Efficient statistical sample designs in a GIS for monitoring the landscape changes

Elisabetta Carfagna[1]
elisabetta.carfagna@unibo.it

Patrizia Tassinari[2], Maroussa Zagoraiou[1],
Stefano Benni[2], Daniele Torreggiani[2]

[1]Department of Statistical Sciences, University of Bologna

[2]Department of Agricultural Economics and Engineering
Spatial Engineering Division, University of Bologna
in Statistical Methods for the analysis of large data-sets, Book of short papers, Cleup, ISBN:9788861294257, SIS Conference 2009, University G. D'annunzio Chieti-Pescara, Pescara, 23-25 September 2009

Study area - sampling frame

- **Study area:**
  - 787 km^2 in eastern part of Bologna province which comprises the group of ten municipalities known as the "Nuovo Circondario Imolese"
- **Within a GIS environment:**
  - Analysis of the whole area
  - Detailed analysis of change of buildings on areal photos on a sample of areal units - acquisition (for old photos digitisation), elaboration and so on
- **Sampling frame:**
  - Division borders of the most recent population and housing census (Istat, 2001)
  - Area frame with irregular physical boundaries

Parameter to be estimated

- **Parameter to be estimated: change of building cover density:**
  - difference between the area covered by buildings in 2005 and in 1975 divided by the land area
- **Data acquisition and in depth data analysis time consuming**
  - Thus very efficient sample design
  - Stratification
  - Optimal allocation

Strata combinations of land-use/land-cover class and land suitability classes

Land-use/land-cover classes

1. Fabric of human settlement
2. Arable crops
3. Orchards, vineyards, vegetable gardens, plant nurseries, greenhouses
4. Forest-pasture land, areas with sparse or absent vegetation, wetlands
5. Water bodies and water courses

Land suitability classes

a Level land well suited for agricultural use

b Level land less suited for agricultural use than class "a"

c Hilly regions intermediately suitable for agricultural use

d Hilly regions moderately suitable for agricultural use

e Areas with low suitability for agricultural use

DEFF very disappointing = 0.9991

Sub-stratification

Land-use/land-cover classes

Land suitability classes

Time of urbanisation:

’ already urbanized in 1975

Kind of final destination:

h predominantly residential (1-a.h)

p productive (1-a.p)

