EUROPEAN FORUM FOR GEOSTATISTICS 2012

Prague Conference

24-26 October 2012
Prague, Czech Republic

„If you cannot describe it, you cannot manage it"

CZECH STATISTICAL OFFICE

EUROPEAN FORUM FOR GEOSTATISTICS

CAGI

EUROPEAN FORUM
FOR GEOSTATISTICS 2012
Prague Conference
24-26 October 2012
Prague, Czech Republic

Gridded register-based data for detail spatio-temporal monitoring and modelling.
Vision of data harmonisation and integration

Jiří HORÁK
VSB-Technical University of Ostrava,
Czech Association for Geoinformation
Czech Republic

Registers

- Increasing number of central and local registers, maintained by public administration bodies and service organisations
- Population, buildings, unemployed, social allowances, economical entities, health, crime
- Similarity - records geocoded by addresses
- Harmonisation, geocoding and aggregation of primary data

Harmonisation Aggregation

- Unification of structures, address description
- Identification and replacement of abnormal names
- Geocoding
- Aggregation - bottom-up approach
- Geocoded data aggregated to various spatial units:
  - address points
  - grid cells
  - administrative zones
  - localities (polygons of interest)

Main issues

„Live“ registers:

- Continuous process of data editing
- Adding new records, modification, deleting of some older records
- Data selection of one date (i.e. data relevant to the end of the year) may differ when we repeat the selection later
- Change of the number of valid records, some addresses etc.

number of registered unemployed	to date	date of export
23500	31. 12. 2011	10. 1. 2012
23700	31. 12. 2011	25. 3. 2012

Main issues

Address selection:

- more than one address per object (i.e. contact address + permanent address)
- selection of the most appropriate address (i.e. for integration with another register)

Address matching:

- differences in address recording
- part of the address is in free text without standardization
- missing location identifiers
- differences in applied code lists

- Almost 100% geocoded, but with different accuracy
- More problems in localities with rapid changes (houses destroyed due to undermining)

Main issues

- **Consistency**
  - differences in time, semantics, and methodology leads to inconsistencies when joining data from more registers
- **Confidentiality**
  - data filtering (i.e. threshold 25 persons)
  - random adding or subtracting small value (Kraak, Ormelling, 1996)
  - data scramble (Batista et al. 2011)
  - data rounding (Quinn 1996)
  - kernel maps
- **Scarcity**
  - i.e. missing data required for calculation of traditional indicators (usually we do not have data about economic activity of individuals in population registers)

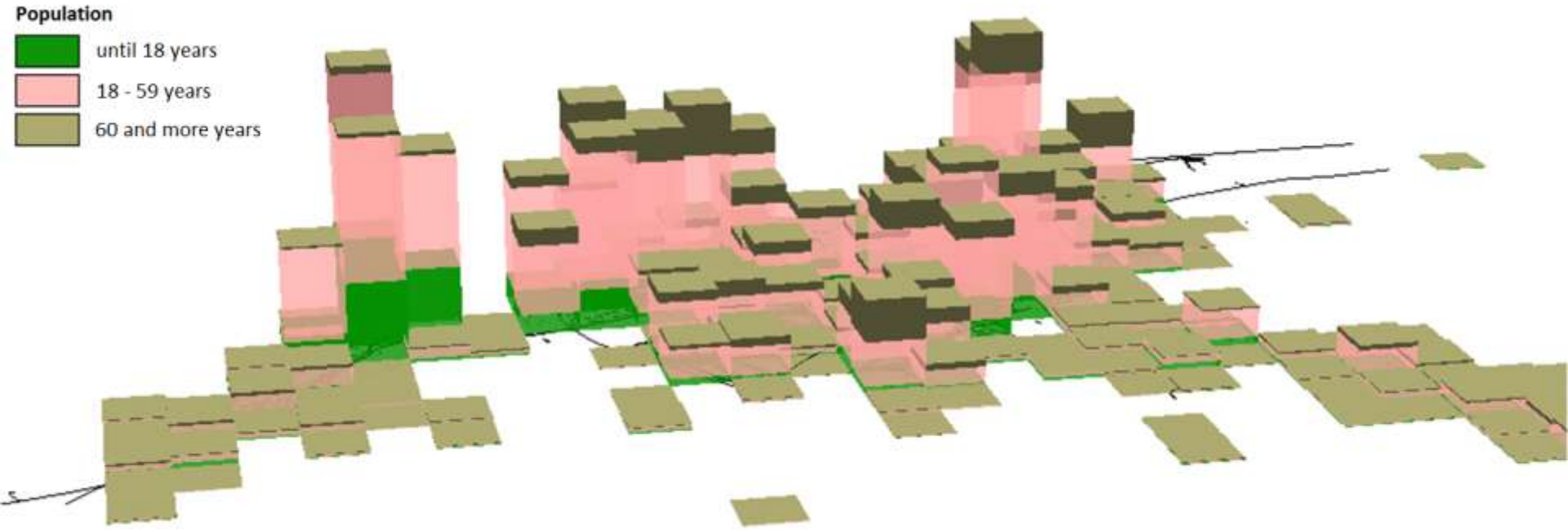
Social inequality

- Rapid growing of social inequality
- Mirroring in settlements, socially excluded localities
- Need for continuous scanning (monitoring) of local situation
- Detailed monitoring of small localities
- Creation of early warnings
- Facilitation of timely and well-targeted actions from local institutions
  - prevention of crime, law enforcement
  - prevention of insolvency
  - health prevention
  - social street workers
  - new social programmes

