Towards a European-wide sampling design for biodiversity monitoring

Dick Brus, Dennis Walvoort

EBONE and BioBio Data Analysis Workshop, 18-19 October 2010, Wageningen

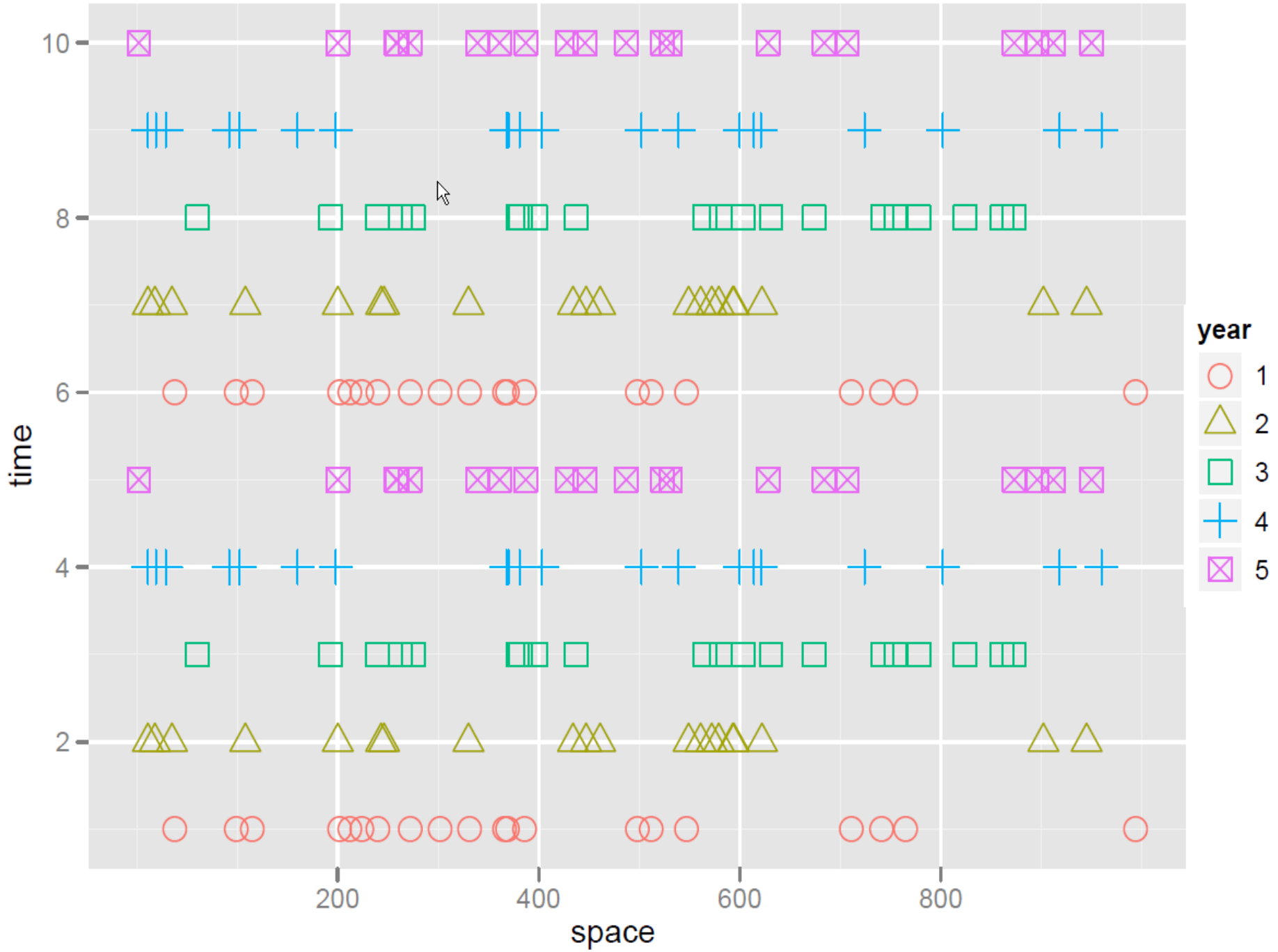
ALTERRA
WAGENINGEN UR

Introduction

- Study area: EU27 (exluding UK and SE), inclusing NO and CH
- We focussed on common habitats
- Sampling units: km-squares (INSPIRE 1 km grid)

Space-time design

- Serially alternating design with periodocity of 5 years
- First five years different sets of km-squares are observed, in sixth year km-squares if first year are revisited et cetera


