

# Sampling for monitoring: on design-based, model-based and mixed approaches

Dick Brus<sup>1</sup>

<sup>1</sup>Alterra, Wageningen University and Research Centre



# Outline

## Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Outline

## Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions



KEN LIVINGSTONE ABOUT THE CLIMATE PROBLEM!

**“We’ve most  
probably  
passed the  
tipping point”**



# Types of monitoring

- ▶ status monitoring
- ▶ trend monitoring
- ▶ compliance monitoring

# Types of monitoring

- ▶ status monitoring
- ▶ trend monitoring
- ▶ compliance monitoring

# Types of monitoring

- ▶ status monitoring
- ▶ trend monitoring
- ▶ compliance monitoring

# Sampling and estimation

- ▶ Statistical literature: strong focus on estimation (prediction); the data are already there
- ▶ Sampling for monitoring deserves more attention

# Sampling and estimation

- ▶ Statistical literature: strong focus on estimation (prediction); the data are already there
- ▶ Sampling for monitoring deserves more attention

# Major design decision: Design-based or model-based approach?

## Definition of design-based and model-based approach

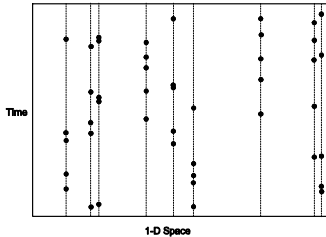
Type of approach	Sampling unit selection	Statistical inference
Design-based	Probability sampling	Design-based
Model-based	Purposive	Model-based

# Four statistical approaches in space–time

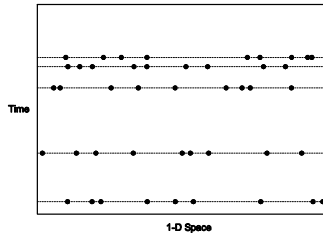
		<i>Space</i>	
	Statistical approach	Design-based	Model-based
<i>Time</i>	Design-based	$D_S D_T$	$M_S D_T$
	Model-based	$D_S M_T$	$M_S M_T$

# Basic sample patterns in space-time

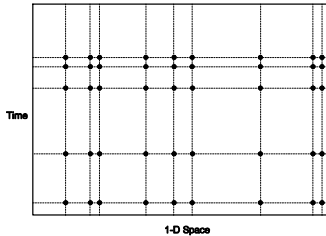
Static



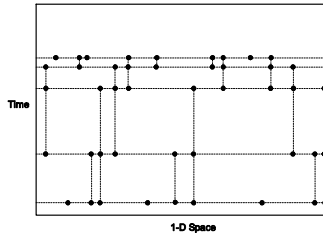
Synchronous



Static-synchronous



Rotational





# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Surface water in Dutch polder



- ▶ Testing quality of surface water against WFD-standards
- ▶ Does *spatio-temporal mean* concentration of N-total and P-total during summer-halfyear comply with MAR-values (N: 2.2 mg/l; P: 0.15 mg/l)?
- ▶  $H_0$ : 'water is dirty' ( $c > c_{MAR}$ )

# Aim

- ▶ Testing quality of surface water against WFD-standards
- ▶ Does *spatio-temporal mean* concentration of N-total and P-total during summer-halfyear comply with MAR-values (N: 2.2 mg/l; P: 0.15 mg/l)?
- ▶  $H_0$ : 'water is dirty' ( $c > c_{MAR}$ )

# Aim

- ▶ Testing quality of surface water against WFD-standards
- ▶ Does *spatio-temporal mean* concentration of N-total and P-total during summer-halfyear comply with MAR-values (N: 2.2 mg/l; P: 0.15 mg/l)?
- ▶  $H_0$ : 'water is dirty' ( $c > c_{MAR}$ )

# Why $D_S D_T$ -approach?

- ▶ space-time mean: global quantity
- ▶ for compliance monitoring validity more important than efficiency
- ▶ too few data for space-time modelling

# Why $D_S D_T$ -approach?

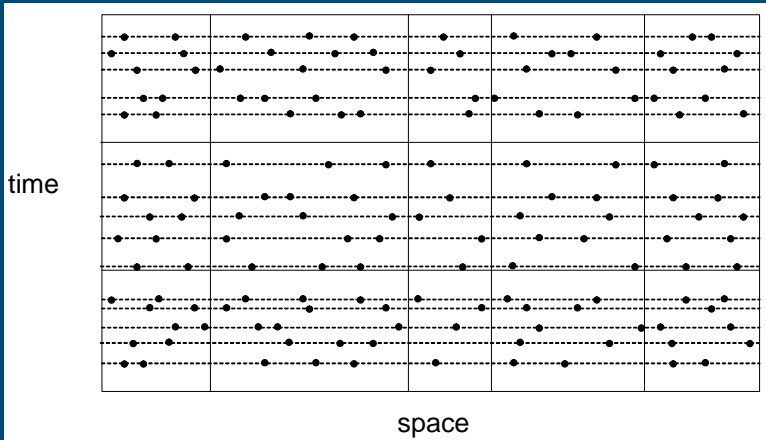
- ▶ space-time mean: global quantity
- ▶ for compliance monitoring validity more important than efficiency
- ▶ too few data for space-time modelling

# Why $D_S D_T$ -approach?

- ▶ space-time mean: global quantity
- ▶ for compliance monitoring validity more important than efficiency
- ▶ too few data for space-time modelling



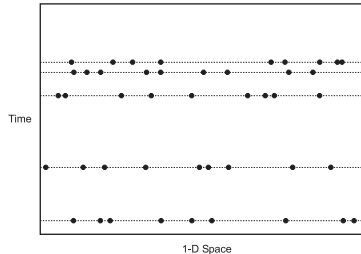
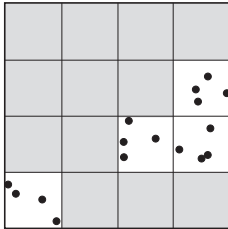
# Synchronous pattern, STSI in space, STSI in time



# Why synchronous pattern?

Unbiased estimation of sampling variance (if samples independent!)

