

Vowel pronunciation in Swedish dialects analyzed with RuG/L04

Therese Leinonen



university of
 groningen

Workshop on Research Infrastructure
for Linguistic Variation

Oslo, 17 September 2009

Outline

- Introduction
- Dialectometric research
- Tools in RuG/L04
- Data
- Acoustic analysis
- Examples of analyses with RuG/L04

Introduction

- RuG/L04: free software for dialectometrics and cartography
- www.let.rug.nl/kleiweg/L04
- developed by Peter Kleiweg, University of Groningen
- Unix, Windows
- no graphical user interface, yet

Dialectometric research

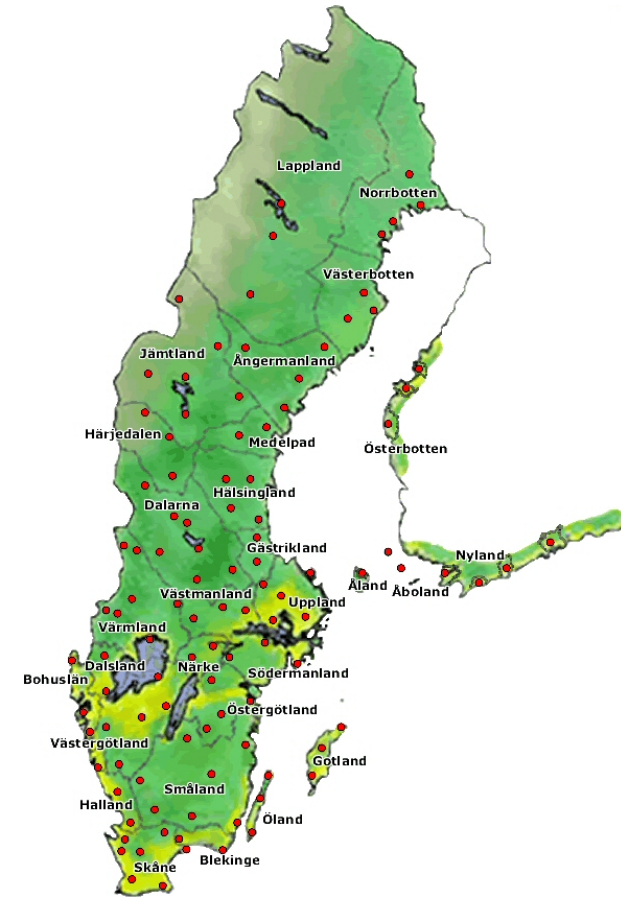
- dialectometry = measuring dialects
- aims: finding dialect areas and describing dialect continua
- dialectometry emphasizes the aggregate analysis and is data-driven
- statistical methods are used for classifying dialects and exploring dialect continua

Tools in RuG/L04

- dialectometric tools, distance measures based on transcribed dialect data:
 - Levenshtein distance (string edit distance)
 - Gewichteter Identitätswert
- statistical tools:
 - hierarchical clustering
 - multidimensional scaling
 - R interface
- cartography:
 - web tool for acquiring geographic data with Google Earth: data points and borders of the studied area
 - tools for displaying dialectometric results

Data

- SweDia (swedia.ling.gu.se): project carried out by the universities of Lund, Stockholm and Umeå 1998-2001
- 105 sites in Sweden and Swedish-language parts of Finland
- 12 speakers from each site: 3 elderly women, 3 elderly men, 3 young women, 3 young men
- vowels elicited with existing mono- or bi-syllabic words with the target vowel in a coronal consonant context
- 19 words of which the vowels cover the standard Swedish vowel space: *dis, disk, dör, dörr, flytta, lass, lat, leta, lett, lott, lus, låt, lär, lös, nät, sot, särk, söt, typ*



