Design-based Generalized Least Squares Estimation of Status and Trend

Dick Brus[1], Jaap de Gruijter [1]

[1]Alterra, Wageningen University and Research Centre

Monitoring

- Important issue in society and scientific research
- Proposition: quality of the monitoring result and the efficiency of whole operation strongly determined by the sampling design
- How many locations and where? How frequent and when?
- Revisit locations or not?

Monitoring

- Important issue in society and scientific research
- Proposition: quality of the monitoring result and the efficiency of whole operation strongly determined by the sampling design
- How many locations and where? How frequent and when?
- Revisit locations or not?

Monitoring

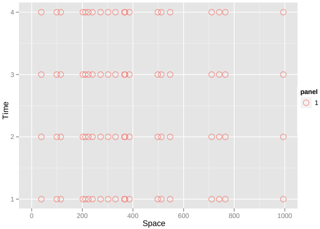
- Important issue in society and scientific research
- Proposition: quality of the monitoring result and the efficiency of whole operation strongly determined by the sampling design
- How many locations and where? How frequent and when?
- Revisit locations or not?

Monitoring

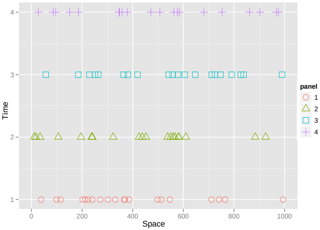
- Important issue in society and scientific research
- Proposition: quality of the monitoring result and the efficiency of whole operation strongly determined by the sampling design
- How many locations and where? How frequent and when?
- Revisit locations or not?

Basic types of space–time design

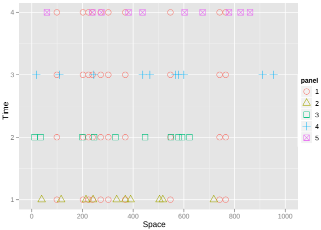
Static-synchronous (pure panel)



Independent synchronous



Supplemented panel



Selection mode of locations and times

Four possible combinations of probability and non-probability sampling

Mode	Space	Time
A	Non-probability	Non-probability
B	Probability	Probability
C	Probability	Non-probability
D	Non-probability	Probability

- Existing networks: mode A, occasionally C
- For estimating space–time means and totals: mode B recommendable (fully design-based inference, see e.g. Brus and Knotters, 2008)
- This presentation: mode C

Selection mode of locations and times

Four possible combinations of probability and non-probability sampling

Mode	Space	Time
A	Non-probability	Non-probability
B	Probability	Probability
C	Probability	Non-probability
D	Non-probability	Probability

- Existing networks: mode A, occasionally C
- For estimating space–time means and totals: mode B recommendable (fully design-based inference, see e.g. Brus and Knotters, 2008)
- This presentation: mode C

Selection mode of locations and times

Four possible combinations of probability and non-probability sampling

Mode	Space	Time
A	Non-probability	Non-probability
B	Probability	Probability
C	Probability	Non-probability
D	Non-probability	Probability

- Existing networks: mode A, occasionally C
- For estimating space–time means and totals: mode B recommendable (fully design-based inference, see e.g. Brus and Knotters, 2008)
- This presentation: mode C

Aim of research

- Central question: how to estimate spatial means and trend of spatial means from partially overlapping sample data obtained by selection mode C?
- To introduce and demonstrate GLS estimation
- Comparison of five space–time designs

Aim of research

- Central question: how to estimate spatial means and trend of spatial means from partially overlapping sample data obtained by selection mode C?
- To introduce and demonstrate GLS estimation
- Comparison of five space–time designs

Aim of research

- Central question: how to estimate spatial means and trend of spatial means from partially overlapping sample data obtained by selection mode C?
- To introduce and demonstrate GLS estimation
- Comparison of five space–time designs

Naive estimation method for partially overlapping samples

- Estimate spatial mean at time t using data of time t only
- Estimate trend by Ordinary Least Squares fitting of simple linear model to the estimated spatial means, using sampling time t as predictor
- Suboptimal for partially overlapping samples: no use is made of temporal correlation of co-located observations

Naive estimation method for partially overlapping samples

- Estimate spatial mean at time t using data of time t only
- Estimate trend by Ordinary Least Squares fitting of simple linear model to the estimated spatial means, using sampling time t as predictor
- Suboptimal for partially overlapping samples: no use is made of temporal correlation of co-located observations

Naive estimation method for partially overlapping samples

- Estimate spatial mean at time t using data of time t only
- Estimate trend by Ordinary Least Squares fitting of simple linear model to the estimated spatial means, using sampling time t as predictor
- Suboptimal for partially overlapping samples: no use is made of temporal correlation of co-located observations