Two-stage sampling in space      Synchronous sampling in space–time



# Estimation of sampling variance

- For **SI** in both stages:

$$\hat{V}(\hat{y}) = \frac{1}{r} \left\{ \hat{V}_T + \frac{1}{n} \hat{V}_S \right\}$$

- $r$ : number of sampling times (psu's)
- $n$ : number of sampling locations per sampling time (ssu's)
- $\hat{V}_T$ : temporal variance of spatial means
- $\hat{V}_S$ : spatial variance of  $y$

# Estimation of sampling variance

- ▶ For **SI** in both stages:

$$\hat{V}(\hat{y}) = \frac{1}{r} \left\{ \hat{V}_T + \frac{1}{n} \hat{V}_S \right\}$$

- ▶  $r$ : number of sampling times (psu's)
- ▶  $n$ : number of sampling locations per sampling time (ssu's)
- ▶  $\hat{V}_T$ : temporal variance of spatial means
- ▶  $\hat{V}_S$ : spatial variance of  $y$

# Estimation of sampling variance

- ▶ For **SI** in both stages:

$$\hat{V}(\hat{y}) = \frac{1}{r} \left\{ \hat{V}_T + \frac{1}{n} \hat{V}_S \right\}$$

- ▶  $r$ : number of sampling times (psu's)
- ▶  $n$ : number of sampling locations per sampling time (ssu's)
- ▶  $\hat{V}_T$ : temporal variance of spatial means
- ▶  $\hat{V}_S$ : spatial variance of  $y$

# Estimation of sampling variance

- ▶ For **SI** in both stages:

$$\hat{V}(\hat{y}) = \frac{1}{r} \left\{ \hat{V}_T + \frac{1}{n} \hat{V}_S \right\}$$

- ▶  $r$ : number of sampling times (psu's)
- ▶  $n$ : number of sampling locations per sampling time (ssu's)
- ▶  $\hat{V}_T$ : temporal variance of spatial means
- ▶  $\hat{V}_S$ : spatial variance of  $y$

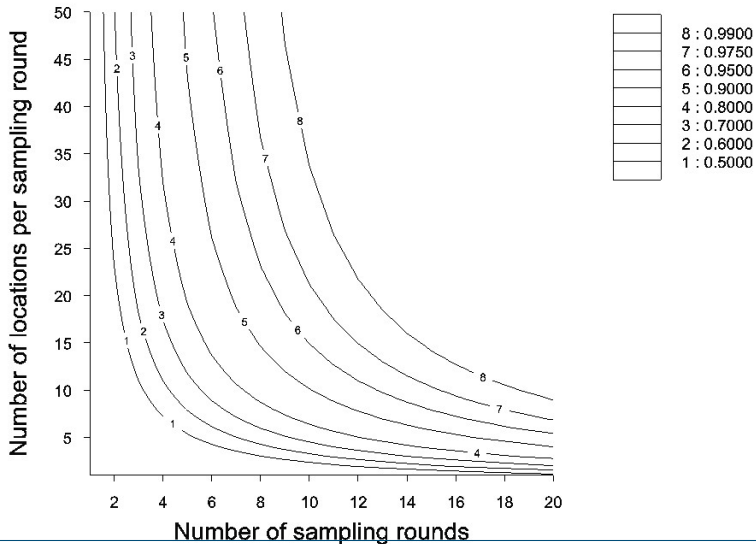
# Estimation of sampling variance

- ▶ For **SI** in both stages:

$$\hat{V}(\hat{y}) = \frac{1}{r} \left\{ \hat{V}_T + \frac{1}{n} \hat{V}_S \right\}$$

- ▶  $r$ : number of sampling times (psu's)
- ▶  $n$ : number of sampling locations per sampling time (ssu's)
- ▶  $\hat{V}_T$ : temporal variance of spatial means
- ▶  $\hat{V}_S$ : spatial variance of  $y$

# Contourplot of power for N





# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Monitoring-networks in the Netherlands

	Number of locations	Sampling interval
Groundwater	371 (1 per 100 km <sup>2</sup> )	4 y (25 m)
Soil and freatic groundwater	200	6 y

- Static-asynchronous pattern
- Purposive sampling of locations

# Monitoring-networks in the Netherlands

	Number of locations	Sampling interval
Groundwater	371 (1 per 100 km <sup>2</sup> )	4 y (25 m)
Soil and freatic groundwater	200	6 y

- ▶ Static-synchronous pattern
- ▶ Purposive sampling of locations

# Monitoring-networks in the Netherlands

	Number of locations	Sampling interval
Groundwater	371 (1 per 100 km <sup>2</sup> )	4 y (25 m)
Soil and freatic groundwater	200	6 y

- ▶ Static-synchronous pattern
- ▶ Purposive sampling of locations

# Research question

“What is the efficiency of the networks compared to alternatives”

