Sampling for mapping and monitoring of soil carbon stocks

Dick Brus, Martin Knotters

Soil Science Centre

Wageningen University and Research Centre - Alterra

Cocos Workshop, July 15-16, 2011, Leuven, Belgium

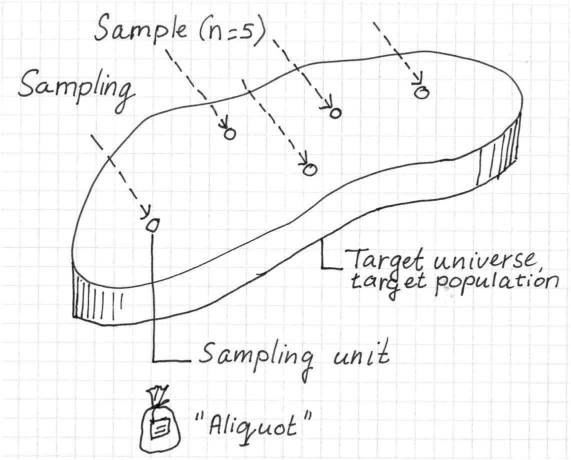
Introduction of the authors and their work

- Brus, D.J. & J.J. de Gruijter, 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). *Geoderma* **80**: 1-59.
- Hoogland, T., M. Knotters, D.J. Brus & P. Kuikman, 2005. *Monitoring of changes in carbon stocks in the Dutch soil. Design of a monitoring strategy* (in Dutch). Wageningen, Alterra report 1354, 56 pp.
- De Gruijter, J.J., D.J. Brus, M.F.P. Bierkens & M. Knotters, 2006. *Sampling for natural resource monitoring*. Berlin, Springer, 332 pp.

Introduction of the authors and their work (continued)

- Brus, D.J. & J.J. de Gruijter, 2011. Design-based generalized least squares estimation of status and trend of soil properties from monitoring data. *Geoderma*, in press.
- Brus, D.J., B. Kempen & G.B.M. Heuvelink, 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* **62**: 394-407.

Clashing terms



Aim

- To discuss sampling aspects of mapping, monitoring and validation

Motivation

- Selection process of sampling locations deserves more attention: ‘representative’ is often ill-defined*
- ‘Start at the end, and reason backward’: integrated planning of data collection and data processing, with respect to the required information

*) Eight different meanings. See for an interesting analysis of the various meanings of ‘representative sampling’: Kruskal, W. and F. Mosteller, 1979. Representative sampling, I, II and III. *International Statistical Review* **47**(1,2,3): 13-24, 111-127, 245-265.

Motivation

- Selection process of sampling locations deserves more attention: 'representative' is often ill-defined*
- 'Start at the end, and reason backward': integrated planning of data collection and data processing, with respect to the required information

*) Eight different meanings. See for an interesting analysis of the various meanings of 'representative sampling': Kruskal, W. and F. Mosteller, 1979. Representative sampling, I, II and III. *International Statistical Review* **47**(1,2,3): 13-24, 111-127, 245-265.

Motivation

- Selection process of sampling locations deserves more attention: ‘representative’ is often ill-defined*
- ‘Start at the end, and reason backward’: integrated planning of data collection and data processing, with respect to the required information

*) Eight different meanings. See for an interesting analysis of the various meanings of ‘representative sampling’: Kruskal, W. and F. Mosteller, 1979. Representative sampling, I, II and III. *International Statistical Review* **47**(1,2,3): 13-24, 111-127, 245-265.

Outline

1. Sampling for *mapping* of soil carbon stocks
2. Sampling for mapping of *changes* in soil carbon stocks (*monitoring*)
3. Sampling for *validation* of maps

Outline

1. Sampling for *mapping* of soil carbon stocks
2. Sampling for mapping of *changes* in soil carbon stocks (*monitoring*)
3. Sampling for *validation* of maps

Outline

1. Sampling for *mapping* of soil carbon stocks
2. Sampling for mapping of *changes* in soil carbon stocks (*monitoring*)
3. Sampling for *validation* of maps

Outline

1. Sampling for *mapping* of soil carbon stocks
2. Sampling for mapping of *changes* in soil carbon stocks (*monitoring*)
3. Sampling for *validation* of maps

1. Sampling for mapping: directed or random?

- Directed (targeted, purposive) sampling:
  - collection of data can be optimized for spatial interpolation (fair spatial coverage).
- Probability sampling:
  - enables model-free estimation of means or totals and their standard errors (design-based inference).
  - collected data are suitable for spatial interpolation if the sampling design guarantees fair spatial distribution of the selected sampling units.