Less naive method: Generalized Least Squares

Spatial mean at time t is estimated as

- weighted average of *all* 'elementary' estimates
- weights determined by the *sampling variances and covariances* of the elementary estimates

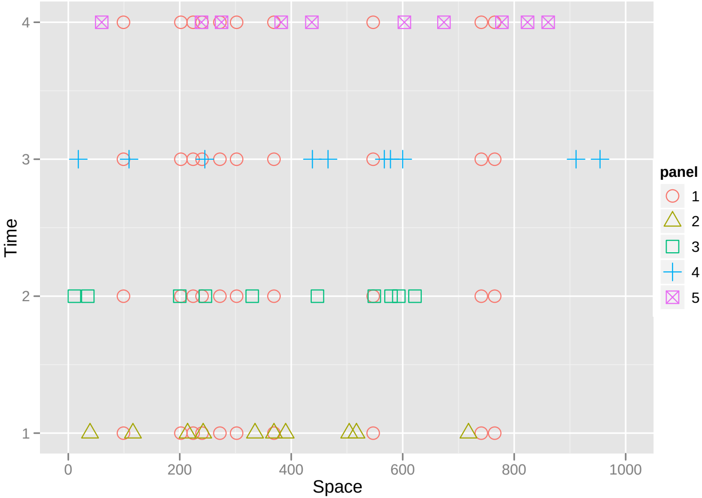
Less naive method: Generalized Least Squares

Spatial mean at time t is estimated as

- weighted average of *all* ‘elementary’ estimates
- weights determined by the *sampling variances and covariances* of the elementary estimates

Panels and elementary estimates

Supplemented panel: 4×2 elementary estimates



GLS estimation of spatial means

- $$\hat{\bar{y}}_i(t_j) = \bar{y}(t_j) + \varepsilon_i(t_j)$$

  with $\varepsilon_i(t_j)$ the *sampling* error of i^{th} elementary estimate of spatial mean at time t_j
- $$\hat{\bar{\mathbf{y}}} = \mathbf{X}\bar{\mathbf{y}} + \boldsymbol{\epsilon}$$

  with $\mathbf{X}$ the $(P \times r)$ design-matrix with 0's and 1's
- $$\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\hat{\bar{\mathbf{y}}}$$

  with $\mathbf{C}$ the *sampling* variance-covariance matrix of elementary estimates
- $$\mathrm{Cov}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}) = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}$$

GLS estimation of spatial means

- $$\hat{\bar{y}}_i(t_j) = \bar{y}(t_j) + \varepsilon_i(t_j)$$

  with $\varepsilon_i(t_j)$ the *sampling* error of i^{th} elementary estimate of spatial mean at time t_j

- $$\hat{\bar{\mathbf{y}}} = \mathbf{X}\bar{\mathbf{y}} + \boldsymbol{\epsilon}$$

  with $\mathbf{X}$ the $(P \times r)$ design-matrix with 0's and 1's

- $$\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\hat{\bar{\mathbf{y}}}$$

  with $\mathbf{C}$ the *sampling* variance-covariance matrix of elementary estimates

- $$\mathrm{Cov}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}) = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}$$

GLS estimation of spatial means

- $$\hat{\bar{y}}_i(t_j) = \bar{y}(t_j) + \varepsilon_i(t_j)$$

  with $\varepsilon_i(t_j)$ the *sampling* error of i^{th} elementary estimate of spatial mean at time t_j

- $$\hat{\bar{\mathbf{y}}} = \mathbf{X}\bar{\mathbf{y}} + \boldsymbol{\epsilon}$$

  with $\mathbf{X}$ the $(P \times r)$ design-matrix with 0's and 1's

- $$\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\hat{\bar{\mathbf{y}}}$$

  with $\mathbf{C}$ the *sampling* variance-covariance matrix of elementary estimates

- $$\mathrm{Cov}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}) = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}$$

GLS estimation of spatial means

- $$\hat{\bar{y}}_i(t_j) = \bar{y}(t_j) + \varepsilon_i(t_j)$$

  with $\varepsilon_i(t_j)$ the *sampling* error of i^{th} elementary estimate of spatial mean at time t_j

- $$\hat{\bar{\mathbf{y}}} = \mathbf{X}\bar{\mathbf{y}} + \boldsymbol{\epsilon}$$

  with $\mathbf{X}$ the $(P \times r)$ design-matrix with 0's and 1's

- $$\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\hat{\bar{\mathbf{y}}}$$

  with $\mathbf{C}$ the *sampling* variance-covariance matrix of elementary estimates