- ▶ Other sampling densities (space) and sampling frequencies (time)
- ▶ Alternative patterns, such as synchronous and serially alternating

# Research question

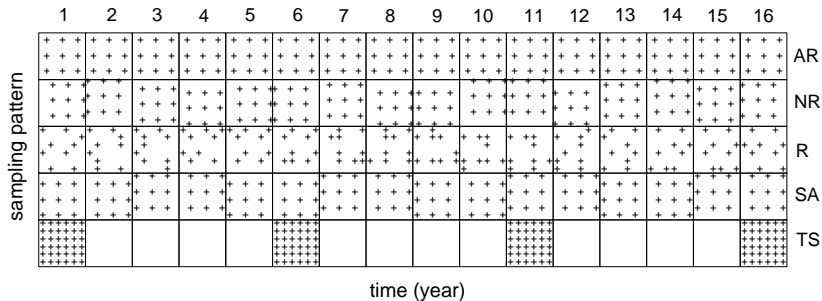
“What is the efficiency of the networks compared to alternatives”

- ▶ Other sampling densities (space) and sampling frequencies (time)
- ▶ Alternative patterns, such as synchronous and serially alternating

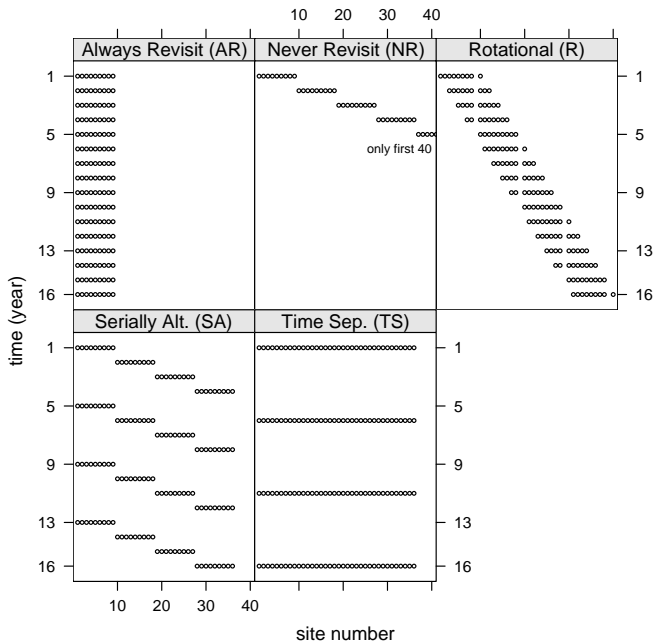
# Research question

“What is the efficiency of the networks compared to alternatives”

- ▶ Other sampling densities (space) and sampling frequencies (time)
- ▶ Alternative patterns, such as synchronous and serially alternating







# Pure model-based ( $M_S M_T$ ) approach

Geostatistical space–time model:

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

$$\mu(\mathbf{s}, t) = \sum_{j=1}^p \beta_j x_j(\mathbf{s}, t)$$

Special case:  $x_1(\mathbf{s}, t) = 1, x_2(\mathbf{s}, t) = t \rightarrow$

$$Y(\mathbf{s}, t) = \beta_1 + \beta_2 t + \varepsilon(\mathbf{s}, t)$$

SMTT, either defined as *model-parameter* or as *population parameter*, can be estimated by **Universal Kriging**

# Pure model-based ( $M_S M_T$ ) approach

Geostatistical space–time model:

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

$$\mu(\mathbf{s}, t) = \sum_{j=1}^p \beta_j x_j(\mathbf{s}, t)$$

Special case:  $x_1(\mathbf{s}, t) = 1, x_2(\mathbf{s}, t) = t \rightarrow$

$$Y(\mathbf{s}, t) = \beta_1 + \beta_2 t + \varepsilon(\mathbf{s}, t)$$

SMTT, either defined as *model-parameter* or as *population parameter*, can be estimated by **Universal Kriging**

# Pure model-based ( $M_S M_T$ ) approach

Geostatistical space–time model:

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

$$\mu(\mathbf{s}, t) = \sum_{j=1}^p \beta_j x_j(\mathbf{s}, t)$$

Special case:  $x_1(\mathbf{s}, t) = 1, x_2(\mathbf{s}, t) = t \rightarrow$

$$Y(\mathbf{s}, t) = \beta_1 + \beta_2 t + \varepsilon(\mathbf{s}, t)$$

SMTT, either defined as *model-parameter* or as *population parameter*, can be estimated by Universal Kriging

# Pure model-based ( $M_S M_T$ ) approach

Geostatistical space–time model:

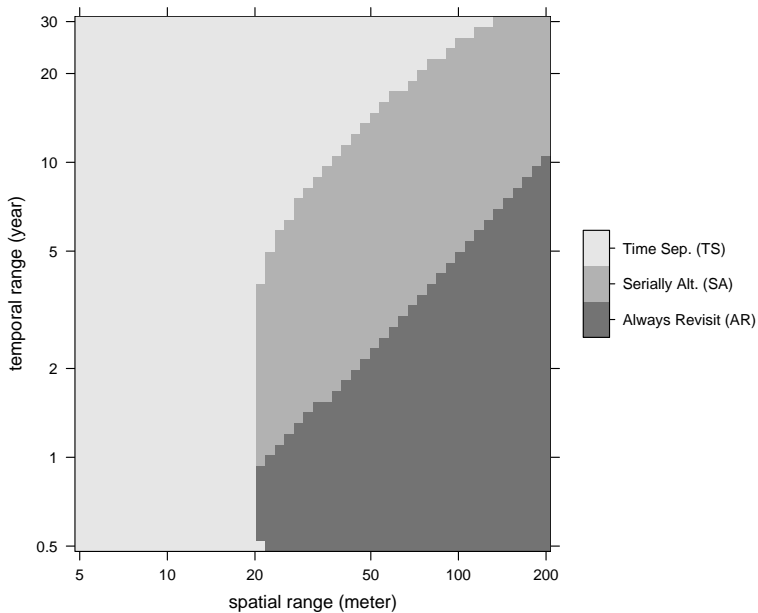
$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