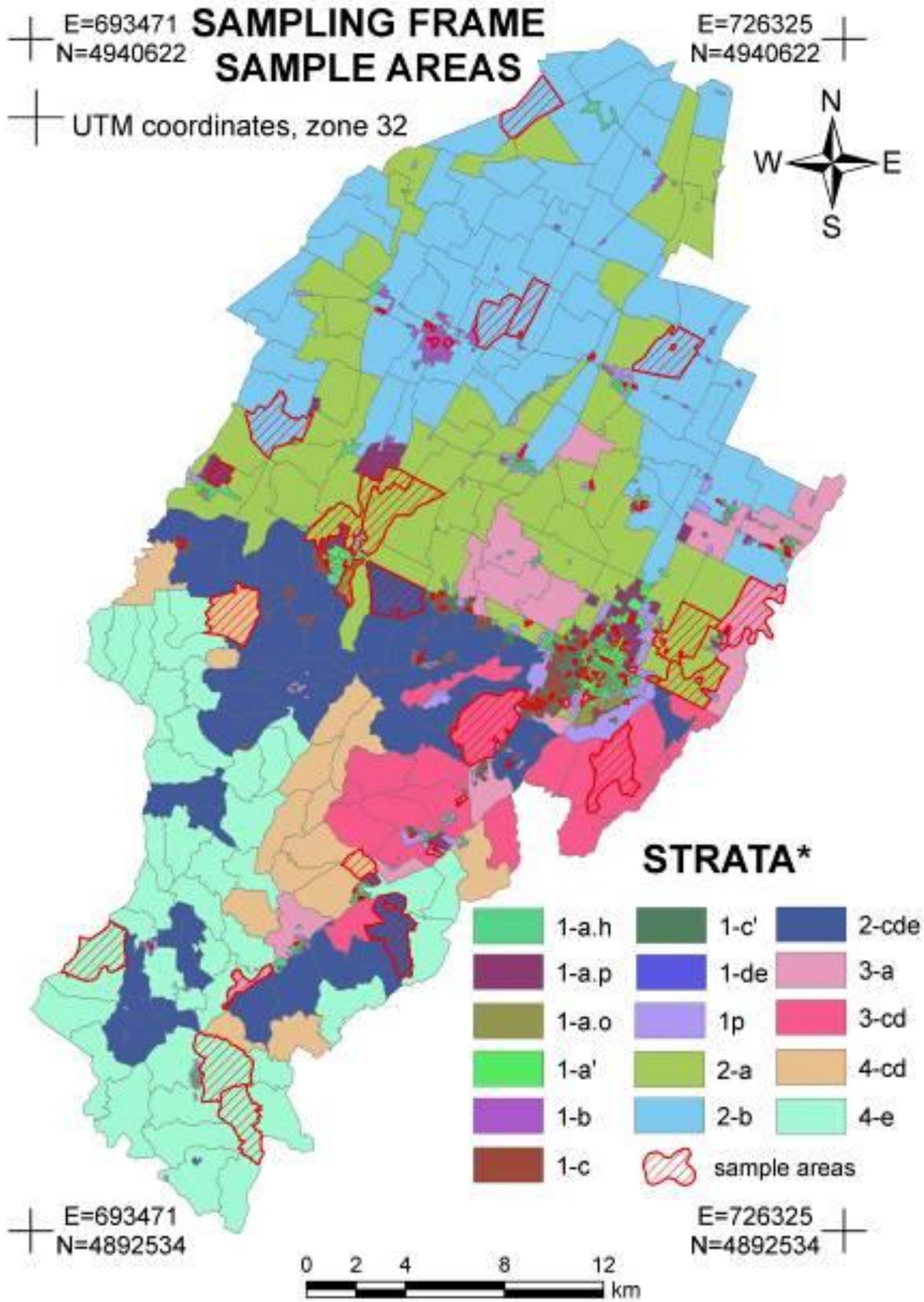
o any other type of urban use (1-a.o)

Sample size 104 **Sampling rate = 7.2%**

CV= 25.5%

DEFF = 0.73 **Relative efficiency = 1.37**

Spatial distribution of pilot sample units in the strata



Sequential sample designs

- **Besides stratification and Neyman's allocation, efficient sample selection procedure**

- **Sequential sample designs very efficient due to:**
  - sample selection dependent on previously selected units
  - stopping rules based on the estimate
  - Biased estimates

Adaptive Sequential Procedure with Permanent Random Numbers (ASPRN)

- Permanent random numbers (Ohlsson‘95) assigned to all units in the population (sections of census 2001)
- A first stratified random sample of sections is selected with probability proportional to stratum size. Call *n* the sample size
- Estimation of standard deviation of change in built area density within each stratum
- Neyman’s allocation is computed with sample size *n+1* and one section is selected in the stratum where the sample size is farthest below the size assigned by Neyman’s allocation
- The change of building cover density and its precision are estimated
- If the precision is acceptable, the process stops; otherwise, Neyman’s allocation is computed with the sample size *n+2*, and so on, until the precision considered acceptable is reached

Advantages of Adaptive Sequential Procedure with Permanent Random Numbers (ASPRN)

- Idea: adopting, in an adaptive procedure, permanent random numbers sample selection method
- Sample size per stratum dependent on previously selected units but sample selection not
- Unbiased estimates under a design based approach
- No need of hypotheses on the distribution of the change of built area density
- Through simulation ASPRN consistent and always more efficient than:
  - stratified sampling with proportional allocation
  - stratified random sampling in two phases (two steps) proposed by Thompson and Seber (1996)

