Statistics Finland

Merging big data and official statistics for modelling statistical commuting

Eurostat Grant Agreement No. 08143.2015.001-2015.713

Pasi Piela

EFGS Conference, Paris 16.11.2016

Content

- Goals and aims of the study
- Data sources
- Technology and methodology
- Results

Goals

- Generally: to promote sustainable development on commuting in enriching the official statistics
- Objective 1: a feasibility study of commuting time and distance statistics with traffic sensor data and public transport data implementation (with actual updates to the Population Statistics Data Warehouse)
- Objective 2: a new commuting time estimate for every citizen based on many data sources both big data oriented and more traditional administrative ones: *Data Integration*.

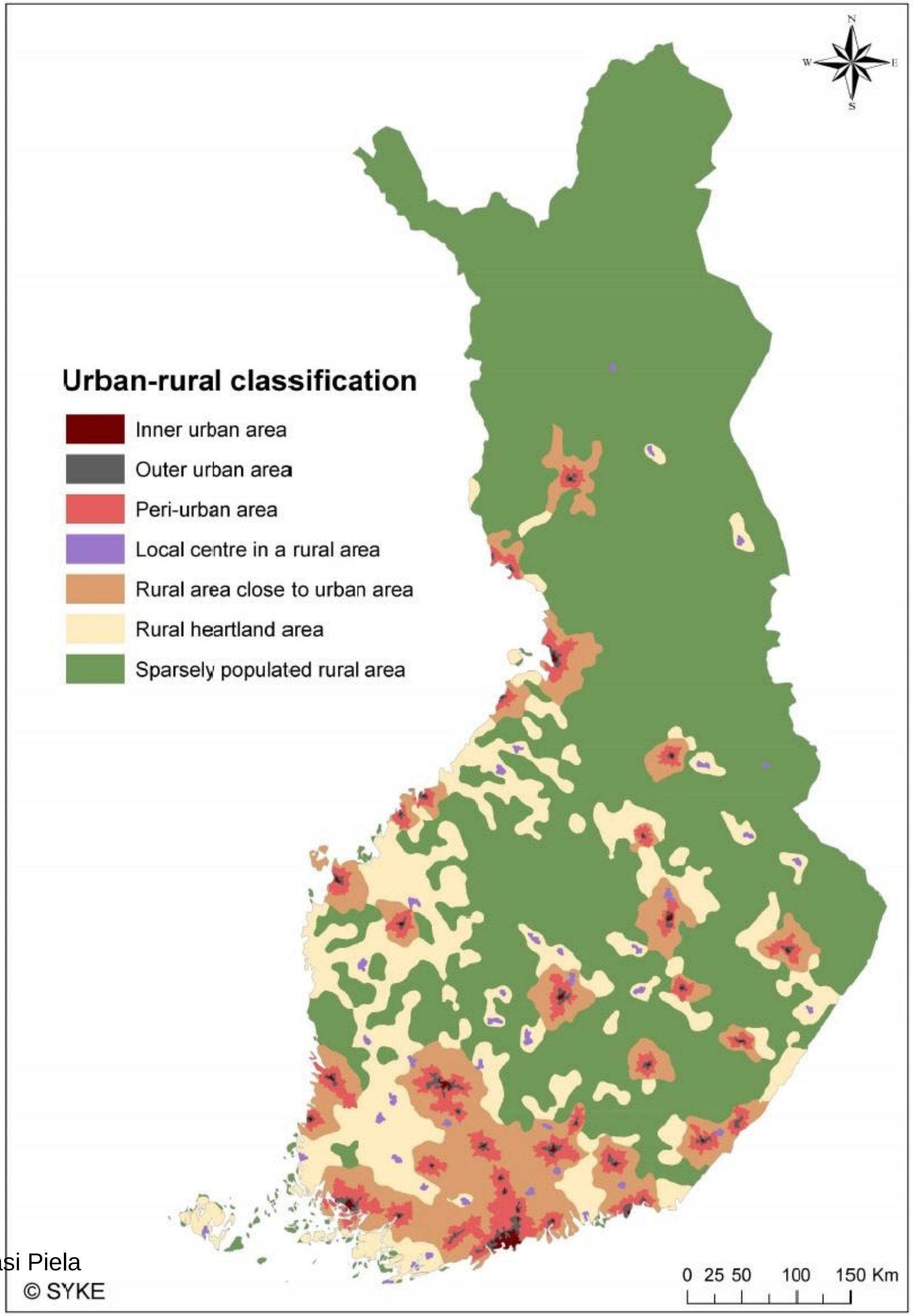
Aims

- Feasibility study on the implementation of traffic sensor data
- General model for a drive speed during the morning rush hour
- Bicycling time model
- Public transport web-service implementation for another commuting option

Note: computational complexity is high; parallel computing environment.

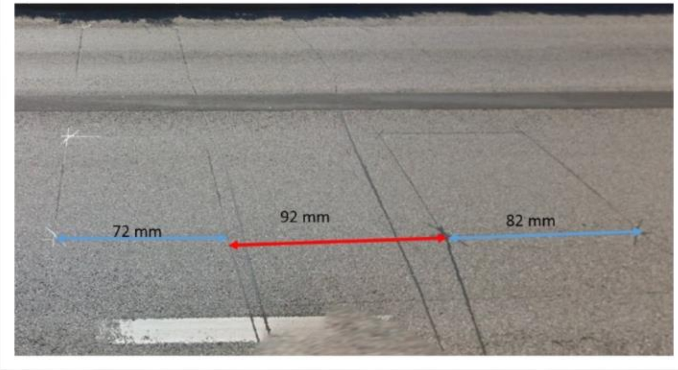
Concept: data sources 1/2

- Digiroad, National Road Database of Finnish Transport Agency
  - accurate data on the location of all roads and streets in Finland
- Social Statistics Data Warehouse of StatFi
  - Dwelling coordinates and work place coordinates along with a variety of demographic features
  - Coordinate coverage of the ”workplace” is 91 % of all employed inhabitants.
- Detailed urban-rural classification by Finnish Environment Institute SYKE



