A sample coordination method suitable for environmental monitoring

Xin Zhao & Anton Grafström

Department of Forest Resource Management
Swedish University of Agriculture Sciences

17 June

Content

Motivation

- High cost in environmental monitoring programs.
- Auxiliary information from remote sensing available.

Motivation

- High cost in environmental monitoring programs.
- Auxiliary information from remote sensing available.
- Efficient sampling strategies to guide the sample selection.

Aim

- Construct a framework for long-term environmental monitoring that has potential to produce superior estimators.

  (i) Current state.

  (ii) Change.

General ideas of the new sampling strategy

- A continuous framework and a double sampling approach are employed.

General ideas of the new sampling strategy

- A continuous framework and a double sampling approach are employed.
- Spatially balanced sampling
- Positive coordination of samples over time

General ideas of the new sampling strategy

- A continuous framework and a double sampling approach are employed.
- Spatially balanced sampling — improve the state estimators.
- Positive coordination of samples over time

General ideas of the new sampling strategy

- A continuous framework and a double sampling approach are employed.
- Spatially balanced sampling — improve the state estimators.
- Positive coordination of samples over time — improve the change estimators.

A continuous population

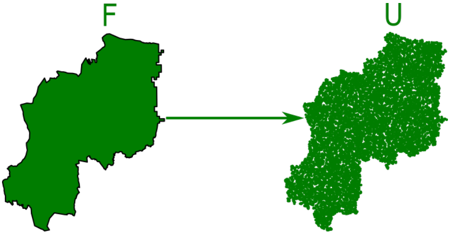
A surface with its surface area $\ell(F)$.

A continuous population

- Response of a target variable for the point $\mathbf{x}$ at time t can be denoted as $y_t(\mathbf{x})$.
- Population total at time t: $Y(t) = \int_F y_t(\mathbf{x})d\mathbf{x}$.
- $\pi_t(\mathbf{x})$: (prescribed) sampling intensity.

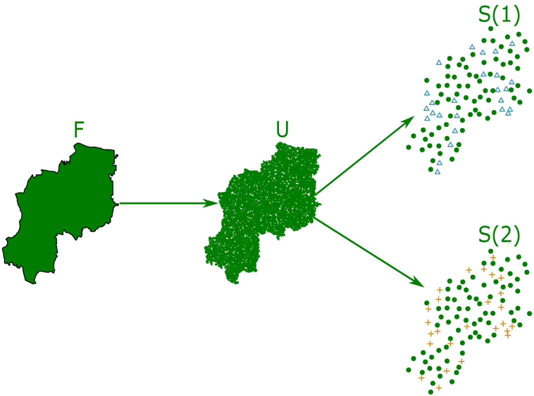
$$\int_F \pi_t(\mathbf{x})d\mathbf{x} = n(t).$$

New sampling strategy



- Select a very large sample of N locations using $\pi(\mathbf{x}) = \frac{N}{\ell(F)}$.
- $U = \{1, ..., i, ..., N\}$.
- Derive auxiliary responses for each unit from U at different time occasions.

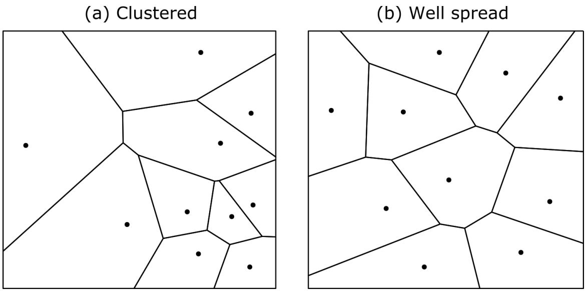
New sampling strategy



- Select positively coordinated and well-spread $S(t)$ from U using auxiliary information by spatially correlated Poisson sampling (SCPS).

Spatial balance

- Vornoi polytopes (Stevens & Olsen, 2004) to describe the spatial balance.



Sample coordination

- Maximize the overlap between samples drawn from overlapping populations.

Sample coordination

- Maximize the overlap between samples drawn from overlapping populations.
- Selection of a new sample depend on the samples previously drawn.

Unbiased Estimator

- The unbiased Horvitz-Thompson (HT) estimator of the population total $Y(t)$ is then defined as

$$\widehat{Y}(t) = \sum_{i \in S(t)} \frac{y_t(\mathbf{x}_i)}{\pi_t(\mathbf{x}_i)}, \tag{1}$$

where $\pi_t(\mathbf{x}_i) = \pi(\mathbf{x}_i) \cdot \pi_i(t)$.

- A list-sequential sampling method and a spatial modification of the correlated Poisson sampling (Bondesson & Thorburn, 2008).

- A list-sequential sampling method and a spatial modification of the correlated Poisson sampling (Bondesson & Thorburn, 2008).
- Update inclusion probabilities of units in U by N steps.

- A list-sequential sampling method and a spatial modification of the correlated Poisson sampling (Bondesson & Thorburn, 2008).
- Update inclusion probabilities of units in U by N steps.
- Assign each unit in the list a permanent random number r_i, with $r_1, r_2, ..., r_N$ i.i.d. $U(0, 1)$.