Data sources

- Register of Enumeration Districts and Buildings (Czech Statistical Office)
- Registers of local governments (population – sex, age; dwellings, bad payers)
- Registers managed by Ministry of Social affairs – unemployment r., social allowances, employers
- Other – crime, health, etc.

Population register



Age structure

Nový Jičín, 1.12.2001

200x200 m

Contrasting social conditions

Age distribution (1.10.2009)

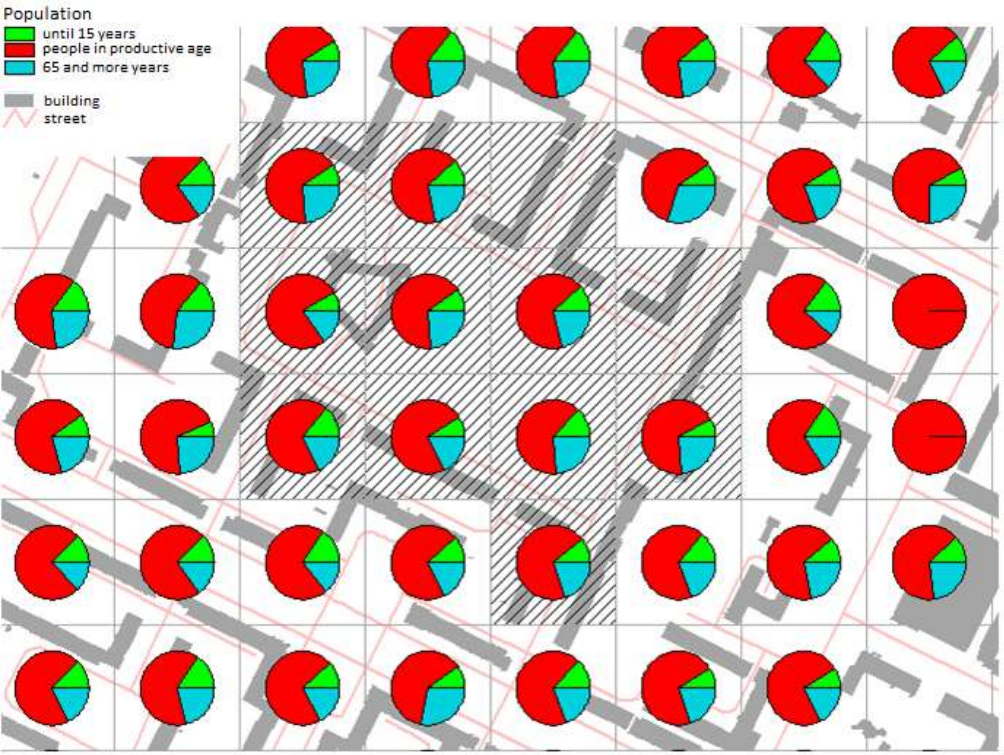
city (Ostrava) x locality (Zadni Privoz)