Statistics Finland

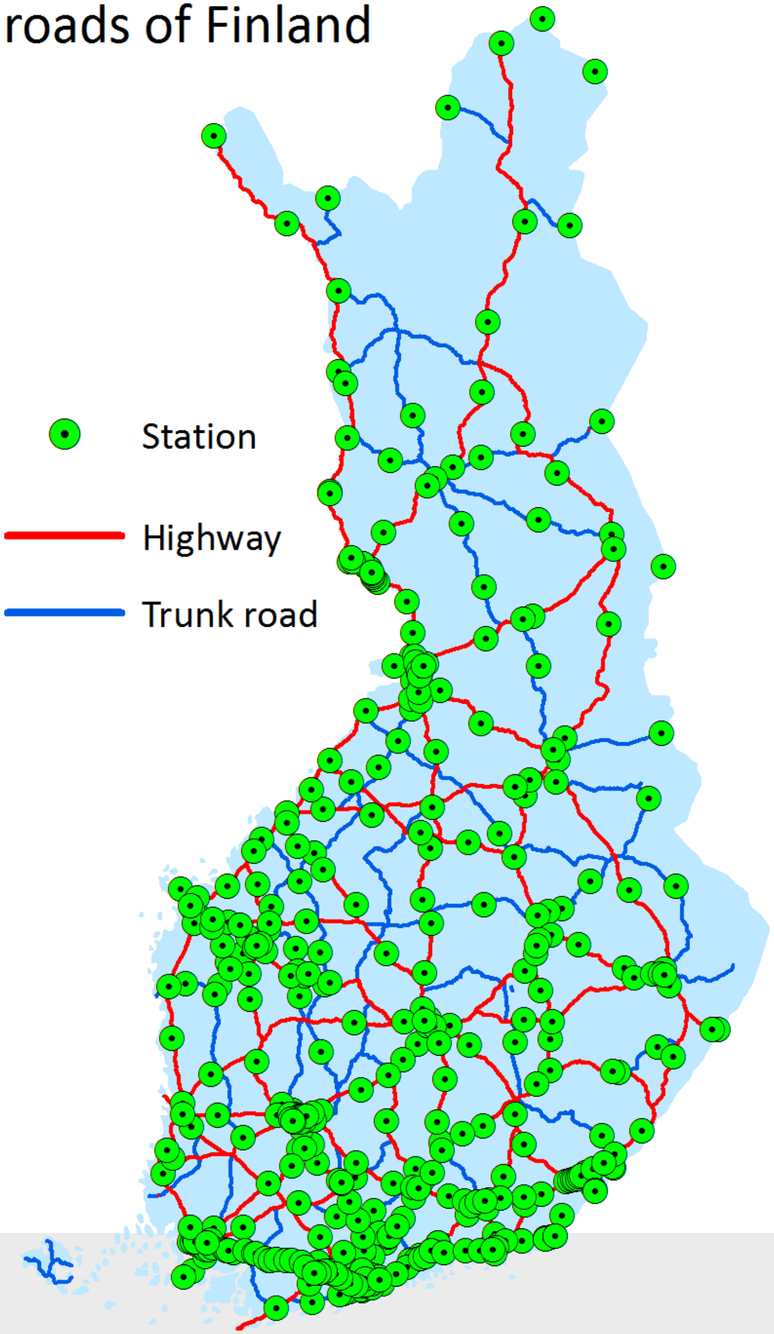
Concept: data sources 2/2



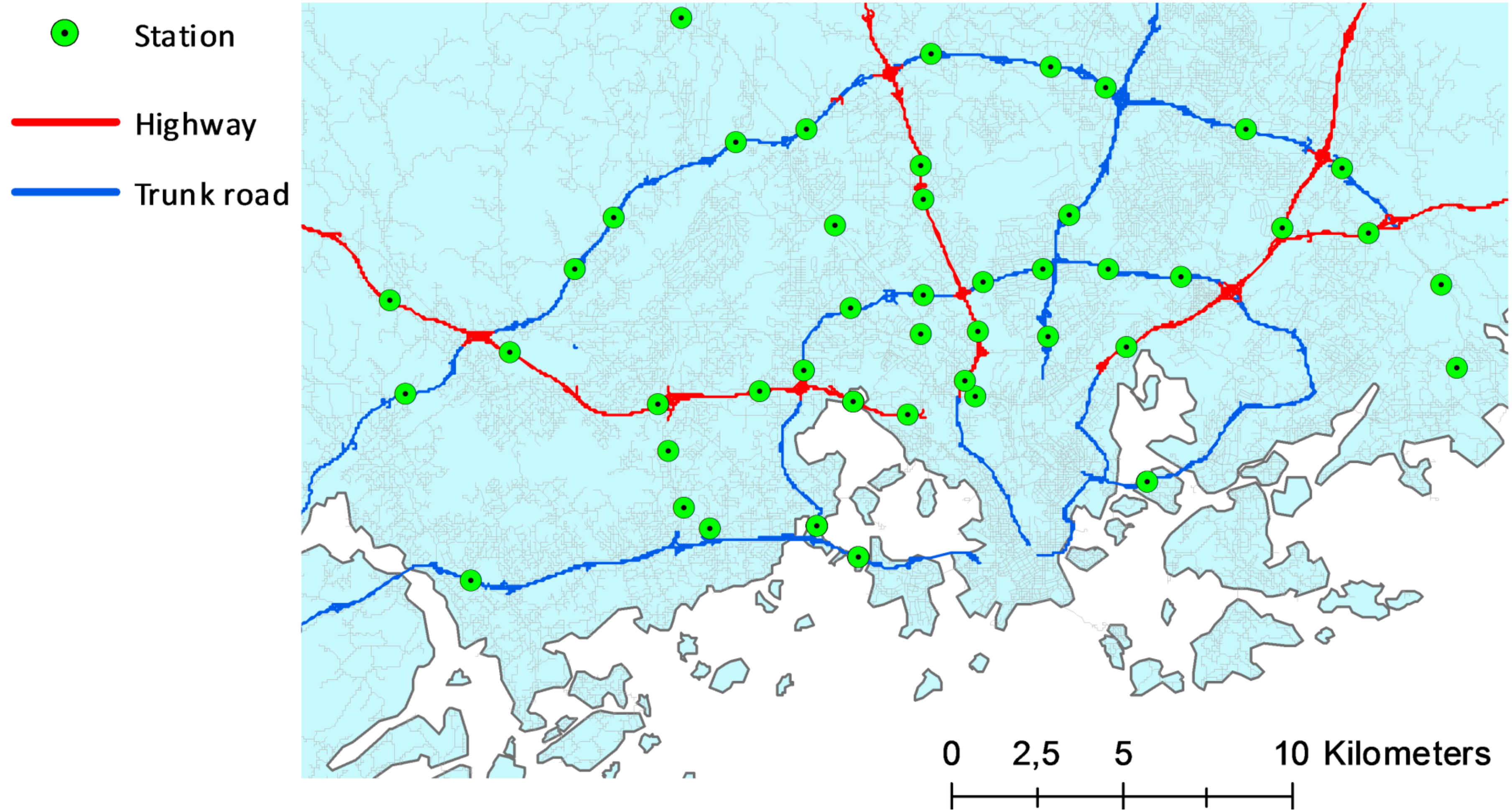
- Traffic sensor data of FTA
  - Currently almost 450 stations (vehicle detection loops) giving information for speed, direction, length and class of a passing vehicle. Target here: speed of passenger cars excluding heavy vehicles.
- Journey Planner for the Helsinki metropolitan area public transport
  - APIs
- Journey Planner by FTA (*Matka.fi*)

Functional possibility: to apply other sources such as Google Maps APIs and Tomtom's anonymised GPS data for comparison.

