Soil Monitoring from a Sample Designers Perspective

Dick Brus

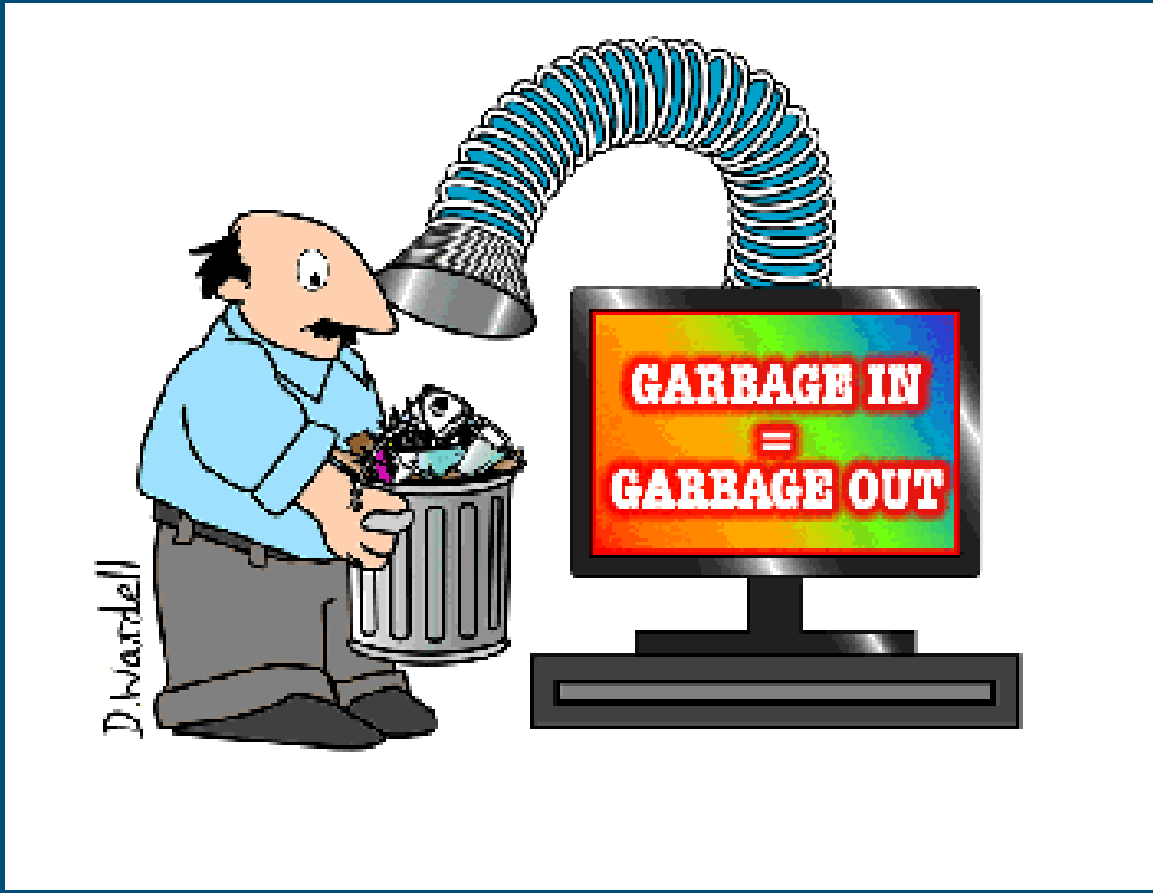
Soil Science Centre
Wageningen University and Research Centre - Alterra

ALTERRA
WAGENINGEN UR

Natural resource monitoring as a research topic

- Strong emphasis on modeling
- Few papers on the statistical design of sample

Why should we bother about sampling design?

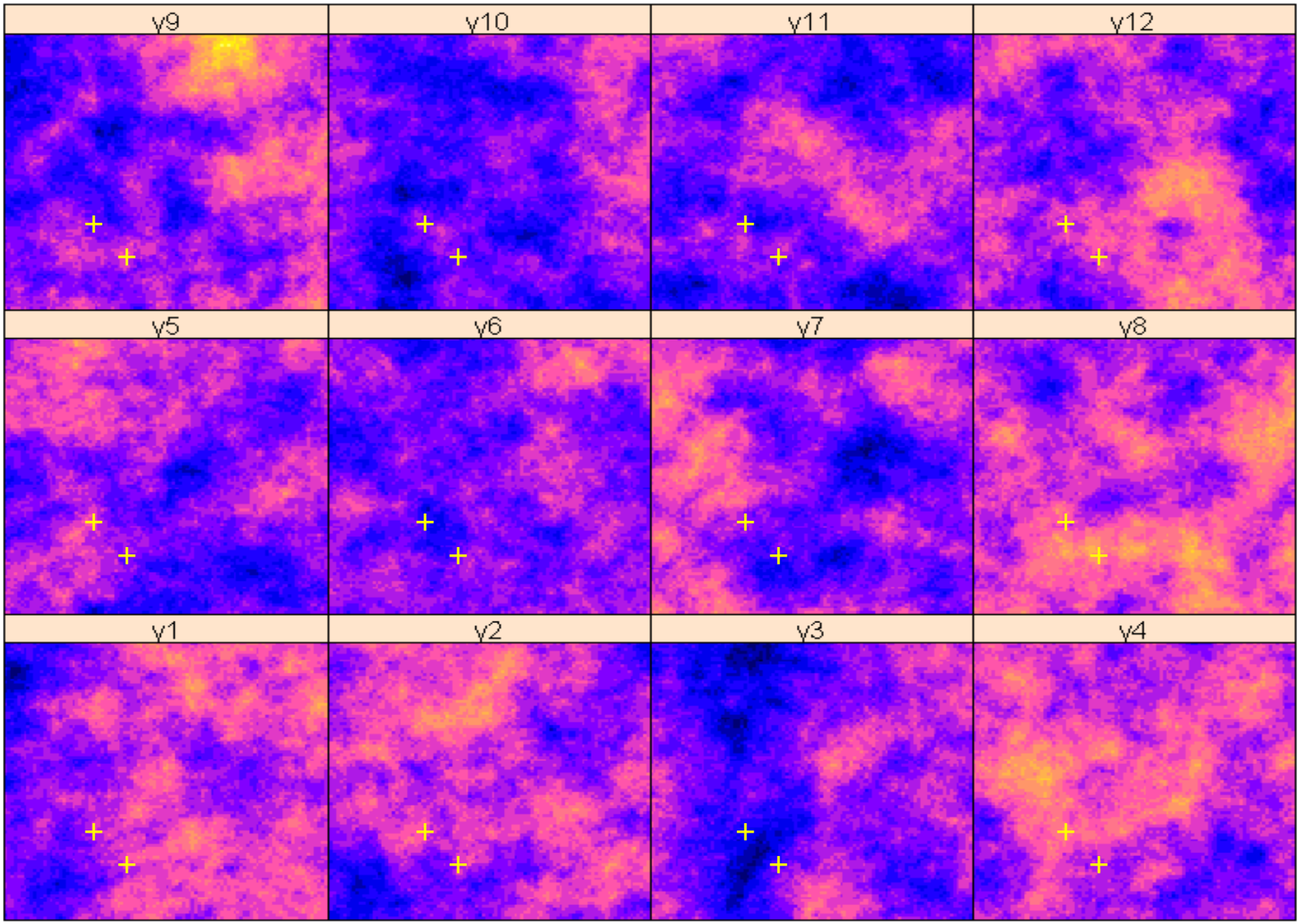


Two major design decisions

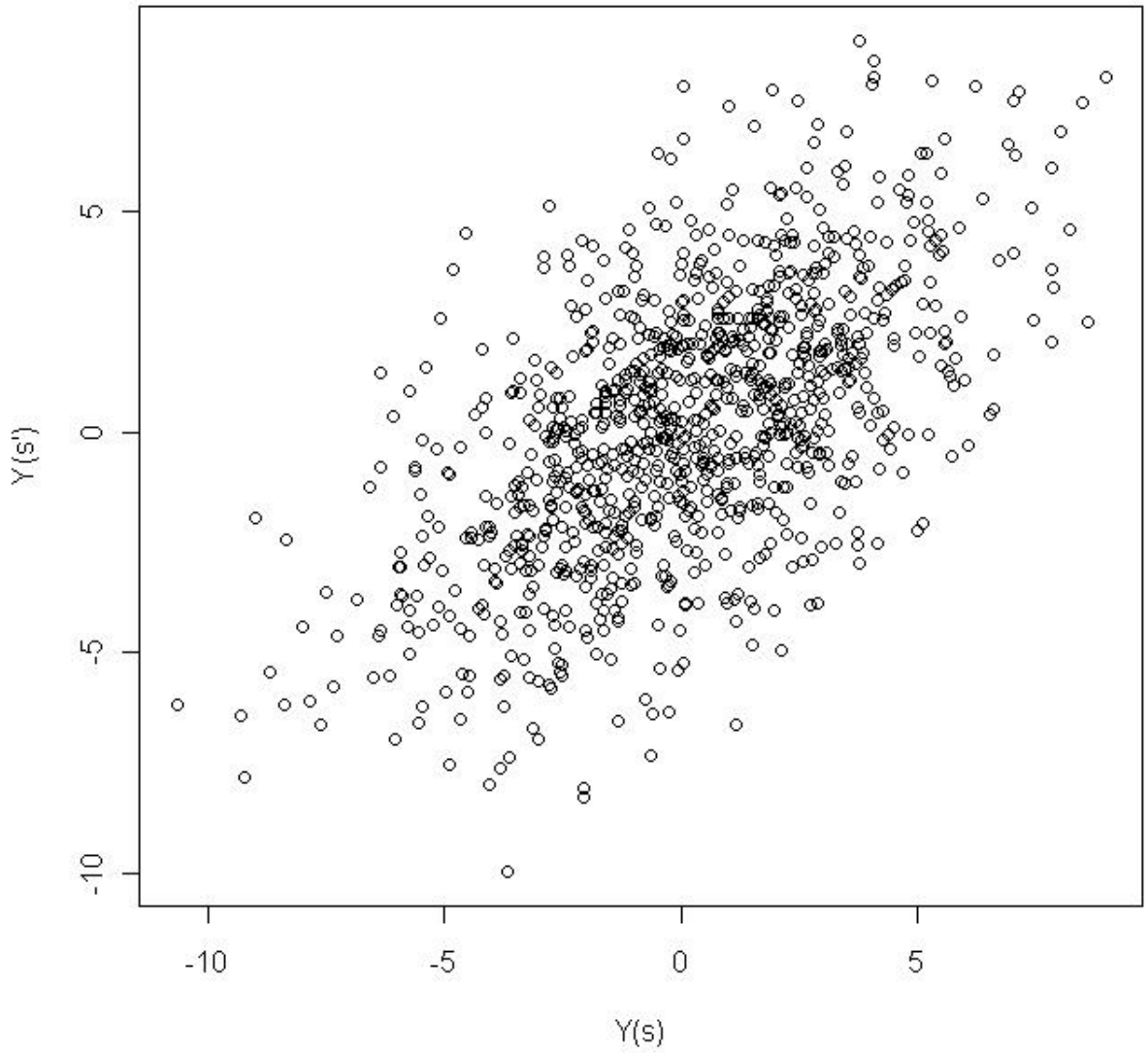
- Design-based or model-based approach?
- Choice of basic type of space-time design

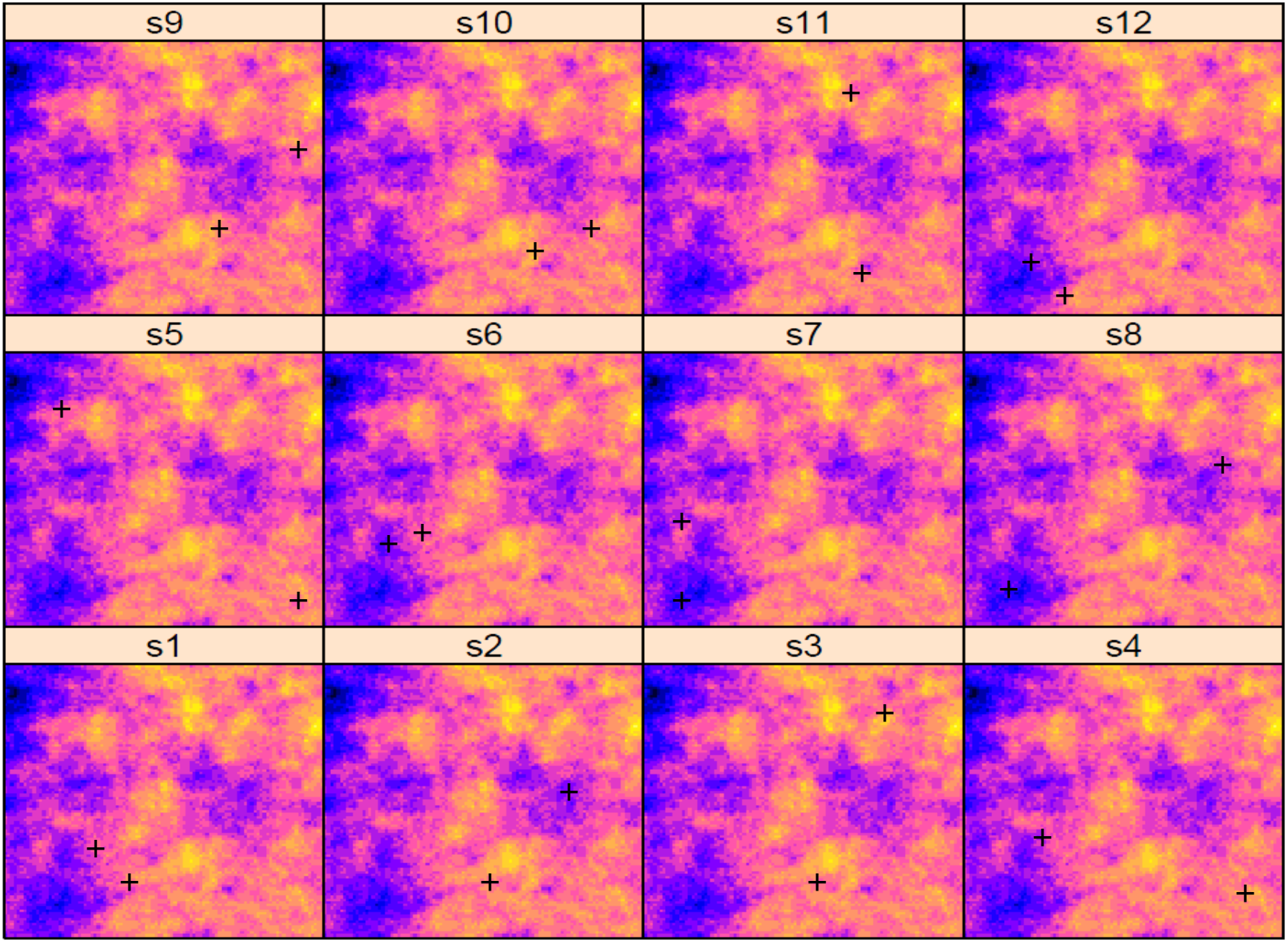
Question

- Two locations were selected randomly, independently from each other. The distance between the points is 15 cm. The measured SOM contents are 3.0% and 3.5%
- Are these two measurements correlated?
- Answers:
  - Yes (confidence > 80%)
  - No (confidence > 80%)
  - Do not know (confidence < 80%)