Allocation of pilot sample and of ASPRN sample

STRATA	Nh	nh pilot	nh ASPRN
1-a.res	142	10	11
1-a.prod	72	5	72
1-a.dis_ver	18	2	2
1-a'	415	29	29
1-b	50	4	4
1-c	156	11	18
1-c'	136	9	9
1-de	23	2	2
1p	169	12	23
2-a	60	4	18
2-b	71	5	14
2-cde	36	2	2
3-a	15	2	2
3-cd	18	2	2
4-cd	17	2	2
4-e	43	3	3
	1441	104	213

Comparison of proportional, ASPRN and Neyman's allocation

STRATA	N_h	n_h proport. allocation	n_h pilot sample	S_h	n_h ASPRN	n_h Neyman's allocation
1-a.res	142	21	10	1688	11	**11**
1-a.prod	72	11	5	24527	72	**83**
1-a.dis_ver	18	3	2	3471	2	**3**
1-a'	415	61	29	1448	29	**28**
1-b	50	7	4	1323	4	**3**
1-c	156	23	11	2366	18	**17**
1-c'	136	20	9	1443	9	**9**
1-de	23	3	2	3	2	**0**
1p	169	25	12	2848	23	**23**
2-a	60	9	4	6566	18	**18**
2-b	71	10	5	4103	14	**14**
2-cde	36	5	2	996	2	**2**
3-a	15	2	2	608	2	**0**
3-cd	18	3	2	1286	2	**1**
4-cd	17	3	2	355	2	**0**
4-e	43	6	3	341	3	**1**
Total	**1441**	**213**	**104**	**53372**	**213**	**213**

Operational disadvantages of ASPRN

TSPRN

Operational disadvantages of ASPRN:

- Organisation
- Continuous interaction between statisticians and environmental engineers for choosing, at each step, the next area sample unit to be selected
- Segmentation of data acquisition and pre-processing
- Cost and time

Carfagna (2007) two-steps selection procedure with permanent random numbers (TSPRN)

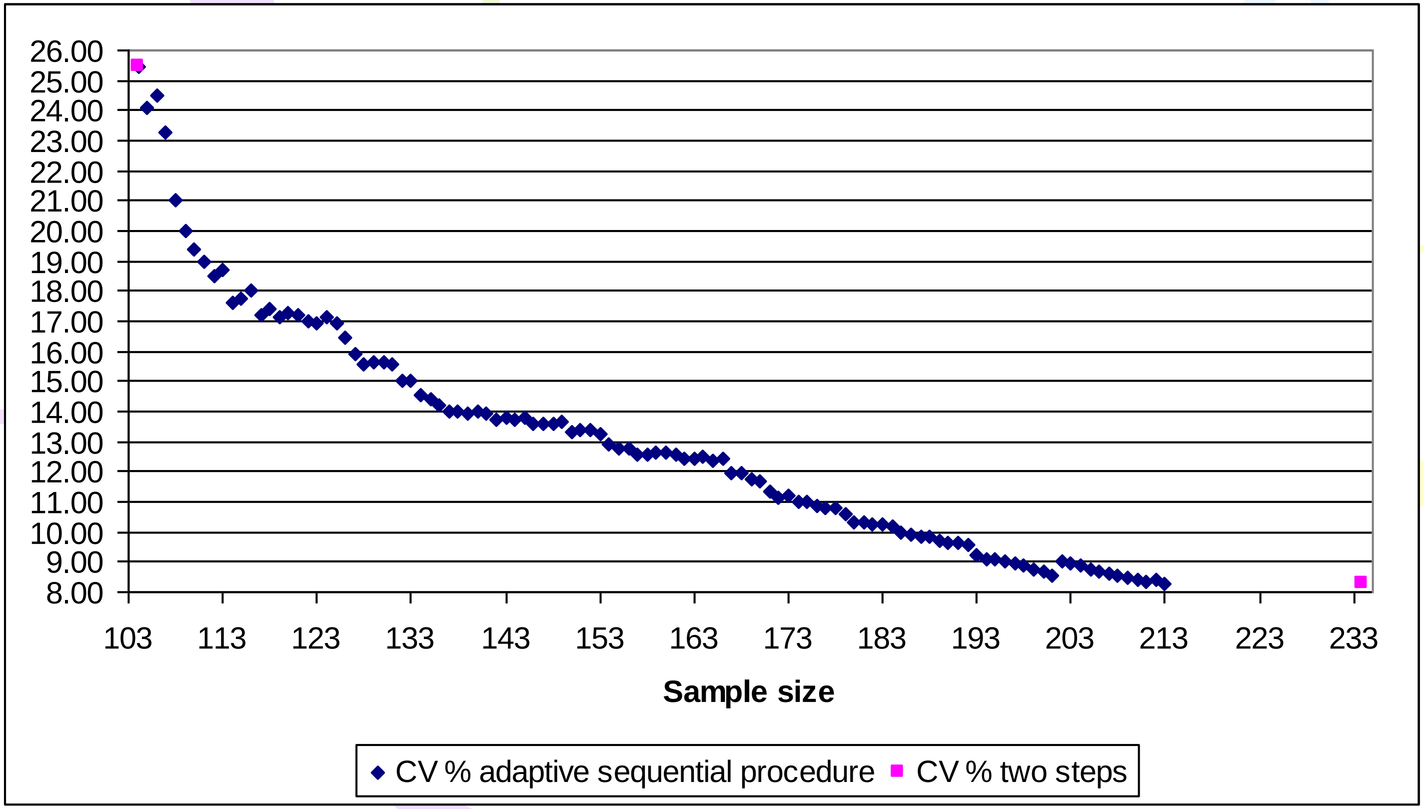
- Overcomes operational drawbacks of ASPRN:
- Less efficient than ASPRN (although more efficient than Thompson and Seber's)