Automatic traffic measurement stations and main roads of Finland



Automatic traffic measurement stations in the metropolitan area of Helsinki



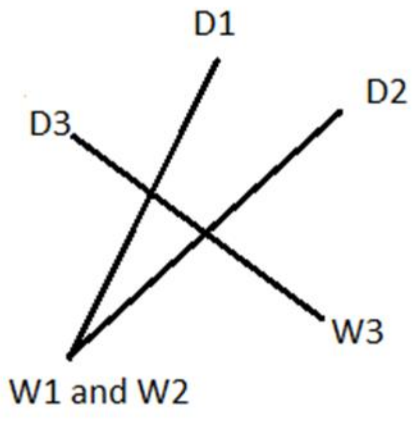
Technology and methodology

Statistics Finland

Concept: Pairwise computing of commuting times and distances

- 2.1 million coordinate pairs i: $((x_{id}, y_{id}), (x_{iw}, y_{iw}))$
- Computational complexity is high
  - ”takes several weeks”

- Helsinki Journey Planner coverage: 490 000 coordinate pairs



Technologies and Standards used

- ESRI ArcGIS® and Python™
  - Network Analyst Route Solver
    - Documentation is not satisfactory but it works: common Dijkstra’s algorithm
  - Cars:
    - with hierarchical routing (6-class functional classification: Class I main road, Class II main road, Regional road etc.)
    - impedance attribute is the travel time (of a smallest road element); accumulation attribute is the length of a route.
  - Bicycling:
    - Non-hierarchical routing, cycle paths taken into account when available

Technologies and Standards used

- SAS® programming for statistical analysis and data handling
  - Helsinki and FTA Journey Planners:
    - remote computing calls are made via APIs
- A combination of desktop computers, remote computers and remote web service platforms for computation and data transfer

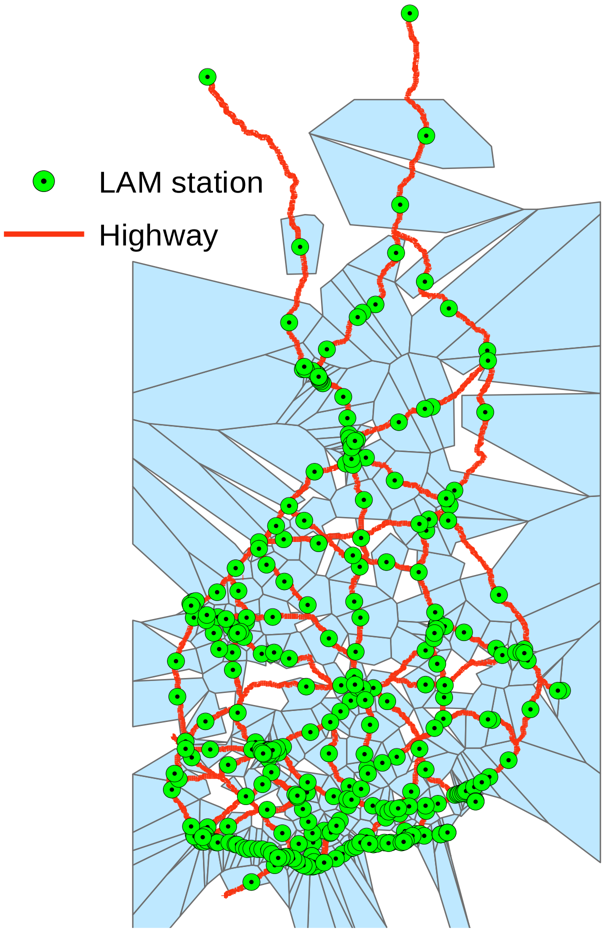
- Note: data protection measures for remote computing outside Statistics Finland in the case of Journey Planners.

Travel-time model for a drive speed

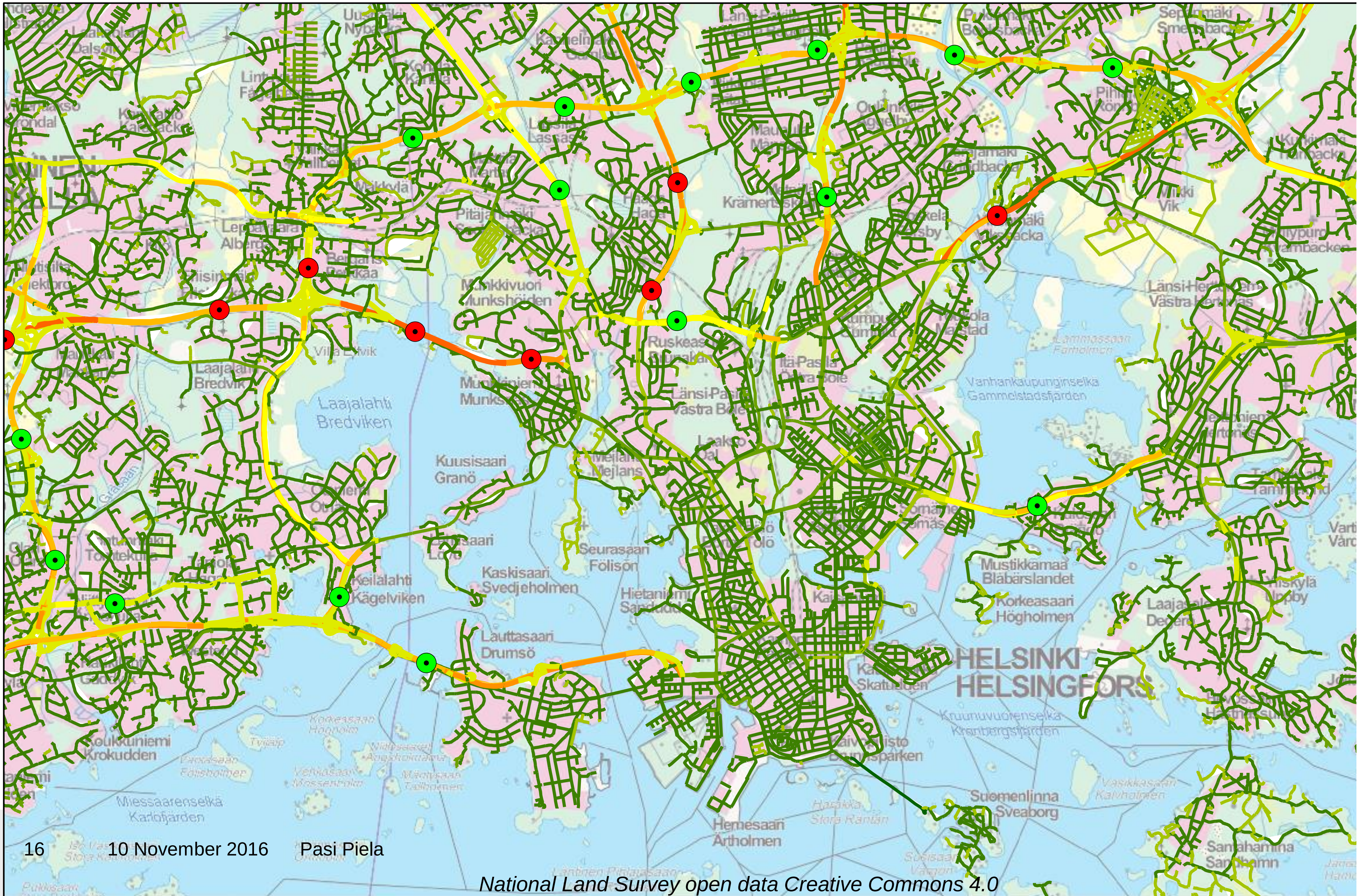
1. Traffic sensor data for a 4 weeks autumn period: <u>Mon-Fri</u> 1 Sep – 26 Sep in this <u>case of 2014</u> for both directions separately between 7:00 and 8:00 only (busiest morning traffic hour).
2. generalised sensor data is compared to the speed limit of the corresponding road element (avg length < 200 meters). The lowest speed is taken.
3. The speed of all the other roads is estimated by using a specific road functional classification (14 classes, *strict* computational speeds). The speed classes are between 20 and 105 km/h.
   – The algorithm uses hierarchical routing moving a car away from a slow street when possible.