- A list-sequential sampling method and a spatial modification of the correlated Poisson sampling (Bondesson & Thorburn, 2008).
- Update inclusion probabilities of units in U by N steps.
- Assign each unit in the list a permanent random number r_i, with $r_1, r_2, ..., r_N$ i.i.d. $U(0,1)$.
- At each step and time, $I_j(t) = 1$, if $r_j < \pi_j^{(j-1)}(t)$.

- The updating can be illustrated as

$$
\begin{array}{rccccccc}
\boldsymbol{\pi}^{(0)}(t): & \pi_1(t) & \pi_2(t) & \pi_3(t) & \pi_4(t) & \cdots & \pi_N(t) \\
\boldsymbol{\pi}^{(1)}(t): & I_1(t) & \pi_2^{(1)}(t) & \pi_3^{(1)}(t) & \pi_4^{(1)}(t) & \cdots & \pi_N^{(1)}(t) \\
\boldsymbol{\pi}^{(2)}(t): & I_1(t) & I_2(t) & \pi_3^{(2)}(t) & \pi_4^{(2)}(t) & \cdots & \pi_N^{(2)}(t) \\
\boldsymbol{\pi}^{(3)}(t): & I_1(t) & I_2(t) & I_3(t) & \pi_4^{(3)}(t) & \cdots & \pi_N^{(3)}(t) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{\pi}^{(N)}(t): & I_1(t) & I_2(t) & I_3(t) & I_4(t) & \cdots & I_N(t)
\end{array}
$$

Illustration of the algorithm for two time occasions

- Example: $N = 4$ and $n(t) = 2$. The visiting order is 1, 2, 3, 4.

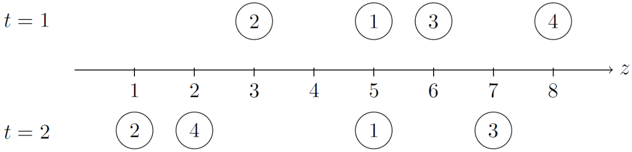


Illustration of the algorithm for two time occasions

	Time occasion 1	Time occasion 2
	$\boxed{\boldsymbol{\pi}^{(0)}(1) : \frac{3}{4}, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}}$	$\boxed{\boldsymbol{\pi}^{(0)}(2) : \frac{3}{4}, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}}$
Step 1:	$r_1 = 0.9821 > \pi_1(1) \Longrightarrow I_1(1) = 0$	$r_1 = 0.9821 > \pi_1(2) \Longrightarrow I_1(2) = 0$
	$\downarrow$	$\downarrow$
	$w_3^{(1)}(1) = \frac{2}{3} \Longrightarrow \pi_3^{(1)}(1) = 1$	$w_3^{(1)}(2) = \frac{2}{3} \Longrightarrow \pi_3^{(1)}(2) = 1$
	$\downarrow$	$\downarrow$
	$w_2^{(1)}(1) = \frac{1}{3} \Longrightarrow \pi_2^{(1)}(1) = \frac{3}{4}$	$w_4^{(1)}(2) = \frac{1}{3} \Longrightarrow \pi_4^{(1)}(2) = \frac{1}{2}$
	$\downarrow$	$\downarrow$
	$w_4^{(1)}(1) = 0 \Longrightarrow \pi_4^{(1)}(1) = \frac{1}{4}$	$w_2^{(1)}(2) = 0 \Longrightarrow \pi_2^{(1)}(2) = \frac{1}{2}$
	$\downarrow$	$\downarrow$
	$\boxed{\boldsymbol{\pi}^{(1)}(1) : 0, \frac{3}{4}, 1, \frac{1}{4}}$	$\boxed{\boldsymbol{\pi}^{(1)}(2) : 0, \frac{1}{2}, 1, \frac{1}{2}}$
Step 2:	$r_2 = 0.6782 < \pi_2(1) \Longrightarrow I_2(1) = 1$	$r_2 = 0.6782 > \pi_2(2) \Longrightarrow I_2(2) = 0$
	$\downarrow$	$\downarrow$
	$w_3^{(2)}(1) = 0 \Longrightarrow \pi_3^{(2)}(1) = 1$	$w_4^{(2)}(2) = 1 \Longrightarrow \pi_4^{(2)}(2) = 1$
	$\downarrow$	$\downarrow$
	$w_4^{(2)}(1) = 1 \Longrightarrow \pi_4^{(2)}(1) = 0$	$w_3^{(2)}(2) = 0 \Longrightarrow \pi_3^{(2)}(2) = 1$
	$\downarrow$	$\downarrow$
	$\boxed{\boldsymbol{\pi}^{(2)}(1) : 0, 1, 1, 0}$	$\boxed{\boldsymbol{\pi}^{(2)}(2) : 0, 0, 1, 1}$
	$\downarrow$	$\downarrow$
	$S(1)$: 2, 3	$S(2)$: 3, 4

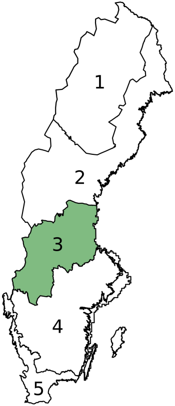
Variance estimator for spatially balanced samples