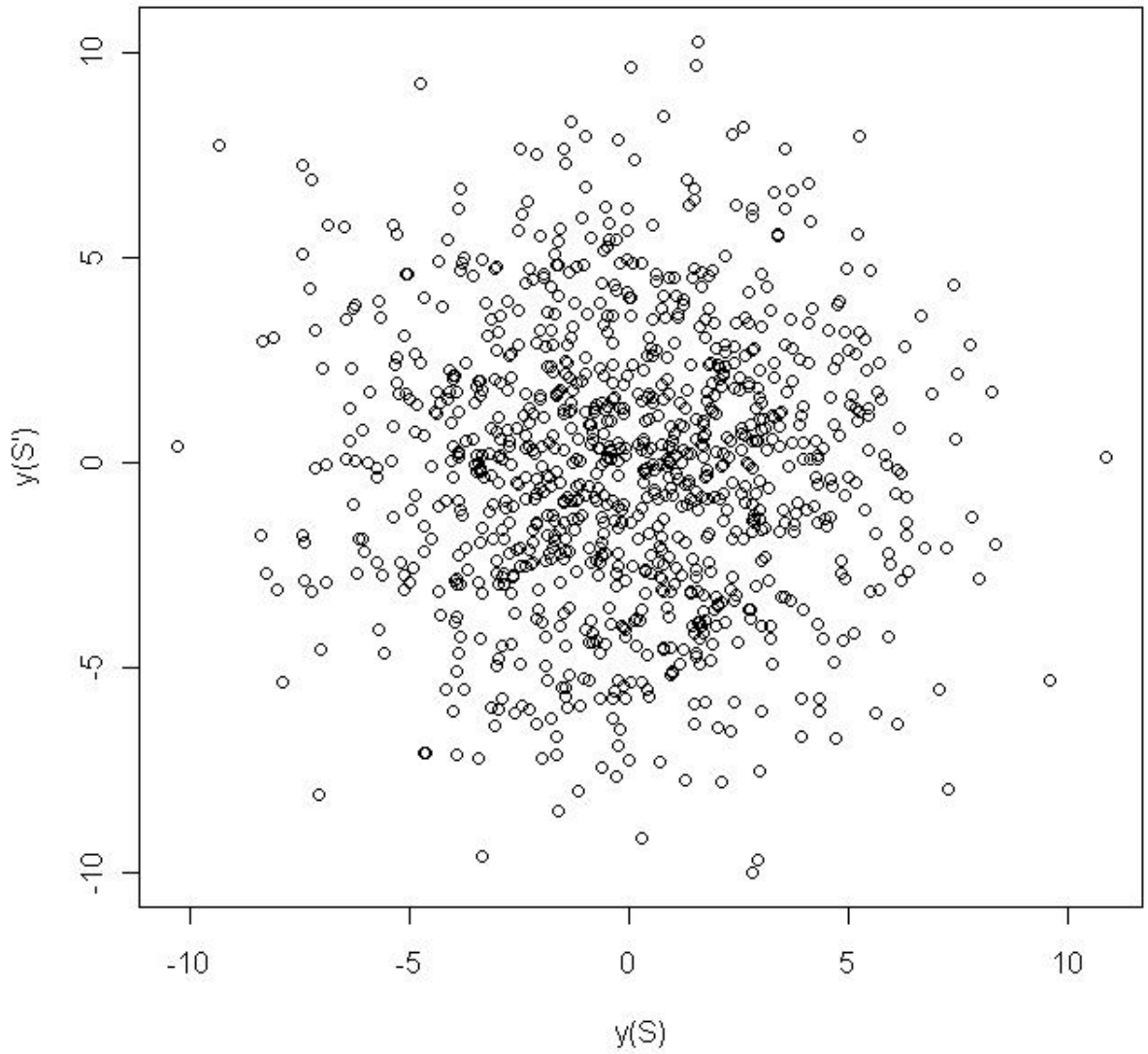


ALTERRA
WAGENINGEN UR


Y(s')
5
0
-5
-10
-10
-5
0
5
Y(s)



ALTERRA
WAGENINGEN UR


y(S')
10
5
0
-5
-10
-10
-5
0
5
10
y(S)

Sampling and inference

	Selection of sampling units	Statistical inference
Design-based	Probability sampling	Uses inclusion probabilities

Sampling and inference

Design-ba

Model-ba

ELSEVIER

Geoderma 80 (1997) 1–44

GEODERMA

Discussion Paper

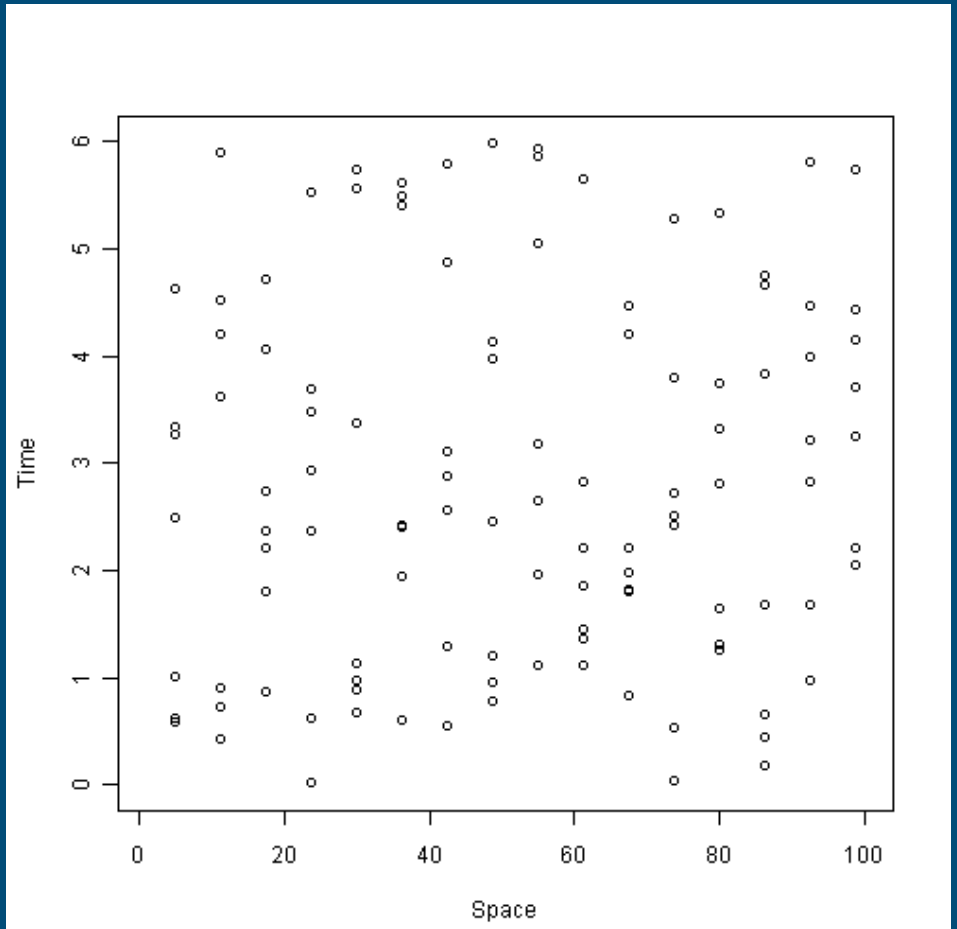
Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion) [1]

D.J. Brus *, J.J. de Gruijter

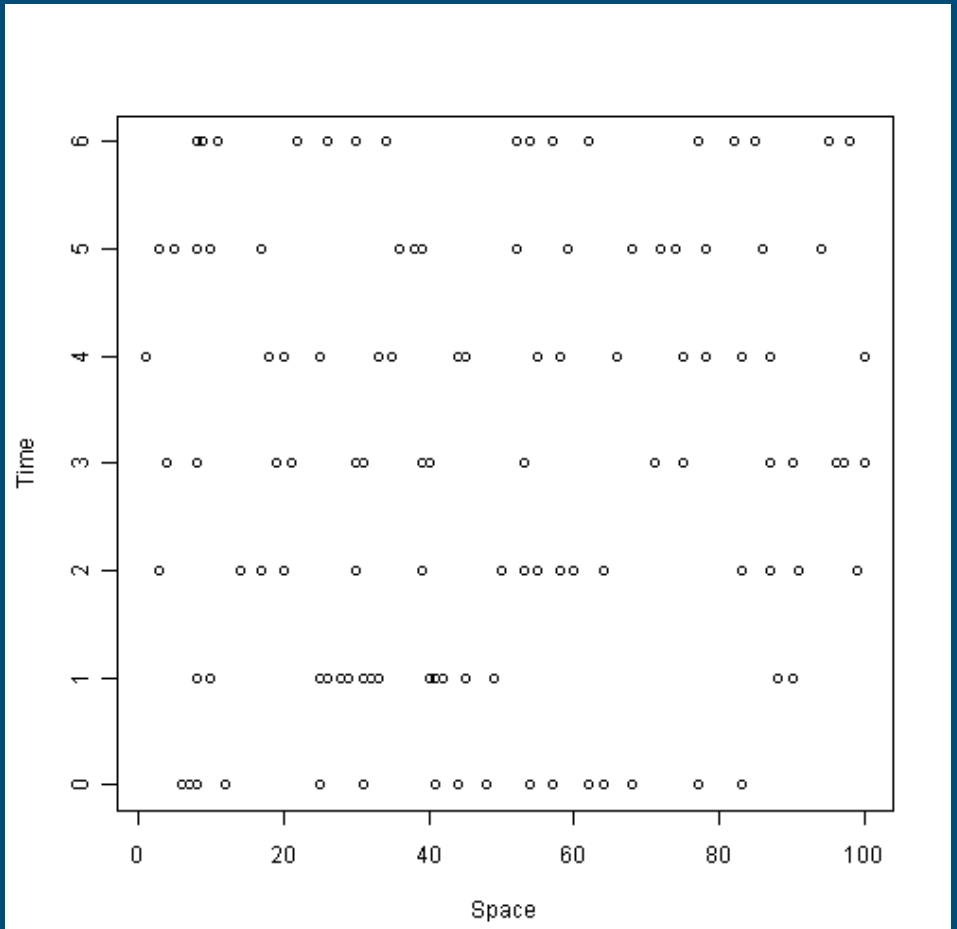
Random sampling is not always probability sampling