Travel-time model for a drive speed

– See the Feasibility Study report 2016

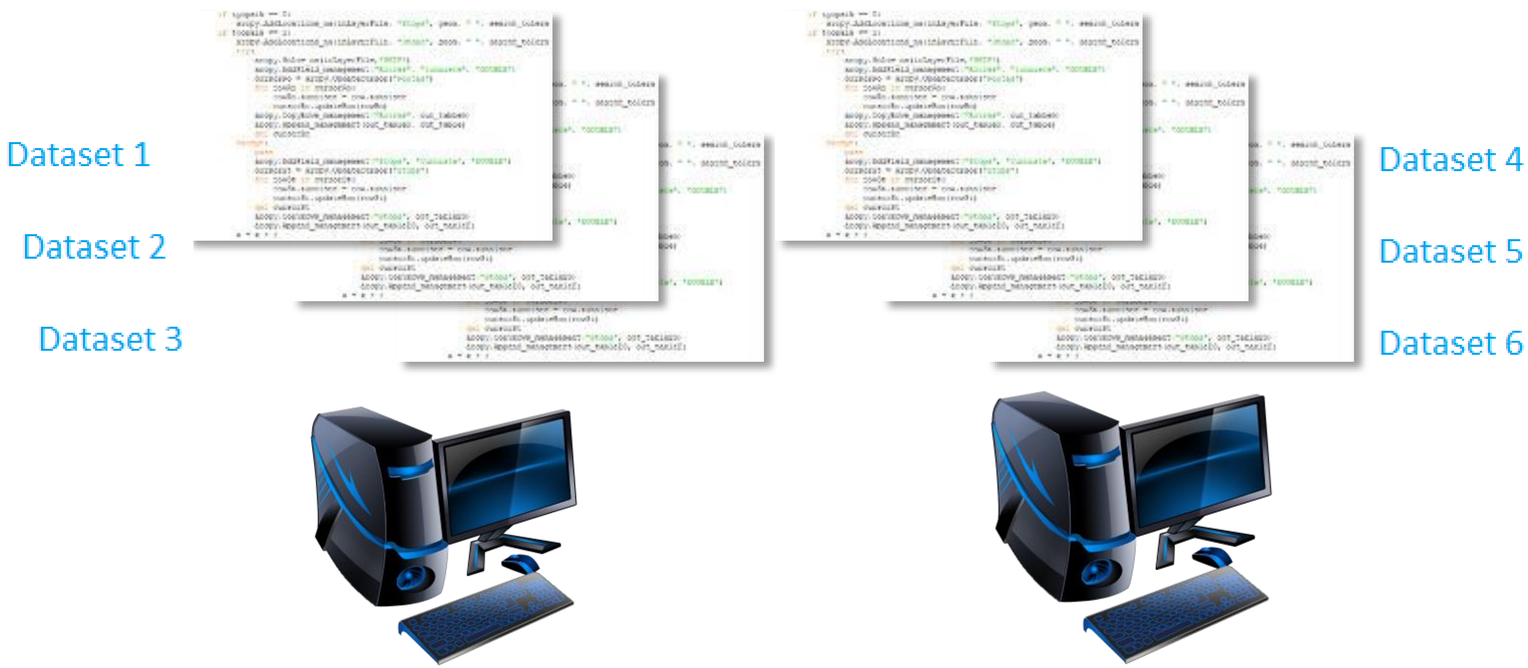


Examble: The capital area speeds



Pairwise computing of commuting distances and time: Network Analyst Route Solver

- The current solution: 100,000 pairs of points (200,000 datarows) for one program run. 3 parallel runs per one standard computer. For 2 computers (2 licenses): *600,000 distances in about 6 days*.

