May AI Help You? Automatic Settlement Selection for Small-Scale Maps Using Selected Machine Learning Models

Izabela Karsznia*, Robert Weibel, Stefan Leyk

* **i.karsznia@uw.edu.pl**, robert.weibel@geo.uzh.ch, stefan.leyk@colorado.edu

UNIVERSITY OF WARSAW

University of Zurich UZH

University of Colorado Boulder

This research was funded by the National Science Centre, Poland, grant number UMO2020/37/B/HS4/02605, “Improving Settlement and Road Network Design for Maps of Small Scales Using Artificial Intelligence and Graph Theory”.

http://geoinformatics.uw.edu.pl/generalizationai/

NATIONAL SCIENCE CENTRE POLAND

Cartographic generalization

modelling and **abstracting** geographic information while **maintaining** or even **emphasizing** important **relations** and **patterns** existing in the data

(Mackaness et al. 2017)

to make it **efficient** and **automatic** the **deep knowledge** encapsulated in well-designed maps and human cartographer mind need to be made explicit and implemented in the software supporting map generalization

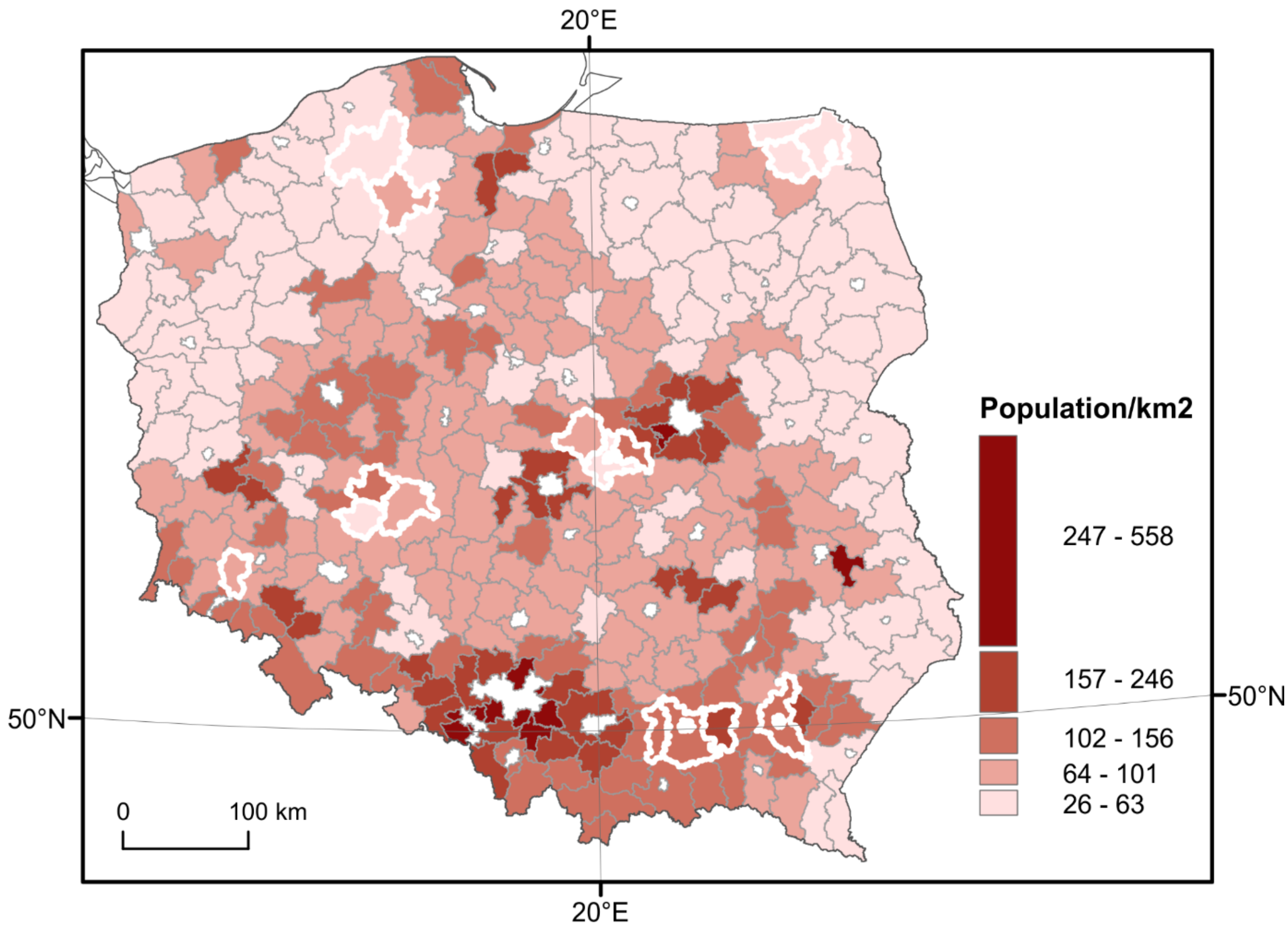
(Müller & Mouwes 1990; Karsznia & Weibel 2018)