- $$\mathsf{Cov}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}) = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}$$

Linear trend defined as a population parameter

- $$b = \frac{\sum_{j=1}^{r}(t_j - \bar{t})(\bar{y}_j - \bar{\bar{y}})}{\sum_{j=1}^{r}(t_j - \bar{t})^2} = \frac{\sum_{j=1}^{r}(t_j - \bar{t})\bar{y}_j}{\sum_{j=1}^{r}(t_j - \bar{t})^2} = \sum_{j=1}^{r} w_j \bar{y}_j$$

  with

  $$w_j = \frac{t_j - \bar{t}}{\sum_{j=1}^{r}(t_j - \bar{t})^2}$$

- $$\hat{b} = \mathbf{w}'\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}$$

- $$\mathrm{Var}(\hat{b}) = \mathbf{w}'\mathsf{Cov}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}})\mathbf{w}$$

Linear trend defined as a population parameter

▶

$$b = \frac{\sum_{j=1}^{r}(t_j - \bar{t})(\bar{y}_j - \bar{\bar{y}})}{\sum_{j=1}^{r}(t_j - \bar{t})^2} = \frac{\sum_{j=1}^{r}(t_j - \bar{t})\bar{y}_j}{\sum_{j=1}^{r}(t_j - \bar{t})^2} = \sum_{j=1}^{r} w_j \bar{y}_j$$

with

$$w_j = \frac{t_j - \bar{t}}{\sum_{j=1}^{r}(t_j - \bar{t})^2}$$

▶

$$\hat{b} = \mathbf{w}'\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}$$

▶

$$\mathrm{Var}(\hat{b}) = \mathbf{w}'\mathsf{Cov}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}})\mathbf{w}$$

Linear trend defined as a population parameter

- $$b = \frac{\sum_{j=1}^{r}(t_j - \bar{t})(\bar{y}_j - \bar{\bar{y}})}{\sum_{j=1}^{r}(t_j - \bar{t})^2} = \frac{\sum_{j=1}^{r}(t_j - \bar{t})\bar{y}_j}{\sum_{j=1}^{r}(t_j - \bar{t})^2} = \sum_{j=1}^{r} w_j \bar{y}_j$$

  with

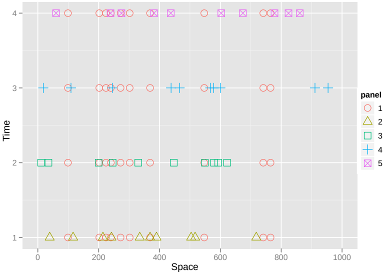
  $$w_j = \frac{t_j - \bar{t}}{\sum_{j=1}^{r}(t_j - \bar{t})^2}$$

- $$\hat{b} = \mathbf{w}'\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}$$

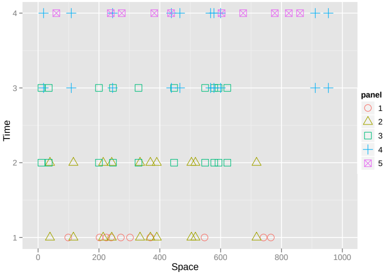
- $$\mathrm{Var}(\hat{b}) = \mathbf{w}'\mathsf{Cov}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}})\mathbf{w}$$

Two space-time designs with 50% overlap

Supplemented panel

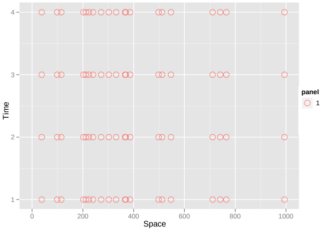


Rotating panel (in-for-2)

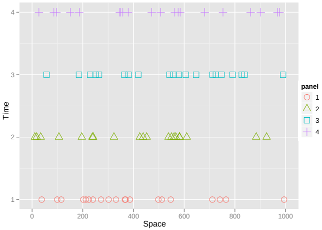


Three space–time designs with 100% or 0% overlap

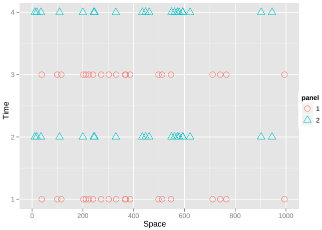
Static-synchronous



Independent synchronous



Serially alternating



GLS estimator for SS, IS and SA

- $$\mathbf{X} = \mathbf{I}$$
- $$\hat{\hat{\mathbf{y}}}_{\mathrm{GLS}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\hat{\hat{\mathbf{y}}} = \hat{\hat{\mathbf{y}}}$$
- $$\mathrm{Var}(\hat{\hat{\mathbf{y}}}_{\mathrm{GLS}}) = \mathrm{Var}(\hat{\hat{\mathbf{y}}}) = \mathbf{C}$$

