Design-based maps for two-phase environmental surveys

S. Franceschi

University of Calabria, Italy

Florence, June 6 2019

Environmental surveys generally address

- design-based estimation of descriptive summaries, such as forest cover and total biomass
- conservative estimation of the design-based variances
- mapping for a visual overview of the spatial pattern of an attribute of interest, usually performed in a model-based approach

Environmental surveys generally address

- design-based estimation of descriptive summaries, such as forest cover and total biomass
- conservative estimation of the design-based variances
- mapping for a visual overview of the spatial pattern of an attribute of interest, usually performed in a model-based approach

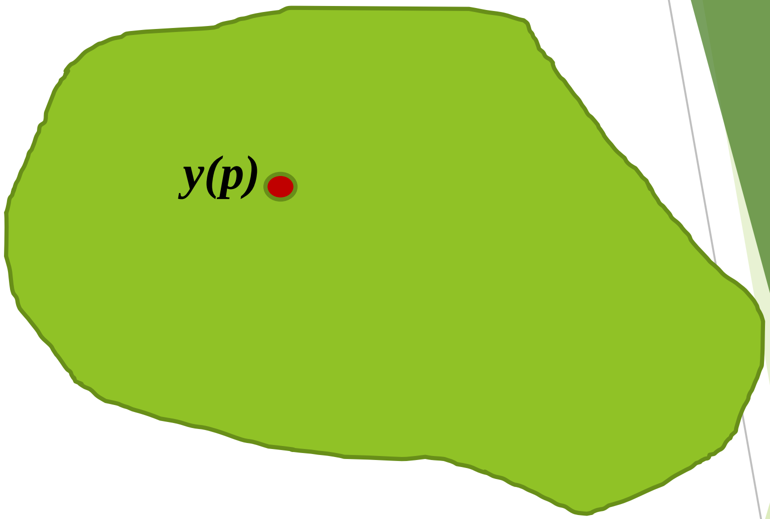
In a **design-based framework**

Some general notation and statement of the problem

$\mathcal{A}$ study region of size A
(connected and compact set of $\mathbb{R}^2$)

$p \in \mathcal{A}$ population unit

$y(p)$ amount of the survey variable
Y at $p \in \mathcal{A}$



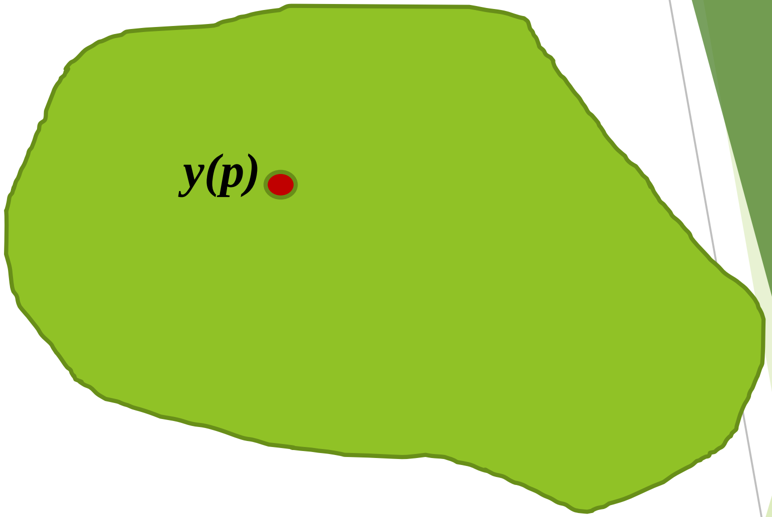
Some general notation and statement of the problem

$\mathcal{A}$ study region of size A
(connected and compact set of R^2)