Motivation

- most of the research concerns large scales (Feng et al. 2019, Zheng et al. 2021, Courtial et al. 2022)
- existing solutions are mainly implemented in raster mode, with machine learning models imported from computer vision, treating maps as images, rather than using the vector format commonly used in cartography to represent maps (Touya et al. 2019, Courtial et al. 2020)
- **extending the toolbox for small-scale mapping in vector mode** (Karsznia & Weibel 2018; Karsznia & Sielicka 2020, Karsznia et al. 2022)


AUTOCARTO 2022

Source data

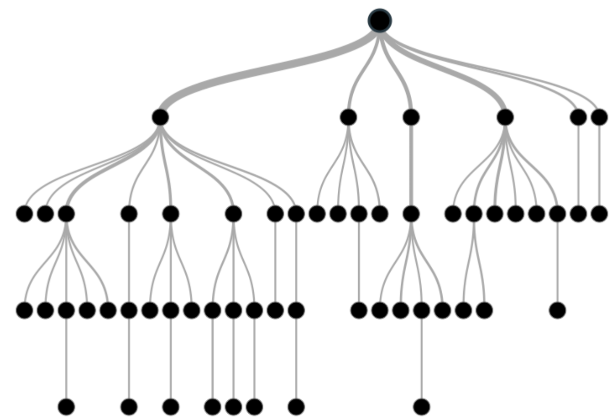

20°E
50°N
50°N
20°E
0
100 km
Population/km2
247 - 558
157 - 246
102 - 156
64 - 101
26 - 63

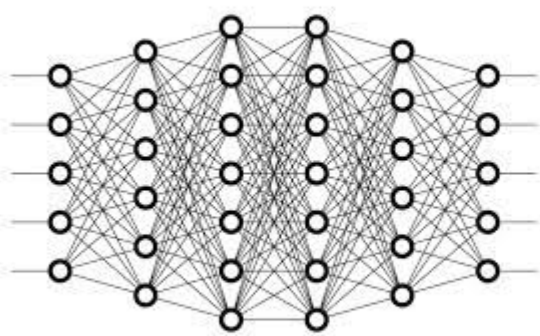
Scope of the research

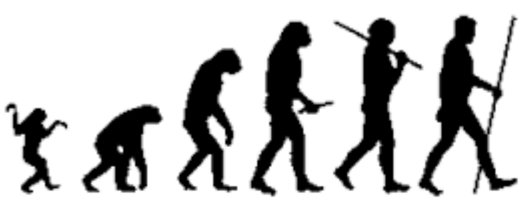
- implementing selected machine learning models as potential tools for automated settlement selection to create an optimal design for small scale maps
- testing **deep learning** (DL), **random forest** (RF), **decision tree** (DT), and **decision tree supported with genetic algorithms** (DT-GA) models to automatically classify settlements into selected and omited
- validation against the selection status acquired from the atlas map (taken as reference for evaluation)
- comparison of performance statistics across implemented models
- automatic settlement selection from GGOD 1:250 000 to 1:500 000