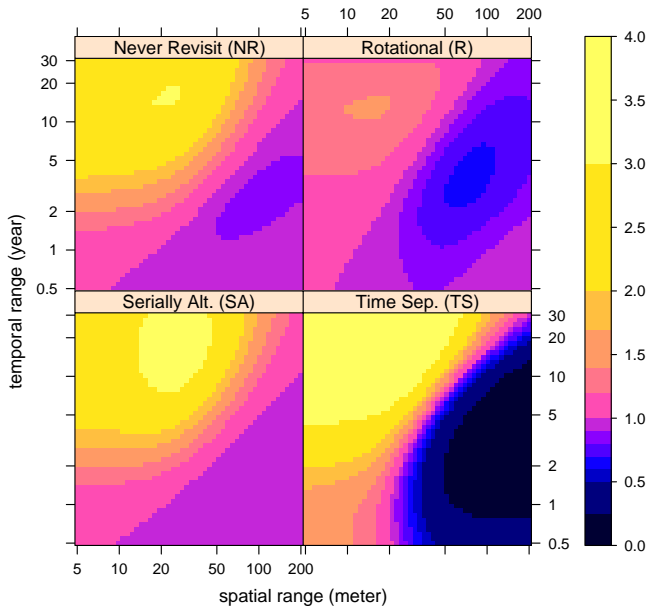
$$\mu(\mathbf{s}, t) = \sum_{j=1}^p \beta_j x_j(\mathbf{s}, t)$$

Special case:  $x_1(\mathbf{s}, t) = 1, x_2(\mathbf{s}, t) = t \rightarrow$

$$Y(\mathbf{s}, t) = \beta_1 + \beta_2 t + \varepsilon(\mathbf{s}, t)$$

SMTT, either defined as *model-parameter* or as *population parameter*, can be estimated by **Universal Kriging**





# Some characteristics of AR, TS and SA

	# of time series	Sampling density	Sampling frequency
AR	few	low	high
TS	many	high	low
SA	many	low	low



# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Mixed $D_S M_T$ approach

- ▶ **Probability sampling** in space, **design-based estimation** of spatial means
- ▶ **Purposive sampling** in time, constant time interval: first sampling round at start ( $T_1$ ), last round at end ( $T_2$ ) of monitoring period
- ▶ **Linear Regression Model** for spatial means at  $t = T_1 \cdots T_2$

$$m_A\{Y(t)\} = \beta_1 + \beta_2 t + \eta(t) + \varepsilon(t)$$

$$\eta(t) \sim N(0, C_\eta)$$

$$\varepsilon(t) \sim N(0, C_\varepsilon)$$

# Mixed $D_S M_T$ approach

- ▶ **Probability sampling** in space, **design-based estimation** of spatial means
- ▶ **Purposive sampling** in time, constant time interval: first sampling round at start ( $T_1$ ), last round at end ( $T_2$ ) of monitoring period
- ▶ **Linear Regression Model** for spatial means at  $t = T_1 \cdots T_2$

$$m_A\{Y(t)\} = \beta_1 + \beta_2 t + \eta(t) + \varepsilon(t)$$

$$\eta(t) \sim G(\mathbf{0}, \mathbf{C}_\xi)$$

$$\varepsilon(t) \sim G(\mathbf{0}, \mathbf{C}_p)$$

# Mixed $D_S M_T$ approach

- ▶ **Probability sampling** in space, **design-based estimation** of spatial means
- ▶ **Purposive sampling** in time, constant time interval: first sampling round at start ( $T_1$ ), last round at end ( $T_2$ ) of monitoring period
- ▶ **Linear Regression Model** for spatial means at  $t = T_1 \cdots T_2$

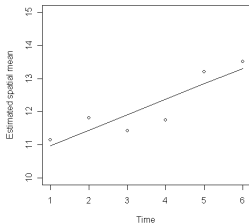
$$m_A\{Y(t)\} = \beta_1 + \beta_2 t + \eta(t) + \varepsilon(t)$$

$$\eta(t) \sim G(\mathbf{0}, \mathbf{C}_\xi)$$

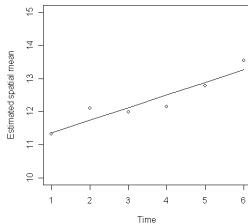
$$\varepsilon(t) \sim G(\mathbf{0}, \mathbf{C}_p)$$