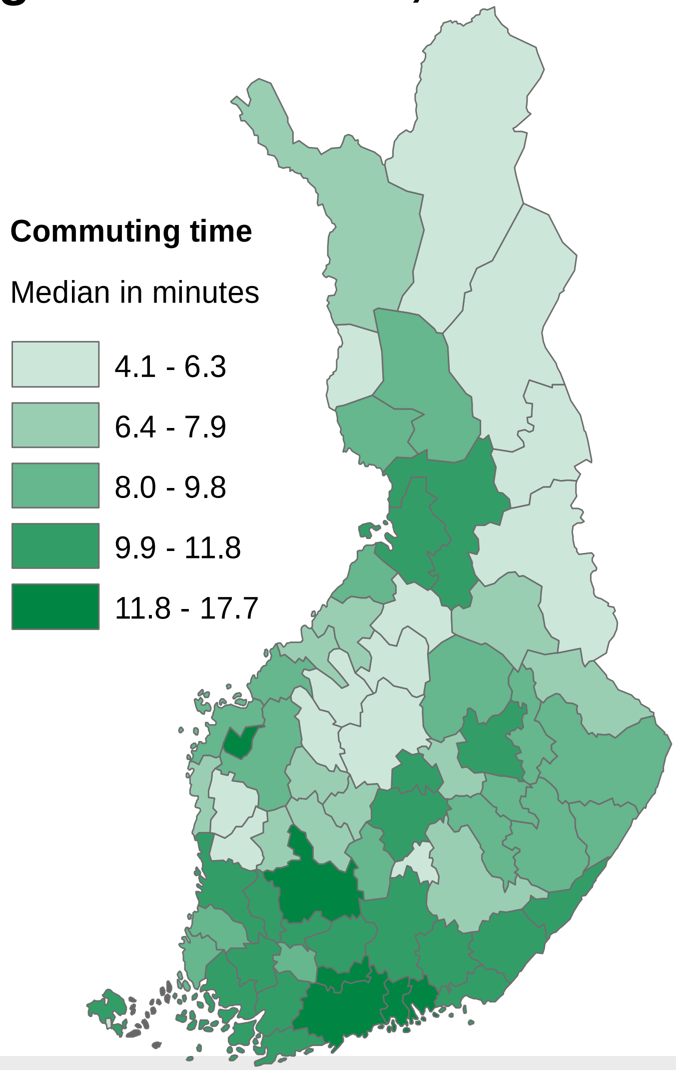


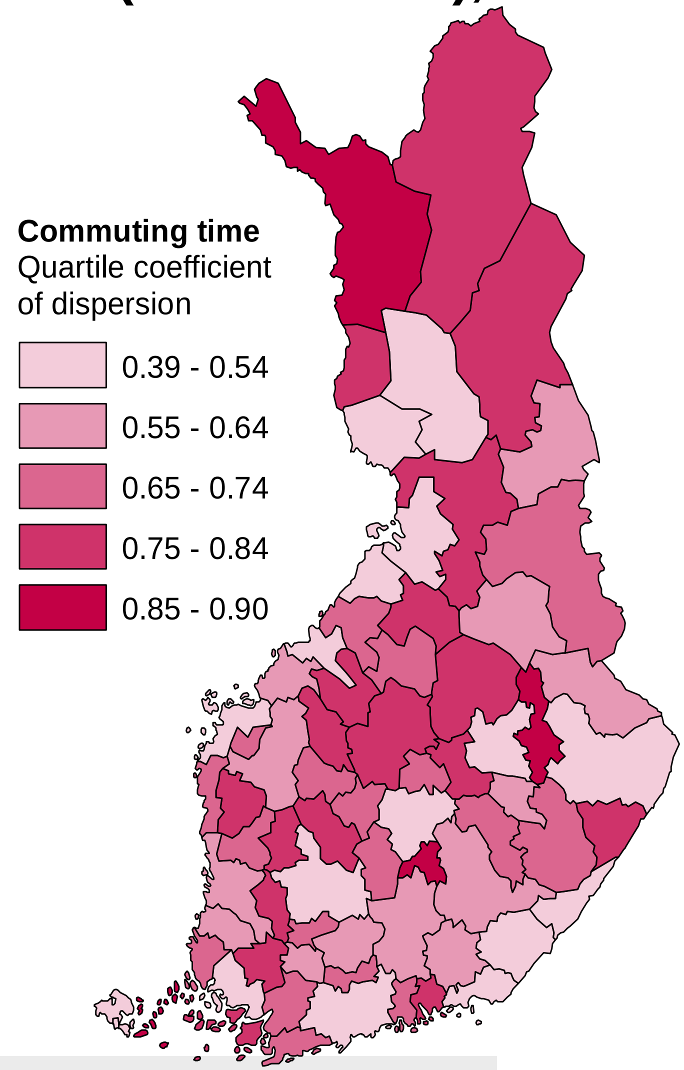
Results

Results: Commuting drive time and distance statistics in kilometres and minutes

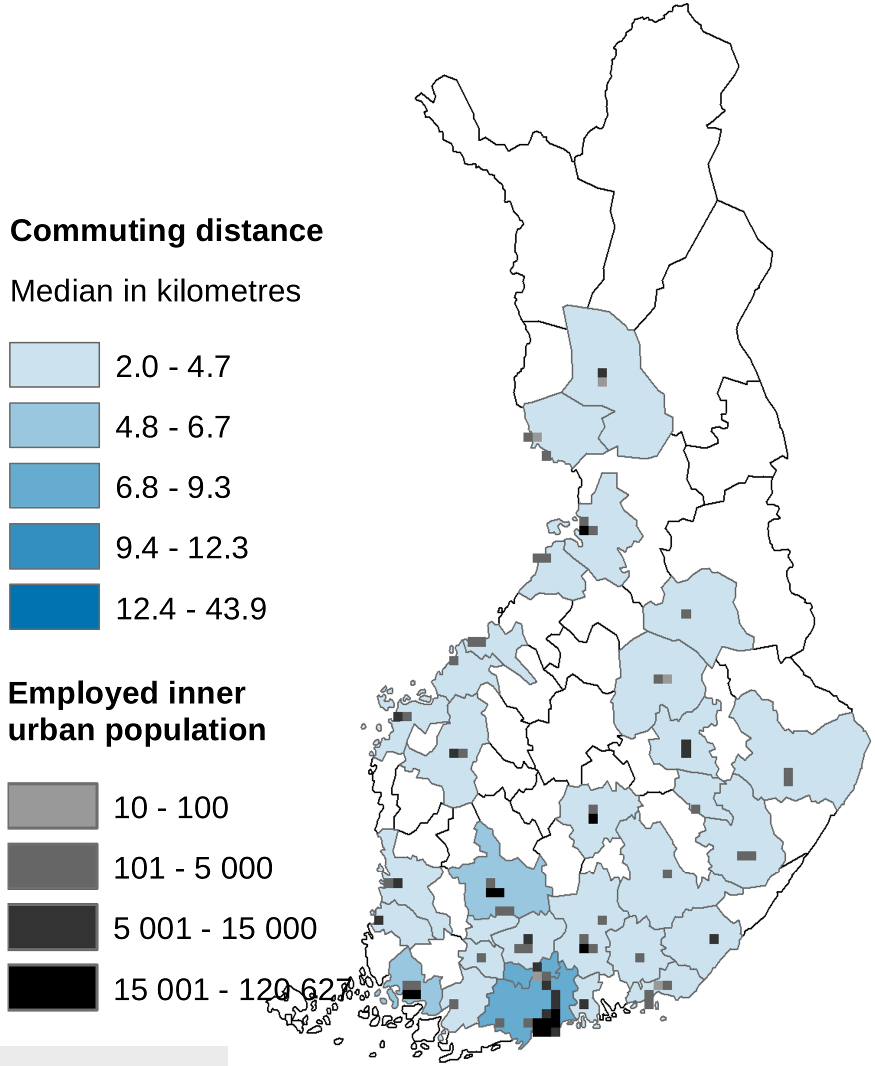
Method	Median	Std	Mean	Q1	Q3
Linear, km	5.97	21.67	13.20	2.07	14.62
Route, km	8.77	25.96	17.14	3.11	19.66
Time, min.	11.47	18.76	16.36	5.64	20.13

The median and the quartile coefficient of dispersion of commuting time along the route network for populations of the sub-regions of Finland, at LAU 1 level (*seutukunnat*), in 2013