Acoustic analysis

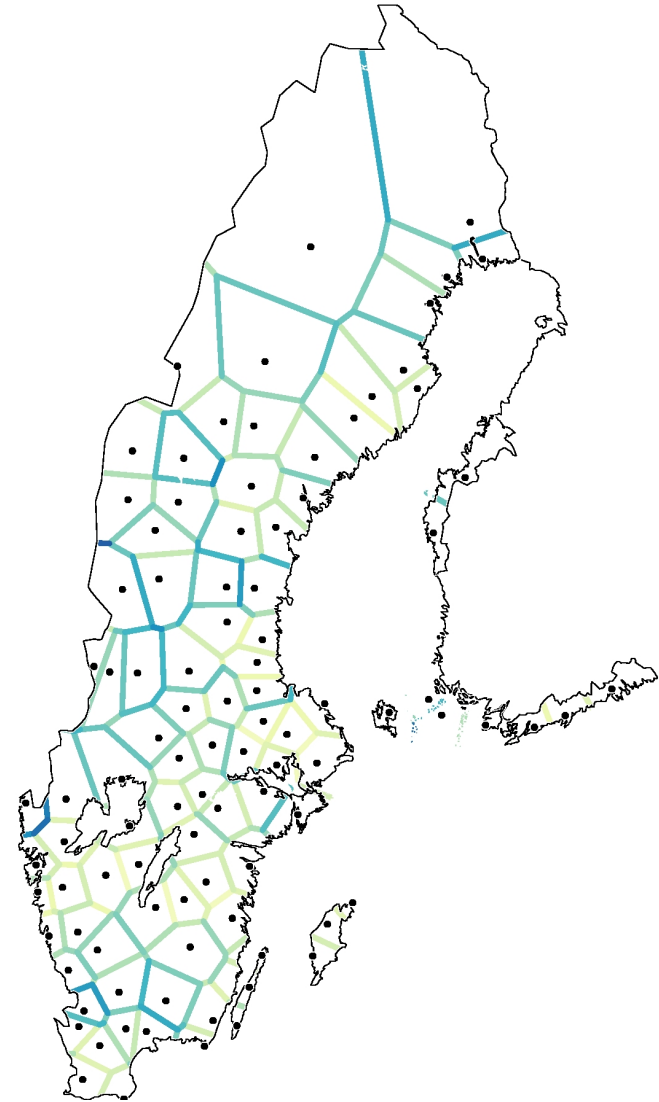
- Principal component analysis (PCA) of Bark-filtered vowel spectra (Pols et al., 1973; Jacobi, 2009)
- two components used as acoustic measure of vowel quality, high correlation with formants
- each vowel measured at nine points within every vowel segment (starting at 25 % and ending at 75 % of the vowel duration)
- the linguistic distance per vowel between any two varieties is calculated as the Euclidean distance of the acoustic parameters
- Euclidean distance, where i ranges over the nine sampling points:

$$distance(x, y) = \sqrt{\sum_{i=1}^9 ((PC1_{xi} - PC1_{yi})^2 + (PC2_{xi} - PC2_{yi})^2)}$$

- the distance between varieties is the average distance of the 19 vowels

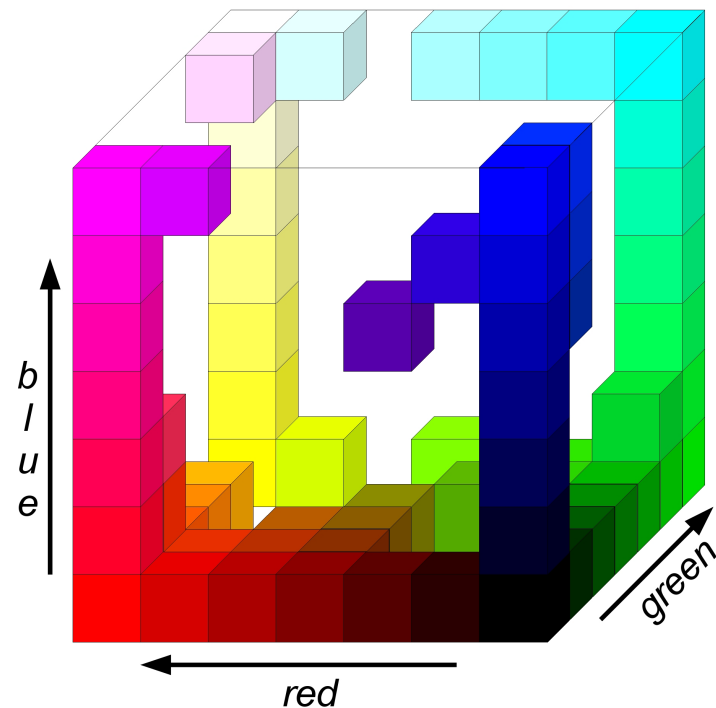
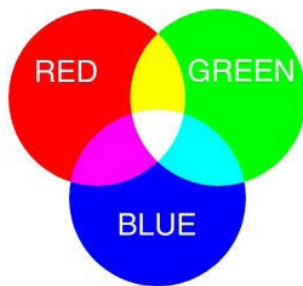
RuG/L04: mapdiff