Research methods

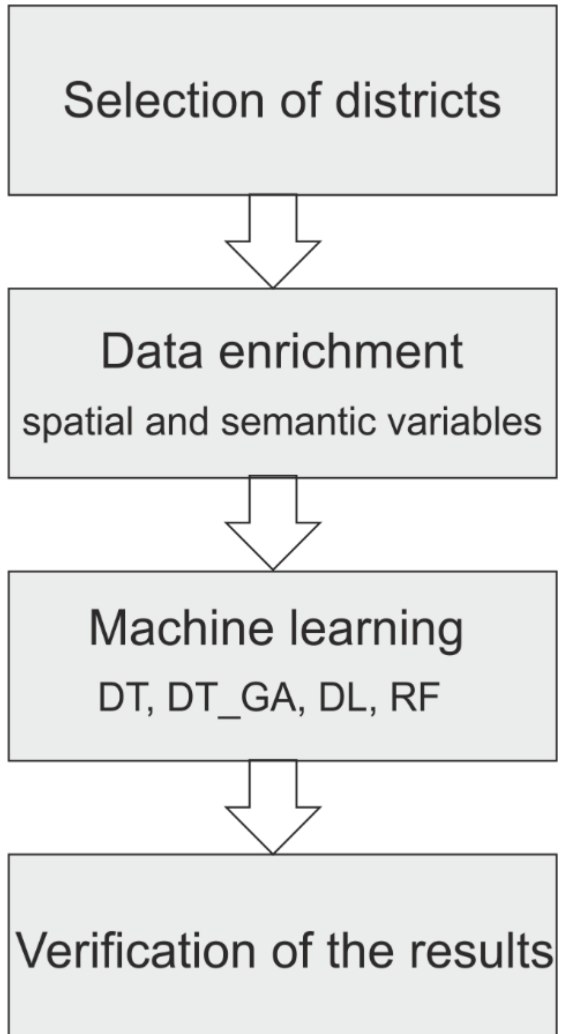
- Data Enrichment
- Machine Learning







Schema of the research methodology



Variables

Semantic

- population
- administrative status
- settlement type

Spatial

- population density per district

Semantic variables. Data Enrichment

- administrative function
- educational function
- cultural function
- health function
- industrial function
- financial function
- sacred function
- commercial function
- accommodation function
- monumental function
- other function

- administrative area
- built-up area
- residential areas
- service and commercial area
- industrial and storage area
- population density in residential areas
- number of railway stops
- number of airports
- number of ports
- number of at least district roads leaving the settlement
- number of communication nodes
- number of crossings of higher categories.

Karsznia I., Weibel R, 2018. Improving Settlement Selection for Small-scale Maps Using Data Enrichment and Machine Learning. Cartography and Geographic Information Science 45(2), pp. 111-127. DOI: 10.1080/15230406.2016.1274237

Spatial variables. Data Enrichment

- distance to nearest neighbor

Density of settlement:

- in district
- in square base field
- in hexagonal base field

- area of Voronoi diagram

Population density:

- in square base field
- in hexagonal base field.

Karsznia I., Sielicka K., 2020. When Traditional Selection Fails: How to Improve Settlement Selection for Small-Scale Maps Using Machine Learning. ISPRS International Journal of Geo-Information 9(4), 230; DOI: 10.3390/ijgi9040230

Performance

Learning model	Overall Accuracy in %	Precision	Recall	F1 score
Random Forest (RF)	**83.27**	82,96	92,85	**87.63**
Deep Learning (DL)	83.13	**83,85**	91,17	87.21
Decision tree with genetic algorithms (DT-GA)	81.69	80.67	93.58	86.65
Decision Tree (DT)	78.86	76.98	**95.62**	85.29