1. Sampling for mapping: directed or random?

- Directed (targeted, purposive) sampling:
  - collection of data can be optimized for spatial interpolation (fair spatial coverage).
- Probability sampling:
  - enables model-free estimation of means or totals and their standard errors (design-based inference).
  - collected data are suitable for spatial interpolation if the sampling design guarantees fair spatial distribution of the selected sampling units.

1. Sampling for mapping: directed or random?

- Directed (targeted, purposive) sampling:
  - collection of data can be optimized for spatial interpolation (fair spatial coverage).
- Probability sampling:
  - enables model-free estimation of means or totals and their standard errors (design-based inference).
  - collected data are suitable for spatial interpolation if the sampling design guarantees fair spatial distribution of the selected sampling units.

1. Sampling for mapping: directed or random?

- Directed (targeted, purposive) sampling:
  - collection of data can be optimized for spatial interpolation (fair spatial coverage).
- Probability sampling:
  - enables model-free estimation of means or totals and their standard errors (design-based inference).
  - collected data are suitable for spatial interpolation if the sampling design guarantees fair spatial distribution of the selected sampling units.

1. Sampling for mapping: directed or random?

- Directed (targeted, purposive) sampling:
  - collection of data can be optimized for spatial interpolation (fair spatial coverage).
- Probability sampling:
  - enables model-free estimation of means or totals and their standard errors (design-based inference).
  - collected data are suitable for spatial interpolation if the sampling design guarantees fair spatial distribution of the selected sampling units.

1. Sampling for mapping: directed or random?

- Directed (targeted, purposive) sampling:
  - collection of data can be optimized for spatial interpolation (fair spatial coverage).
- Probability sampling:
  - enables model-free estimation of means or totals and their standard errors (design-based inference).
  - collected data are suitable for spatial interpolation if the sampling design guarantees fair spatial distribution of the selected sampling units.

Probability sampling

- *random* selection of elements
- selection probabilities are known, and > 0
- inference based on selection probabilities

Probability sampling

- *random* selection of elements
- selection probabilities are known, and > 0
- inference based on selection probabilities

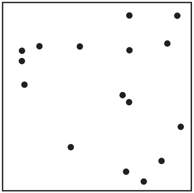
Probability sampling

- *random* selection of elements
- selection probabilities are known, and > 0
- inference based on selection probabilities

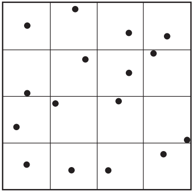
Probability sampling

- *random* selection of elements
- selection probabilities are known, and > 0
- inference based on selection probabilities

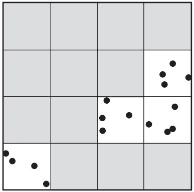
Design types for probability sampling



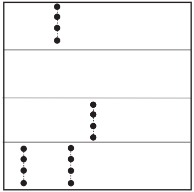
Simple random sampling



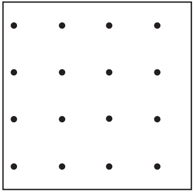
Stratified simple random sampling



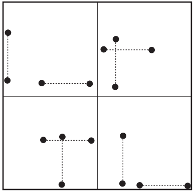
Two-stage random sampling



Cluster random sampling



Systematic random sampling



Stratified cluster random sampling

Design-based or model-based approach?

Design-based method best choice when:

- we want to estimate the distribution function or parameters thereof (mean, median, P90 etc.) for the area as a whole or for subareas;
- objective estimates of target properties are required, i.e. no subjective judgement on 'representativeness', no subjective model choices;
- objective estimates of *estimation variance* or *confidence intervals* are required, i.e. validity.

Design-based or model-based approach?

Design-based method best choice when:

- we want to estimate the distribution function or parameters thereof (mean, median, P90 etc.) for the area as a whole or for subareas;
- objective estimates of target properties are required, i.e. no subjective judgement on 'representativeness', no subjective model choices;
- objective estimates of *estimation variance* or *confidence intervals* are required, i.e. validity.

Design-based or model-based approach?

Design-based method best choice when:

- we want to estimate the distribution function or parameters thereof (mean, median, P90 etc.) for the area as a whole or for subareas;
- objective estimates of target properties are required, i.e. no subjective judgement on ‘representativeness’, no subjective model choices;
- objective estimates of *estimation variance* or *confidence intervals* are required, i.e. validity.