Basic space-time sampling designs

Static

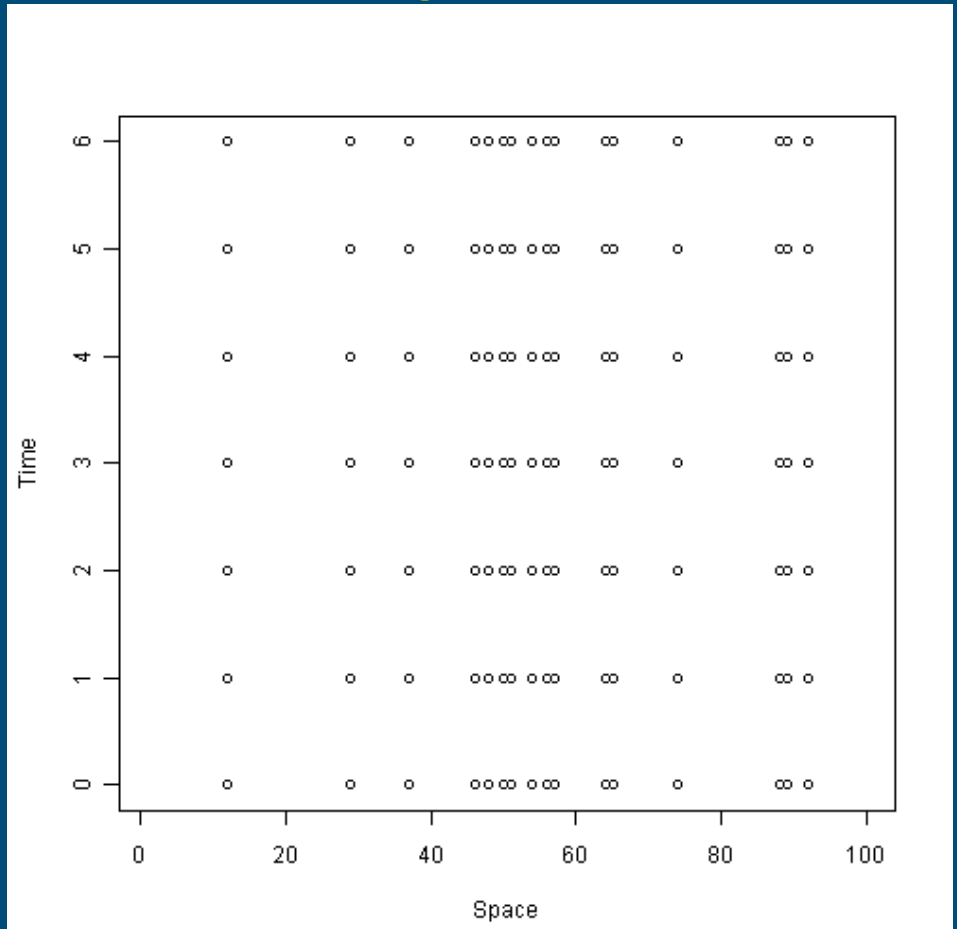


Synchronous

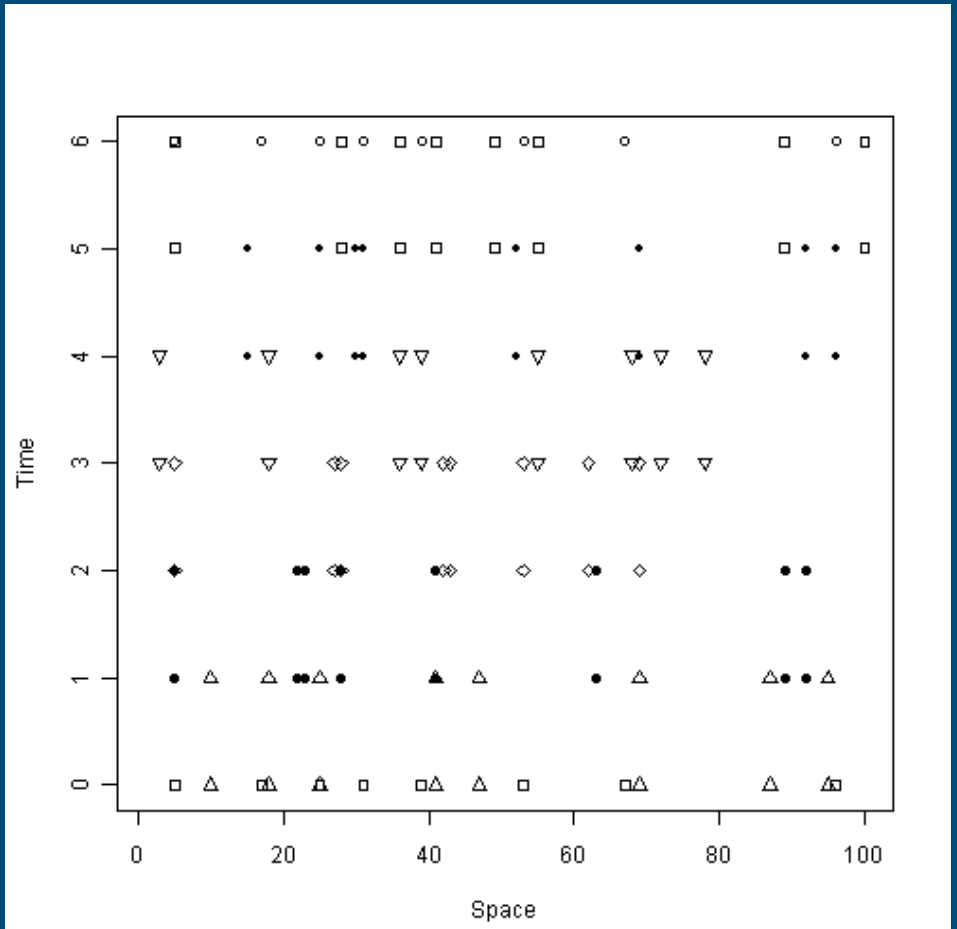


Sampling designs in space-time

Static-synchronous



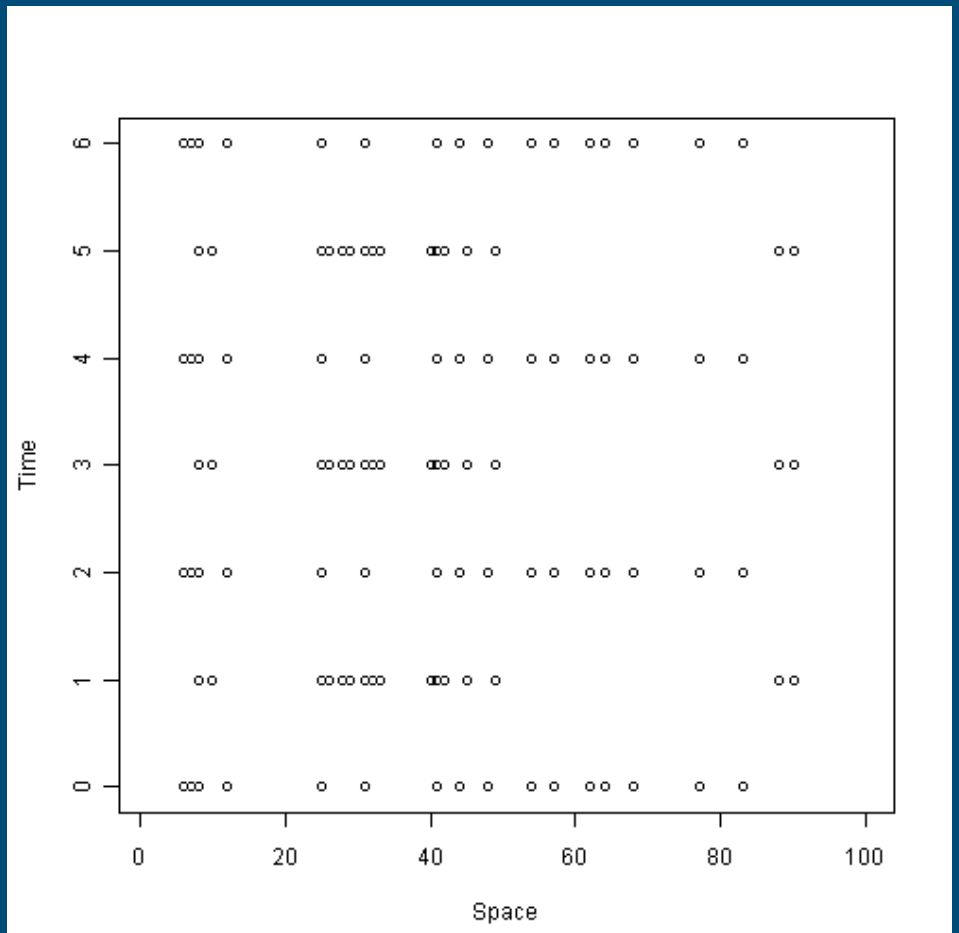
Rotational



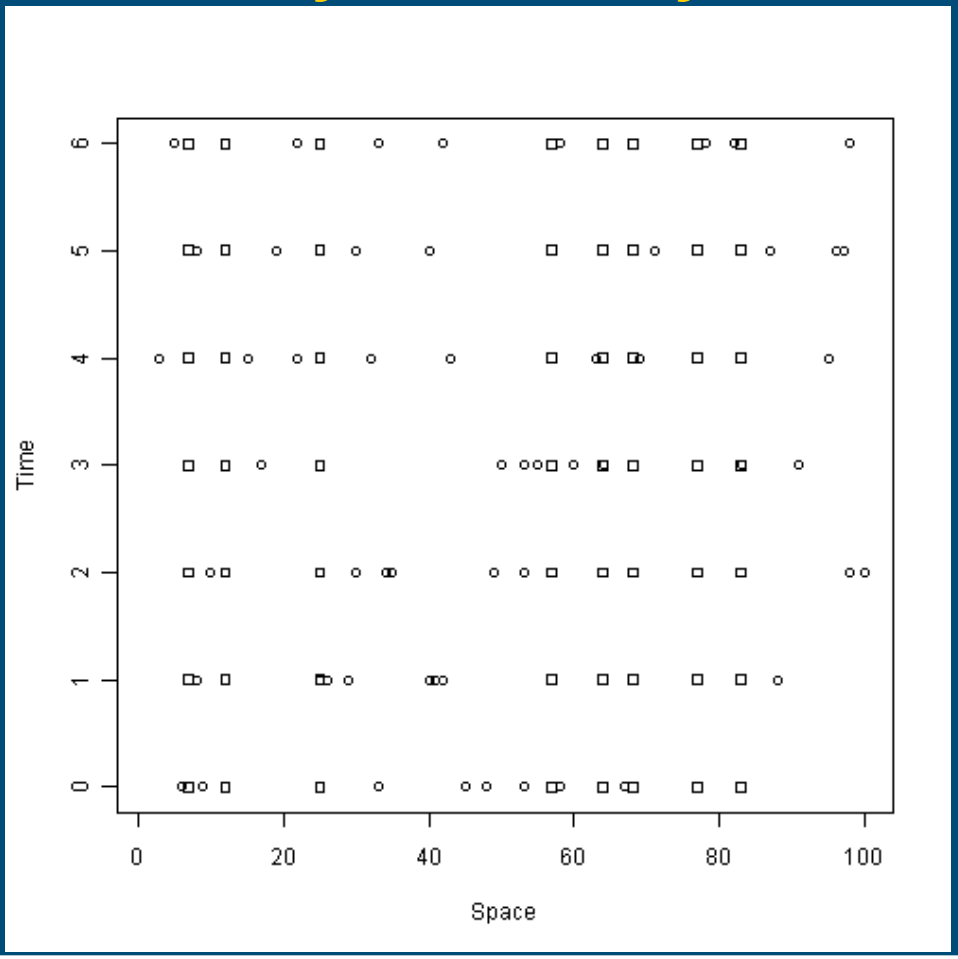
ALTERRA
WAGENINGEN UR

Sampling designs in space-time

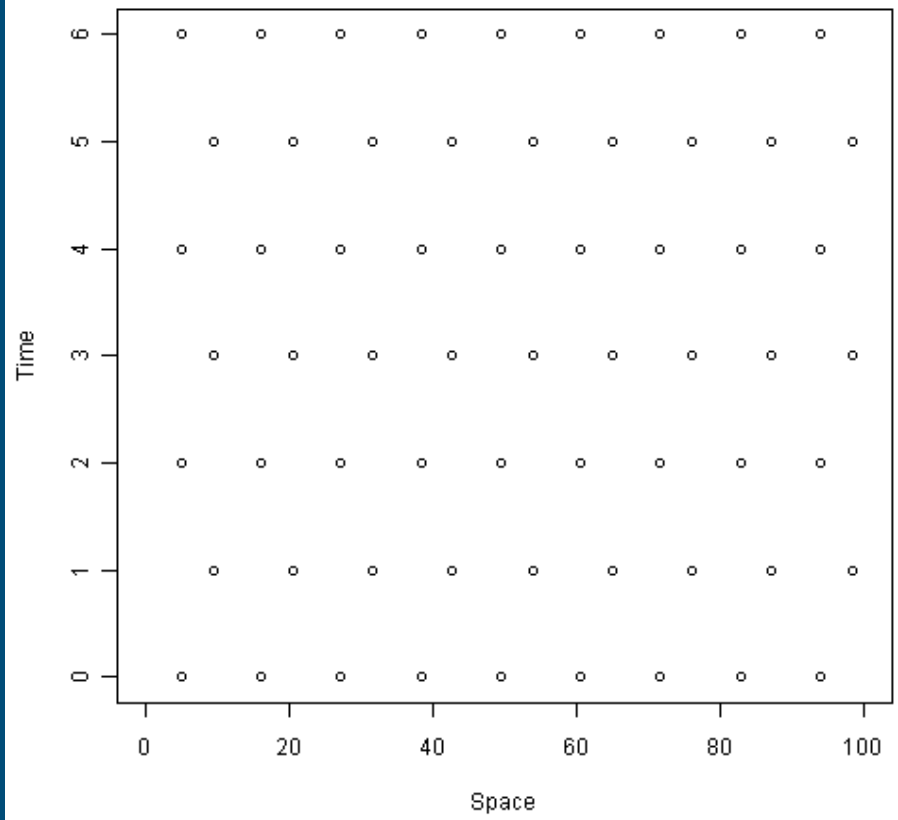
Serially alternating



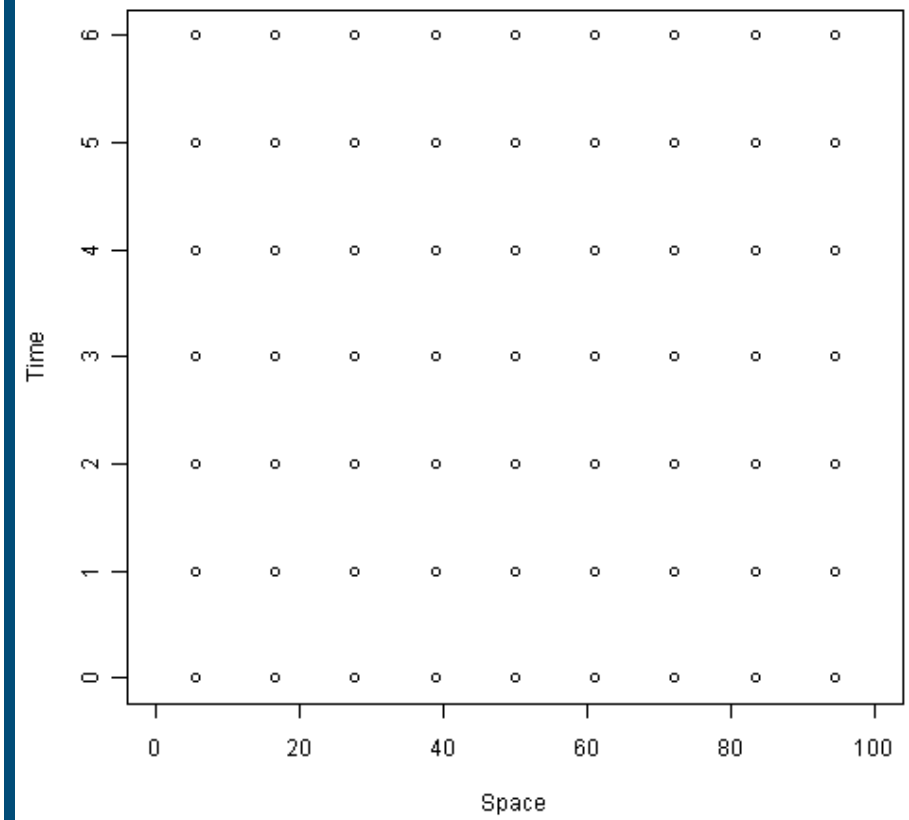
Static-synchr + Synchr.