$p \in \mathcal{A}$ population unit

$y(p)$ amount of the survey variable Y at $p \in \mathcal{A}$

$\boxed{y(p) \text{ is a surface onto } \mathcal{A}}$

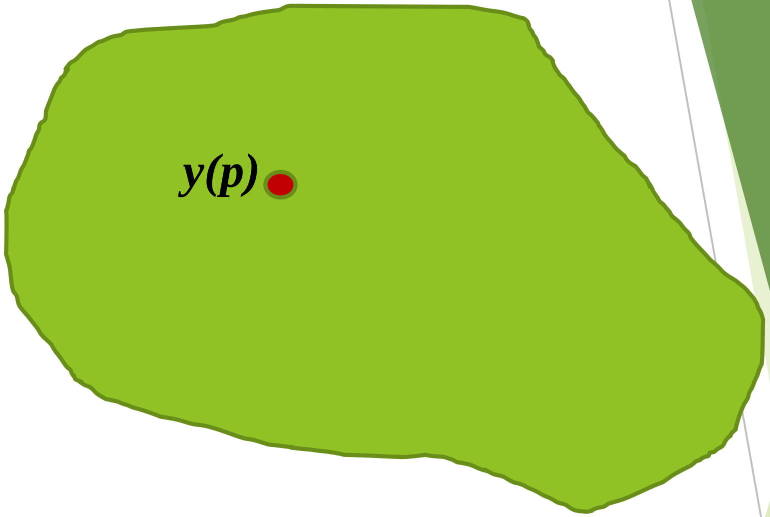


Some general notation and statement of the problem

$\mathcal{A}$ study region of size A
(connected and compact set of R^2)

$p \in \mathcal{A}$ population unit

$y(p)$ amount of the survey variable
Y at $p \in \mathcal{A}$

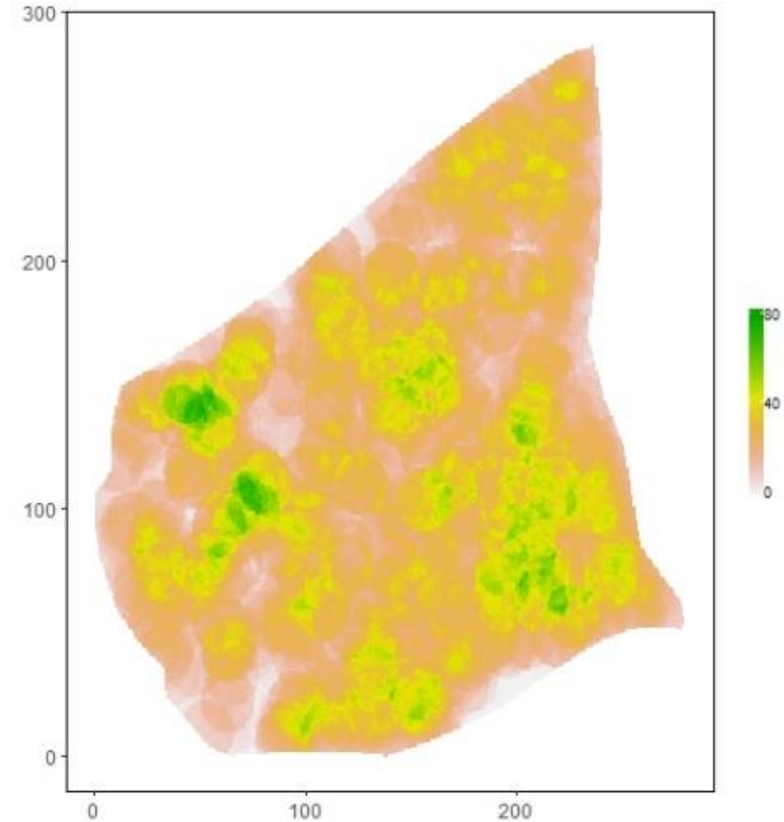


$\boxed{y(p) \text{ is a surface onto } \mathcal{A}}$

Aim:
to obtain a whole map of the population,
i.e. estimating $y(p)$ for any $p \in \mathcal{A}$
on the basis of a sample of points selected from $\mathcal{A}$

Example

$y(p)$ is the basal area (m^2) within a circular plot of radius $13m$ in Val di Sella (northeastern Italy, southern Alpine area)