# Number of sampling locations per round: 10

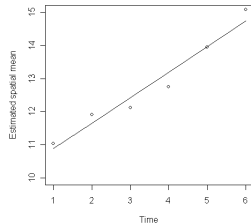
0.469



0.380

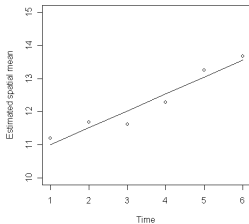


0.768

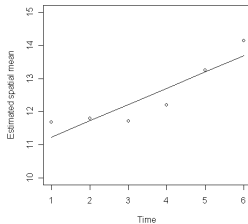


# Number of sampling locations per round: 50

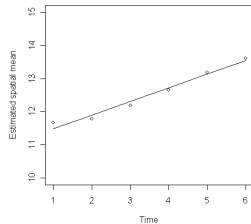
0.508



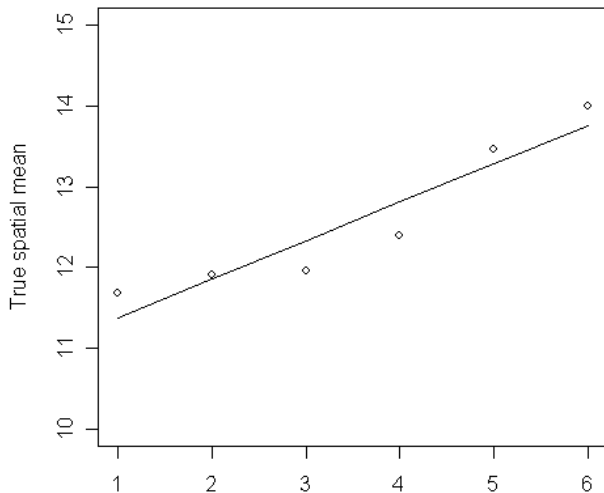
0.490



0.412



## All locations; temporal trend: 0.475



# Estimators for trend and variance of trend

$$\mathbf{b} = (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{y})$$

$$\text{Var}(\mathbf{b}) = (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{X})^{-1}$$

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



# Estimators for trend and variance of trend

$$\mathbf{b} = (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{y})$$

$$\text{Var}(\mathbf{b}) = (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{X})^{-1}$$

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

# Estimators for trend and variance of trend

$$\mathbf{b} = (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{y})$$

$$\text{Var}(\mathbf{b}) = (\mathbf{X}^T \mathbf{C}_{\xi p}^{-1} \mathbf{X})^{-1}$$

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

# Simple cost model

$$C = c_{\text{round}} \cdot r + c_{\text{site}} \cdot r \cdot n$$

- ▶  $c_{\text{round}}$ : costs per sampling round
- ▶  $c_{\text{site}}$ : costs per sampling site
- ▶  $r$ : number of sampling rounds
- ▶  $n$ : number of sampling sites per sampling round

# Simple cost model

$$C = c_{\text{round}} \cdot r + c_{\text{site}} \cdot r \cdot n$$

- ▶  $c_{\text{round}}$ : costs per sampling round
- ▶  $c_{\text{site}}$ : costs per sampling site
- ▶  $r$ : number of sampling rounds
- ▶  $n$ : number of sampling sites per sampling round

# Simple cost model

$$C = c_{\text{round}} \cdot r + c_{\text{site}} \cdot r \cdot n$$

- ▶  $c_{\text{round}}$ : costs per sampling round
- ▶  $c_{\text{site}}$ : costs per sampling site
- ▶  $r$ : number of sampling rounds
- ▶  $n$ : number of sampling sites per sampling round

# Simple cost model

$$C = c_{\text{round}} \cdot r + c_{\text{site}} \cdot r \cdot n$$

- ▶  $c_{\text{round}}$ : costs per sampling round
- ▶  $c_{\text{site}}$ : costs per sampling site
- ▶  $r$ : number of sampling rounds
- ▶  $n$ : number of sampling sites per sampling round

# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period



# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

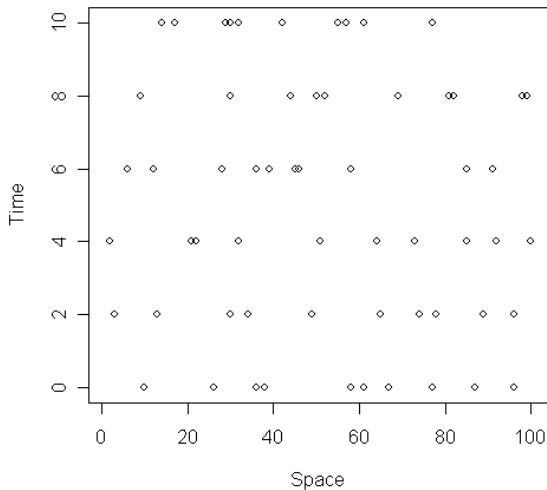
# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

# Experiment

- ▶ Total costs: 1000
- ▶ Number of sampling rounds  $r = 2 \dots 50$
- ▶  $c_{\text{round}}$ : 10, 100;  $c_{\text{site}}$ : 1
- ▶ Model variance of residuals  $\sigma_{\xi}^2\{\eta(t)\} = 0.01 \dots 1$
- ▶ Exponential correlogram for residuals  $\eta(t)$
- ▶ Spatial variance  $v_A\{y(t)\} = 0.25 \dots 25$
- ▶ Constant spatial variance:  $v_A\{y(t_i)\} = v_A\{y(t_j)\}$
- ▶ Simple random sampling in space: sampling variance of estimated spatial mean  $\sigma_p^2\{\hat{m}_A(y)\} = v_A\{y(t)\}/n$
- ▶ Systematic sampling in time: first round at start, last round at end of monitoring period