Serially alternating *random* grid in space

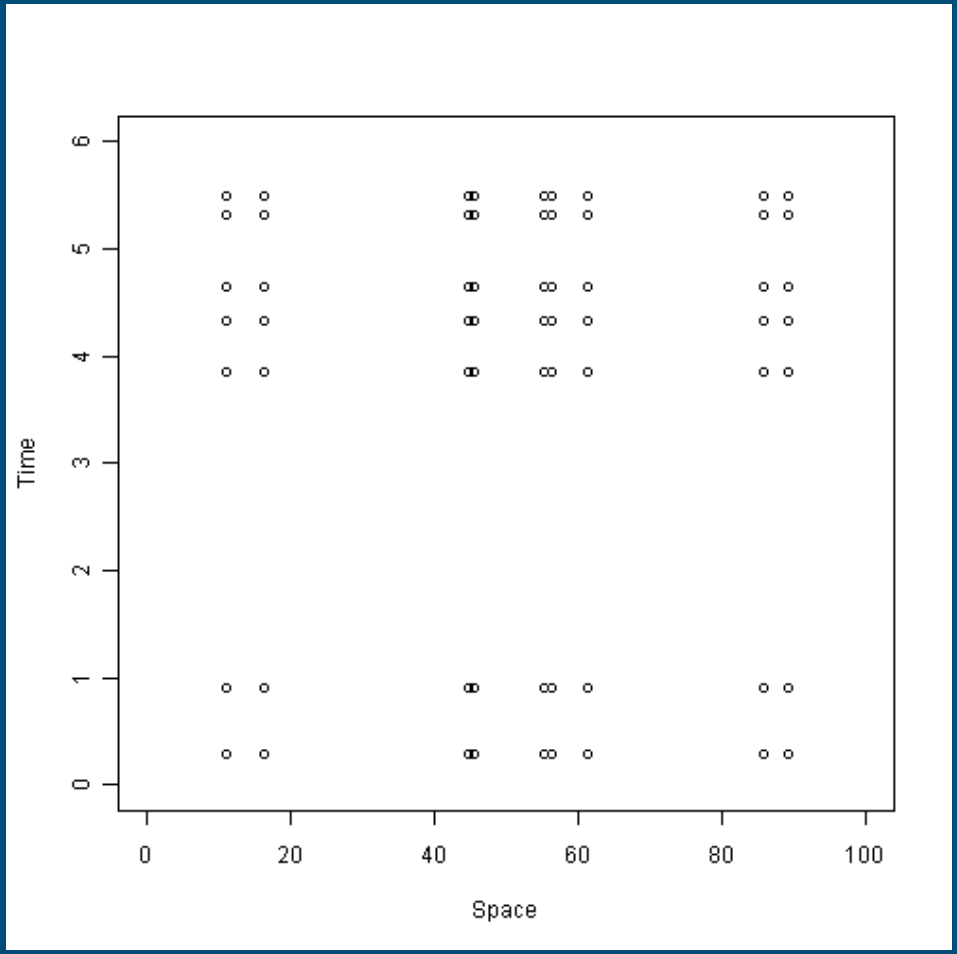


Static synchronous spatial grid *centered*



ALTERRA
WAGENINGEN UR

Static-synchronous, SI of locations, SI of times



ALTERRA
WAGENINGEN UR

Choice of statistical approach and space-time design

- Partly determined by statistical parameter of interest
- Global quantities:
  - Current mean and change of mean (total)
  - Space-time mean (total)
  - Temporal trend of spatial mean (total)

- For current mean and change of mean selection of sampling times not part of design process

Estimation of means and change of means

- Design-based approach: probability sampling of locations

- Fuller and Breidt (1999) proposed to estimate the means and the change of the means by Generalized Least Squares:

$$\hat{\bar{\mathbf{y}}} = \mathbf{X}\boldsymbol{\mu} + \mathbf{e}$$

$\hat{\bar{\mathbf{y}}}$: estimated spatial means at times 1...r

$\mathbf{X}$: design matrix

$\boldsymbol{\mu}$: spatial means at times 1...r

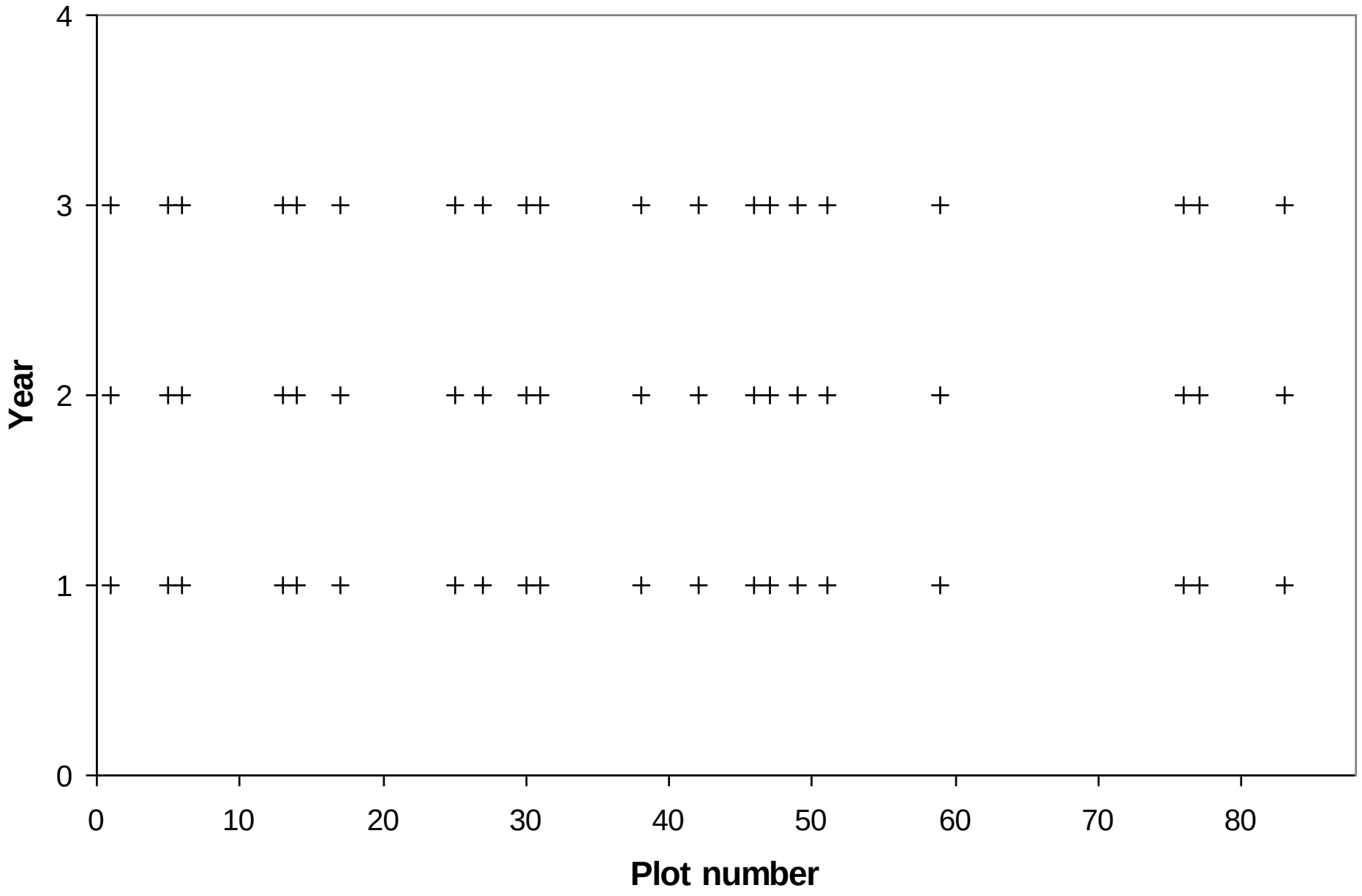
$\mathbf{e}$: sampling errors in $\hat{\bar{\mathbf{y}}}$

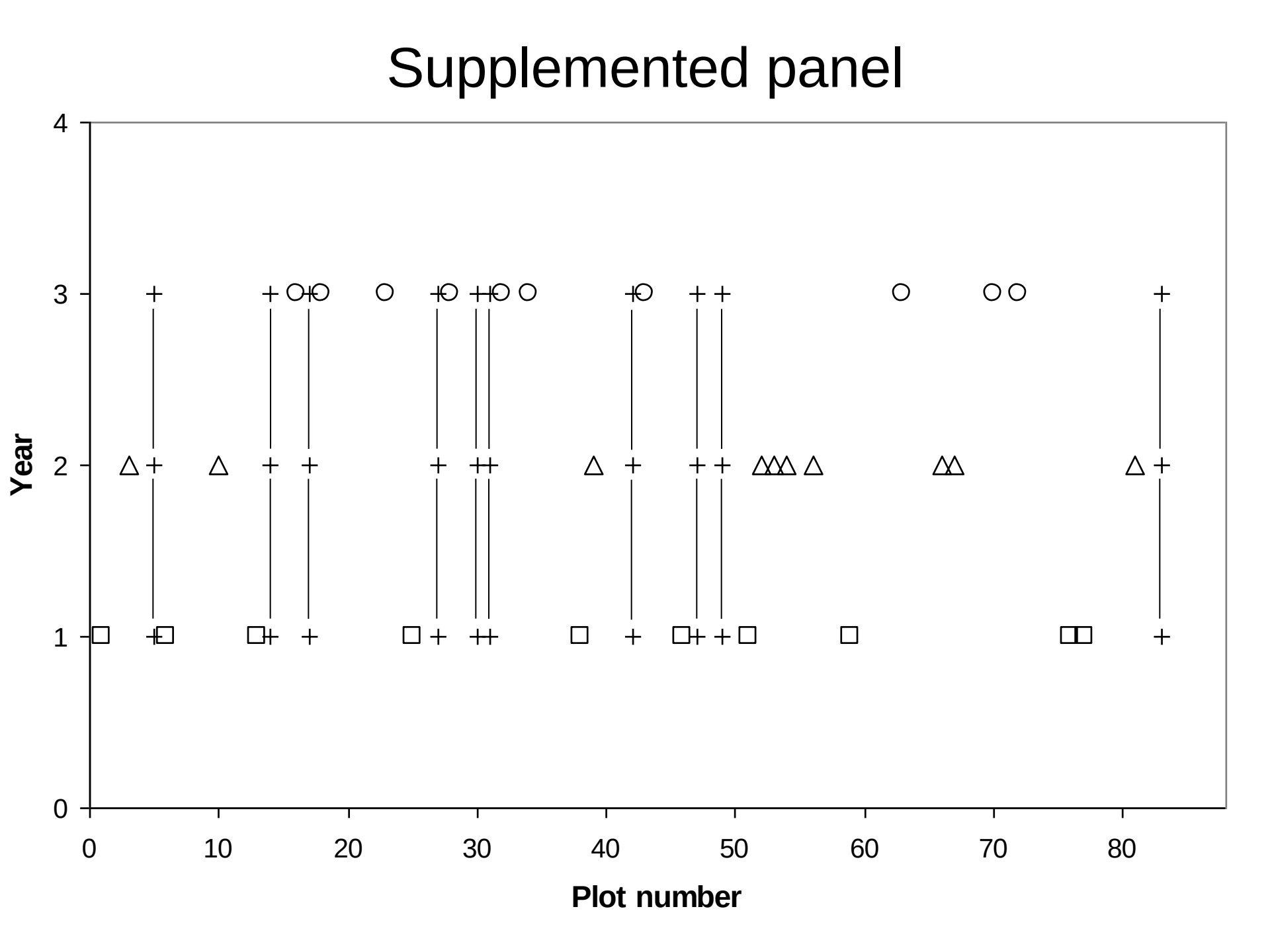
GLS estimation of means and change of means

$$\hat{\mu} = (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C}^{-1} \hat{\hat{\mathbf{y}}}$$