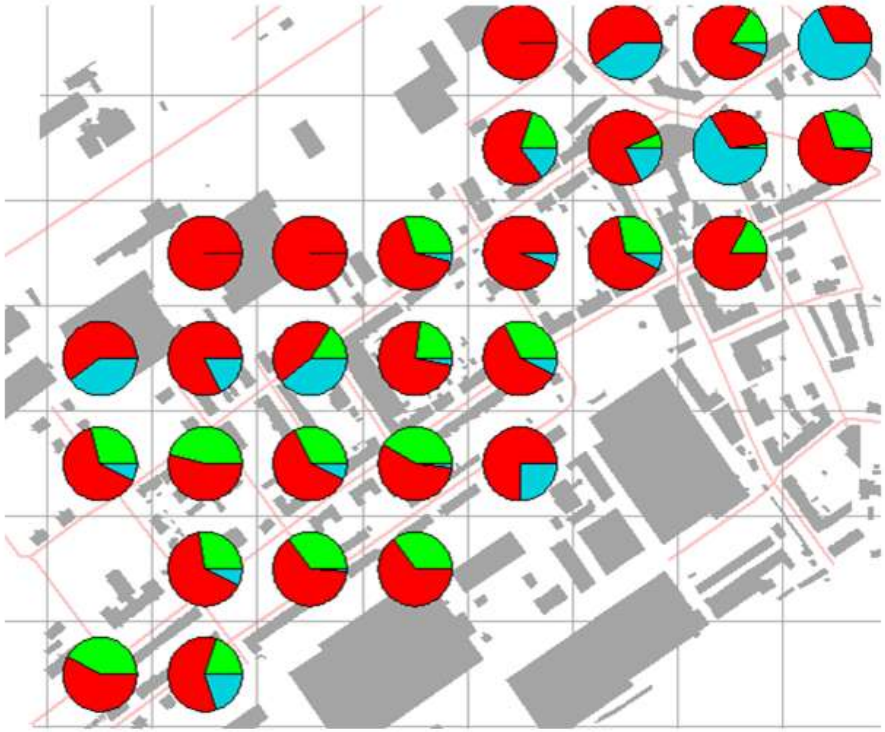
Age Index 148

Age Index 11

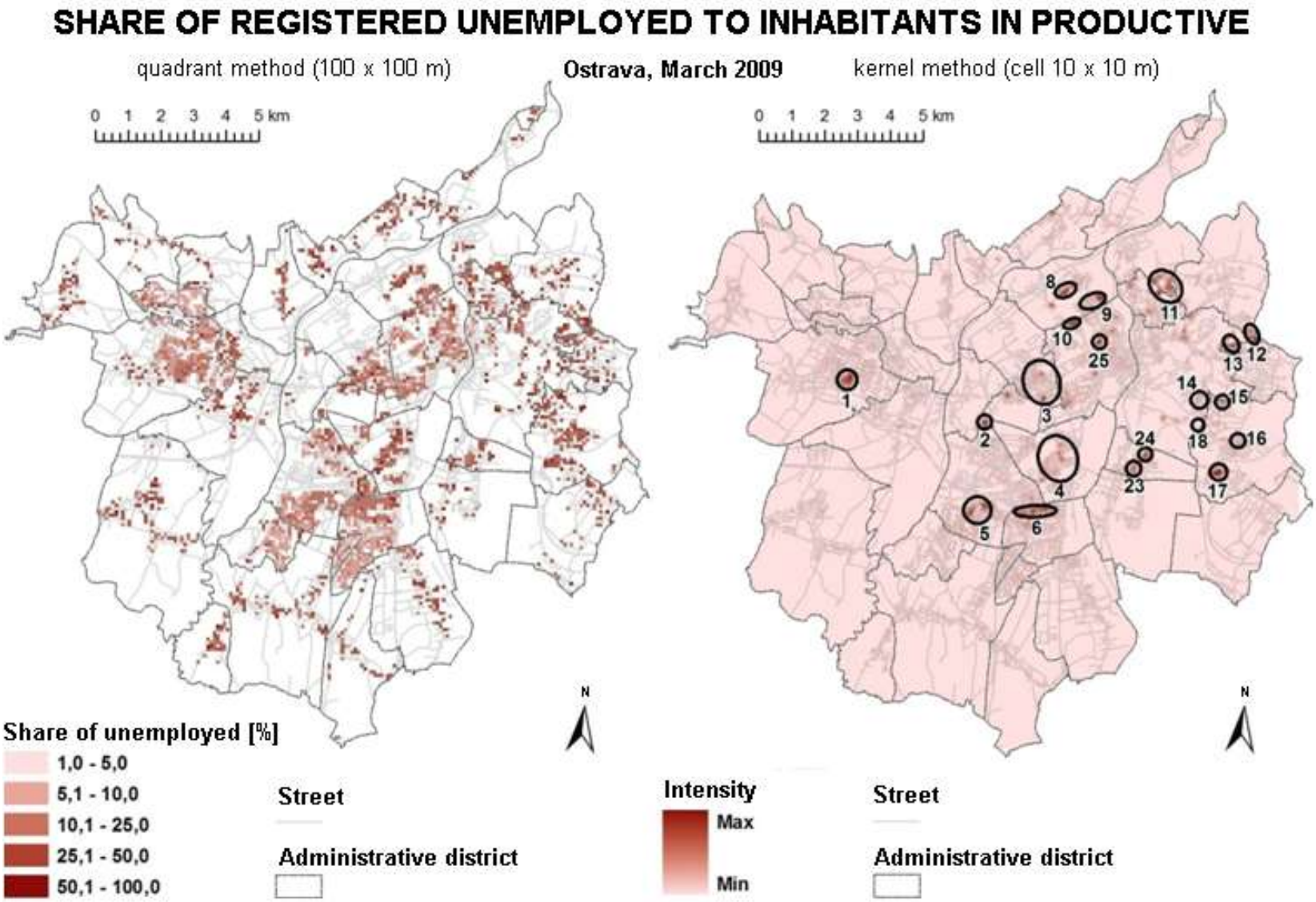
Gridded age distribution

Homogeneous and variable



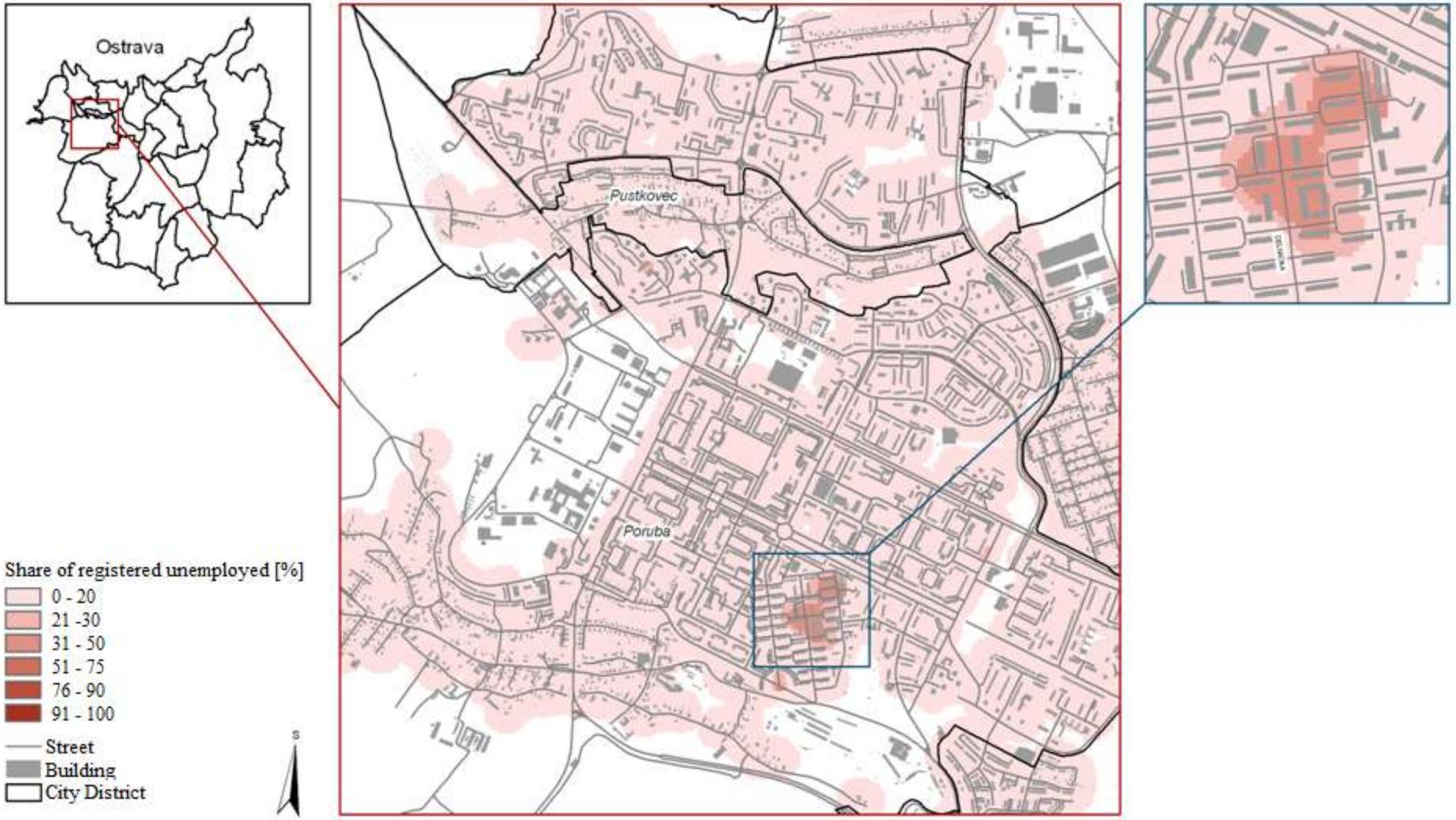


Hot spots


SHARE OF REGISTERED UNEMPLOYED TO INHABITANTS IN PRODUCTIVE
quadrant method (100 x 100 m)
Ostrava, March 2009
kernel method (cell 10 x 10 m)
0 1 2 3 4 5 km
0 1 2 3 4 5 km
N
N
Share of unemployed [%]
1,0 - 5,0
5,1 - 10,0
10,1 - 25,0
25,1 - 50,0
50,1 - 100,0
Street
Administrative district
Intensity
Max
Min
Street
Administrative district
1
2
3
4
5
6
8
9
10
11
12
13
14
15
16
17
18
23
24
25