10
8
6
4
2
time
200
400
600
800
space
year
1
2
3
4
5

Motivation of serially alternating space-time design

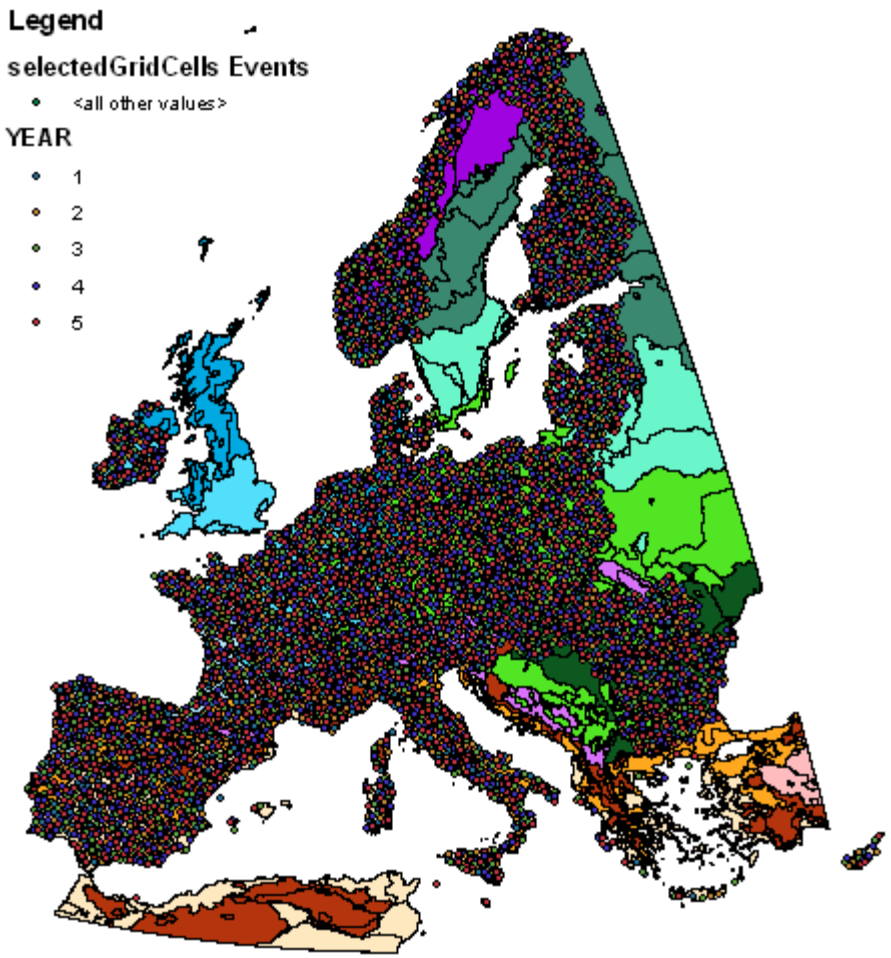
- Precise estimates of temporal trend of spatial means
  - Urquhart, N. S., and Kincaid, T. M. (1999), "Designs for Detecting Trend from Repeated Surveys of Ecological Resources," *Journal of Agricultural, Biological and Environmental Statistics*, 4, 404–414.
  - C.J.F. ter Braak, C.J.F., Brus, D.J. and Pebesma, E. (2008) "Comparing sampling patterns for kriging the spatial mean temporal trend" *Journal of Agricultural, Biological and Environmental Statistics* 13, 159-176
  - Brus, D.J. and de Gruijter J.J. (2011) "Generalized Least Squares estimation of status and trend of soil properties from repeated surveys" *Geoderma* (in prep.)
- NILS is also serially alternating space-time design

