents

1	Changelog	2
2	Data format.....	3
2.1	sampers.stockholm.....	3
2.2	demopan.od_odsthlm_monthu_stockholm.....	4
2.3	demopan.od_odsthlm_monthhourly_stockholm	5
2.4	demopan.od_home_stockholm	6
3	Data processing.....	7
3.1	Raw dataset.....	7
3.2	Data processing steps	7
3.3	Ping-pong definition.....	7
3.4	Stop definition	7
3.5	Home definition	8
3.6	Mapping from antennas to zones	8
3.7	Technical information	8
4	Statistics.....	9

1 Changelog

The table below indicates changes to previous datasets. The pipeline IDs allow to later find logs and statistics as well as the exact code used for the data processing.

Dataset	Changes to the dataset	Pipelines
2020-11-19	<ul style="list-style-type: none">- Fixed that matrixes actually only included Monday-Thursday- Renamed columns mean, count, std to obs_mean, obs_count, obs_std- Added new column mean that gives the average flow per time period (not only over time periods with observations)	#21088 (Import, Trips) #21524 (User, Filter, OD) #23718 (Export)
2020-11-08	Initial dataset	

2 Data format

Explains the data format of the tables in the database.

2.1 sampers.stockholm

Defines the zones used for home count sand OD-matrixes. It is based on Trafikverkets national Sampers zones (available as sampers.sverige in the DB) with the following modifications:

1. Only zones in Stockholms län are included
2. 9 external zones (id 90001-90009) have been added to represent flows from/to outside of Stockholms län. These zones have a point geometry instead of a polygon.
3. For internal zones, population data has been added from a file called "TAB_1_SAMSSYSS_David.xlsx"

Column	Type	Description
<i>id</i>	int	Zone ID as referenced from other tables (1-1423 internal, 90001-90009 external) (for internal zones same as TAB_1_SAMSSYSS_David, column ID)
<i>geom</i>	geometry (Polygon or Point)	The zone geometry as polygon (internal zones) or point (external zones) using SWEREF99 (SRID 3006)
<i>id_zon</i>	int	Area as in Sampers shapefile (Sverige_slutversion1)
<i>area</i>	double precision	Area as in Sampers shapefile (Sverige_slutversion1)
<i>id2</i>	bigint	id2 as in Sampers shapefile (Sverige_slutversion1)
<i>id_sams</i>	character varying	Sams ID (for internal zones same as TAB_1_SAMSSYSS_David, column SAMSID)
<i>population</i>	double precision	Number of inhabitants in the zone (from TAB_1_SAMSSYSS_David, column BefSum)
<i>bef_förv</i>	double precision	Number of working population (?) (from TAB_1_SAMSSYSS_David, column bef_förv)
<i>bef_ejförv</i>	double precision	Number of non-working population (?) (from TAB_1_SAMSSYSS_David, column bef_ejförv)
<i>name</i>	text	Name of the zone (only external zones)
<i>external</i>	boolean	TRUE if the zone is an external zone, FALSE if the zone is an internal zone (inside Stockholms län)