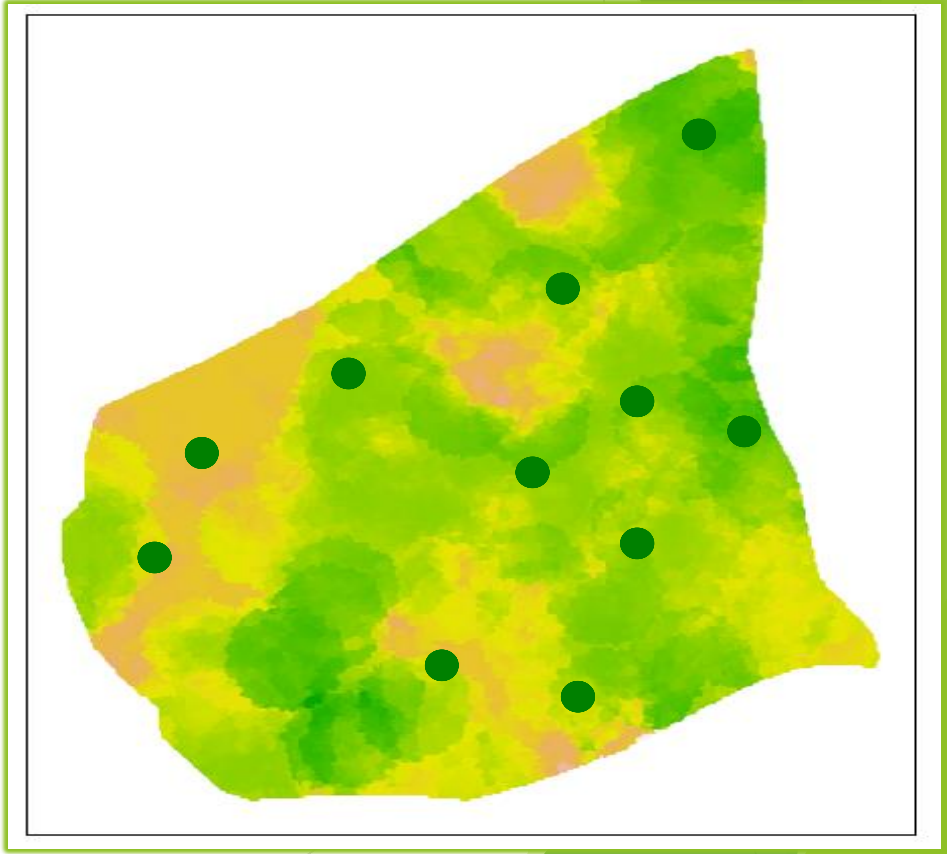
Sampling

in most cases, the available resources render impractical to completely census the entire area

Sampling

in most cases, the available resources render impractical to completely census the entire area

$y(p)$ is recorded only for a sample of n points $p_1, p_2, \ldots, p_n$ ●

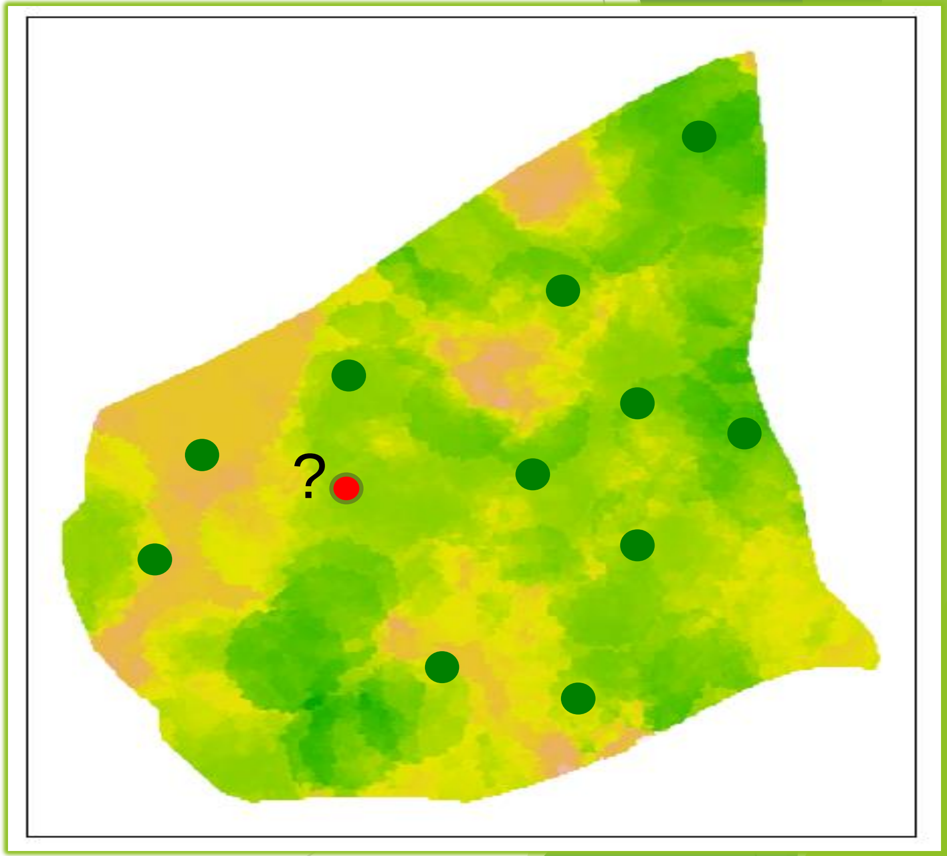


Sampling

in most cases, the available resources render impractical to completely census the entire area