Spatial sampling design

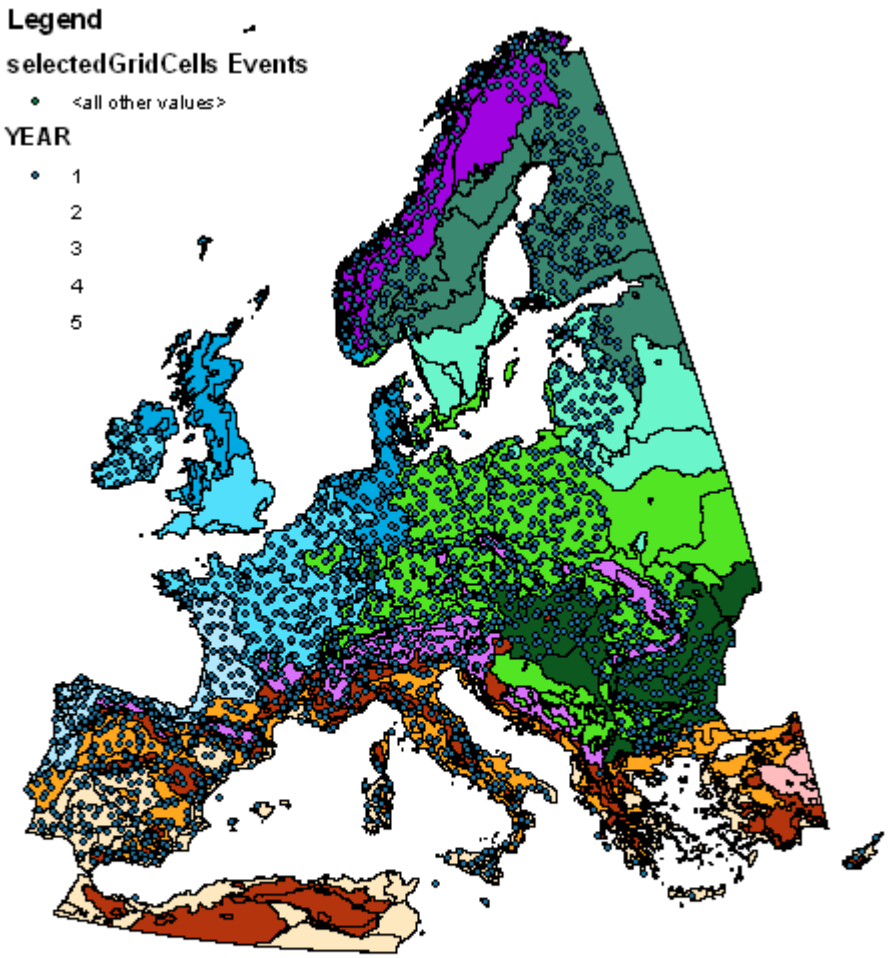
- Set of km-squares observed in a given year should cover entire study area ('EU')
- Each year unbiased estimates of target parameters (total area of GHC et cetera) for EU
- Sampling design should be *simple* and *flexible*
- Flexibility in adaptation of number of selected km-squares