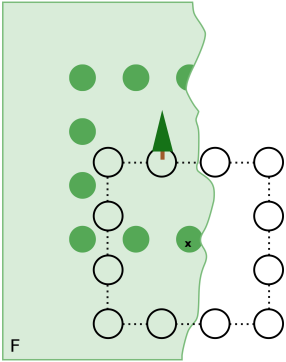
- Grafström & Schelin (2014) derived an approximate variance estimator under spatially balanced sampling. It can be expressed as

$$\widehat{V}(\widehat{Y}_t) = \frac{1}{2}\sum_{i \in S_t}\left[\frac{y_i(t)}{\pi_i(t)} - \frac{y_{i'}(t)}{\pi_{i'}(t)}\right]^2, \qquad (2)$$

where i' is the nearest neighbour to i in the random sample with n_t locations selected at time t.

The study region





Auxiliary variables

- Geographic coordinates of the cluster center.

Auxiliary variables

- Geographic coordinates of the cluster center.
- Mean elevation of the cluster.

Auxiliary variables

- Geographic coordinates of the cluster center.
- Mean elevation of the cluster.
- Cluster mean tree height and mean basal area.

Five strategies

- Strategy 1: Positive coordination using SCPS.

Five strategies

- Strategy 1: Positive coordination using SCPS.
- Strategy 2: Permanent geographic-spread sample over time.

Five strategies

- Strategy 1: Positive coordination using SCPS.
- Strategy 2: Permanent geographic-spread sample over time.
- Strategy 3: Permanent sample selected by SCPS which is only well spread at the first time occasion.

Five strategies

- Strategy 1: Positive coordination using SCPS.
- Strategy 2: Permanent geographic-spread sample over time.
- Strategy 3: Permanent sample selected by SCPS which is only well spread at the first time occasion.
- Strategy 4: Independent well-spread samples selected by SCPS over time.

Five strategies

- Strategy 1: Positive coordination using SCPS.
- Strategy 2: Permanent geographic-spread sample over time.
- Strategy 3: Permanent sample selected by SCPS which is only well spread at the first time occasion.
- Strategy 4: Independent well-spread samples selected by SCPS over time.
- Strategy 5: Split panel designs to split the sample into two parts (permanent geographic-spread + well-spread), it stands for the current strategy of the Swedish NFI.

Simulation result

Strategy	SB	Overlap	$V\left(\widehat{\overline{Z}}_h(1)\right)$	$V\left(\widehat{\overline{Z}}_h(2)\right)$	$V\left(\widehat{\Delta}_{\overline{Z}_h(1,2)}\right)$	$V\left(\widehat{\overline{Z}}_b(1)\right)$	$V\left(\widehat{\overline{Z}}_b(2)\right)$	$V\left(\widehat{\Delta}_{\overline{Z}_b(1,2)}\right)$
1	0.127	62	0.809	0.744	0.978	0.017	0.017	0.022
2	0.238	100	10.311	10.165	2.130	0.233	0.242	0.048
3	0.171	100	0.809	2.475	1.969	0.017	0.058	0.043
4	0.128	1	0.809	0.776	1.608	0.017	0.018	0.035
P_{28}	0.167	20	3.157	1.115	4.169	0.071	0.026	0.094
P_{55}	0.209	50	4.694	2.122	5.986	0.106	0.050	0.138
P_{73}	0.228	70	6.486	4.111	7.370	0.148	0.098	0.171
P_{82}	0.236	80	7.869	6.125	7.411	0.177	0.144	0.170

- We can achieve great improvements for the state estimators and some improvements for the estimator of change for the auxiliary variables.

- We can achieve great improvements for the state estimators and some improvements for the estimator of change for the auxiliary variables.
- If we use spatially balanced samples without doing positive coordination, we can only improve the state estimators.

- We can achieve great improvements for the state estimators and some improvements for the estimator of change for the auxiliary variables.
- If we use spatially balanced samples without doing positive coordination, we can only improve the state estimators.
- Potential to reduce the variance for the target variables related to the auxiliary variables.

- We can achieve great improvements for the state estimators and some improvements for the estimator of change for the auxiliary variables.
- If we use spatially balanced samples without doing positive coordination, we can only improve the state estimators.
- Potential to reduce the variance for the target variables related to the auxiliary variables.
- Potential to change the complex strategy of the current Swedish NFI.

Bondesson, L. & Thorburn, D. (2008). A List Sequential Sampling Method Suitable for Real-Time Sampling. *Scandinavian Journal of Statistics*, 35, 466-483.

Deville, J.-C. & Tillé, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika*, 85, 89-101.

Grafström, A. (2012). Spatially correlated Poisson sampling. *Journal of Statistical Planning and Inference*, 142(1), 139-147.

Grafström, A. & Schelin, L. (2014). How to Select Representative Samples. *Scandinavian Journal of Statistics*, 41(2), 277-290.

Stevens, D.L. & Olsen, A.R. (2004). Spatially Balanced Sampling of Natural Resources. *Journal of the American Statistical Association*, 99(465), 262-278.