- draws a map of differences between neighbors
- darker lines indicate a larger difference

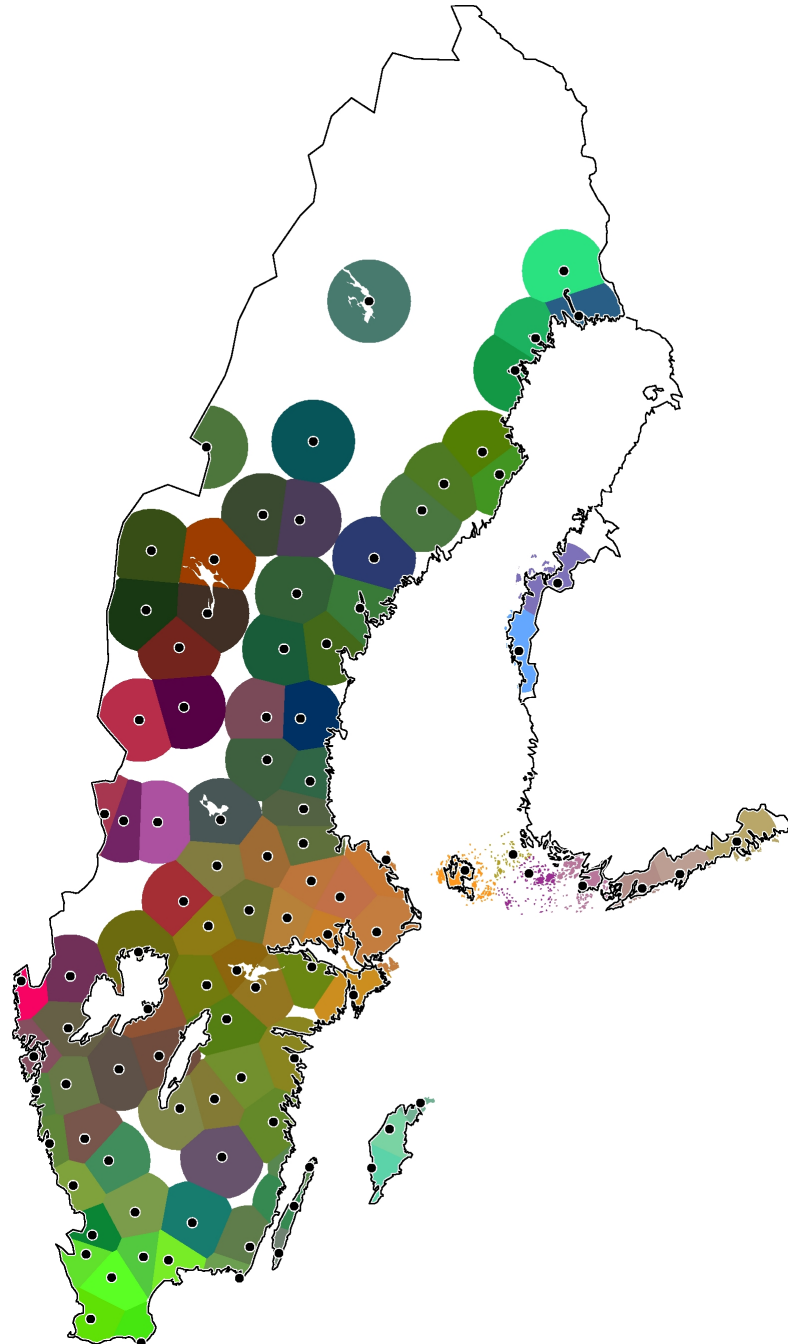


Multidimensional scaling

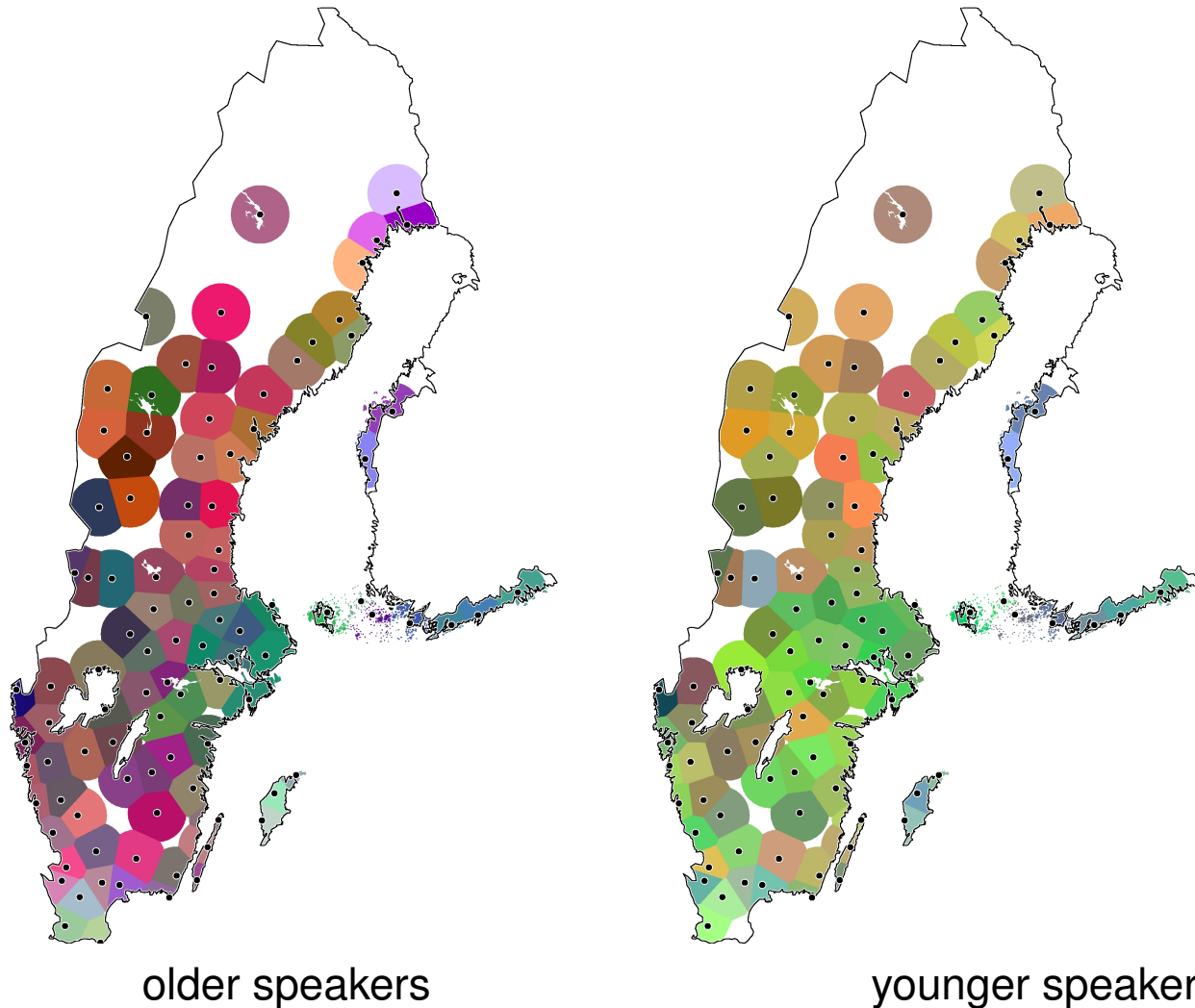
- method for visualizing and exploring similarities/dissimilarities in data
- with given pair-wise distances positions in a low-dimensional space can be assigned to data points
- 3 dimensions visualized in **red**, **green** and **blue** → maps where the language varieties form a continuum (Heeringa, 2004)