Spatial sampling design

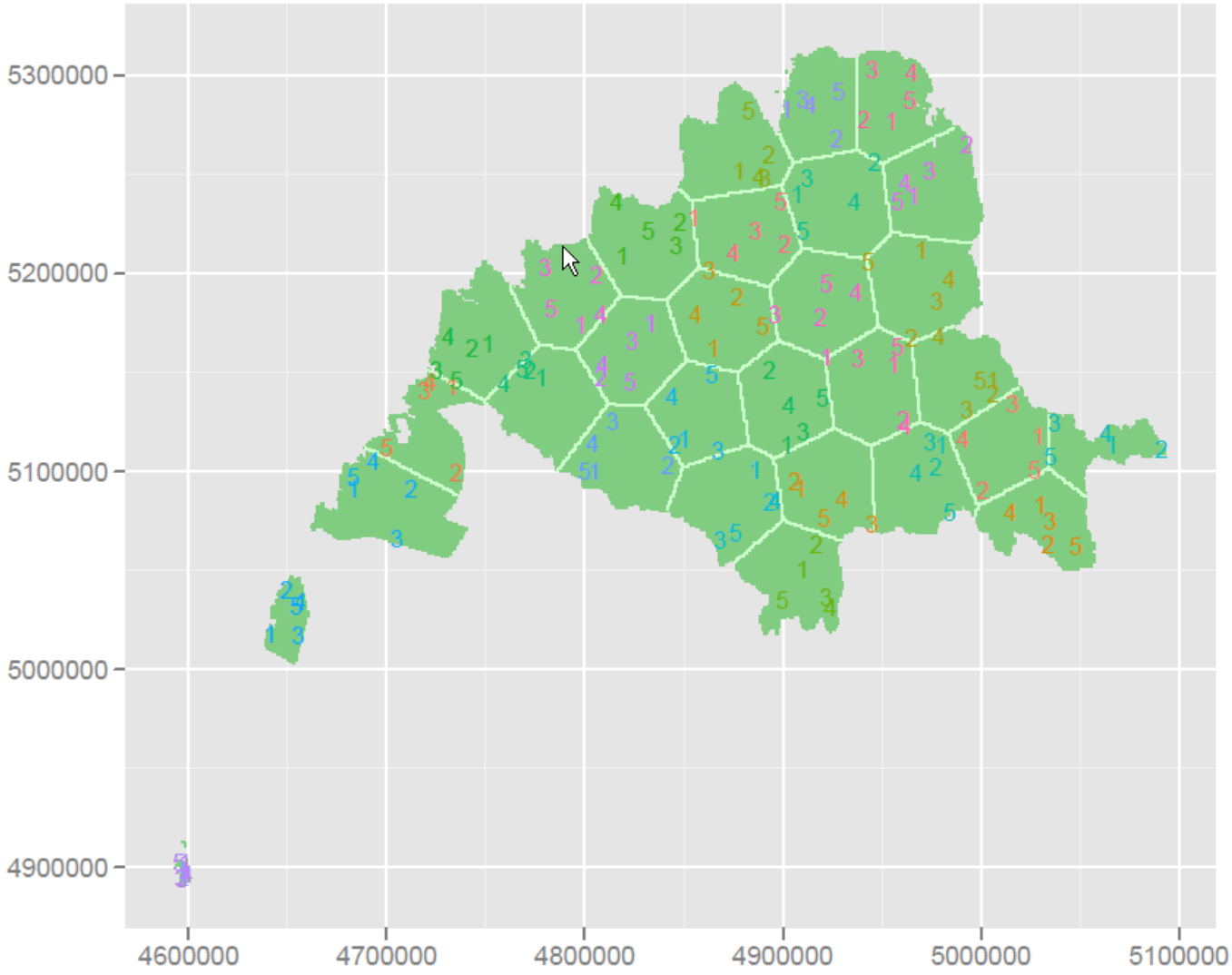
- Stratified simple random sampling
- Strata
  - Main strata: EnS strata (81 within EU)
  - Substrata: compact geographical strata (geostrata), formed by k-means clustering of 1 km x 1 km cells (Walvoort et al, 2010)
- Within each geostratum five km-squares are selected by simple random sampling *without replacement*
- Number of geostrata proportional to area of EnS-stratum
- Number of km-squares: 2000 per year; this gives 10 000 km-squares in total)