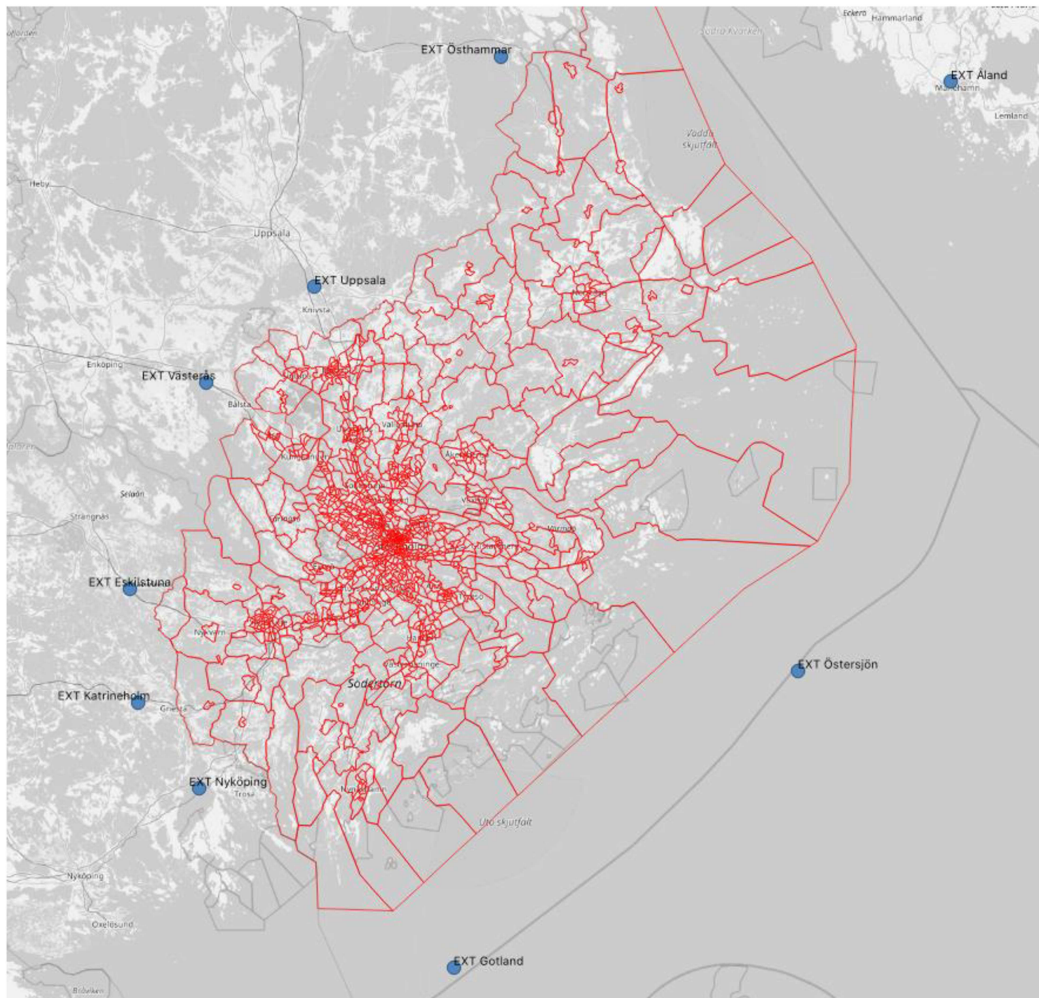


Figure 1. Zones in sampers.stockholm. Internal zones (red) and external zones (blue). Background: OpenStreetMap

2.2 demopan.od_odsthlm_monthu_stockholm

This table is containing an OD matrix for typical weekday in Stockholm (Monday-Thursday). OD-pairs that do not fulfill the anonymity criterion (total ≥ 7) are excluded and summarized in total in an extra row with origin and destination = NULL.

Column	Type	Description
<i>origin</i>	int	Origin zone ID referring to sampers.stockholm

<i>destination</i>	int	Destination zone ID referring to sampers.stockholm
<i>daytype</i>	text	Always “weekday”
<i>odtype</i>	text	“external”: origin and destination outside Stockholm “in”: origin outside Stockholm and destination inside “out”: origin in Stockholm and destination outside “internal”: origin and destination inside Stockholm
<i>mean</i>	double precision	Average flow per day (Monday-Thursday)
<i>total</i>	double precision	Total flow (Monday-Thursday)
<i>obs_mean</i>	double precision	Average flow per days with observations
<i>obs_count</i>	double precision	Number of days with observations
<i>obs_std</i>	double precision	Standard deviation of flow among days with observations

There is a corresponding view `demopan.odsthlm_monthu_stockholm_geom` that adds geometries (origin centroid, destination centroid, line between origin and destination centroid).

2.3 demopan.od_odsthlm_monthuhourly_stockholm

This table is containing an OD matrix for the hourly flow of a typical weekday in Stockholm (Monday-Thursday). OD-pairs that do not fulfil the anonymity criterion ($total \geq 7$) are excluded and summarized in total in an extra row with origin and destination = NULL.

Column	Type	Description
<i>origin</i>	int	Origin zone ID referring to sampers.stockholm
<i>destination</i>	int	Destination zone ID referring to sampers.stockholm
<i>daytype</i>	text	Always “weekday”
<i>odtype</i>	text	“external”: origin and destination outside Stockholm “in”: origin outside Stockholm and destination inside “out”: origin in Stockholm and destination outside “internal”: origin and destination inside Stockholm
<i>h</i>	int	Hour, for example 6 includes all trips starting 6:00-6:59)

<i>mean</i>	double precision	Average flow per hour (Monday-Thursday)
<i>total</i>	double precision	Total flow (number of trips) during that hour as sum over Monday-Thursday
<i>obs_mean</i>	double precision	Average flow per hour among days with observations
<i>obs_count</i>	double precision	Number of days with observations for that hour
<i>obs_std</i>	double precision	Standard deviation of flow among days with observations for that hour

There is a corresponding view `demopan.odsthlm_monthhourly_stockholm_geom` that adds geometries (origin centroid, destination centroid, line between origin and destination centroid).

2.4 demopan.od_home_stockholm

This table contains the number of users with a home in each zone.