$y(p)$ is recorded only for a sample of n points $p_1, p_2, \ldots, p_n$

a criterion is adopted to estimate $y(p)$ at an unsampled location p



Model-dependent inference

under model-dependent approaches the properties of the resulting maps are determined by the

super-population (model)

which has been supposed to generate the surface

Model-dependent inference

under model-dependent approaches the properties of the resulting maps are determined by the

super-population (model)

which has been supposed to generate the surface

The most common techniques are

- the kriging predictors (e.g. Cressie 1993)
- other model-dependent methods exploiting the auxiliary information
  - regression kriging (Christensen, 2001)
  - cokriging (Cressie 1993)
  - locally weighted regression (Cleveland and Devlin 1988)
  - *k*-nearest neighbour (McRoberts et al. 2007)

Design-based framework

Design-based framework

the properties of the resulting maps are determined by the probabilistic sampling scheme adopted to select the n points

Design-based framework

the properties of the resulting maps are determined by the probabilistic sampling scheme adopted to select the n points

drawbacks and merits of design-based and model-based approaches are well delineated in statistical literature:

e.g. Smith 2001, Gregoire 1998, Thompson 2002,
Schreuder et al. 1993

"Design-based inference is objective, nobody can challenge that the sample was really selected according to the given sampling design. The probability distribution associated with the design is real, not modelled or assumed" Särndal et al. (1992, p. 21)

Environmental surveys are usually performed in two phases

Environmental surveys are usually performed in two phases

- **first phase**: a collection of points are located throughout the study area by means of a suitable sampling scheme (systematic grid sampling, tessellation stratified sampling, Gräfstrom balanced sampling schemes)

Environmental surveys are usually performed in two phases

- **first phase**: a collection of points are located throughout the study area by means of a suitable sampling scheme (systematic grid sampling, tessellation stratified sampling, Gräfstrom balanced sampling schemes)
- **second phase**: points are classified by means of satellite or aerial information and a sample of points is selected (stratified design)

Environmental surveys are usually performed
in two phases

- **first phase**: a collection of points are located throughout the study area by means of a suitable sampling scheme (systematic grid sampling, tessellation stratified sampling, Gräfstrom balanced sampling schemes)
- **second phase**: points are classified by means of satellite or aerial information and a sample of points is selected (stratified design)

↓

any second-phase point is visited
and survey variable is recorded

First-phase estimation

(Fattorini et al. 2017, Biometrika)

If all the first-phase n points $P_1, P_2, \ldots, P_n$ were visited on the ground

↓

$y(P_1)$, $y(P_2), \ldots,$ $y(P_n)$ were recorded

First-phase estimation
(Fattorini et al. 2017, Biometrika)

If all the first-phase n points $P_1, P_2, \ldots, P_n$ were visited on the ground

$\Downarrow$

$y(P_1),\ y(P_2), \ldots,\ y(P_n)$ were recorded

Exploiting Tobler's first law of geography, the authors adopt the inverse distance weighting (IDW) interpolator for estimating values at the unsampled locations by a weighted sum of the sampled values, with weights $w_i(p, P_i)$ inversely decreasing with distances

First-phase estimation

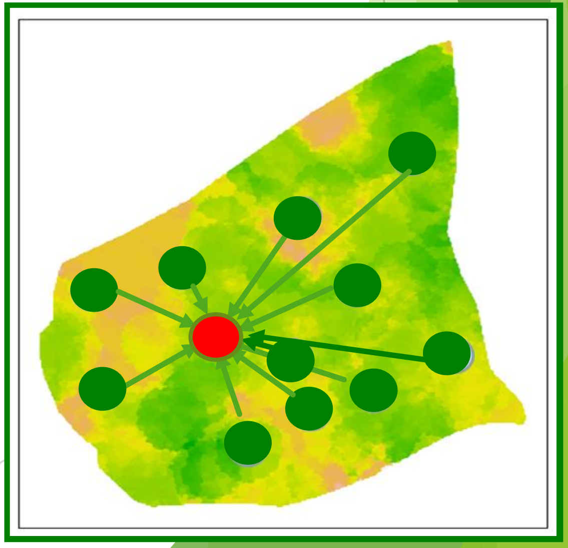
(Fattorini et al. 2017, Biometrika)