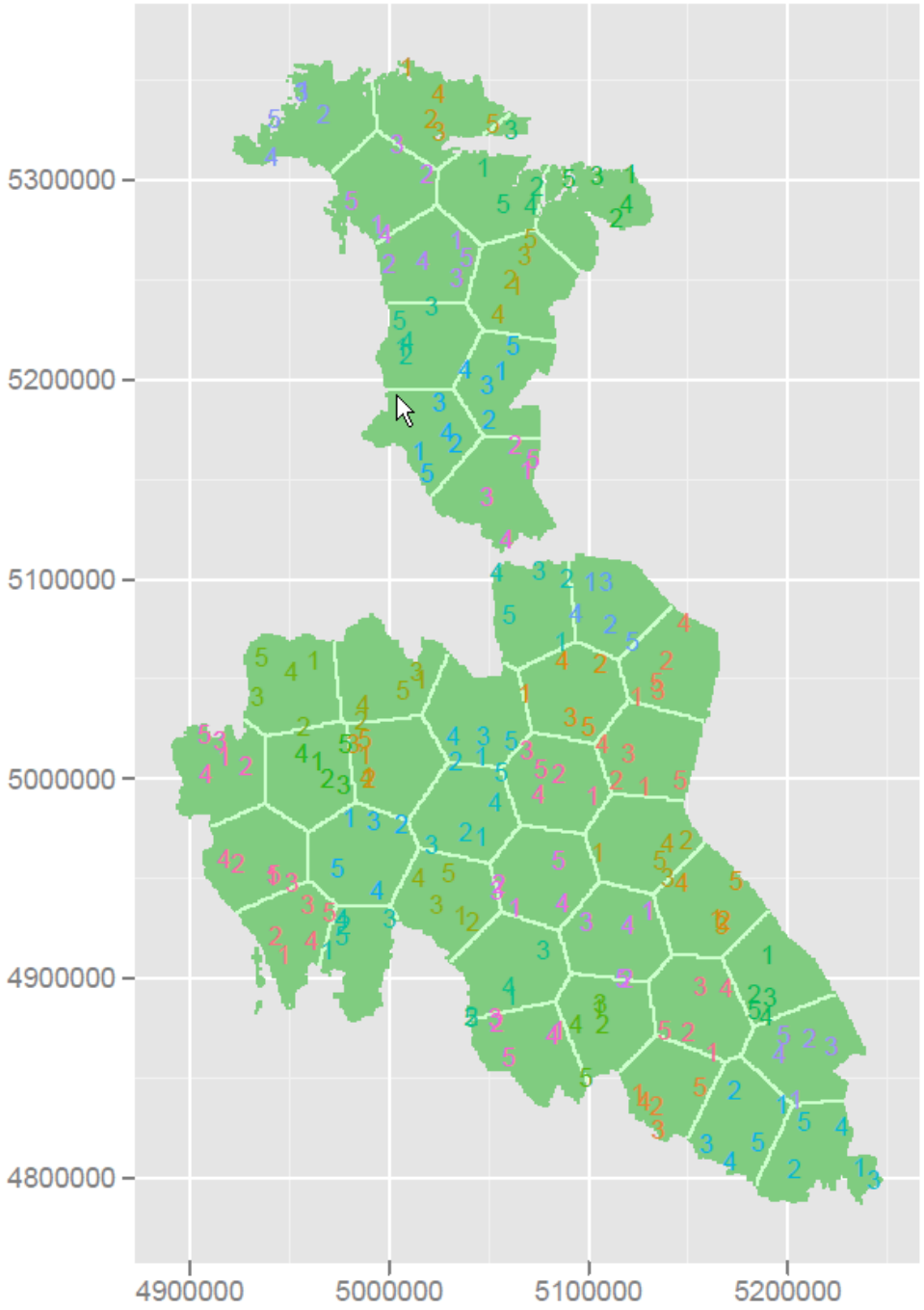


5 years x 2000



1 year x 2000


5300000
5200000
5100000
5000000
4900000
4600000
4700000
4800000
4900000
5000000
5100000


5300000
5200000
5100000
5000000
4900000
4800000
4900000
5000000
5100000
5200000

Motivation of stratified simple random sampling

- EnS-stratification:
  - Control of number of km-squares per EnS-stratum
  - Simple to obtain estimates per EnS stratum
  - More precise estimates of some target parameters for EU?
- Geographical substratification of EnS strata:
  - Takes care of spatial coverage (avoids spatial clustering of km-squares within EnS strata): increased precision

Statistical inference

- How to obtain estimates for EU, 12 EnZones and 81 EnStrata?

- Estimates of
  - Spatial means (totals, areal proportions) for a given year
  - Change of spatial means between successive years
  - Temporal trend of spatial means
  - Uncertainty of these estimates!

Estimation of spatial means for a given year

- Design-based estimation:
  - We must take the sampling design into account
- Area of geostrata not constant!
- Due to this inclusion probabilities not equal for all km-squares
- Unweighted average is *biased* estimate
- *Unbiased* estimates can be obtained as weighted average of geostratum-means

Unbiased estimation of spatial man for a given year

- Weighted average of geostratum-means:

$$\hat{\bar{y}} = \sum_{h=1}^{L} \frac{A_h}{A} \hat{\bar{y}}_h$$

$$\hat{\bar{y}}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}; n_h = 1 \rightarrow \hat{\bar{y}}_h = y_{hi}$$

- A_h: area of geostratum h; A: total area of EU (EnS-stratum)

Sampling variance of estimated mean

$$\mathrm{Var}(\hat{\bar{y}}) = \sum_{h=1}^{L} \left(\frac{A_h}{A} \right)^2 \mathrm{Var}(\hat{\bar{y}}_h)$$

$$\mathrm{Var}(\hat{\bar{y}}_h) \approx \frac{S_h^2(y)}{n_h}; S_h^2(y) = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \hat{\bar{y}}_h)^2$$

- Problem: for a given year we have only 1 km-square per geostratum; spatial variance within geostratum cannot be estimated
- *Solution*: collapsed strata method (Cochran, 1977)