## Synchronous pattern, independent SI samples





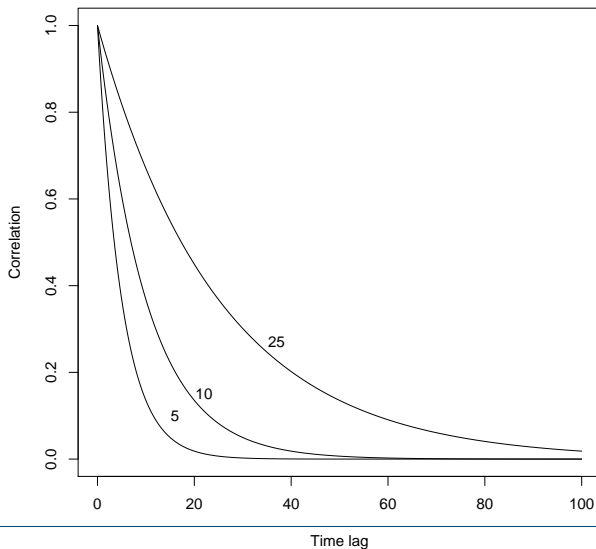
# Synchronous pattern, independent samples

- Synchronous pattern (Never Revisit), independent samples →

$$\text{Cov}_p\{\varepsilon(t_i), \varepsilon(t_j)\} = 0, i \neq j$$

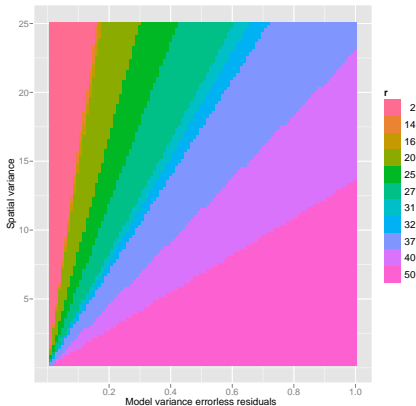
# Correlogram of model residuals $\eta(t)$

Exponential correlogram

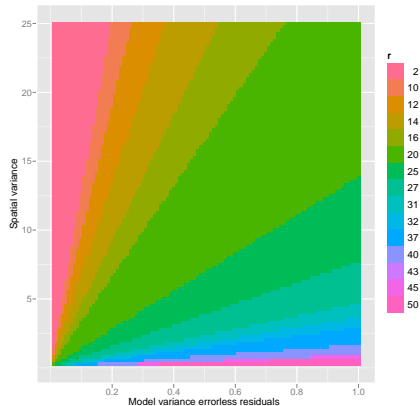


# Optimal number of sampling rounds; $c_{\text{round}} = 10$

Pure nugget

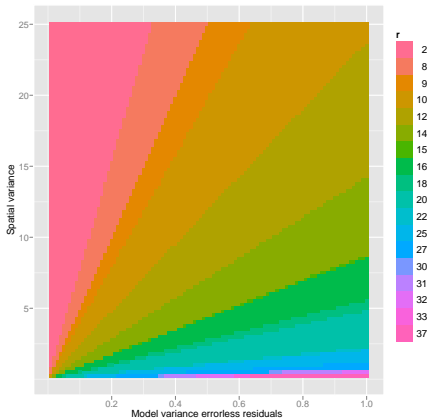


Exponential(5)

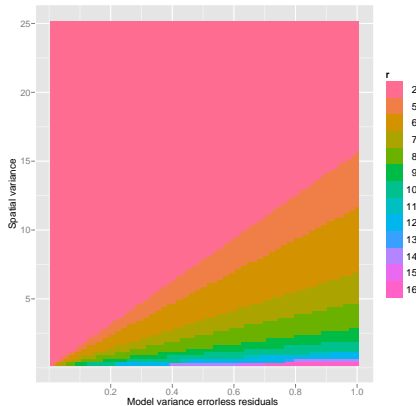


# Optimal number of sampling rounds; $c_{\text{round}} = 10$

## Exponential(10)

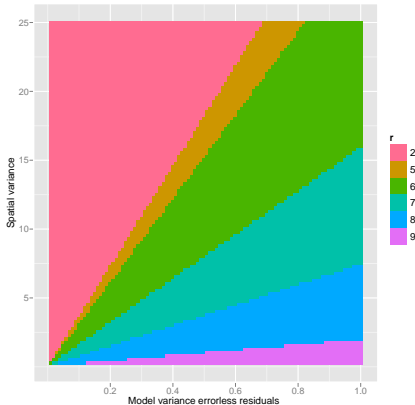


## Exponential(25)

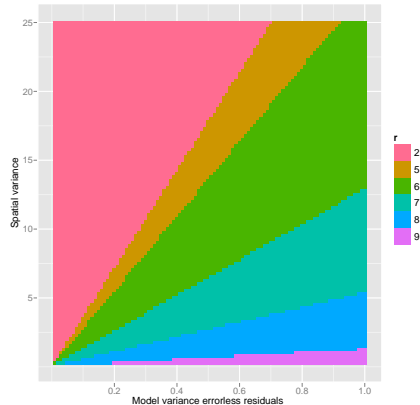


# Optimal number of sampling rounds; $c_{\text{round}} = 100$

Pure nugget

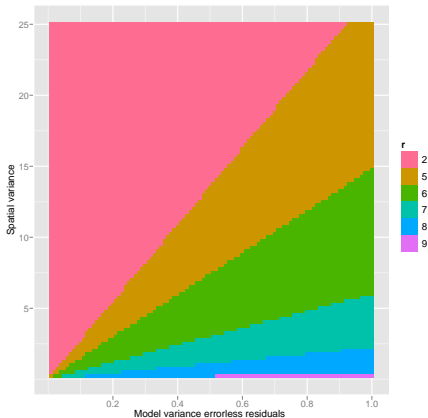


Exponential(5)