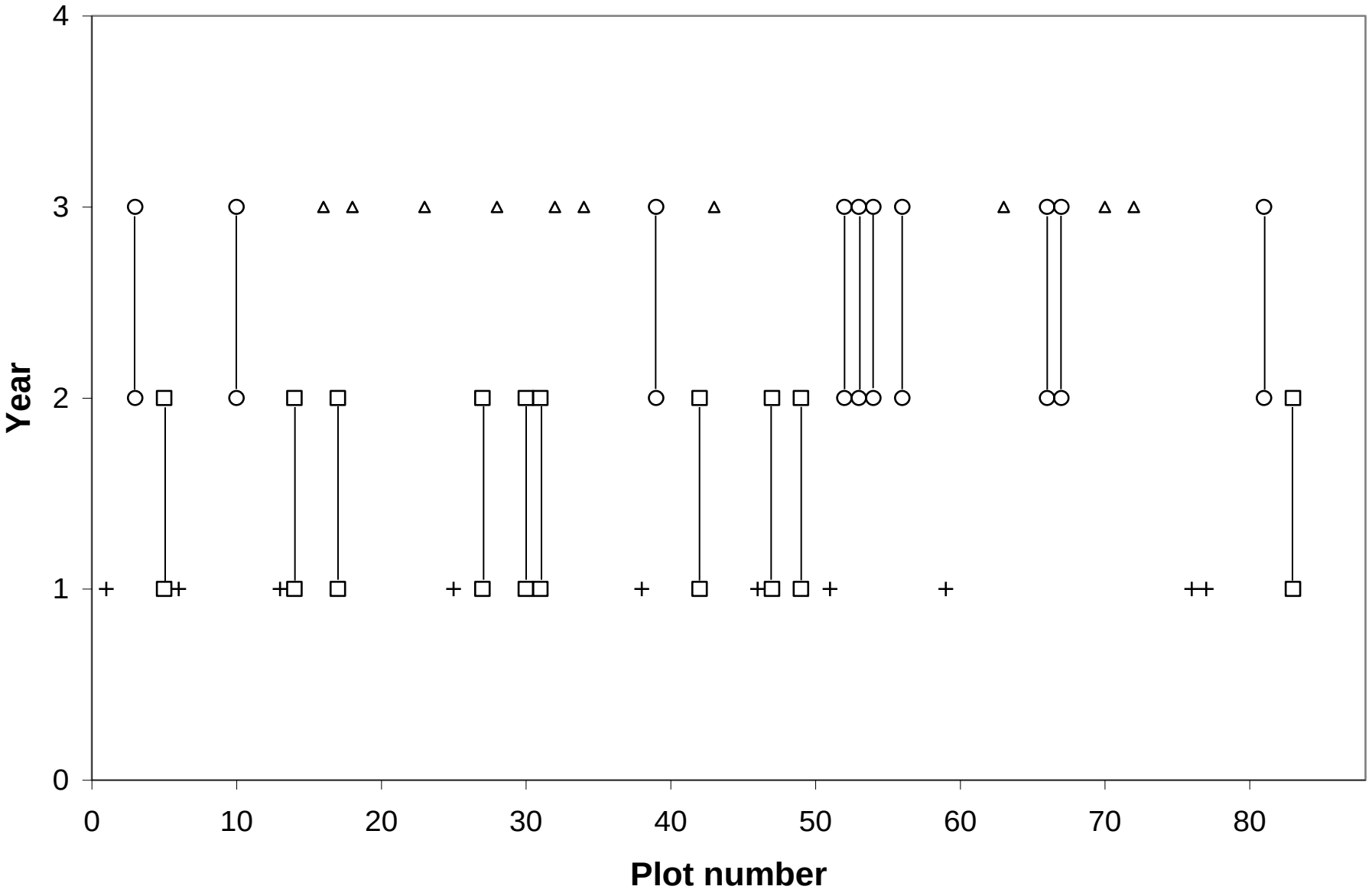
$$\mathrm{Var}(\hat{\mu}) = (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X})^{-1}$$

Pure panel




Supplemented panel
Year
Plot number
0
1
2
3
4
0
10
20
30
40
50
60
70
80

Rotational design



First order autoregressive process

$$\rho_j = \rho^j$$

$$\mathbf{X}_{pure} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{C}_{pure} = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \frac{2\sigma^2}{n}$$

$$\mathbf{X}_{sup} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

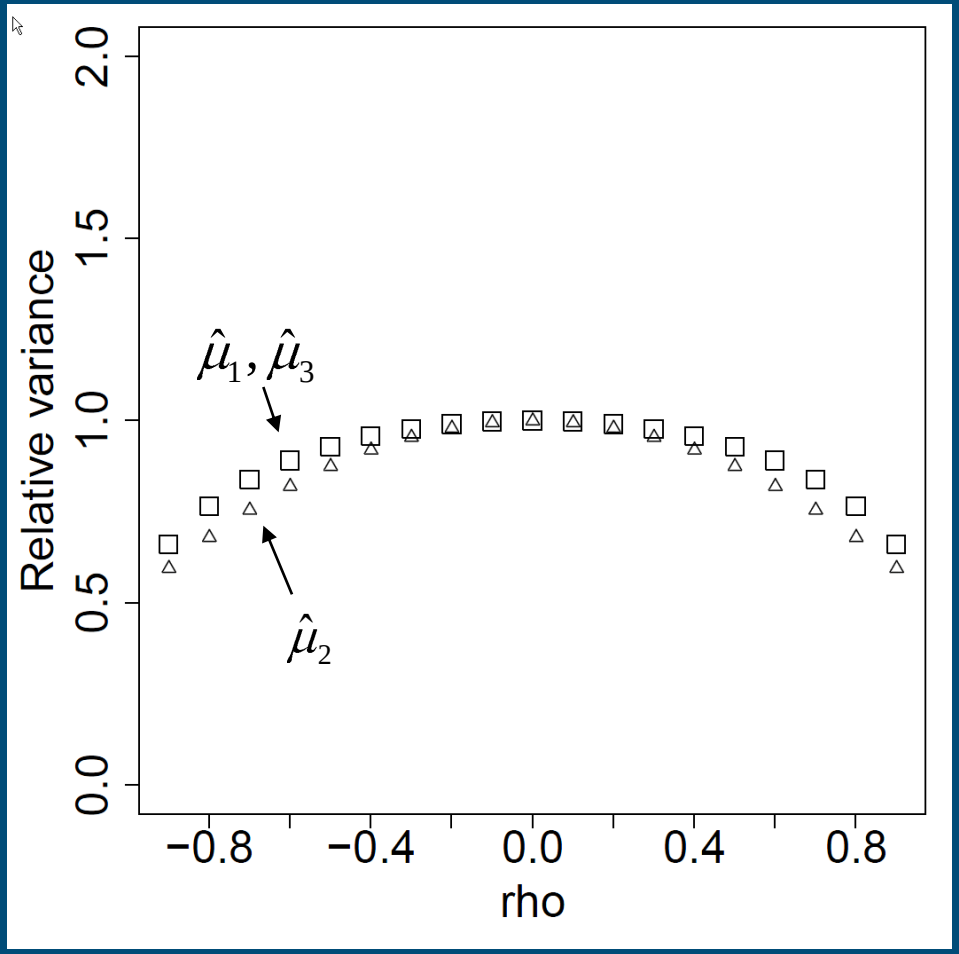
$$\mathbf{C}_{sup} = \begin{pmatrix} 1 & \rho & \rho^2 & 0 & 0 & 0 \\ \rho & 1 & \rho & 0 & 0 & 0 \\ \rho^2 & \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \frac{4\sigma^2}{n}$$

$$\mathbf{X}_{rot} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

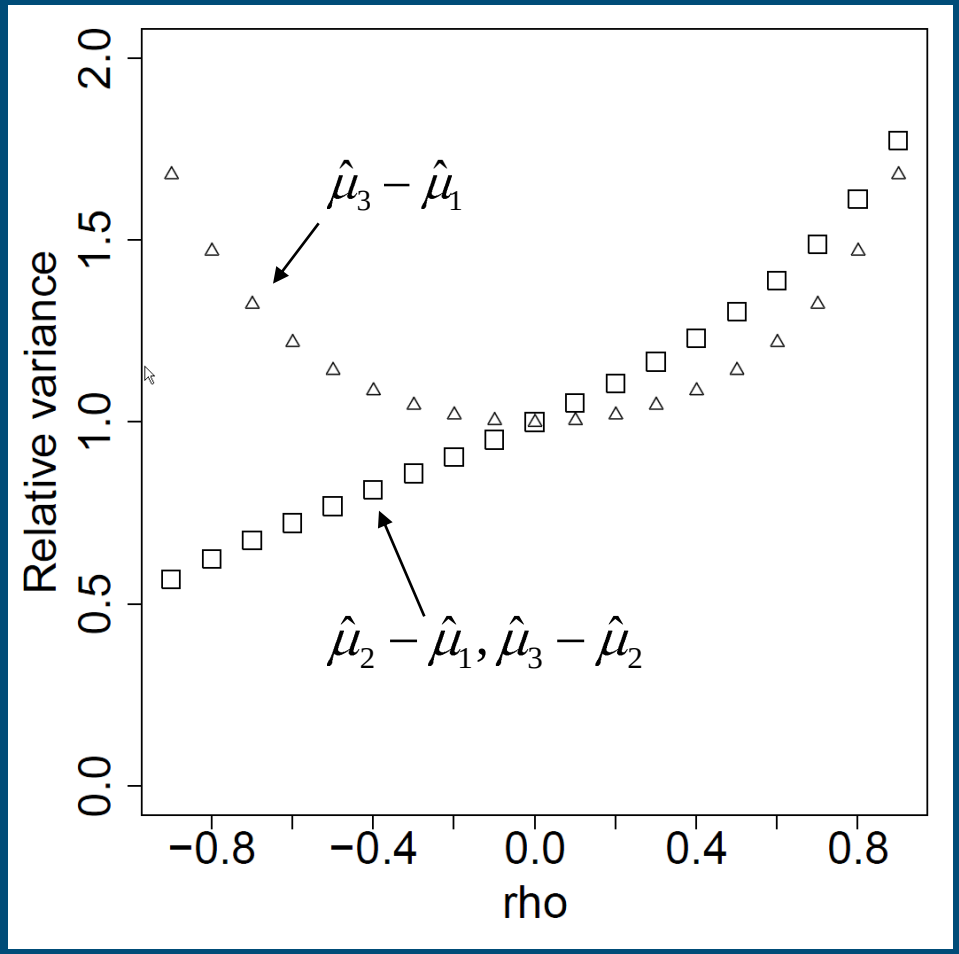
$$\mathbf{C}_{rot} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \rho & 0 & 0 & 0 \\ 0 & \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho & 0 \\ 0 & 0 & 0 & \rho & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \frac{4\sigma^2}{n}$$

ALTERRA
WAGENINGEN UR

Mean



Change of mean



ALTERRA
WAGENINGEN UR

Space-time mean & temporal trend of spatial mean

- How to select sampling locations *and* sampling times?
- Both can be selected either by probability sampling or non-probability sampling
- This leads to four basic statistical sampling approaches

Four approaches for space-time sampling

SPACE

TIME	Design-based	Model-based
Design-based	$D_S D_T$	$(M_S D_T)$
Model-based	$D_S M_T$	$M_S M_T$

Space-time mean (total)

- Space-time mean concentration of nutrients in soil, groundwater or surface water
- Space-time total of greenhouse gas emissions in an area
- Fully design-based approach can be attractive option
- D.J. Brus & M. Knotters, 2008. Sampling design for compliance monitoring of surface water quality: A case study in a polder area. Water Res. Res. 44, W11410

Temporal trend of spatial mean

- For fully model-based approacH: ter Braak et al (2008), JABES 13.
- Hybrid sampling approach
- Probability sampling in space at all sampling times
- Model-free design-based estimation of spatial means
- Systematic sampling in time, constant interval, first time at start, last time at end
- Stochastic time-series model for spatial mean

Linear mixed model for spatial means

$$\bar{Y}(t) = \sum_{j=1}^{q} \beta_j x_j(t) + \eta(t)$$

with $\eta(t)$ the model inadequacy error (process error), mean 0 and covariance matrix $\mathbf{C}_\xi$

Linear mixed model for *estimated* means

$\bar{Y}(t)$ unknown, must be estimated $\hat{\bar{Y}}(t) = \frac{1}{\|\mathcal{A}\|} \sum_{i=1}^{n} \frac{Y_i(t)}{\pi_i}$:

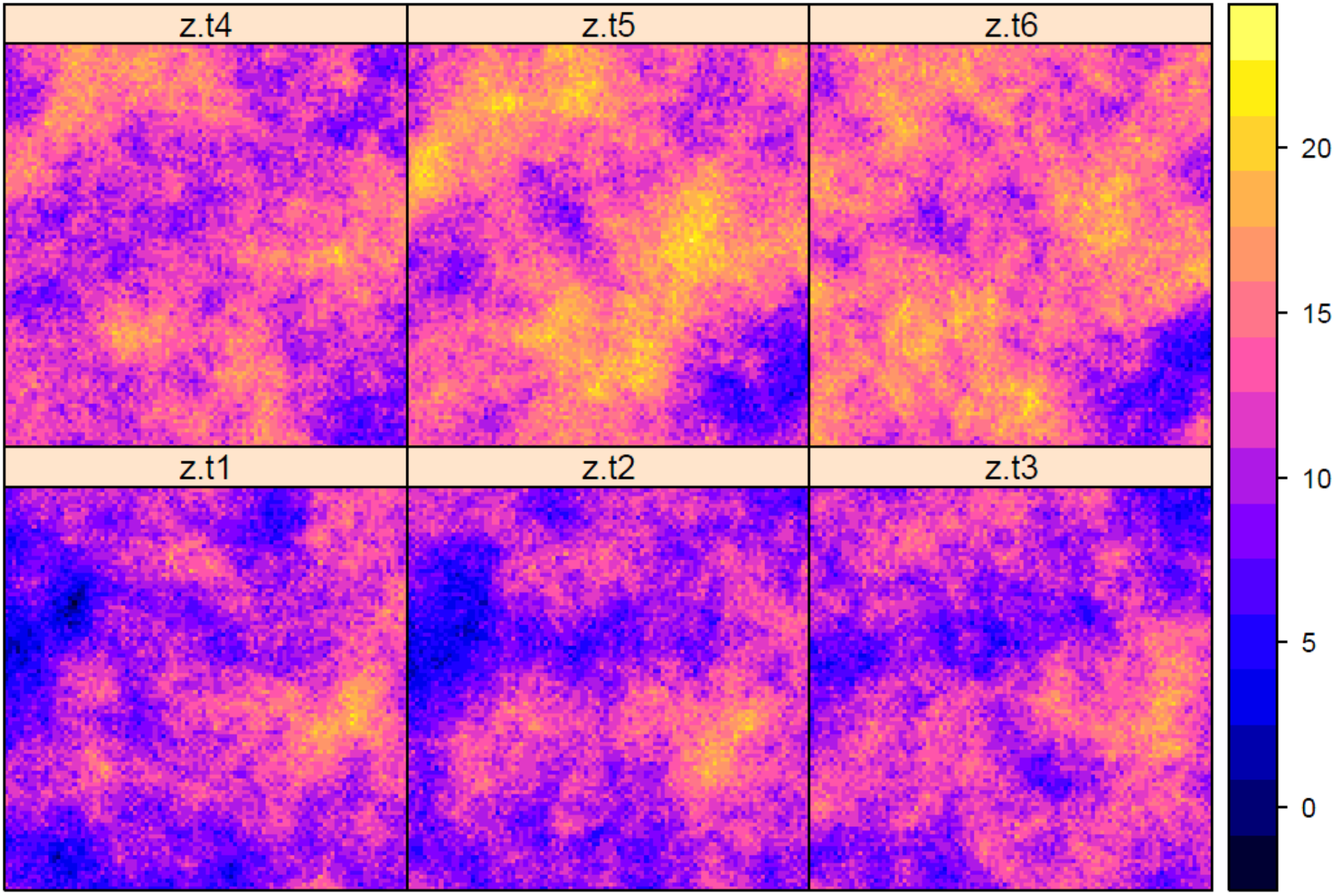
$$\hat{\bar{Y}}(t) = \sum_{j=1}^{q} \beta_j x_j(t) + \eta(t) + \varepsilon(t)$$

with $\varepsilon(t)$ the sampling error, mean 0 and covariance matrix $\mathbf{C}_p$, and $\varepsilon(t)$ and $\eta(t)$ independent

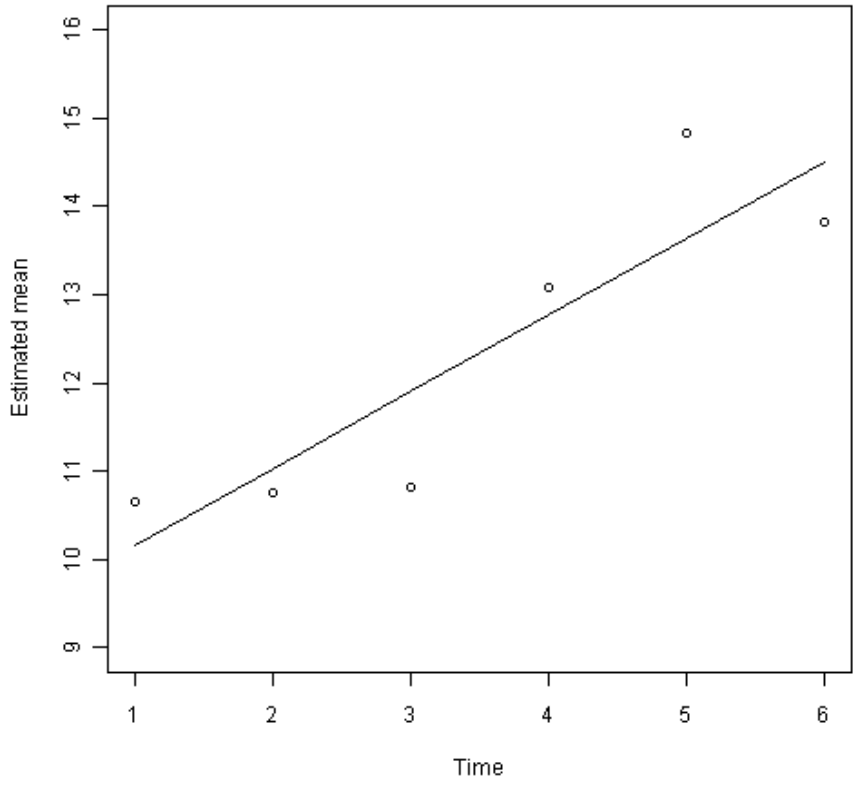
If we take $x_1(t) = 1$ and $x_2(t) = t$, then

$$\hat{\bar{Y}}(t) = \beta_1 + \beta_2 \cdot t + \eta(t) + \varepsilon(t)$$

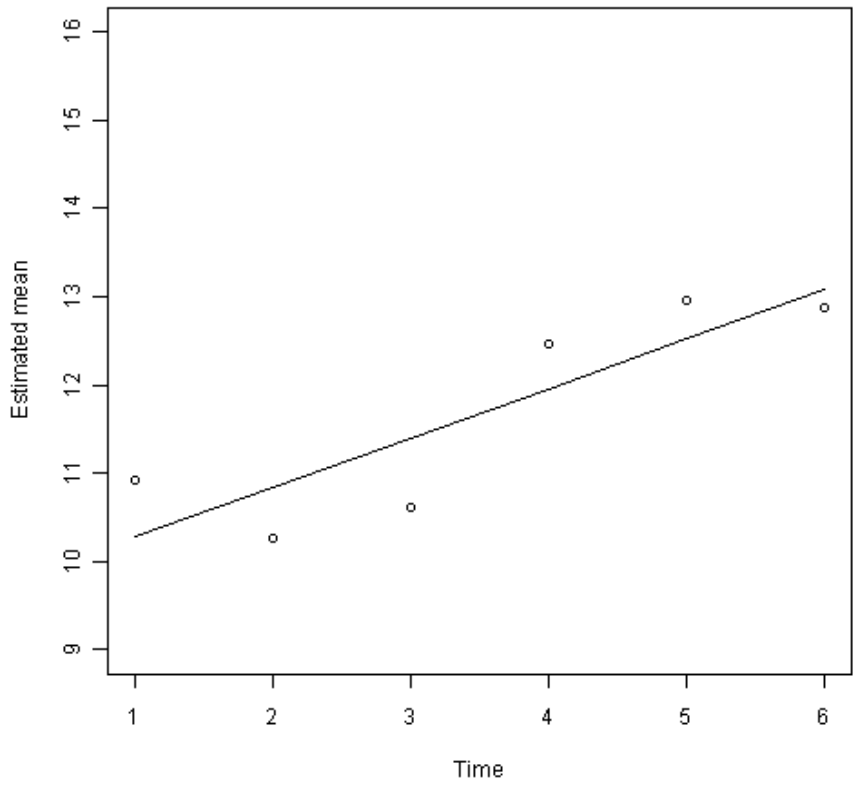
with β_2 the linear trend parameter to be estimated


z.t4
z.t5
z.t6
z.t1
z.t2
z.t3
20
15
10
5
0

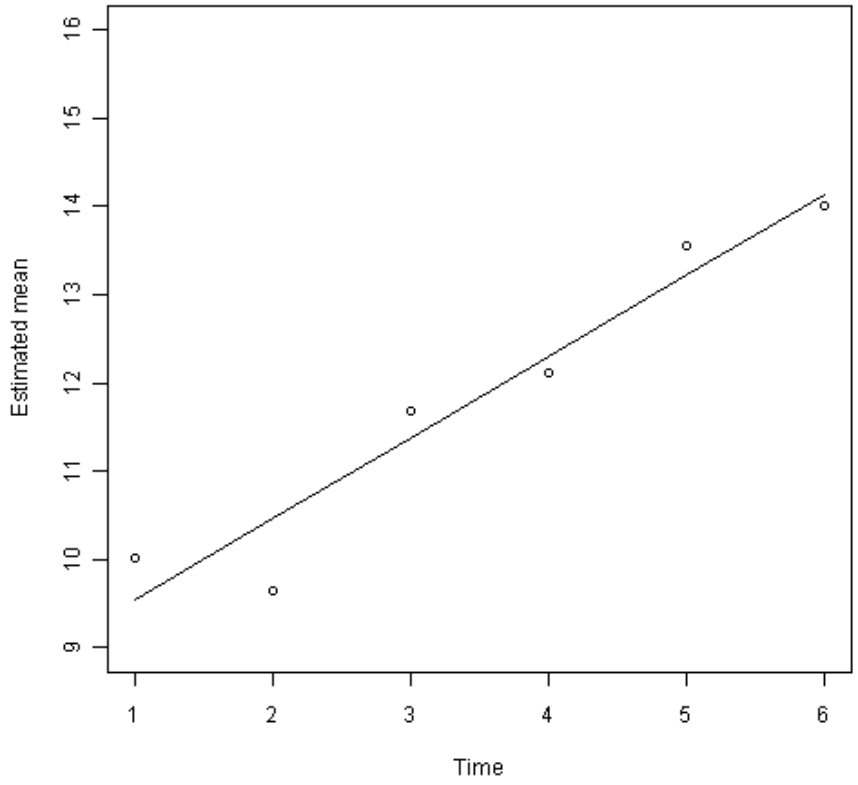
Space-time sample 1 (static-synchronous, SI in space)