Complication

- Land area within selected km-squares can be < 100 ha, and will be different between km-squares
- Spatial means (areal proportions) cannot be estimated directly as a weighted average of geostratum means
- *Solution*:
  - first estimate totals (e.g. total area of GHC within EU)
  - then estimate spatial mean (areal proportion)

$$\hat{\bar{y}} = \frac{\hat{t}(y)}{A}$$

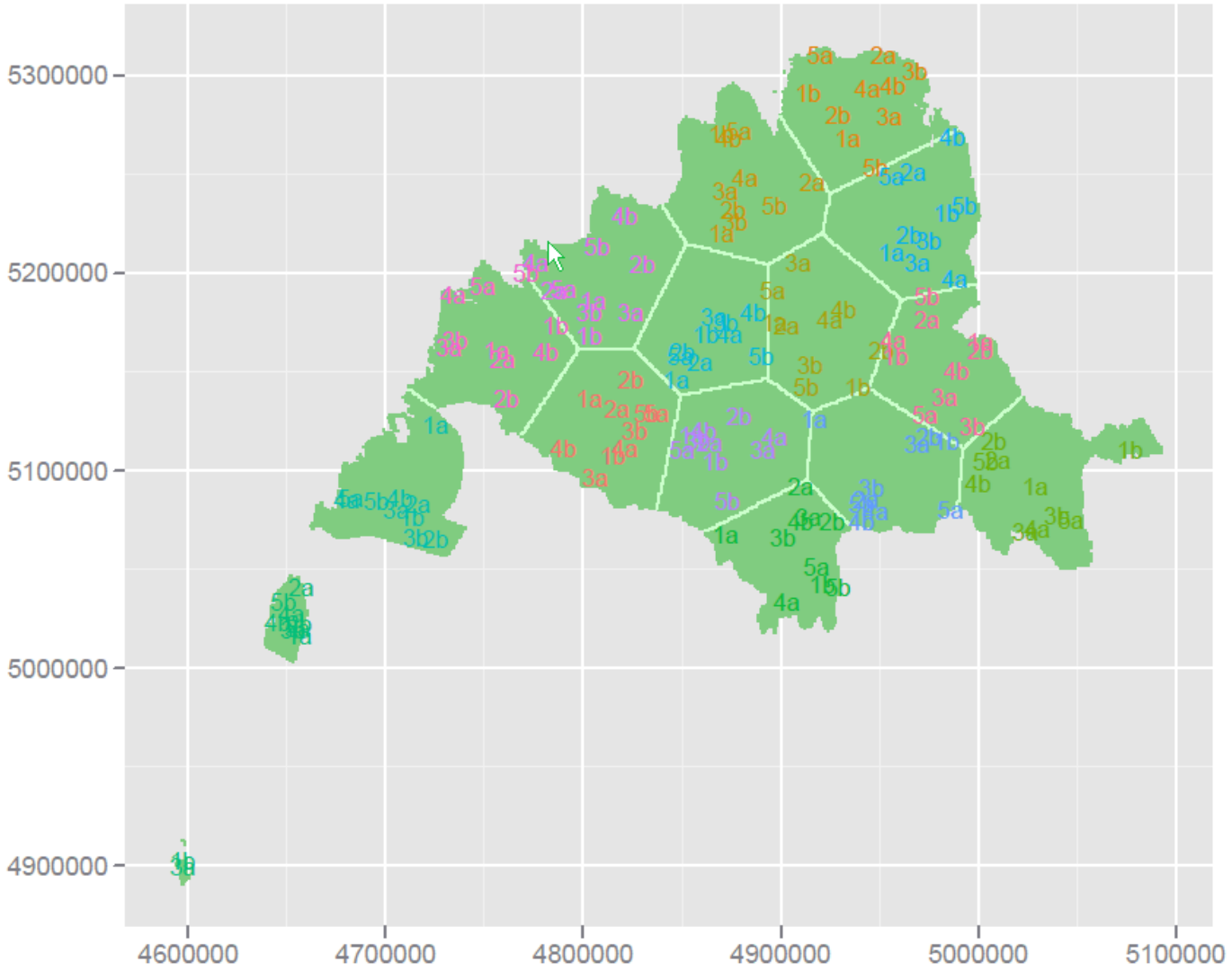
Ratio-estimate

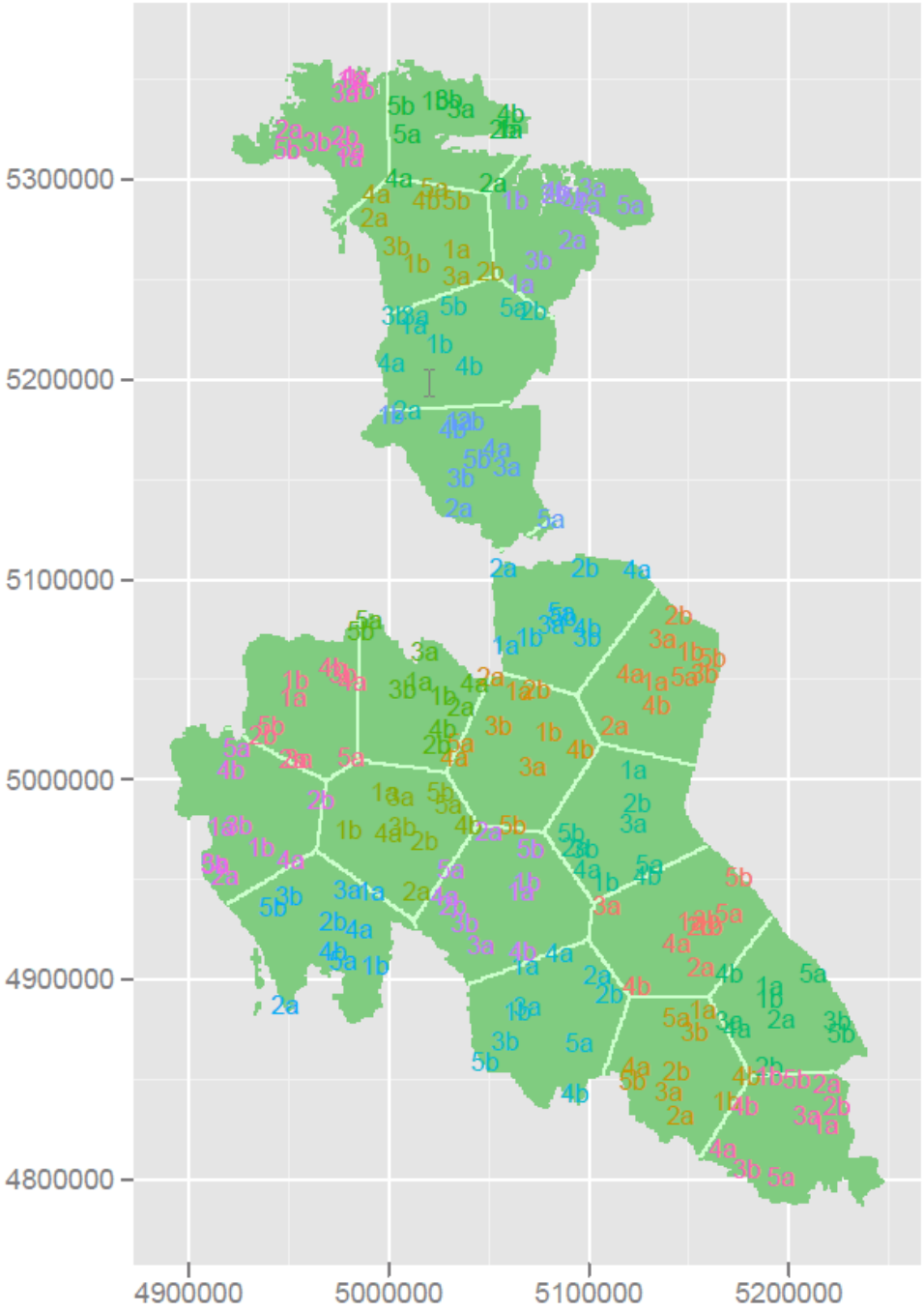
- Alternative, more precise estimate of spatial mean (areal proportion):

$$\hat{\bar{y}} = \frac{\hat{t}(y)}{\hat{A}}$$

Flexibility of spatial sampling design

- Increase of number of km-squares: ok
- Decrease of number of km-squares:
  - with 1 km-square per geostratum (per year) EnS not fully covered by sample anymore, *biased* estimates
  - alternative: 2 km-squares per geostratum


5300000
5200000
5100000
5000000
4900000
4600000
4700000
4800000
4900000
5000000
5100000


5300000
5200000
5100000
5000000
4900000
4800000
4900000
5000000
5100000
5200000

To be done

- Evaluation of alternative space-time designs, for instance static-synchronous (pure panel) with sampling interval of five years and 5*n* km-squares per sampling round (Countryside Survey)
- Evaluation of alternative spatial designs (stratified *systematic* random sampling, LUCAS design)