RuG/L04: maprgb



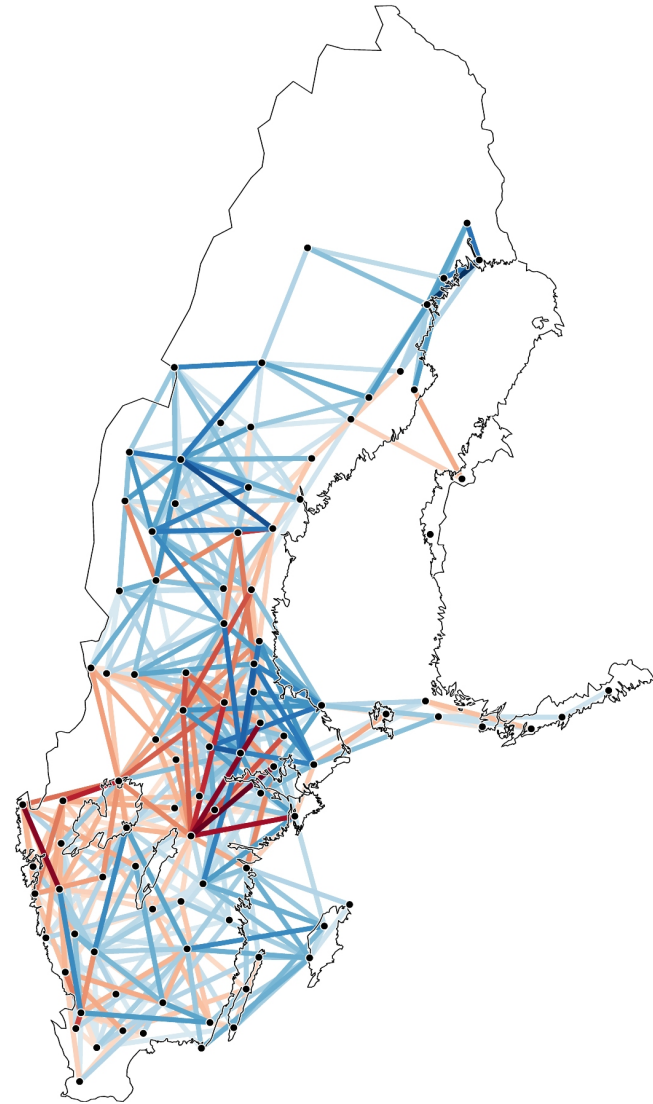
RuG/L04: maprgb



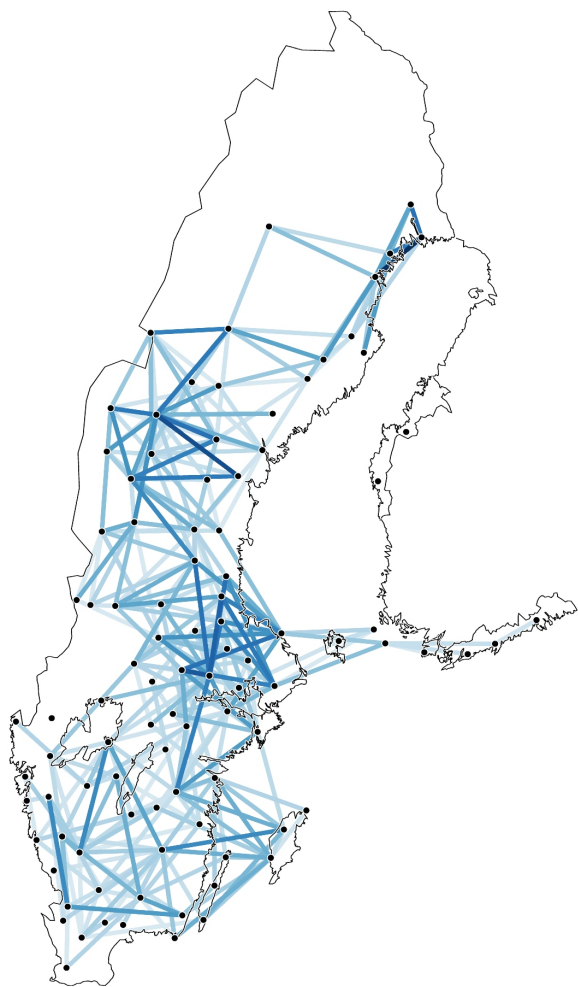
- significantly shorter distances between geographic varieties among younger speakers than between older speakers ($t(96) = 8.4, p < 0.001$)