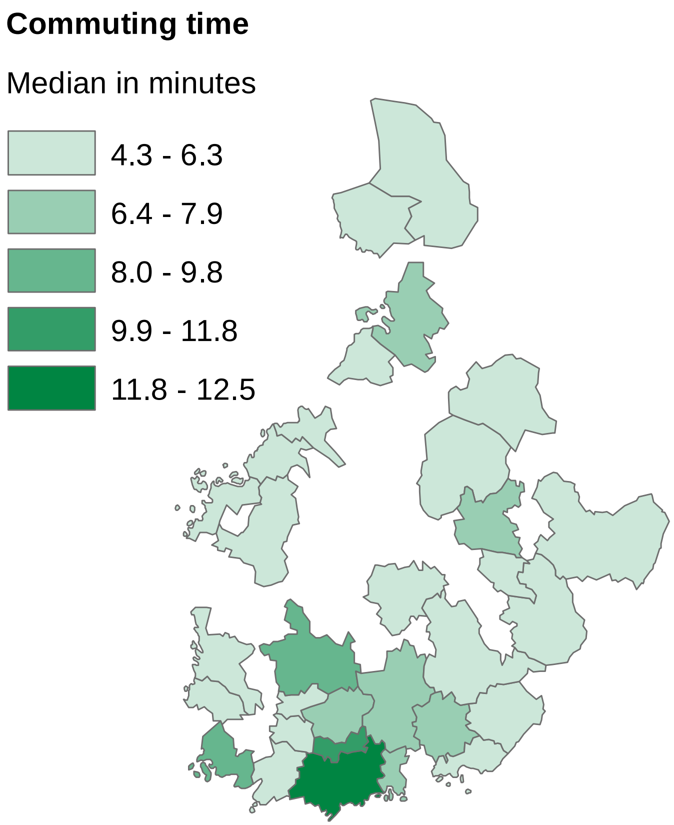


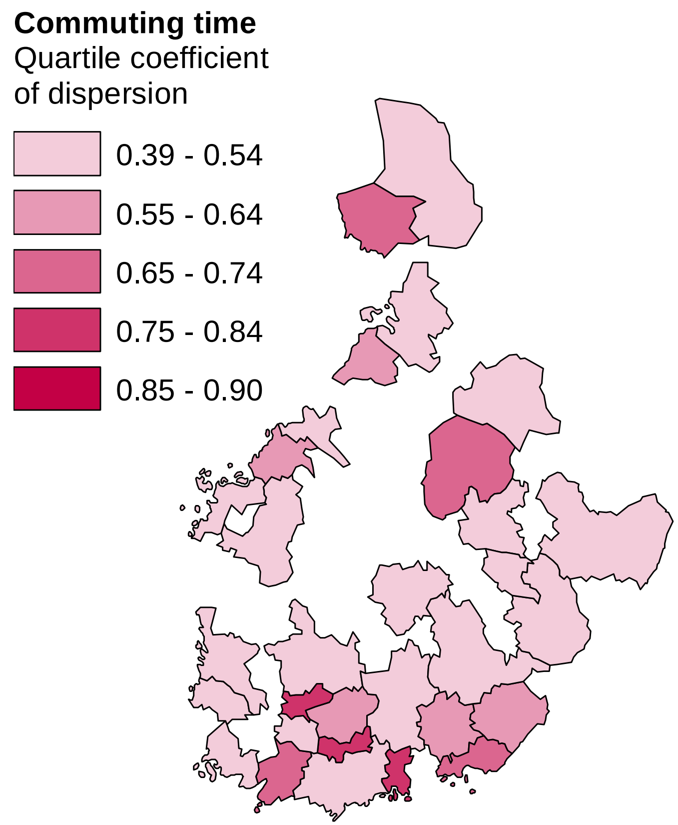


The median of commuting distance of the inner urban populations by the sub-regions and the location and the number of the corresponding populations in 10 square kilometer grids