Column	Type	Description
<i>id</i>	int	Row number in this table
<i>zone_id</i>	int	Zone ID referring to <code>sampers.stockholm</code>
<i>homes</i>	double precision	Number of homes (number of users that have their home antenna in this zone)

3 Data processing

3.1 Raw dataset

- From Telenor, one week during fall 2018
- Location updates include periodic, location area (LA), routing area (RA), tracking area (TA), and cell updates
- Events by “data devices” have not been used

3.2 Data processing steps

1. Import: cellular network data from one week during 2018 (whole Sweden)
2. Data cleaning: Removed ping-pong events
3. Trip extraction: Detect stops. Everything between two stops is a trip.
4. Home identification: Using the definitions below a home antenna is identified for the user (if there is any according to the definition)
5. Filter trips: Keep only trips with at least one update inside Stockholms län
6. Aggregate OD matrix: Map each trip to an origin and destination Sampers zone (see Section 3.6). Aggregate by counting the number of trips for each OD-pair and time period.
7. Export OD: Export only OD-pairs with where the total flow is ≥ 7 (k), aggregate the remaining trips in single extra row contain the total number of trips not included

The definitions used in the process are given below.

3.3 Ping-pong definition

In the following cases, the event at B is removed:

- Ping pong strict (A-B-A)
 - o If the time between A and A is less than 15 min
- Ping pong (A-B-A)
 - o If the time between B and A is less than 15 min
- Oscillation (A-B-C)
 - o If the speed between A and B as well as B and C is faster than 200km/h and the distance between A and C is less than 50km

3.4 Stop definition

- In the same area (max 2km)
- For at least 2 hours

3.5 Home definition

1. Select all stops of the user:
 - a. start after 14:00
 - b. end before 6:00
2. Aggregate above stops per antenna:
 - a. Total time spent
3. Home location is:
 - a. the antenna with the highest total time

3.6 Mapping from antennas to zones

- Trips with origin and destination inside Stockholms län are mapped to the zone of the origin/destination antenna of the trip
- Trips with origin outside Stockholms län are mapped to the external zone that is closest to the last update before entering Stockholms län
- Trips with destination outside Stockholms län are mapped to the external zone that is closest to the first update after leaving Stockholms län

3.7 Technical information

The code used for computation is found here:

<https://gitlab.liu.se/ktsdemopan/mode/-/tree/691d846717700c8c84a0af8de4878e570251df01>

The Pipeline has been run with the variable PROFILE="demopan"

4 Statistics

The subfolder “Statistics” contains a number of reports with data statistics on the data used in the process to create the dataset.

Report (html)	Description
<i>invalid_event_stat</i>	Report on the events removed from the raw data during data cleaning
<i>data_stat</i>	Raw data statistics after data cleaning
<i>trip_stat</i>	Statistics about all detected trips (whole Sweden)
<i>trip_stat_odsthlm</i>	Statistics about the trips included in the OD matrix (at least one update during the trip in Stockholms län)
<i>log_stat-X</i>	Detailed logs from the pipelines run to create the dataset

Some important info from the logs:

process	args	step_name	message
<i>import_antenna_cov.py</i>	demopan	count antennas	Imported ~138462 antennas.
<i>import_antenna_cov.py</i>	demopan	check duplicate antennas	There are 2 duplicate cellnames. This will cause duplicate events.
<i>setup_event.py</i>	demopan	Verify event count	Number of proper events: 1478703993. Expected 1483317090 (number of raw events)
<i>import_event.py</i>	demopan	Import events	Skipped dd events: 169617950
<i>import_event.py</i>	demopan	Import events	Imported ~1431990000 events.
<i>validate_events.py</i>	demopan	Validate events	Found ~495444000 invalid events.
<i>filter_events.py</i>	demopan	Count filtered events	Filtered: ~909589000 events (61.51% of total 1478700000)
<i>extract_trips.py</i>	demopan 120	Extract trips	Extracted ~21174100 trips.
<i>home.py</i>	demopan stockholm	Counting homes	Found ~1554880 homes.

<i>home_zone.py</i>	demopan stockholm	Verify home count	✓ The sum of homes in zones equals the total number of homes
<i>export_home_zone.py</i>	demopan stockholm	Calculate exportable homes	Exportable homes: 1687131 (99.96% of total 1687746)
<i>filter_trips.py</i>	demopan odsthlm	Counting data	Filtered: ~661854 users (24.16% of total 2739940)
<i>filter_trips.py</i>	demopan odsthlm	Counting data	Filtered: ~5432470 trips (23.29% of total 23324600)
<i>od.py</i>	demopan odsthlm dow stockholm	Verify OD flow	✓ The total flow is similar to the total number of trips
<i>od.py</i>	demopan odsthlm hourly stockholm	Verify OD flow	✓ The total flow is similar to the total number of trips
<i>export_od.py</i>	demopan odsthlm monthu stockholm	Calucate exportable flow	Exportable flow: ~2777556 (82.94% of total 3348737)
<i>export_od.py</i>	demopan odsthlm monthhourly stockholm	Calucate exportable flow	Exportable flow: ~1405114 (42.63% of total 3295737)