RuG/L04: maplink

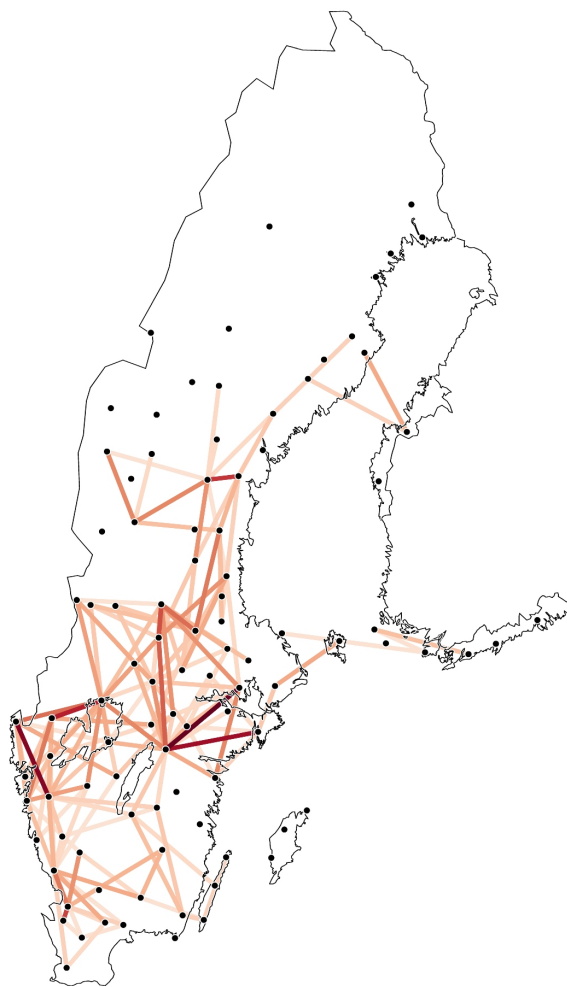
- for each pair of sites: measure the distance of older and younger speakers separately
- $distance(older_i, older_j) > distance(younger_i, younger_j)$
convergence(blue)
- $distance(older_i, older_j) < distance(younger_i, younger_j)$
divergence(red)
- darker lines indicate larger differences