SHARE OF REGISTERED UNEMPLOYED TO INHABITANTS IN PRODUCTIVE AGE

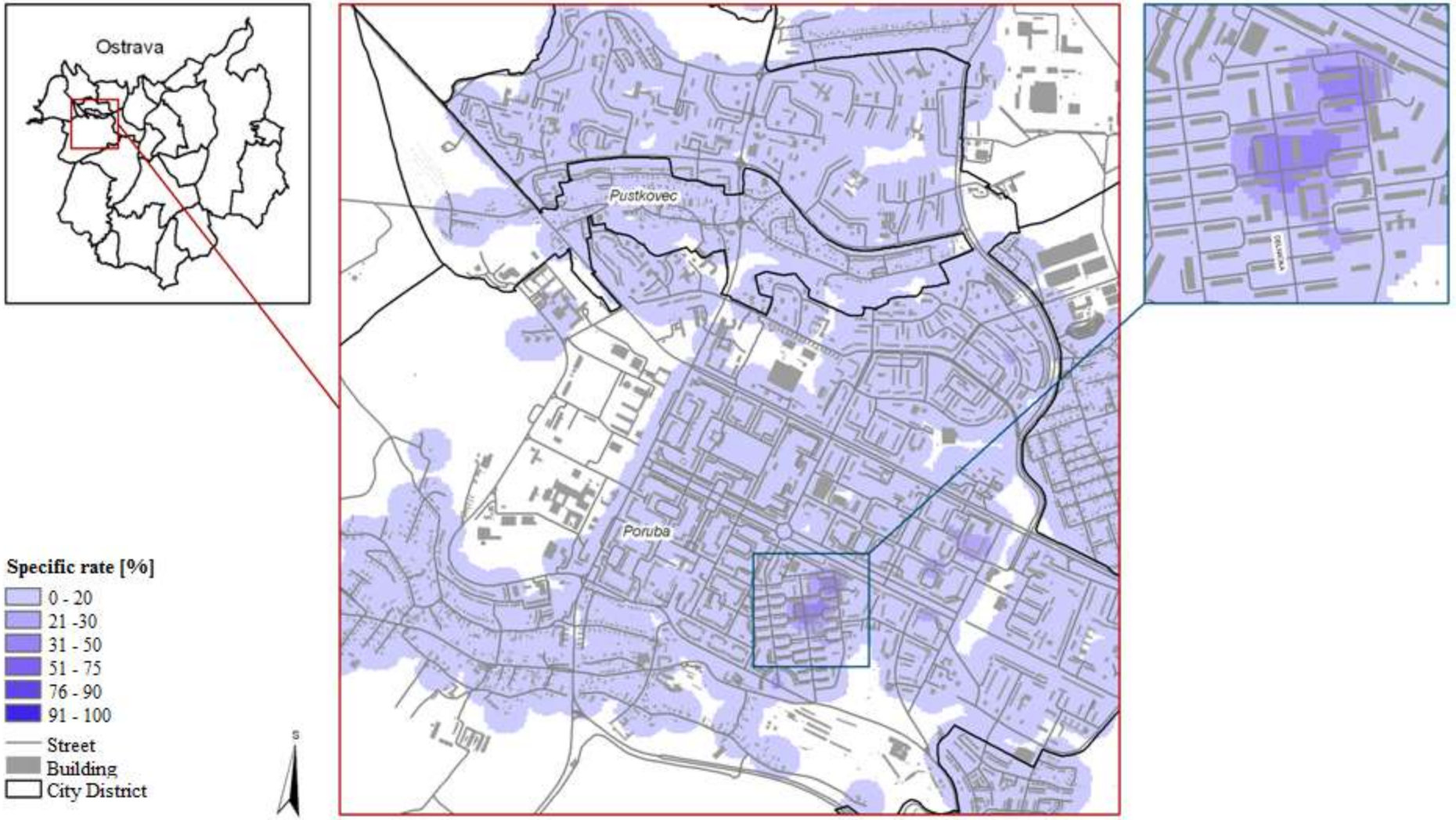
City District Poruba and Pustkovec, 31. 10. 2006