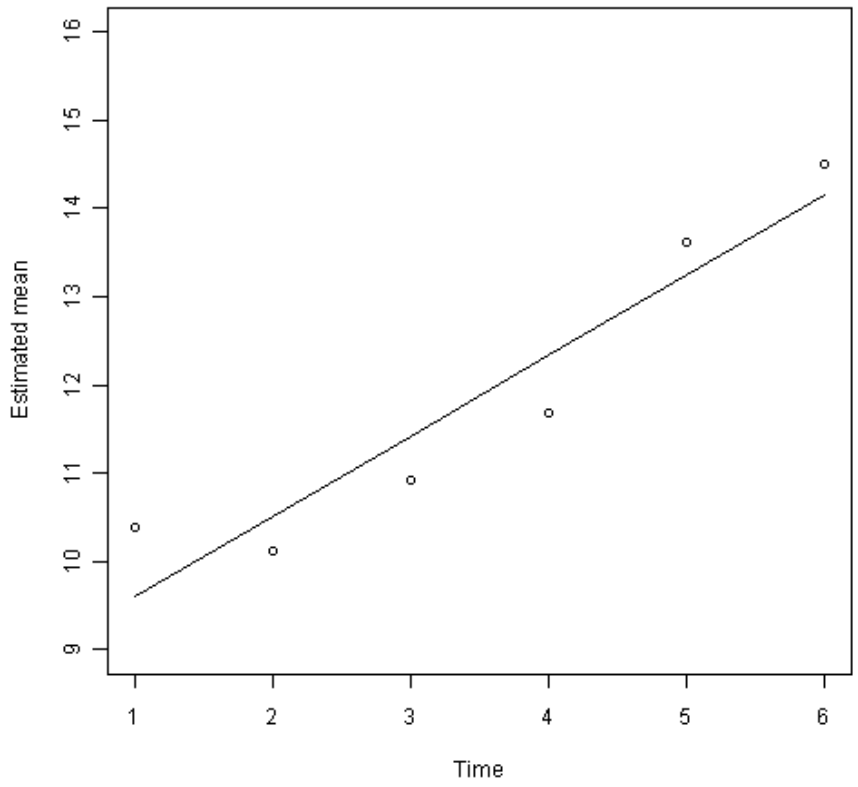
Space-time sample 2



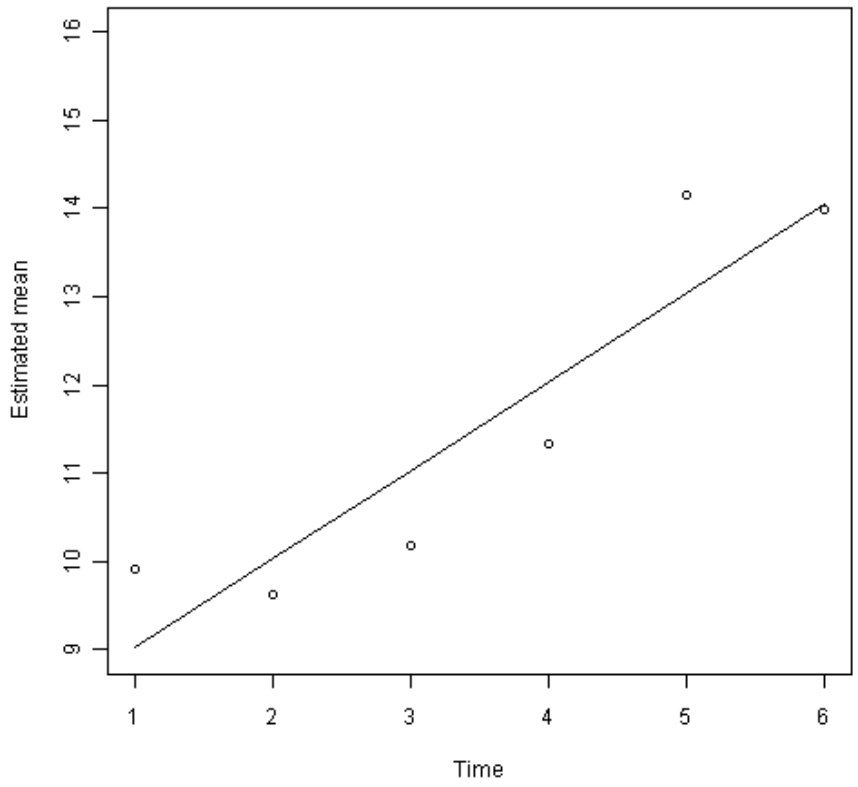
Space-time sample 3



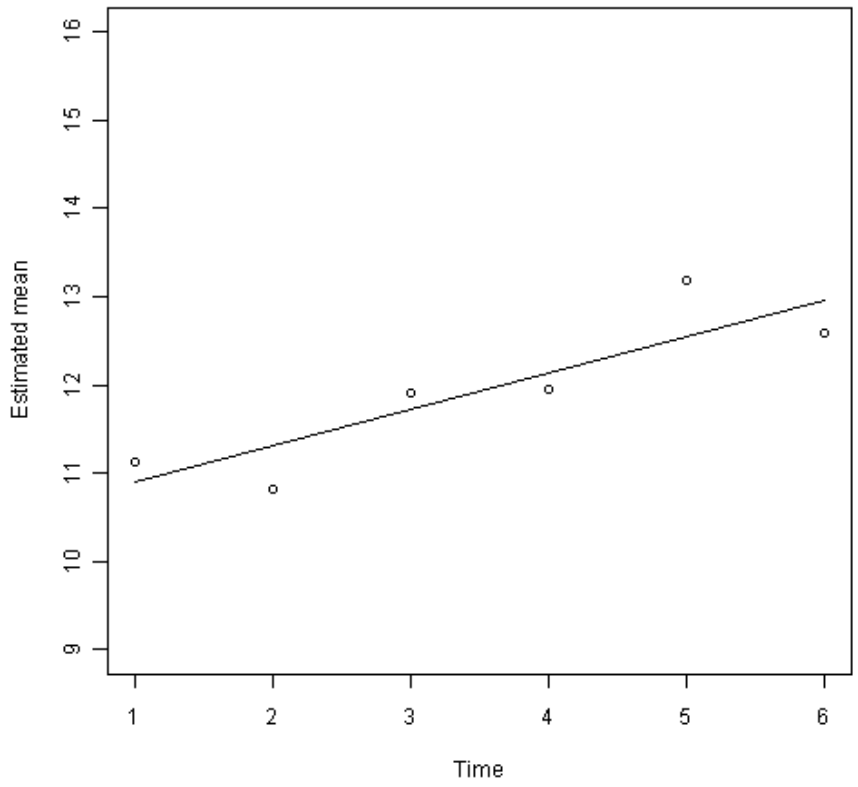
Space-time sample 4



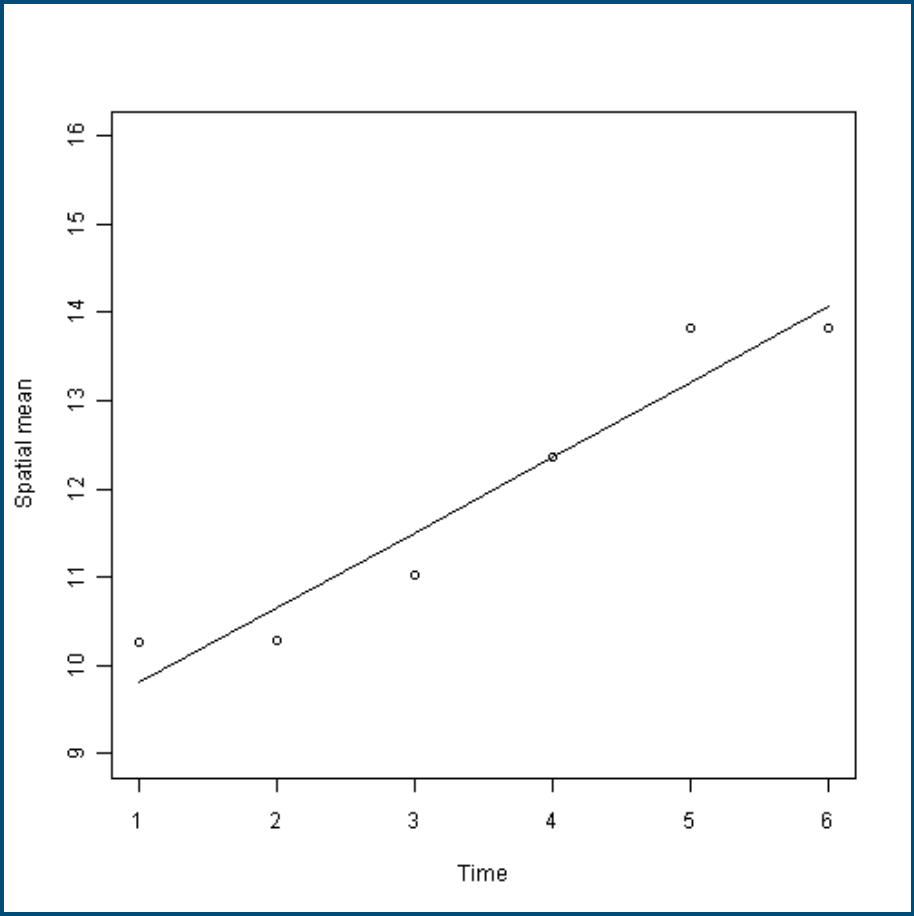
Space-time sample 5



Space-time sample 1000



Exhaustive space-time sample



GLS estimation of trend

- $$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}_{\xi p}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{C}_{\xi p}^{-1}\hat{\mathbf{y}})$$

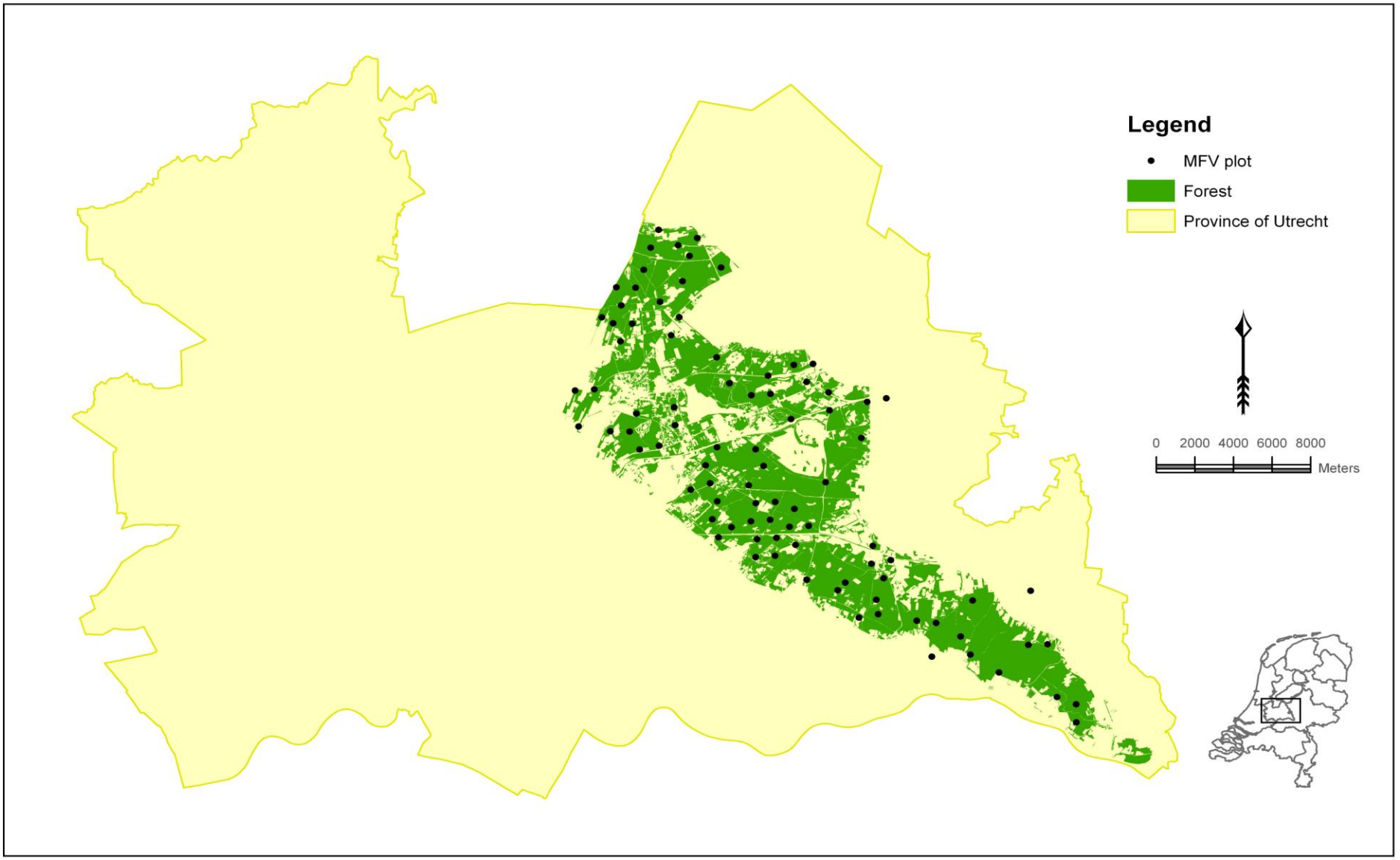
with

$$\mathbf{C}_{\xi p} = \mathbf{C}_{\xi} + \mathbf{C}_{p},$$

- $$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{C}_{\xi p}^{-1}\mathbf{X})^{-1}$$

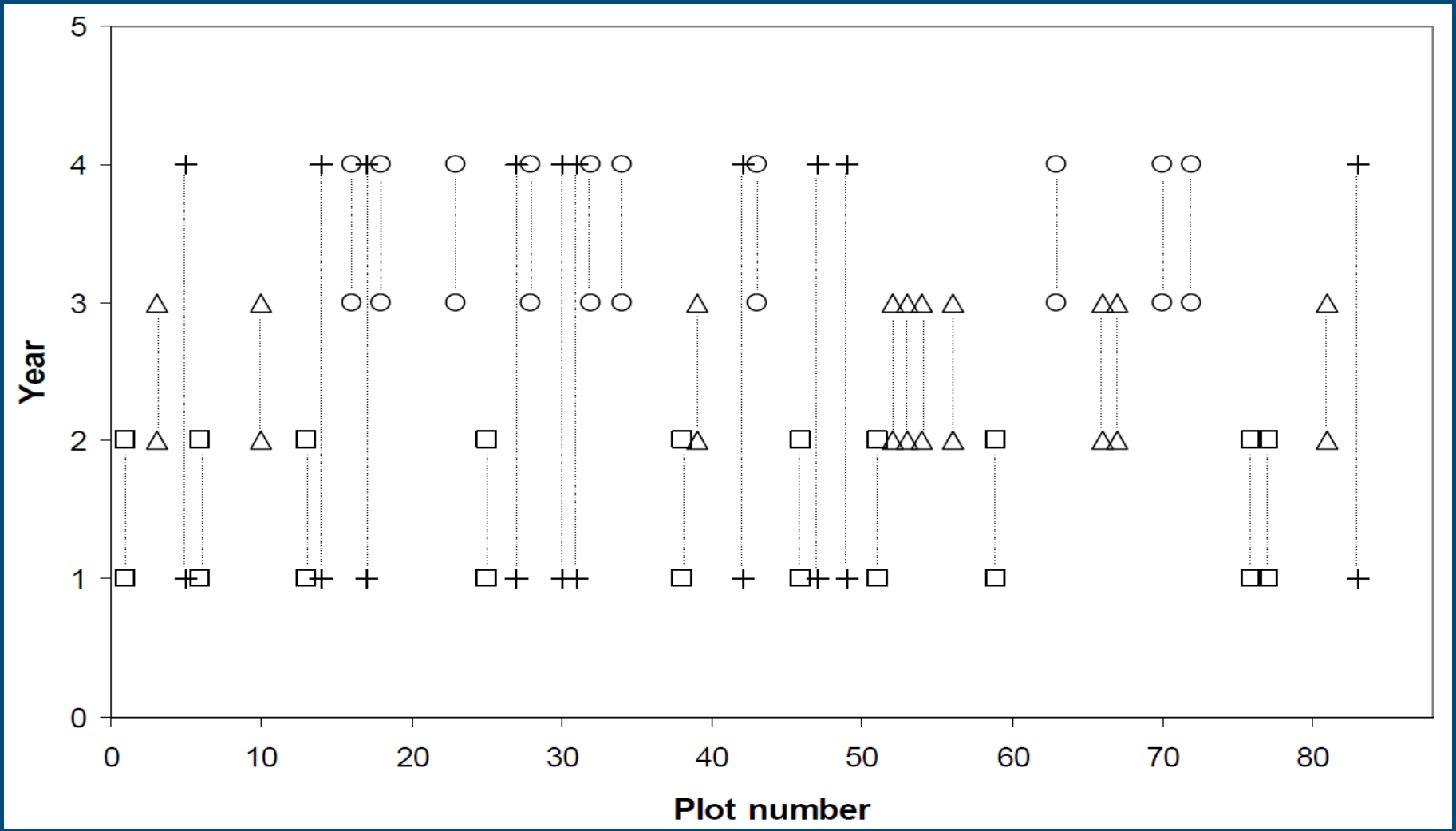
- $$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \mathrm{Var}_{\xi}\left\{\mathrm{E}_{p}(\hat{\boldsymbol{\beta}})|\xi_0\right\} + \mathrm{E}_{\xi}\left\{\mathrm{Var}_{p}(\hat{\boldsymbol{\beta}})|\xi_0\right\}$$

- $$\mathrm{Var}_{\xi}\left\{\mathrm{E}_{p}(\hat{\boldsymbol{\beta}})|\xi_0\right\} = (\mathbf{X}'\mathbf{C}_{\xi}^{-1}\mathbf{X})^{-1}\ ,$$