RuG/L04: maplink



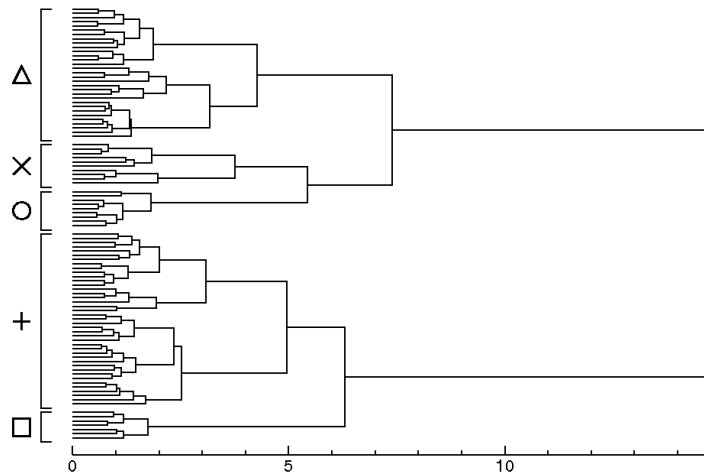
convergence



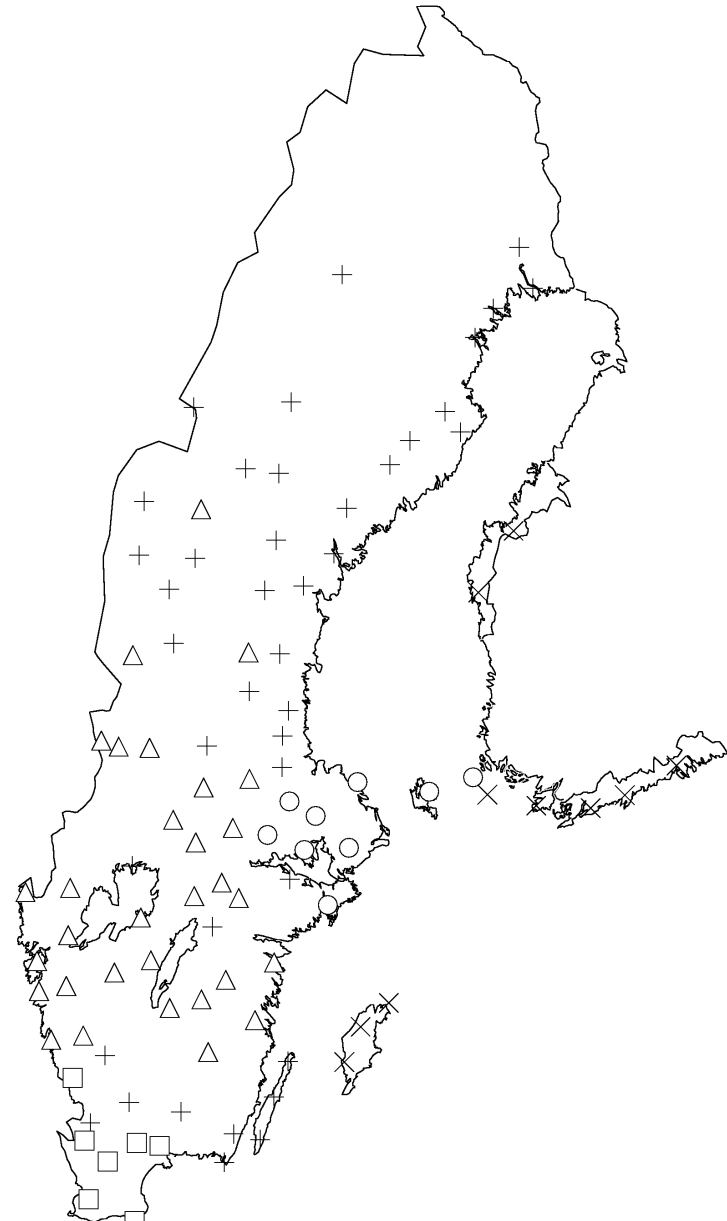
divergence

RuG/L04: mapclust

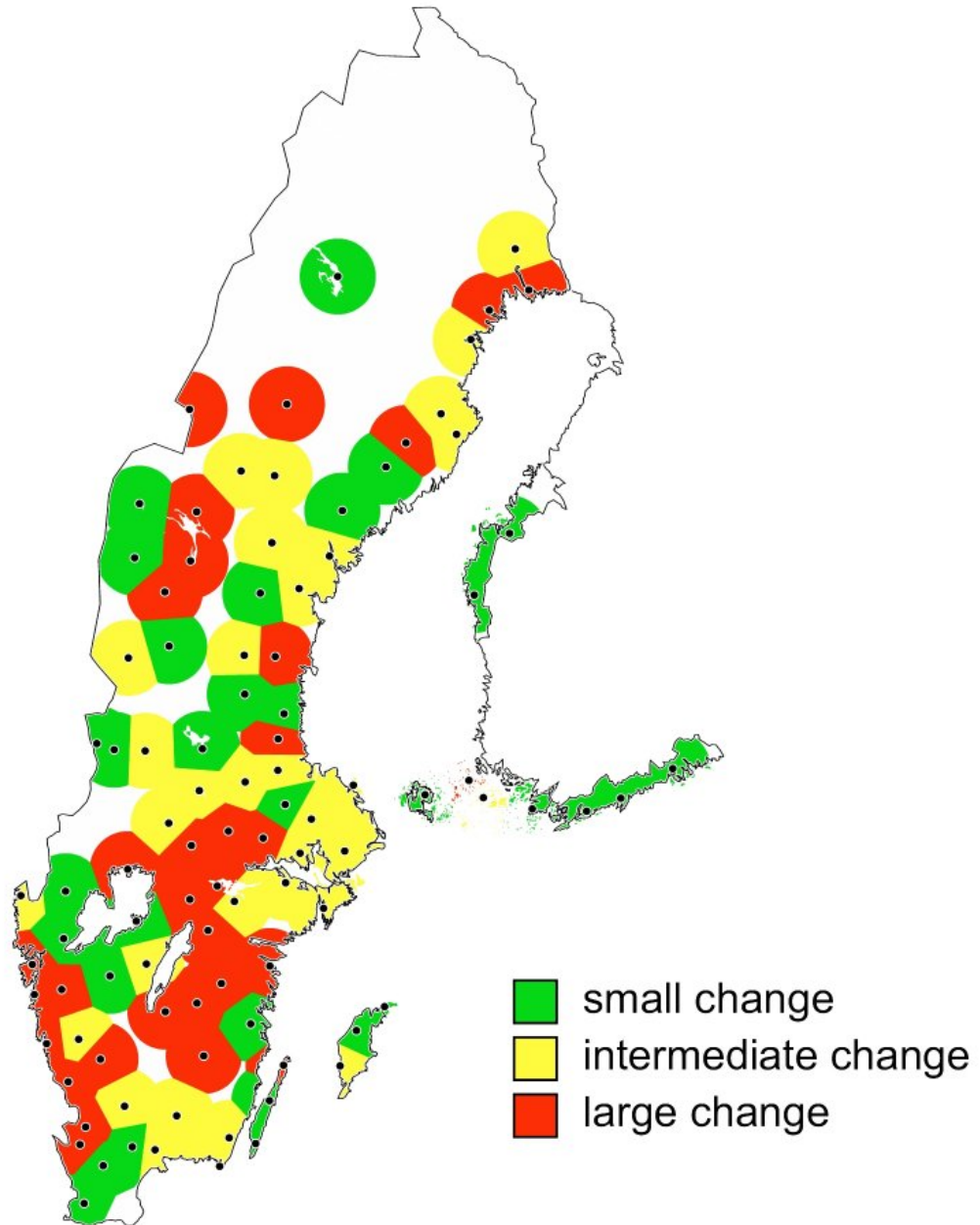
- displays groupings in data by using colors, patterns, numbers or symbols
- groupings based on hierarchical clustering (RuG/L04) or manually indexed data



5 clusters using Ward's method



RuG/L04: mapclust



Thanks to:

Peter Kleiweg for making the software available:

<http://www.let.rug.nl/kleiweg/L04/>

The SweDia project for making the data available

The dialectometric research group in Groningen
for comments and discussion

YOU
for listening!

References:

Bruce, G., Elert, C.-C., Engstrand, O. and Eriksson, A. (1999), Phonetics and phonology of the Swedish dialects: a project presentation and a database demonstrator, *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS 99)*, San Francisco, pp. 321–324.

Heeringa, W. (2004), *Measuring Dialect Pronunciation Differences using Levenshtein Distance*, PhD thesis, Rijksuniversiteit Groningen.

Jacobi, I. (2009), *On Variation and Change in Diphthongs and Long Vowels of Spoken Dutch*, PhD thesis, Universiteit van Amsterdam.

Nerbonne, J. (2009), Data-driven dialectology, *Language and Linguistics Compass* **3**(1), 175–198.

Pols, L. C. W., Tromp, H. R. C. and Plomp, R. (1973), Frequency analysis of Dutch vowels from 50 male speakers, *Journal of the Acoustical Society of America* **53**, 1093–1101.

Tabachnik, B. G. and Fidell, L. S. (2007), *Using Multivariate Statistics*, 5th edn, Pearson.