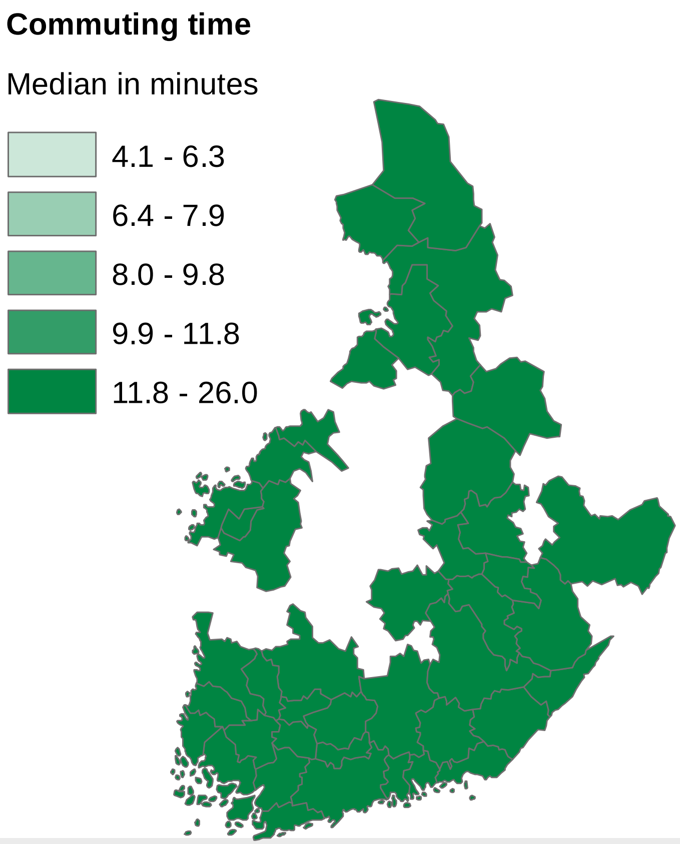


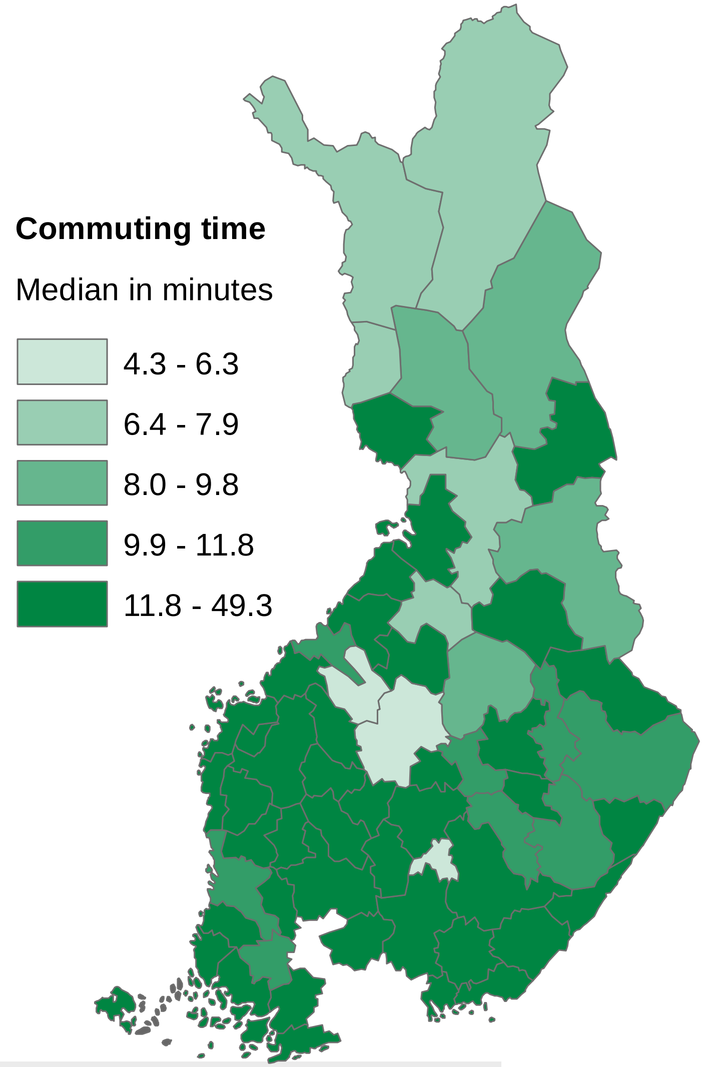
The median of commuting time and the corresponding quartile coefficient of dispersion in Inner urban areas



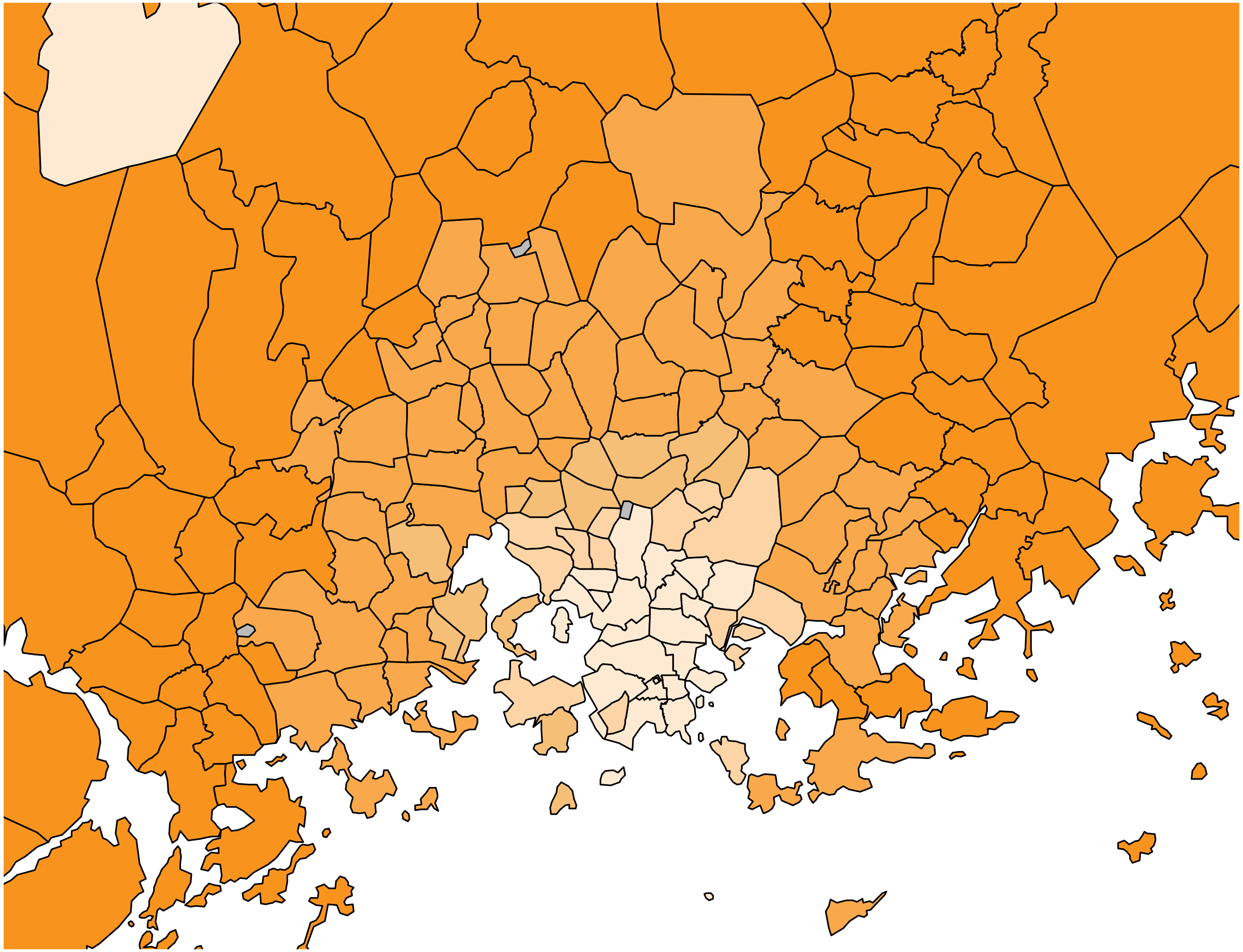


The median of commuting time in rural areas close to urban areas (left) and in sparsely populated rural areas (right).