TSPRN for estimating the change of building cover density

- Selection of the pilot or first step sample (104 census sections)
- Estimation of standard deviation of change in built area density within each stratum
- Computation of number of sample units needed for reaching a coefficient of variation of the estimate equal to 10% (234)
- Selection of 130 more units with allocation as near as possible to Neyman's one
- CV obtained: 8.3%, lower than the target CV
- The first phase sample overestimated the standard deviation in some strata

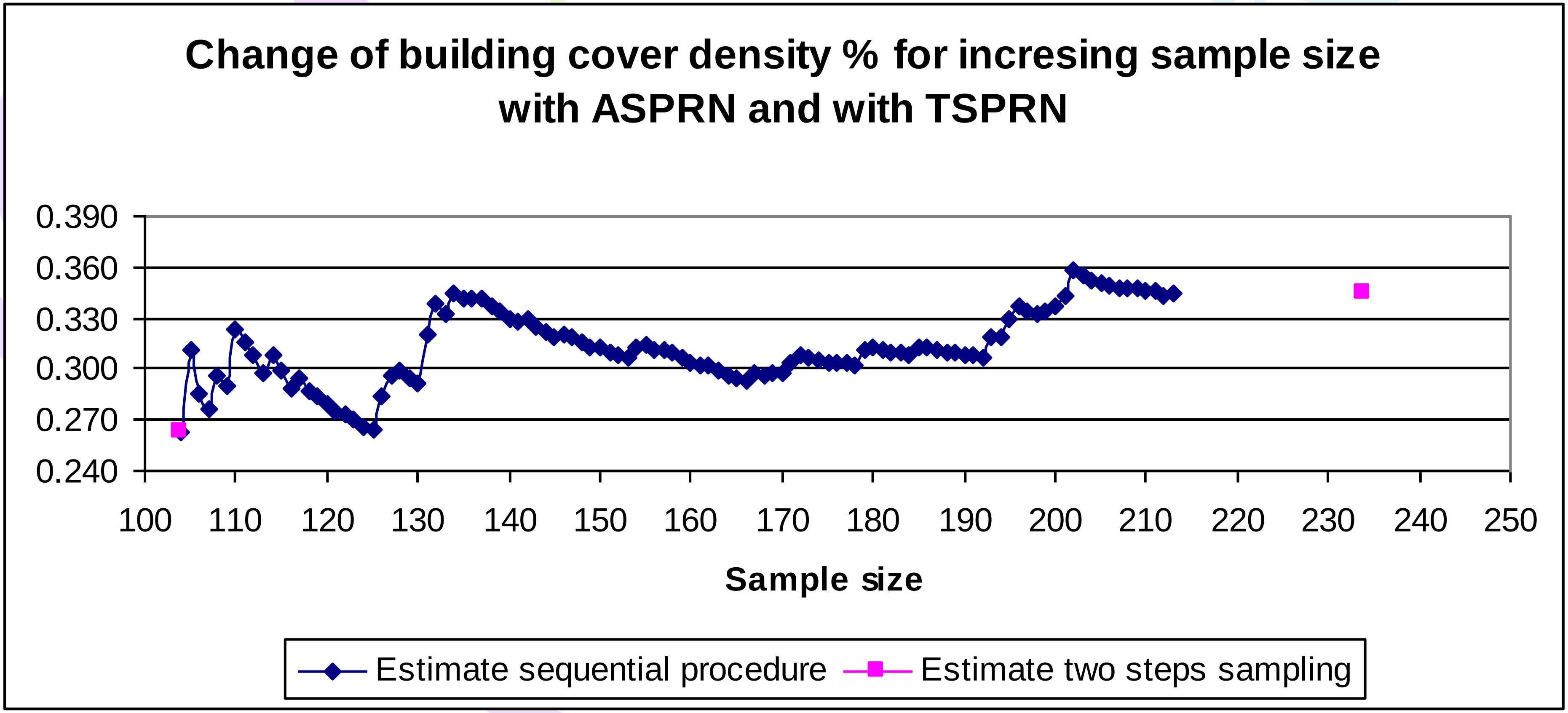
ASPRN much more efficient than TSPRN

- With ASPRN, CV equal to 10% reached with 180 sample units
- CV equal to 8.3% reached with 213 sample units (with TSPRN 234, 21 units more, 10% more)



Comparison of estimates given by ASPRN and TSPRN when the CV for both estimates is about 8%

- Difference of estimates: 81 m^2 (2.9%) - 0.01% of the study area
- Difference between estimates with pilot sample and ASPRN with CV equal to 8.3%: 642 square meters (23.7%), 0.08% of study area



Conclusions

- ASPRN and TSPRN estimates tend to converge
- With real data ASPRN confirms to be much more efficient than TSPRN
- Operational disadvantages of ASPRN are important and in some cases cannot be overcome
- Use ASPRN whenever possible
- Otherwise, use TSPRN
- More efficient than Thompson and Seber’s
- More efficient than proportional allocation if stratification efficient