Design-based or model-based approach?

Design-based method best choice when:

- we want to estimate the distribution function or parameters thereof (mean, median, P90 etc.) for the area as a whole or for subareas;
- objective estimates of target properties are required, i.e. no subjective judgement on 'representativeness', no subjective model choices;
- objective estimates of *estimation variance* or *confidence intervals* are required, i.e. validity.

Design-based or model-based approach?, continued

Model based method best choice when:

- we want to map the target property (making predictions for unvisited locations);
- sample size large enough for calibrating a model of variation (e.g. variogram: $n > 100$);
- strong autocorrelation exists, from which we may profit in mapping;

Further reading: D.J. Brus and J.J. de Gruijter (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma, 80: 1-59.

Design-based or model-based approach?, continued

Model based method best choice when:

- we want to map the target property (making predictions for unvisited locations);
- sample size large enough for calibrating a model of variation (e.g. variogram: $n > 100$);
- strong autocorrelation exists, from which we may profit in mapping;

Further reading: D.J. Brus and J.J. de Gruijter (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma, 80: 1-59.

Design-based or model-based approach?, continued

Model based method best choice when:

- we want to map the target property (making predictions for unvisited locations);
- sample size large enough for calibrating a model of variation (e.g. variogram: $n > 100$);
- strong autocorrelation exists, from which we may profit in mapping;

Further reading: D.J. Brus and J.J. de Gruijter (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma, 80: 1-59.

Design-based or model-based approach?, continued

Model based method best choice when:

- we want to map the target property (making predictions for unvisited locations);
- sample size large enough for calibrating a model of variation (e.g. variogram: $n > 100$);
- strong autocorrelation exists, from which we may profit in mapping;

Further reading: D.J. Brus and J.J. de Gruijter (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma, 80: 1-59.

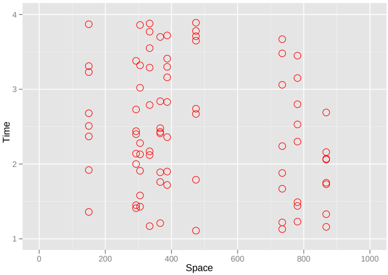
Design-based or model-based approach?, continued

Model based method best choice when:

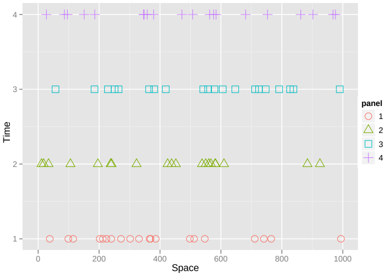
- we want to map the target property (making predictions for unvisited locations);
- sample size large enough for calibrating a model of variation (e.g. variogram: $n > 100$);
- strong autocorrelation exists, from which we may profit in mapping;

Further reading: D.J. Brus and J.J. de Gruijter (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma, 80: 1-59.