Karsznia I., 2023. Machine Learning and Geospatial Technologies. In: Alexander J. Kent and Doug Specht (Eds.) The Routledge Handbook of Geospatial Technologies and Society. ISBN 9780367428877

Variable weights

Administrative Area

Residential Area

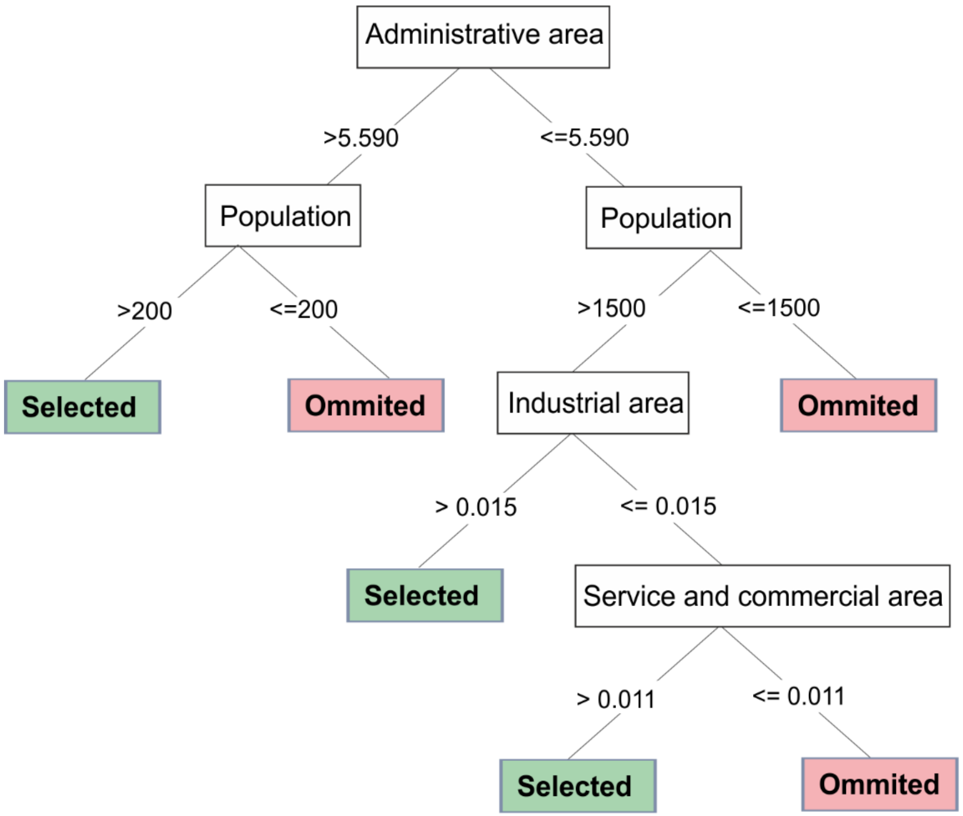
Population

Service and commercial area

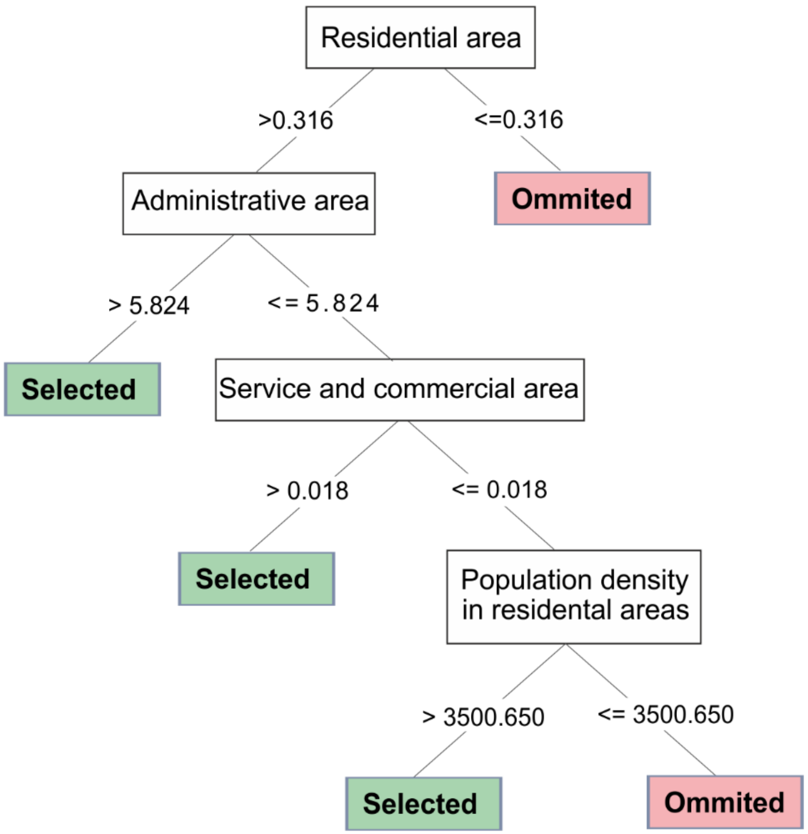
Population density

……..