If all the first-phase n points $P_1, P_2, \ldots, P_n$ were visited on the ground

$y(P_1)$, $y(P_2), \ldots,$ $y(P_n)$ were recorded

Exploiting Tobler's first law of geography, the authors adopt the inverse distance weighting (IDW) interpolator for estimating values at the unsampled locations by a weighted sum of the sampled values, with weights $w_i(p, P_i)$ inversely decreasing with distances

$$\hat{y}(p) = \sum_{i=1}^{n} w_i(p, P_i)\, y(P_i)$$



sampled first-phase points

unsampled location

$\mathrm{E}\left(\hat{y}(p)\right) \neq y(p)$ Bias how much?

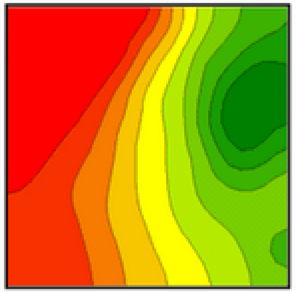
$\mathrm{V}\left(\hat{y}(p)\right) = ?$ Precision?

$\mathrm{E}\big(\hat{y}(p)\big) \neq y(p)$ Bias how much?

$\mathrm{V}\big(\hat{y}(p)\big) = ?$ Precision?

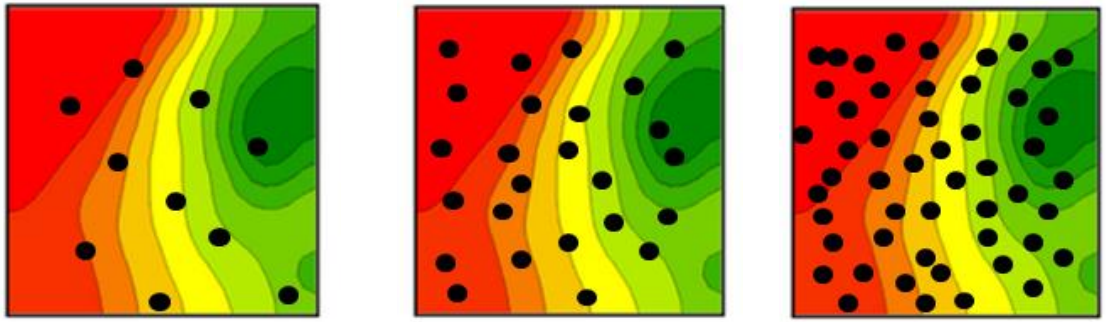
Investigations are needed
for DBAU&C

Asymptotic framework



The surface *y*(*p*) is fixed

Suppose a sequence of designs $\{D_k\}$ to select a sample of locations $S_k = (P_1^{(k)}, \ldots, P_{n_k}^{(k)})$ of size n_k from $\mathcal{A}$, with $n_k \to \infty$ as *k* increases



…

Properties of the first-phase IDW interpolator

Under suitable assumptions they proved that, as the number of sampled locations increases, the estimated map converges to the true map unless a set of zero measure, that is

$$\lim_{n\to\infty} \int_{\mathcal{A}} \mathrm{E}\{|\hat{y}(p) - y(p)|\}\, dp = 0$$

Fattorini et al. (2017)

The IDW interpolator is almost everywhere **DAU&C**

- **D**esign **A**symptotically **U**nbiased
- **C**onsistent

Conditions for DAU&C of the first-phase IDW interpolator

C1. $y(p)$ is a piecewise continuous function

Conditions for DAU&C
of the first-phase IDW interpolator

C1. $y(p)$ is a piecewise continuous function

In real world there are parts of the study region in which density changes smoothly throughout space, well approaching the theoretical condition of continuity

Conditions for DAU&C
of the first-phase IDW interpolator

C1. $y(p)$ is a piecewise continuous function

In real world there are parts of the study region in which density changes smoothly throughout space, well approaching the theoretical condition of continuity

Even when density changes abruptly, that usually occurs along borders delineating sudden variations in the characteristic of the study region

These borders may be realistically approximated by curves well approaching the theoretical condition of discontinuity over a region of zero measure