Ing. Olga KOTLÍKOVÁ
Ostrava 2007
Data z Úřadu práce a Magistrátu města Ostravy

INTENSITY OF SPECIFIC RATE OF UNEMPLOYMENT UNTIL 25 YEARS

City District Poruba and Pustkovec, 31. 10. 2006



Ing. Olga KOTLÍKOVÁ
Ostrava 2007
Data z Úřadu práce a Magistrátu města Ostravy

Register of unemployed

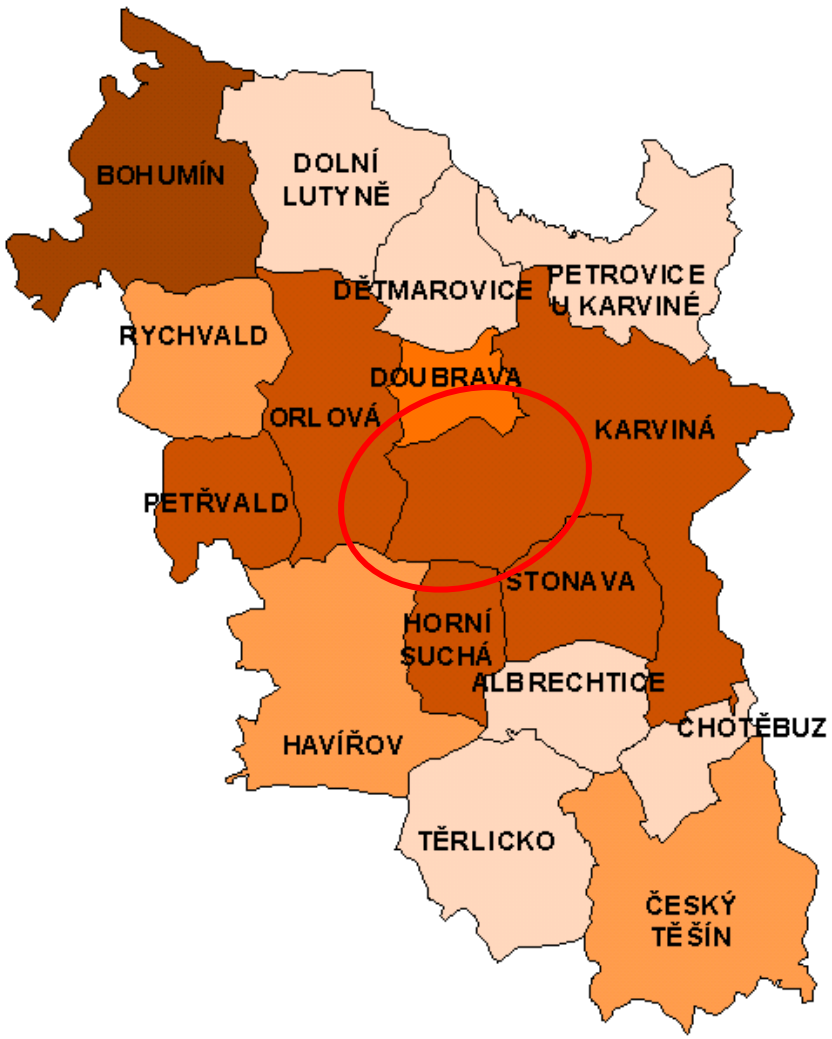
- Share of registered unemployed to inhabitants in productive age

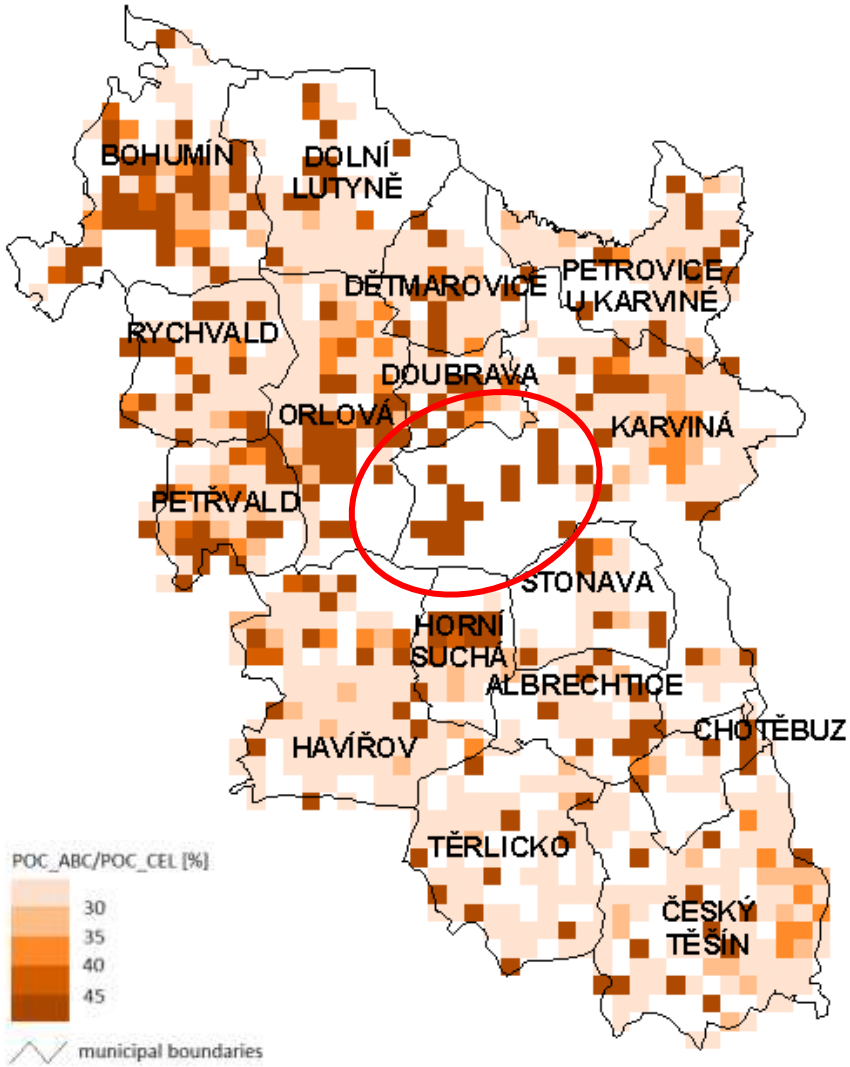
Lower consistency

- Rate of unemployment for young people (under 25)
- Rate of unemployment for older people (above 50)
- Share of unemployed with basic education
- Share of health-handicapped unemployed
- Share of long-term unemployed persons