Decision tree

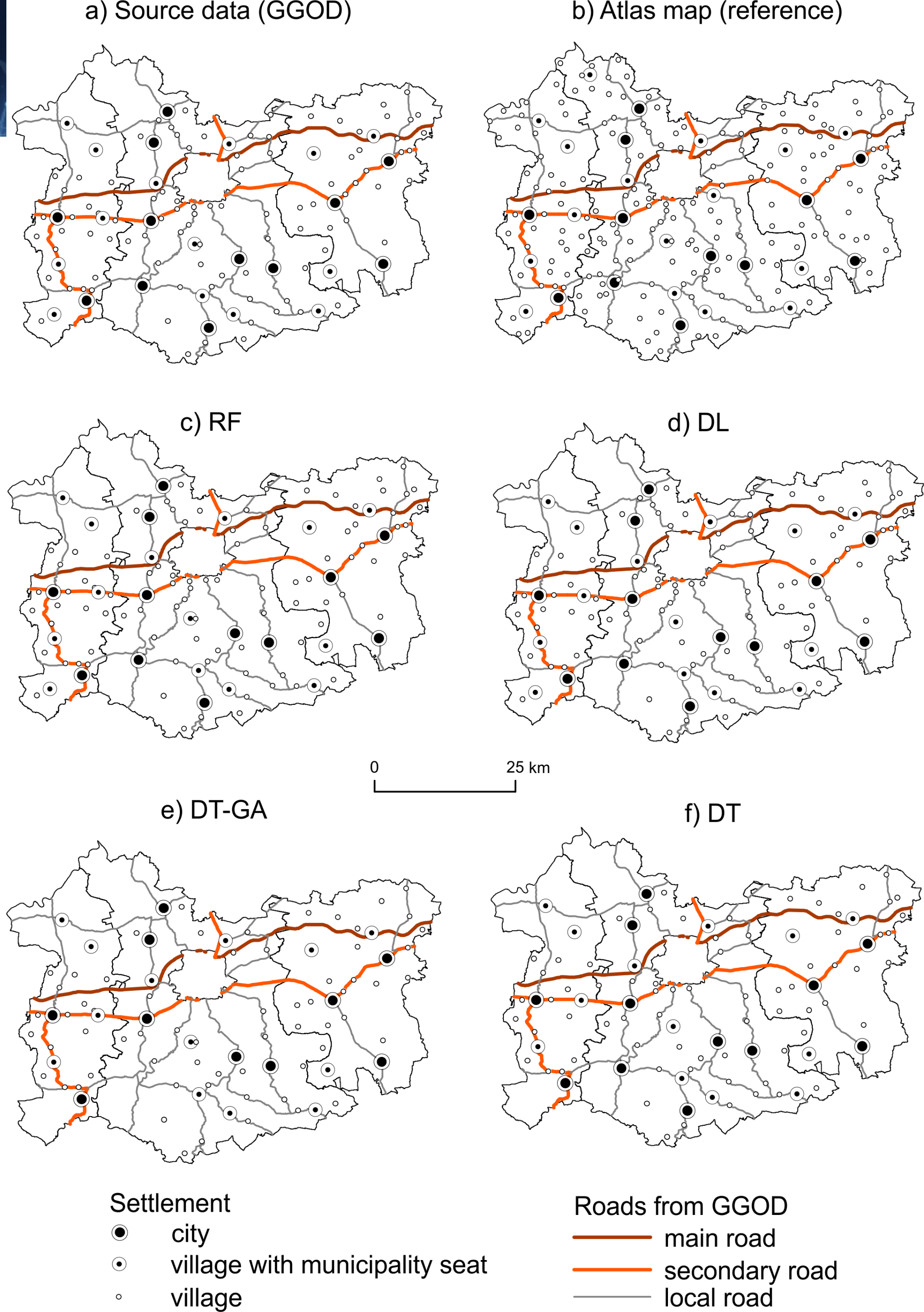


DT-GA

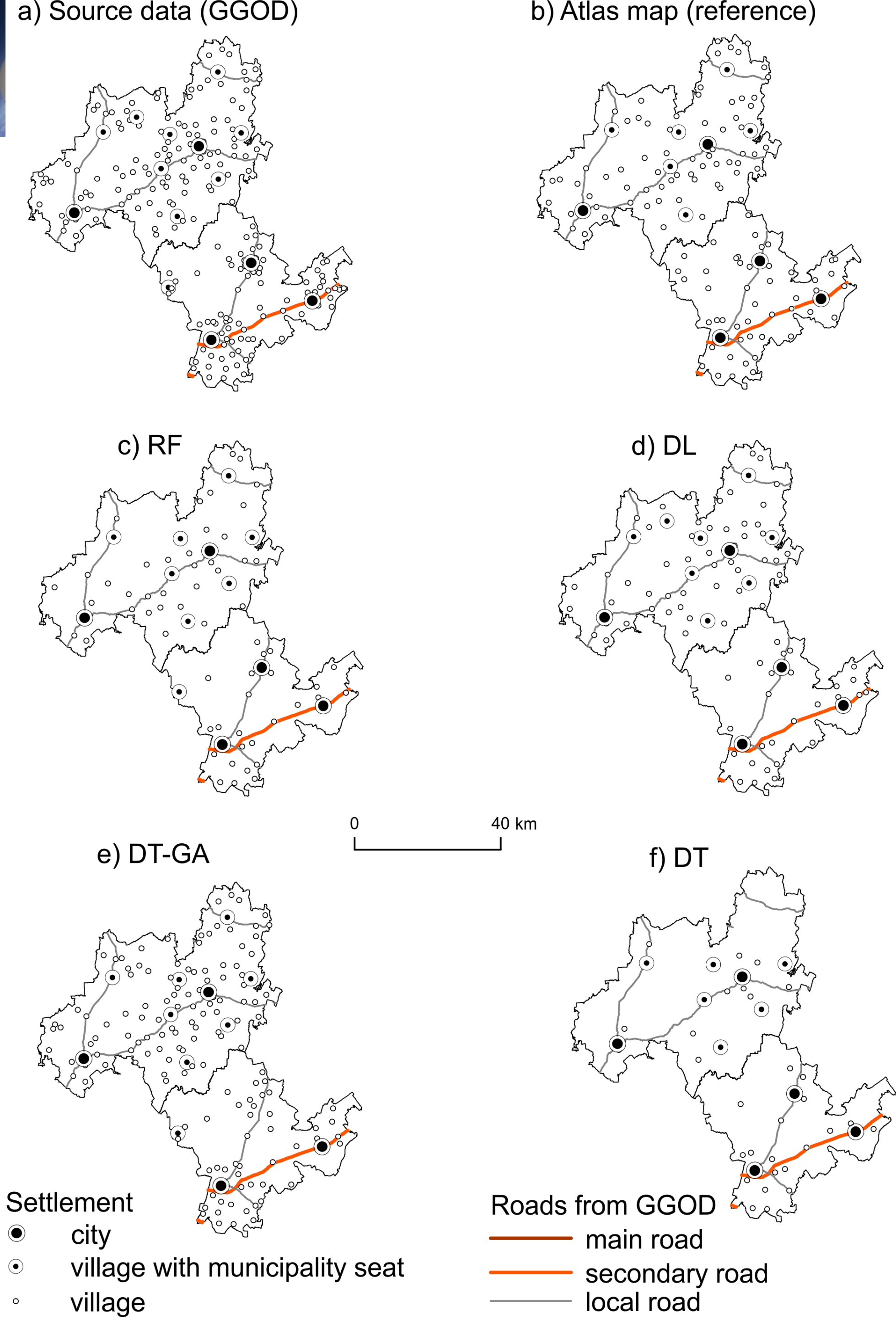


DT

Results


a) Source data (GGOD)
b) Atlas map (reference)
c) RF
d) DL
0
25 km
e) DT-GA
f) DT
Settlement
city
village with municipality seat
village
Roads from GGOD
main road
secondary road
local road

Results


a) Source data (GGOD)
b) Atlas map (reference)
c) RF
d) DL
e) DT-GA
f) DT
0
40 km
Settlement
city
village with municipality seat
village
Roads from GGOD
main road
secondary road
local road

Discussion

- The difference between the best and least performing model is 4,4%.
- RF and DL provide the results that are closest to the manual atlas map but the drawback is we do not get an intuitive and straightforward description of the decision process.
- DT or DT_GA do not give high performance results, but holistic decision trees are generated with explicit decision process.
- Population, population density and different area of settlement are decisive variables.