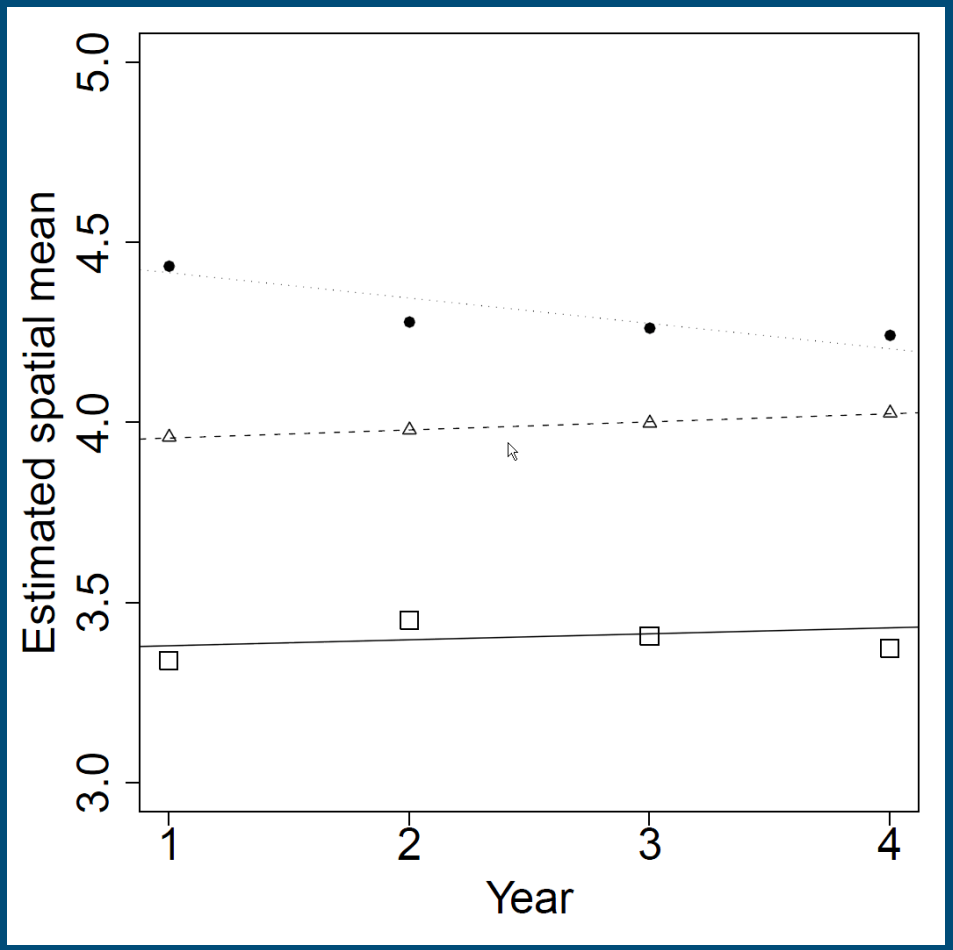


ALTERRA
WAGENINGEN UR

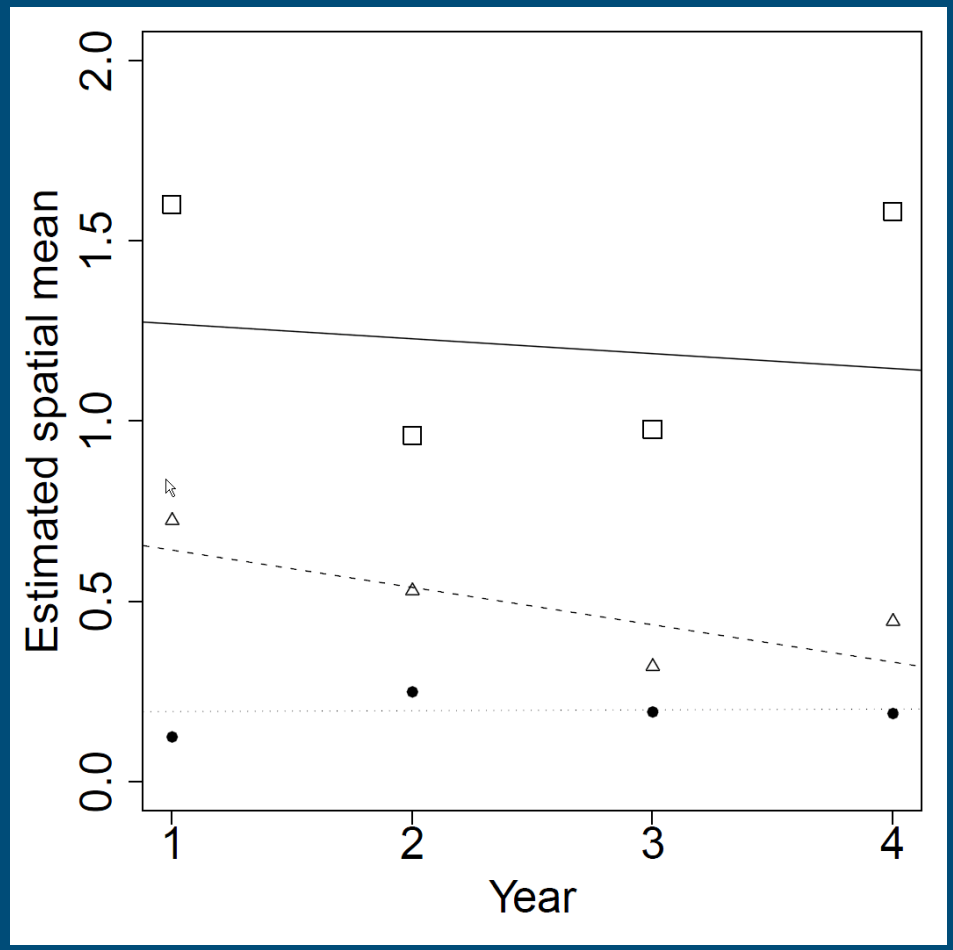
Rotational subsample



pH

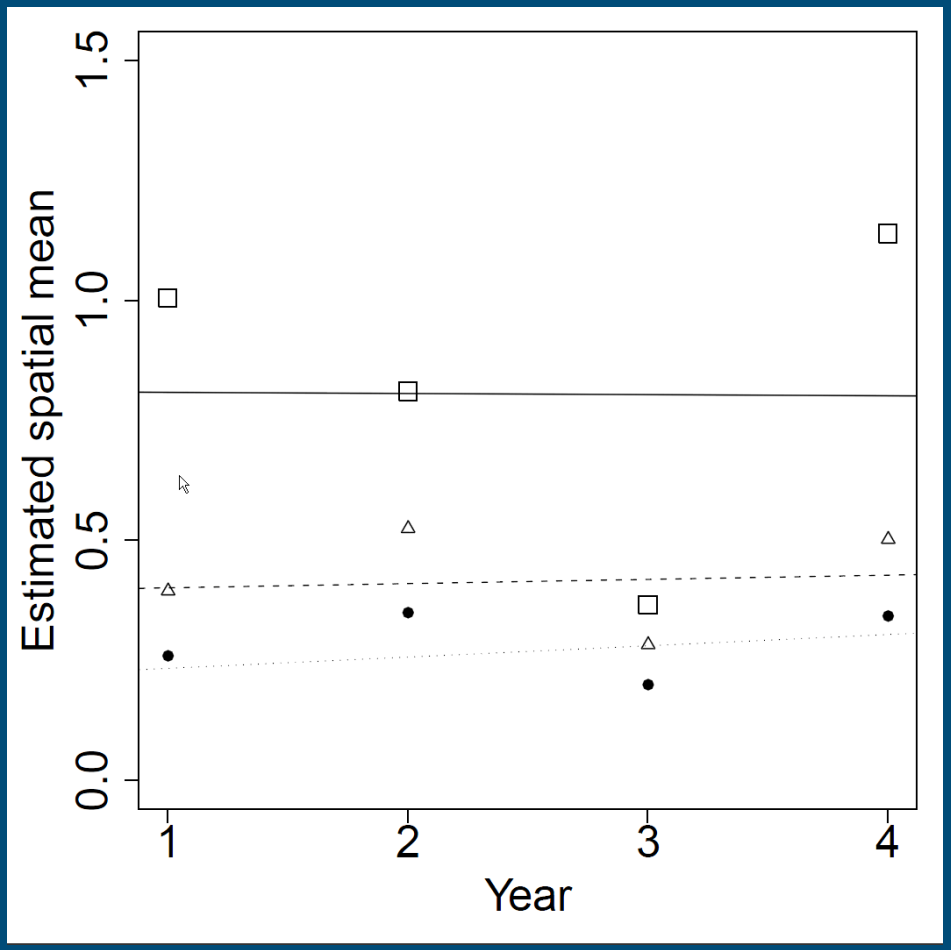


NH_4

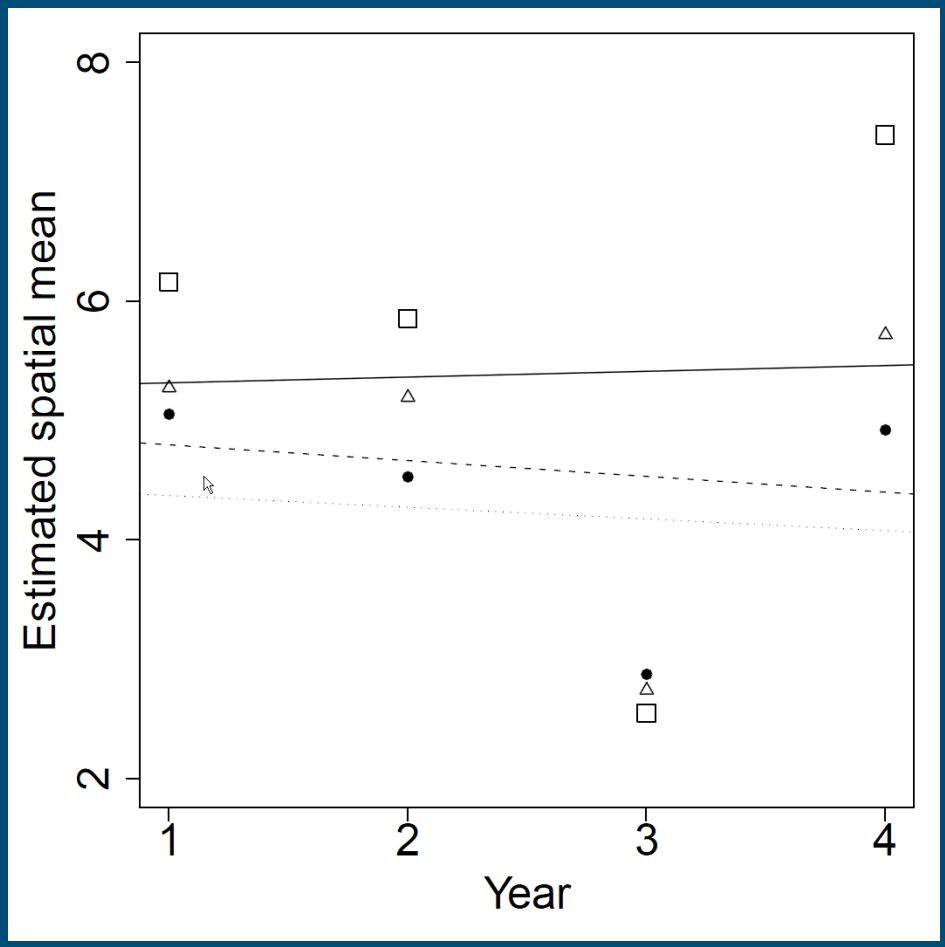


ALTERRA
WAGENINGEN UR

NO_3



NO_3 soil solution



ALTERRA
WAGENINGEN UR

Estimated trend and standard errors. Numbers multiplied by 10^4

Comp	Depth	trend	se	se_p	se_ξ
pH	top	170	280	280	0
pH	mid	230	240	240	0
pH	sub	-700	210	150	130
NO_3	top	-24	1900	700	1800
NO_3	mid	88	740	470	550
NO_3	sub	240	460	300	310
NH_4	top	-410	2300	880	1700
NH_4	mid	*-1000*	630	360	470
NH_4	sub	23	280	160	200
NO_3 ss	top	480	12000	380	11000
NO_3 ss	mid	-1300	8200	4500	6800
NO_3 ss	sub	-980	6400	4200	4600

Conclusions (1)

- There are many, many possibilities for sampling in space and time
- For global quantities probability sampling of locations recommendable; it enables model-free unbiased estimation of spatial means
- Probability sampling of times can be attractive for estimating space-time means (totals), but undesirable for estimating temporal trends

Conclusions (2)

- Suitability of basic type of space-time design depends on statistical parameter of interest
- Estimates of sampling (co)variances of estimated spatial means can be used in GLS estimation of means, change of means and of temporal trends
- Still much to be explored and sorted out

References

- J.J. de Gruijter, D.J. Brus, M.F.P. Bierkens & M. Knotters, 2006. Sampling for Natural Resource Monitoring. Springer
- D.J. Brus & M. Knotters, 2008. Sampling design for compliance monitoring of surface water quality: A case study in a polder area. Water Res. Res. 44, W11410
- D.J. Brus and I. Noij. 2008. Designing sampling schemes for effect monitoring of nutrient leaching from agricultural soils. EJSS 59
- C.J ter Braak, D.J Brus & E. Pebesma (2008) Comparing Sampling Patterns for Kriging the Spatial Mean Temporal Trend, JABES 13.
- M. Knotters & D.J. Brus. 2010. Estimating space-time mean concentrations of nutrients in surface-waters of variable depth, Wat. Res. Res. (in press).
- D.J. Brus & J.J. de Gruijter. A hybrid design-based and model-based sampling approach to estimate temporal trends of spatial means JABES (tentatively accepted)