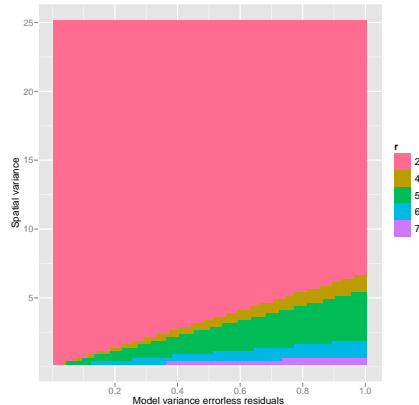


# Optimal number of sampling rounds; $c_{\text{round}} = 100$

## Exponential(10)



## Exponential(25)



# Results

- ▶ Optimal number of sampling rounds not always 2!
- ▶ Optimal number of sampling rounds is determined by ratio of spatial variance of target variable and model variance of residuals (given the cost model parameters)
- ▶ The smaller the ratio, the larger the optimal number of sampling rounds
- ▶ Given a ratio, the stronger the temporal correlation, the smaller the optimal number of sampling rounds

# Results

- ▶ Optimal number of sampling rounds not always 2!
- ▶ Optimal number of sampling rounds is determined by ratio of spatial variance of target variable and model variance of residuals (given the cost model parameters)
- ▶ The smaller the ratio, the larger the optimal number of sampling rounds
- ▶ Given a ratio, the stronger the temporal correlation, the smaller the optimal number of sampling rounds



# Results

- ▶ Optimal number of sampling rounds not always 2!
- ▶ Optimal number of sampling rounds is determined by ratio of spatial variance of target variable and model variance of residuals (given the cost model parameters)
- ▶ The smaller the ratio, the larger the optimal number of sampling rounds
- ▶ Given a ratio, the stronger the temporal correlation, the smaller the optimal number of sampling rounds

# Results

- ▶ Optimal number of sampling rounds not always 2!
- ▶ Optimal number of sampling rounds is determined by ratio of spatial variance of target variable and model variance of residuals (given the cost model parameters)
- ▶ The smaller the ratio, the larger the optimal number of sampling rounds
- ▶ Given a ratio, the stronger the temporal correlation, the smaller the optimal number of sampling rounds

# Conclusions

- ▶ Model in  $D_S M_T$  approach less complicated than in  $M_S M_T$  approach, fewer assumptions required → better validity properties
- ▶ Mixed  $D_S M_T$  approach with probability sampling in space seems promising for estimating Temporal Trend of Spatial Means
- ▶ Further research:
  - other spatial sampling designs
  - optimization of  $c$  and  $\alpha$  in static-synchronous and rotational patterns
  - optimal matching proportion in rotational sampling
  - in  $D_S M_T$  approach cross-correlation in space-time not used → less precise than  $M_S M_T$  approach?

# Conclusions

- ▶ Model in  $D_S M_T$  approach less complicated than in  $M_S M_T$  approach, fewer assumptions required → better validity properties
- ▶ Mixed  $D_S M_T$  approach with probability sampling in space seems promising for estimating Temporal Trend of Spatial Means
- ▶ Further research:
  - other spatial sampling designs
  - optimization of  $c$  and  $\alpha$  in static-synchronous and rotational patterns
  - optimal matching proportion in rotational sampling
  - in  $M_S M_T$  approach cross-correlation in space-time not used, less precise than  $M_S M_T$  approach

# Conclusions

- ▶ Model in  $D_S M_T$  approach less complicated than in  $M_S M_T$  approach, fewer assumptions required → better validity properties
- ▶ Mixed  $D_S M_T$  approach with probability sampling in space seems promising for estimating Temporal Trend of Spatial Means
- ▶ Further research:
  - ▶ other spatial sampling designs
  - ▶ optimization of  $r$  and  $n$  in static-synchronous and rotational patterns
  - ▶ optimal matching proportion in rotational sampling
  - ▶ in  $D_S M_T$  approach cross-correlation in space-time not used → less precise than  $M_S M_T$  approach?

# Conclusions

- ▶ Model in  $D_S M_T$  approach less complicated than in  $M_S M_T$  approach, fewer assumptions required → better validity properties
- ▶ Mixed  $D_S M_T$  approach with probability sampling in space seems promising for estimating Temporal Trend of Spatial Means
- ▶ Further research:
  - ▶ other spatial sampling designs
  - ▶ optimization of  $r$  and  $n$  in static-synchronous and rotational patterns
  - ▶ optimal matching proportion in rotational sampling
  - ▶ in  $D_S M_T$  approach cross-correlation in space-time not used → less precise than  $M_S M_T$  approach?

# Conclusions

- ▶ Model in  $D_S M_T$  approach less complicated than in  $M_S M_T$  approach, fewer assumptions required → better validity properties
- ▶ Mixed  $D_S M_T$  approach with probability sampling in space seems promising for estimating Temporal Trend of Spatial Means
- ▶ Further research:
  - ▶ other spatial sampling designs
  - ▶ optimization of  $r$  and  $n$  in static-synchronous and rotational patterns
  - ▶ optimal matching proportion in rotational sampling
  - ▶ in  $D_S M_T$  approach cross-correlation in space-time not used → less precise than  $M_S M_T$  approach?