- RF and DL models performed better than DT_GA and DT, especially in terms of maintaining settlement density in highly populated density districts. In low populated ones the outperformance of RF and DL is not so evident.

Conclusions

- Very high accuracies from 78% (DT) to nearly 84% (RF).
- Goal to automatically generate the results that would be nearly equivalent to the manual map design has been achieved.
- DT-GA performed similar to RF and DL models plus it can be transformed into human-readable results – which is of key importance when the ultimate aim is to support rule formation for cartographic practice.
- Future works should evaluate automatic results qualitatively with the support of experienced cartographers (user studies).
- More quantitative evaluation measures should be proposed.

This research was funded by the National Science Centre, Poland, grant number UMO2020/37/B/HS4/02605, “Improving Settlement and Road Network Design for Maps of Small Scales Using Artificial Intelligence and Graph Theory”.

http://geoinformatics.uw.edu.pl/generalizationai/

NATIONAL SCIENCE CENTRE
POLAND

UNIVERSITY OF WARSAW

Research team: Izabela Karsznia, Primary Investigator, Karolina Wereszczyńska, Investigator, Albert Adolf, Phd student, Iga Ajdacka, Phd student.

International cooperation: Professor Robert Weibel, Professor Stefan Leyk, Professor Arzu Çöltekin.

University of Zurich[UZH]

University of Colorado Boulder

University of Applied Sciences and Arts Northwestern Switzerland

Thank you for your attention

Izabela Karsznia*, Robert Weibel, Stefan Leyk

* **i.karsznia@uw.edu.pl**, robert.weibel@geo.uzh.ch, stefan.leyk@colorado.edu

This research was funded by the National Science Centre, Poland, grant number UMO2020/37/B/HS4/02605, "Improving Settlement and Road Network Design for Maps of Small Scales Using Artificial Intelligence and Graph Theory".

http://geoinformatics.uw.edu.pl/generalizationai/

NATIONAL SCIENCE CENTRE
POLAND