Thank you

elisabetta.carfagna@unibo.it

References

- Benni, S., D. Torreggiani, P. Tassinari, 2009. Sistemi integrati per l'analisi delle risorse del territorio rurale: alcune note di metodo. Agribusiness Paesaggio & Ambiente 12, 16-22.
- Carfagna E. (2007) Crop area estimates with area frames in the presence of measurement errors, Proceeding of ICAS-IV, Fourth International Conference on Agricultural Statistic. Invited paper, Beijing, 22-24 October 2007. http://www.stats.gov.cn/english/icas/ICAS4%20abstract/P020071112583693287850.pdf
- Carfagna E. and Marzialetti J. "Sequential Design in Quality Control and Validation of Land Cover Data Bases" Journal of Applied Stochastic Models in Business and Industry Volume 25, Issue 2, 2009, pp. 195-205, DOI: 10.1002/asmb.742, John Wiley & Sons, Ltd.
- Istat, Italian statistic board, 2001. 14th Census of Population and Housing. Istat, Italy.
- Ohlsson E. (1995) Coordination of Samples Using Permanent Random Numbers, in Business survey methods, Cox B. G., Binder D.A., Chinnapa B.N., Christianson A., Colledge M.J., Kott P.S. (Eds.), Wiley, New York, 153-169.
- Tassinari P., Carfagna E., Benni S., Torreggiani D. (2008) Wide-area spatial analysis: A first methodological contribution for the study of changes in the rural built environment, Biosystems Engineering, 100 (3), 435-447.
- Thompson S.K., Seber G.A.F. (1996) Adaptive Sampling, Wiley, New York.