GLS estimator for SS, IS and SA

- $$\mathbf{X} = \mathbf{I}$$

- $$\hat{\hat{\mathbf{y}}}_{\text{GLS}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\hat{\hat{\mathbf{y}}} = \hat{\hat{\mathbf{y}}}$$

- $$\text{Var}(\hat{\hat{\mathbf{y}}}_{\text{GLS}}) = \text{Var}(\hat{\hat{\mathbf{y}}}) = \mathbf{C}$$

GLS estimator for SS, IS and SA

- $$\mathbf{X} = \mathbf{I}$$
- $$\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\hat{\bar{\mathbf{y}}} = \hat{\bar{\mathbf{y}}}$$
- $$\mathrm{Var}(\hat{\bar{\mathbf{y}}}_{\mathrm{GLS}}) = \mathrm{Var}(\hat{\bar{\mathbf{y}}}) = \mathbf{C}$$

- Spatial sampling design: simple random sampling
- Sampling in time at constant interval, number of sampling rounds $r = 4, 5$ or 6, length of monitoring period fixed
- AR(1): $\mathrm{cor}\{y(\mathbf{s},t), y(\mathbf{s},t+\Delta(t))\} = \rho^{\Delta(t)}$

- Spatial sampling design: simple random sampling
- Sampling in time at constant interval, number of sampling rounds $r = 4, 5$ or 6, length of monitoring period fixed
- AR(1): $\mathrm{cor}\{y(\mathbf{s},t), y(\mathbf{s},t+\Delta(t))\} = \rho^{\Delta(t)}$

- Spatial sampling design: simple random sampling
- Sampling in time at constant interval, number of sampling rounds $r = 4, 5$ or 6, length of monitoring period fixed
- AR(1): $\mathrm{cor}\{y(\mathbf{s}, t), y(\mathbf{s}, t + \Delta(t))\} = \rho^{\Delta(t)}$

Covariance matrix $\mathbf{C}$ for static-synchronous design, $r = 4$

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \frac{\sigma^2}{n}$$

with σ^2 the spatial variance, and n number of sampling locations per time

Covariance matrix $\mathbf{C}$ for independent synchronous design

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \frac{\sigma^2}{n}$$

Covariance matrix $\mathbf{C}$ for supplemented panel, $r = 4$

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & 0 & 0 & 0 & 0 \\ \rho & 1 & \rho & \rho^2 & 0 & 0 & 0 & 0 \\ \rho^2 & \rho & 1 & \rho & 0 & 0 & 0 & 0 \\ \rho^3 & \rho^2 & \rho & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \frac{\sigma^2}{n/2}$$

Covariance matrix $\mathbf{C}$ for rotating panel, $r = 4$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \rho & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \rho & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \frac{\sigma^2}{n/2}$$

Evaluation of space–time designs

- Average of sampling variances of r estimated spatial means
- Generalized variance of estimated spatial means (determinant of variance-covariance matrix of estimated means)
- Sampling variance of estimated trend of spatial means

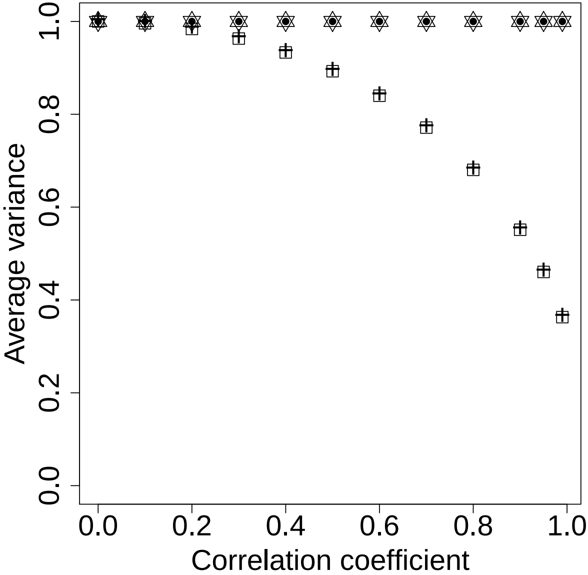
Evaluation of space–time designs

- Average of sampling variances of r estimated spatial means
- Generalized variance of estimated spatial means (determinant of variance-covariance matrix of estimated means)
- Sampling variance of estimated trend of spatial means