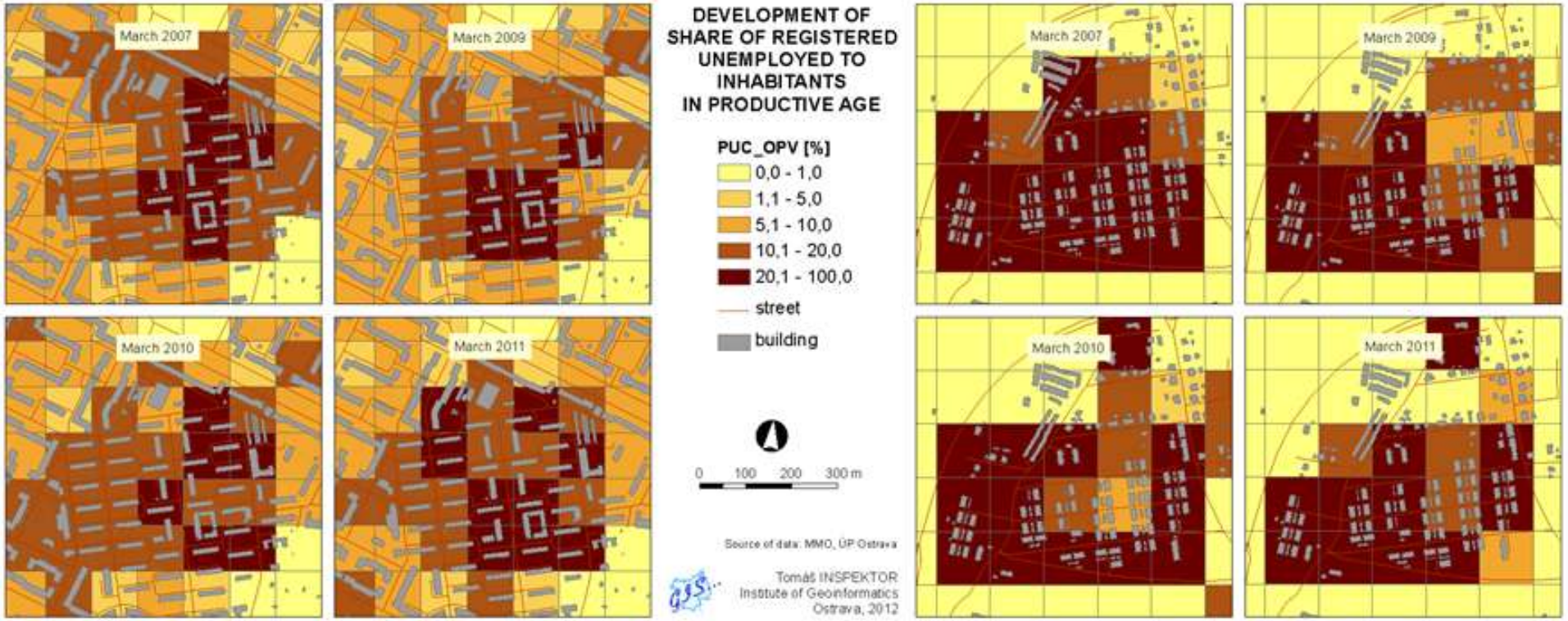
Higher consistency

Unemployment (low edu.) uneven distribution



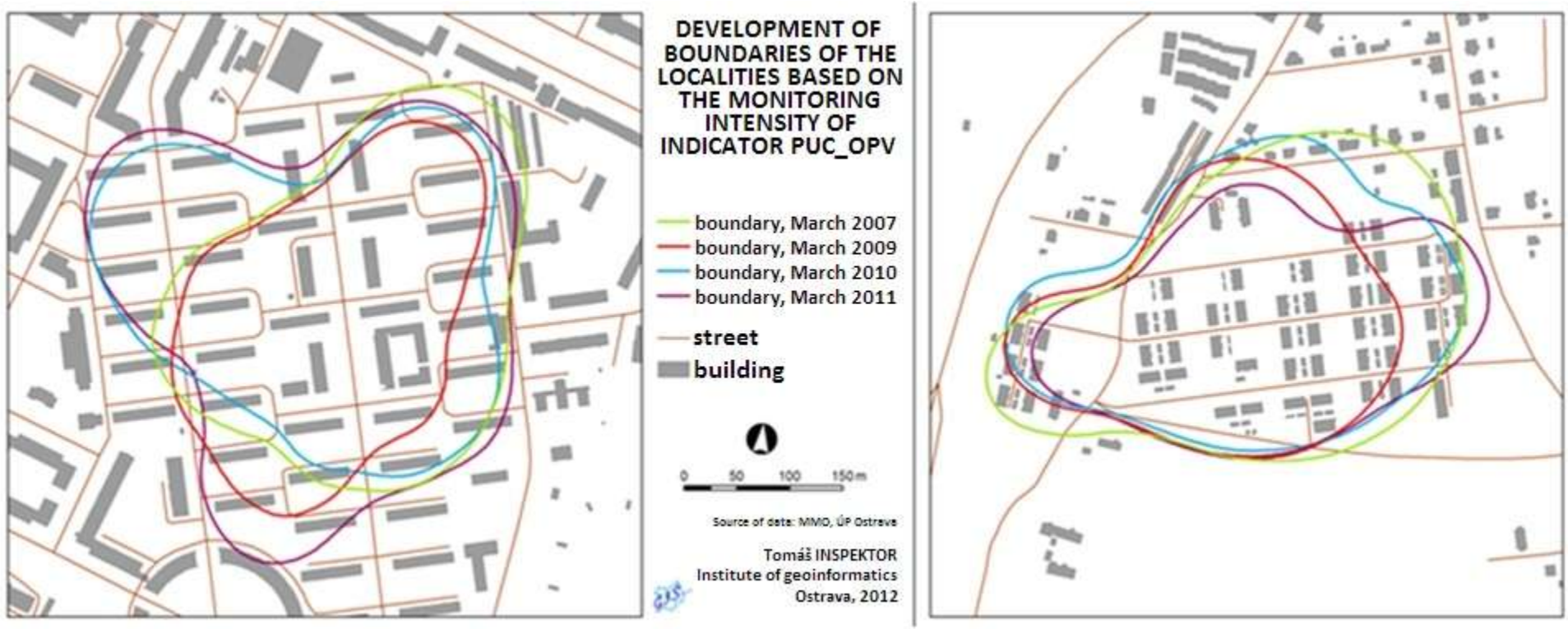


Changes inside localities



March 2007, 2009, 2010 a 2011

Isolinies of intensity

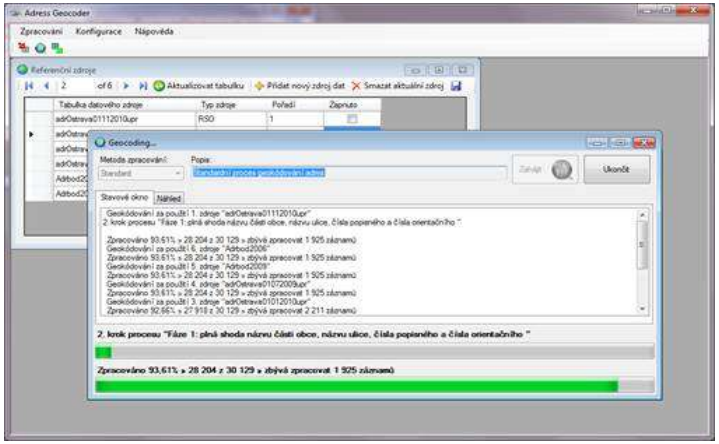


March 2007, 2009, 2010 a 2011

Optimisation of grid parameters

- Necessary for local studies
- Grid size, event. orientation
- According to required minimal preserved details
- Methods helping to quantify the impact of increased data aggregation:
  - segregation measures (Gini etc.) - the maximal segregation is reached in case of perfect separation of different categories
  - correlation measures (R^2, Moran I etc.) - finding equilibrium among influences of ecological fallacy, small bases of data and increased correlation due to the enlarged grid size

Utilization



- Ministry of Social Affairs
  - 2004 first study for data sources
  - 2011 SW for geocoding
  - 2012 reconstruction of IS, assumption of implementation new spatial level for analysis (grid, kernel), i.e. spatial accumulation of various forms of social allowances
- Local governments
  - quantitative measures (indicators) of current situation,
  - identification of trends
- Ministry of Interior
  - ancillary maps for crime mapping etc.

Registers & Census

- Census preparation
- Consistency of census data
- Data disaggregation
- Intercensal estimations
  - the population register monitors the current overall state of population, timing of migration, and enable to discover sudden changes – make critical points suitable to adjust data population and break continual functional development of population in intercensal period

Vision of data harmonisation and integration

- Continuously updated basic registers (especially addresses, buildings) adhering the reality
- All registers of the public sector linked with and based on basic registers to minimise harmonisation problems
- Improved geocoding systems to minimise data loss, minimise location uncertainty, speed up and facilitate processing
- Consistent integration data across registers – making more operational
- Self-control of data consistency inside and across registers
  - primary - integrity restrictions to data
  - secondary - sophisticated solutions of revealed inconsistencies.
- Appropriate control of data confidentiality (not excessively restricting data processing and analysing)

Thank you for your attention!

jiri.horak@vsb.cz