Cycle commuting time by postal areas in Capital Region



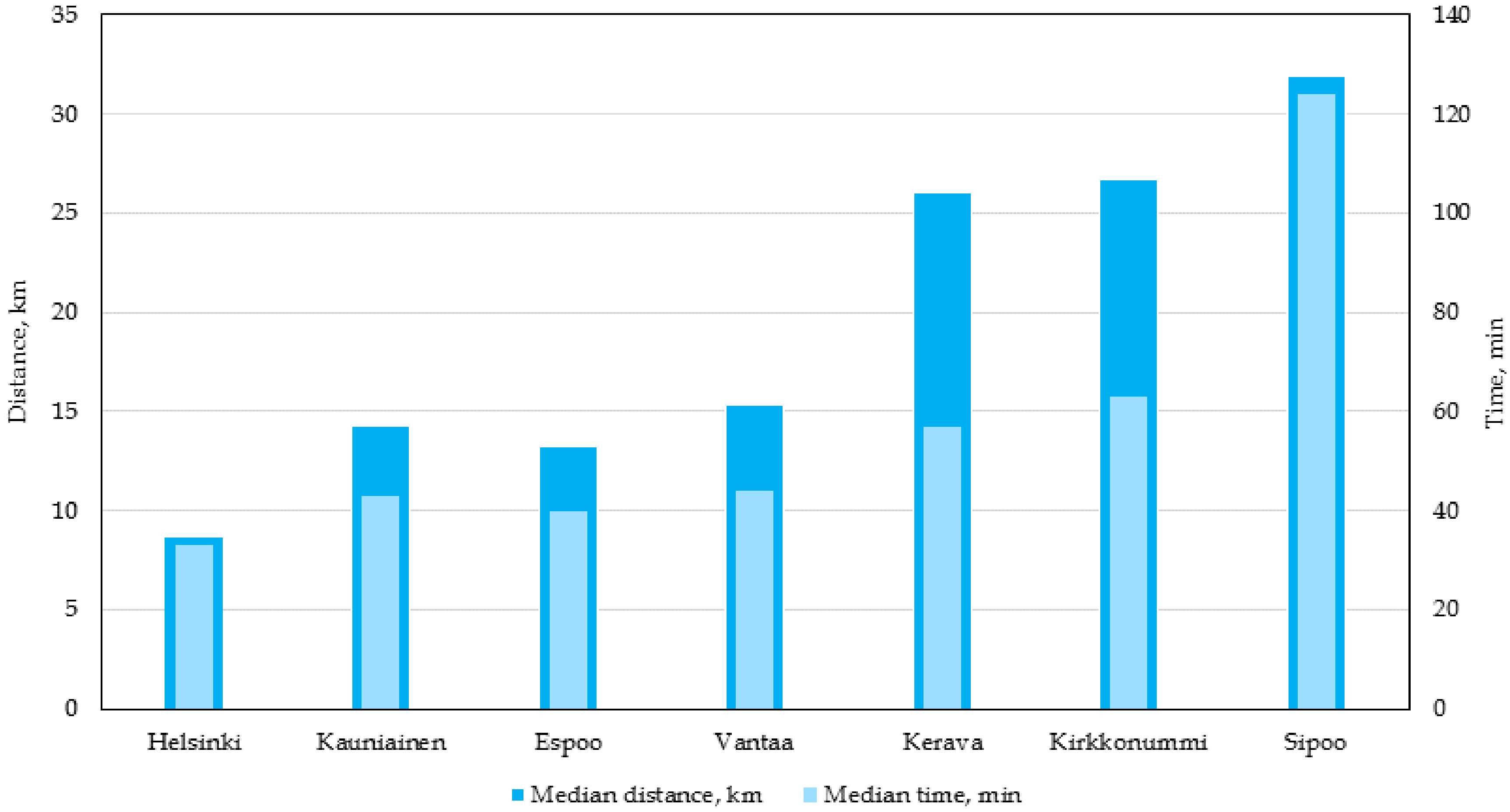
Commuting time

Median in minutes

- 15.5
15.6 - 21.0
21.1 - 26.0
26.1 - 36.5
36.6 -

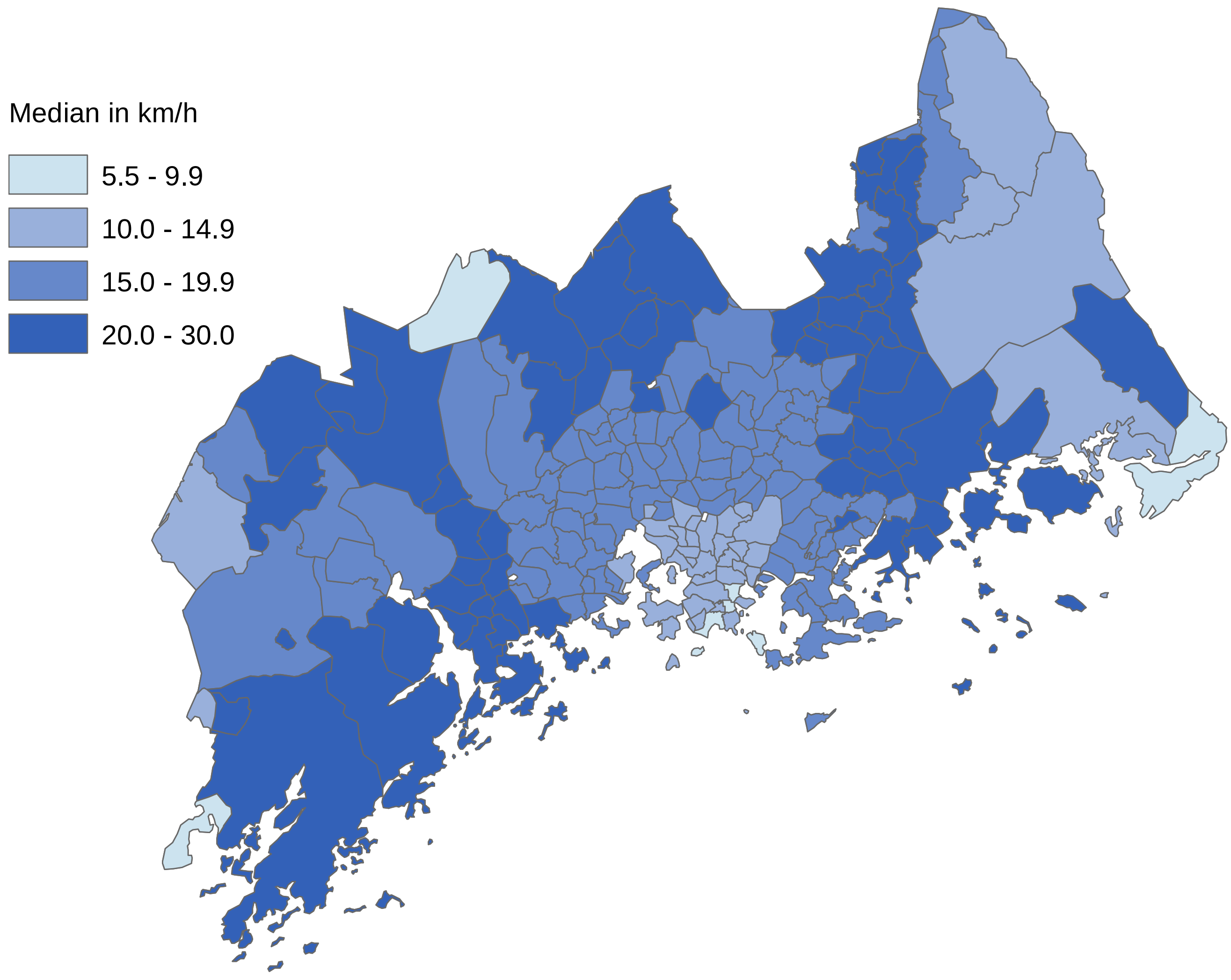
Helsinki Journey Planner: covered area

Helsinki Region Public Transport, *HSL*



Statistics Finland

Postal area medians of commuting speeds by using HSL



Extra slides will be presented

...

Je vous remercie de votre attention!

Pasi Piela pasi.piela(at)stat.fi & Aura Pasila aura.pasila(at)stat.fi

EFGS Paris Conference 2016