# Conclusions

- ▶ Model in  $D_S M_T$  approach less complicated than in  $M_S M_T$  approach, fewer assumptions required → better validity properties
- ▶ Mixed  $D_S M_T$  approach with probability sampling in space seems promising for estimating Temporal Trend of Spatial Means
- ▶ Further research:
  - ▶ other spatial sampling designs
  - ▶ optimization of  $r$  and  $n$  in static-synchronous and rotational patterns
  - ▶ optimal matching proportion in rotational sampling
  - ▶ in  $D_S M_T$  approach cross-correlation in space-time not used → less precise than  $M_S M_T$  approach?



# Conclusions

- ▶ Model in  $D_S M_T$  approach less complicated than in  $M_S M_T$  approach, fewer assumptions required → better validity properties
- ▶ Mixed  $D_S M_T$  approach with probability sampling in space seems promising for estimating Temporal Trend of Spatial Means
- ▶ Further research:
  - ▶ other spatial sampling designs
  - ▶ optimization of  $r$  and  $n$  in static-synchronous and rotational patterns
  - ▶ optimal matching proportion in rotational sampling
  - ▶ in  $D_S M_T$  approach cross-correlation in space-time not used → less precise than  $M_S M_T$  approach?

# Outline

Introduction

Pure design-based approach for compliance monitoring

Pure model-based approach for trend monitoring

Mixed approach for trend monitoring

Conclusions

# Conclusions

- ▶ For compliance monitoring of global quantities a full design-based approach can be advantageous
- ▶ For trend monitoring: full model-based or mixed ( $D_S M_T$ ) approach most appropriate
- ▶ Full model-based approach: Serially alternating pattern good option
- ▶ Mixed ( $D_S M_T$ ) approach: model more simple, better validity, but less efficient?

# Conclusions

- ▶ For compliance monitoring of global quantities a full design-based approach can be advantageous
- ▶ For trend monitoring: full model-based or mixed ( $D_S M_T$ ) approach most appropriate
- ▶ Full model-based approach: Serially alternating pattern good option
- ▶ Mixed ( $D_S M_T$ ) approach: model more simple, better validity, but less efficient?

# Conclusions

- ▶ For compliance monitoring of global quantities a full design-based approach can be advantageous
- ▶ For trend monitoring: full model-based or mixed ( $D_S M_T$ ) approach most appropriate
- ▶ Full model-based approach: Serially alternating pattern good option
- ▶ Mixed ( $D_S M_T$ ) approach: model more simple, better validity, but less efficient?

# Conclusions

- ▶ For compliance monitoring of global quantities a full design-based approach can be advantageous
- ▶ For trend monitoring: full model-based or mixed ( $D_S M_T$ ) approach most appropriate
- ▶ Full model-based approach: Serially alternating pattern good option
- ▶ Mixed ( $D_S M_T$ ) approach: model more simple, better validity, but less efficient?

# Message to take home

Start thinking before you go into the field!

# Thanks for your attention



**JAAP I. DE GRUIJTER** worked as a researcher and statistical consultant at Alterra-Wageningen University and Research Centre for 36 years, until his retirement by November 2005. His research interest focuses on statistical methodology for spatial inventory and monitoring including soil science, groundwater hydrology, geomorphology, plant- and landscape ecology, remote sensing and landscape. Among his publications are articles in *Mathematical Geology* and *Geoderma*, the Editorial Board of which he is a member.

**DICK J. BRUS** is senior researcher at Alterra, Wageningen University and Research Centre. His main activities include research on statistical methods for survey and monitoring of natural resources, propagation of errors in soil maps, as well as statistical consulting. He has published numerous papers in journals including *Geoderma*, *Environmetrics* and *European Journal of Soil Science*.

**MARC R. BIERKENS** is a Professor of Hydrology at Utrecht University. His fields of expertise are groundwater hydrology, stochastic hydrology, hydrological regionalisation and geostatistics. He received Best Paper Awards in the field of pedometrics in 1994 and 2004, and the hydrology Prize of the Netherlands Hydrological Society for the years 1994 to 1997.

**MARTIN KNOTTERS** is a soil scientist and hydrologist at Alterra-Wageningen University and Research Centre since 1985. His fields of interest include spatio-temporal modelling of water table depths, designing monitoring networks, and the utilization of statistical knowledge on uncertainty in environmental policy. He has published relevant papers in the journals *Water Resources Research*, *Journal of Hydrology*, and *Geoderma*.

de Groot - Brus - Bierkens - Knotters  
Sampling for Natural Resource Monitoring

The book presents the statistical knowledge and methodology of sampling and data analysis useful for spatial inventory and monitoring of natural resources. The authors omitted all theory not essential for applications or for basic understanding. This presentation is broader than standard statistical texts, as the authors pay much attention to how statistical methodology can be employed and embedded in real-life spatial inventory and monitoring projects. Thus they discuss in detail how efficient sampling schemes and monitoring systems can be designed in view of the aims and constraints of the project.

ISBN 3-540-22486-6



springer.com

Jaap de Groot  
Dick Brus  
Marc Bierkens  
Martin Knotters

de Groot  
Brus - Bierkens  
Knotters



Sampling for Natural  
Resource Monitoring

## Sampling for Natural Resource Monitoring

 Springer