2. Sampling for mapping of *changes* in soil carbon stocks (*monitoring*)

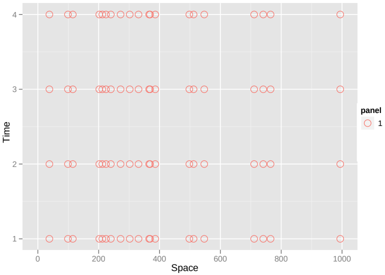


Static

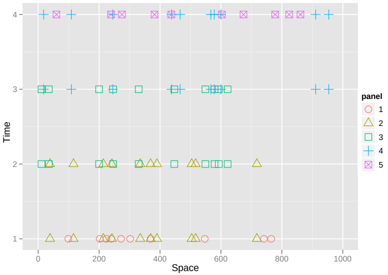


Synchronous

2. Sampling for mapping of *changes* in soil carbon stocks (*monitoring*)

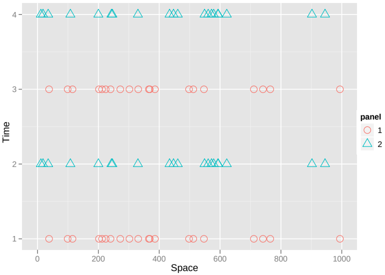


Static-Synchronous

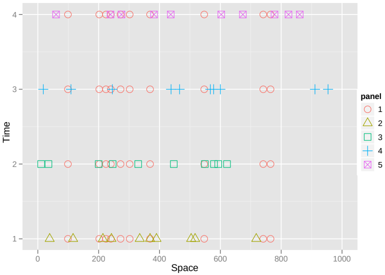


Rotational

2. Sampling for mapping of *changes* in soil carbon stocks (*monitoring*)

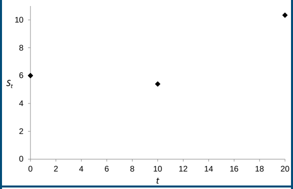


Serially alternating

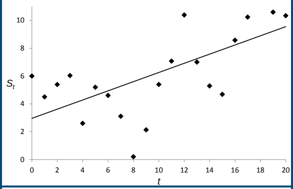


Supplemented

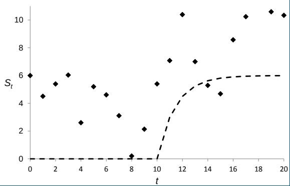
Changes: differences, trends or effects?



difference,
$d_{10-0} = -0.6$ kg m^{-2}
$d_{20-10} = 4.9$ kg m^{-2}



linear trend,
0.33 kg m^{-2} year^{-1}



effect of policy,
intervention at $t = 10$,
step decay model,
effect $=$6.0 kg m^{-2} after 10 years

2. Sampling for monitoring

Possible selection modes for various aims of monitoring

space	time	
	purposive	random
purposive	space-time mapping	–
random	estimation of trends in spatial means, totals, areal proportions*	model-free estimation of space-time means

*) 'Trend' can be defined as a time-series model parameter or as *a linear combination of true but unknown spatial means*, see: Brus, D.J., de Gruijter, J.J., Design-based Generalized Least Squares estimation of status and trend of soil properties from monitoring data, *Geoderma* (2011), doi:10.1016/j.geoderma.2011.06.001

3. Sampling for validation

- Validation: testing whether a map or a model satisfies its purpose
- Objectivity is crucial in validation (so that the validation procedure cannot get the blame for bad results)
- Collection of additional data by probability sampling is therefore recommended

For sampling aspects of map validation, see Brus, D.J., B. Kempen & G.B.M. Heuvelink, 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* **62**: 394-407.

3. Sampling for validation

- Validation: testing whether a map or a model satisfies its purpose
- Objectivity is crucial in validation (so that the validation procedure cannot get the blame for bad results)
- Collection of additional data by probability sampling is therefore recommended

For sampling aspects of map validation, see Brus, D.J., B. Kempen & G.B.M. Heuvelink, 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* **62**: 394-407.

3. Sampling for validation

- Validation: testing whether a map or a model satisfies its purpose
- Objectivity is crucial in validation (so that the validation procedure cannot get the blame for bad results)
- Collection of additional data by probability sampling is therefore recommended

For sampling aspects of map validation, see Brus, D.J., B. Kempen & G.B.M. Heuvelink, 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* **62**: 394-407.

3. Sampling for validation

- Validation: testing whether a map or a model satisfies its purpose
- Objectivity is crucial in validation (so that the validation procedure cannot get the blame for bad results)
- Collection of additional data by probability sampling is therefore recommended

For sampling aspects of map validation, see Brus, D.J., B. Kempen & G.B.M. Heuvelink, 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* **62**: 394-407.

Concluding remarks

- Sampling for mapping (spatial prediction): directed sampling or probability sampling, striving for fair spatial distribution.
- Sampling for monitoring of trends in spatial means: directed sampling in time, probability sampling in space. Revisiting of locations is recommended.
- Sampling for validation: collection of additional data by probability sampling is recommended.

Concluding remarks

- Sampling for mapping (spatial prediction): directed sampling or probability sampling, striving for fair spatial distribution.
- Sampling for monitoring of trends in spatial means: directed sampling in time, probability sampling in space. Revisiting of locations is recommended.
- Sampling for validation: collection of additional data by probability sampling is recommended.

Concluding remarks

- Sampling for mapping (spatial prediction): directed sampling or probability sampling, striving for fair spatial distribution.
- Sampling for monitoring of trends in spatial means: directed sampling in time, probability sampling in space. Revisiting of locations is recommended.
- Sampling for validation: collection of additional data by probability sampling is recommended.

Concluding remarks

- Sampling for mapping (spatial prediction): directed sampling or probability sampling, striving for fair spatial distribution.
- Sampling for monitoring of trends in spatial means: directed sampling in time, probability sampling in space. Revisiting of locations is recommended.
- Sampling for validation: collection of additional data by probability sampling is recommended.

Thank you!