10
8
6
4
2
time
200
400
600
800
space

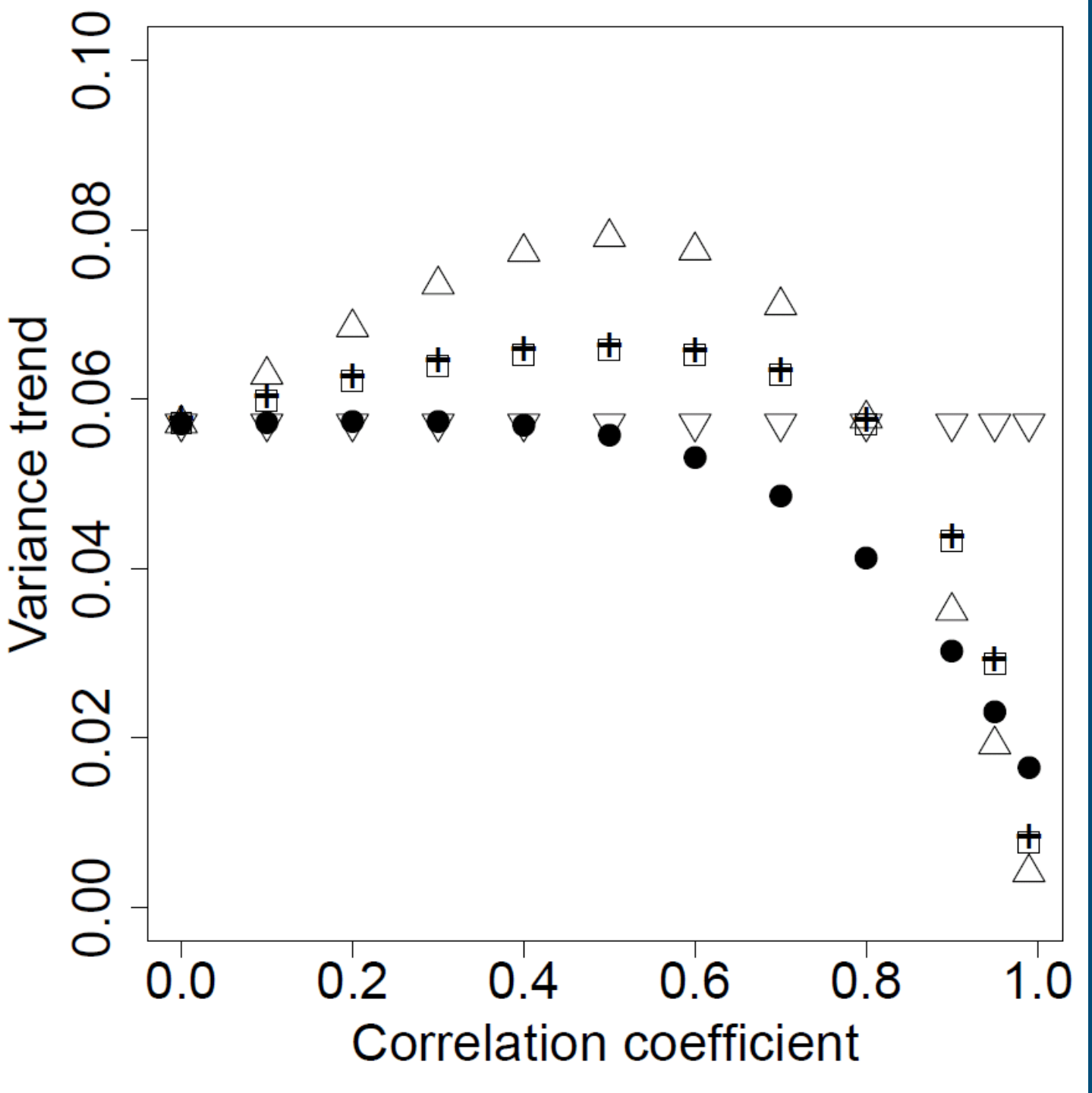
To be done (2)

- Determination of sample size (number of km-squares)
- Two options:
  - compute affordable sample size given yearly budget
  - compute required sample size given quality constraints
- Quality constraints:
  - Sampling variance of estimated means (proportions, totals)
  - Variance of estimated temporal trend of spatial means
  - Power of test for estimated temporal trend

Prior information on spatial variation

- Prior information on spatial variation of target properties (statistics at the level of km-squares) between km-squares within EnS-strata is needed for determination of required number of km-squares given quality constraint

- Available data in EBONE database can be used


Variance trend
0.00
0.02
0.04
0.06
0.08
0.10
0.0
0.2
0.4
0.6
0.8
1.0
Correlation coefficient