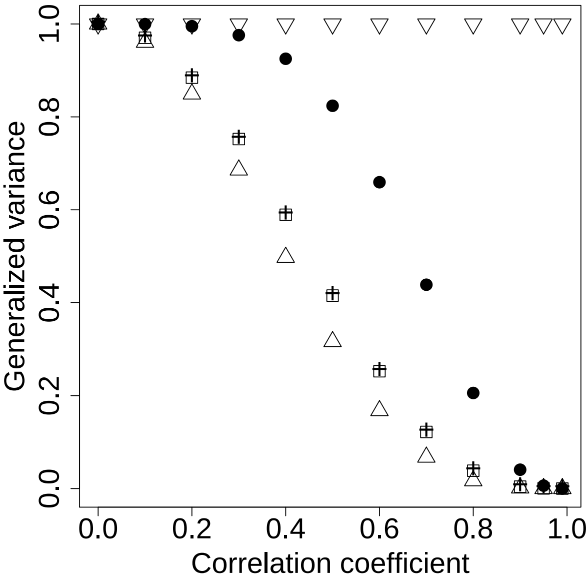
Evaluation of space–time designs

- Average of sampling variances of r estimated spatial means
- Generalized variance of estimated spatial means (determinant of variance-covariance matrix of estimated means)
- Sampling variance of estimated trend of spatial means

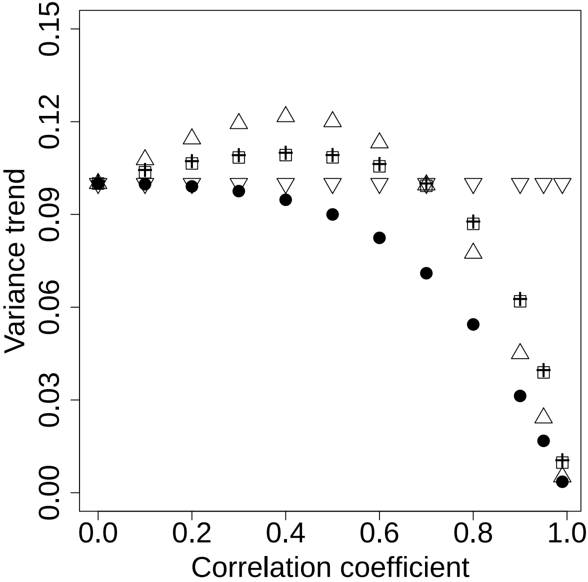
Average sampling variance of estimated means, $r = 5$



Generalized sampling variance of estimated means, $r = 5$



Sampling variance of estimated trend, $r = 5$



Conclusions

- Average variance of means: Supplemented Panel (SP) and Rotating Panel (RP) outperform Static-Synchronous (SS), Independent Synchronous (IS) and Serially Alternating (SA) designs
- Generalized variance of means: SS design performs best, closely followed by SP and R
- Variance of trend:
  - SA performs best
  - SS poor for weak to moderate temporal correlations

Conclusions

- Average variance of means: Supplemented Panel (SP) and Rotating Panel (RP) outperform Static-Synchronous (SS), Independent Synchronous (IS) and Serially Alternating (SA) designs
- Generalized variance of means: SS design performs best, closely followed by SP and R
- Variance of trend:
  - SA performs best
  - SS poor for weak to moderate temporal correlations

Conclusions

- Average variance of means: Supplemented Panel (SP) and Rotating Panel (RP) outperform Static-Synchronous (SS), Independent Synchronous (IS) and Serially Alternating (SA) designs
- Generalized variance of means: SS design performs best, closely followed by SP and R
- Variance of trend:
  - SA performs best
  - SS poor for weak to moderate temporal correlations

Conclusions

- Average variance of means: Supplemented Panel (SP) and Rotating Panel (RP) outperform Static-Synchronous (SS), Independent Synchronous (IS) and Serially Alternating (SA) designs
- Generalized variance of means: SS design performs best, closely followed by SP and R
- Variance of trend:
  - SA performs best
  - SS poor for weak to moderate temporal correlations

Best space–time design for status and trend monitoring

Aim	Best type of space–time design
Status	SP,RP
Trend	SA (SS)
Status *and* Trend	? $\longrightarrow$ prioritize two aims

Thanks for your attention

For more details on method and a case study on soil acidification/eutrophication, see:

- Brus, D.J. and de Gruijter, J.J., (2011), Design-based Generalized Least Squares estimation of status and trend of soil properties from monitoring data, *Geoderma* 164: 172-180.