Conditions for DAU&C
of the first-phase IDW interpolator

C1. $y(p)$ is a piecewise continuous function 🙂

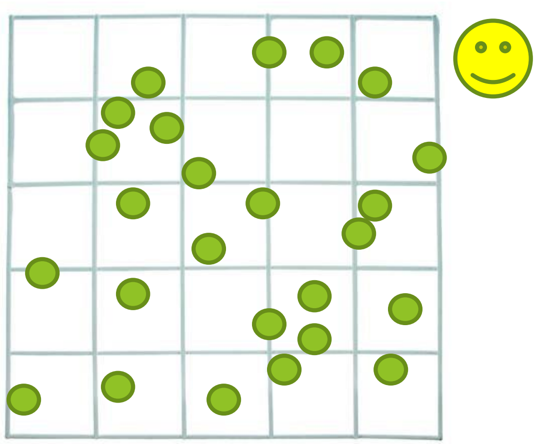
Conditions for DAU&C of the first-phase IDW interpolator

C1. $y(p)$ is a piecewise continuous function

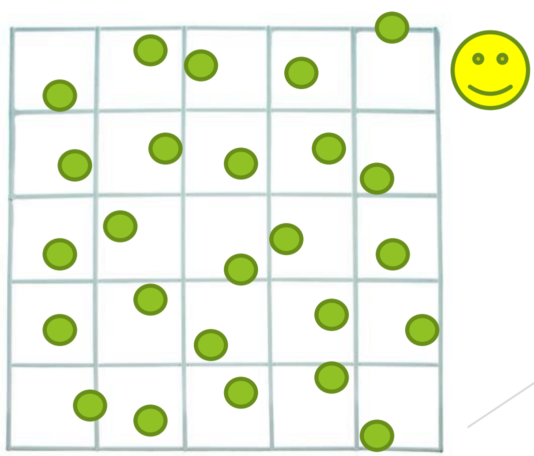
C2. weights should decrease as a power of distance

C3. for large n the sampling scheme is able to evenly spread the sample points in such a way that any unsampled location is likely to have neighbouring points sampled → asymptotical spatial balance

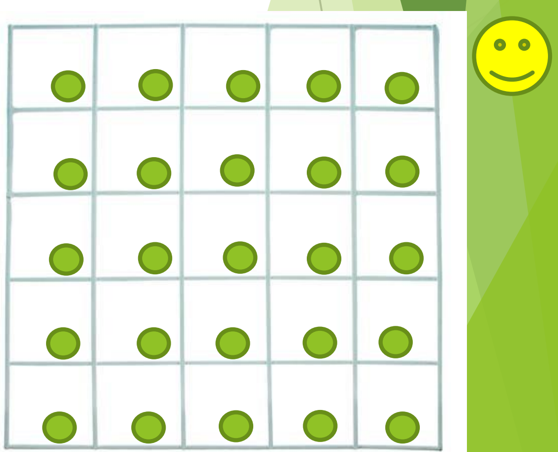
Uniform random sampling



Tesselation stratified sampling



Systematic grid sampling



What about DAU&C
when a two-phase sampling
is implemented?

Second-phase estimation

(work in progress)

Second-phase estimation
(work in progress)

- a sample S of m first-phase points is selected by means of a suitable sampling scheme inducing first-order inclusion probabilities π_i , $i \in S$
- any second-phase point is visited on the ground
- $y(P_i)$ is recorded for any $i \in S$

Second-phase estimation
(work in progress)

- a sample S of m first-phase points is selected by means of a suitable sampling scheme inducing first-order inclusion probabilities π_i , $i \in S$
- any second-phase point is visited on the ground
- $y(P_i)$ is recorded for any $i \in S$

Estimation of values at the unsampled locations
can be performed by
two-phase IDW interpolators

$$\hat{y}(p, S) = \frac{\sum_{i=1}^{n} w_i(p, P_i) y(P_i) G_i}{\sum_{i=1}^{n} w_i(p, P_i) G_i}$$

where G_i are not-negative random variables equal to 0
if point i has not been selected in the second-phase sample S

Second-phase estimation
(work in progress)

- a sample S of m first-phase points is selected by means of a suitable sampling scheme inducing first-order inclusion probabilities $\pi_i\,, i \in S$
- any second-phase point is visited on the ground
- $y(P_i)$ is recorded for any $i \in S$

Estimation of values at the unsampled locations
can be performed by
two-phase IDW interpolators

$$\hat{y}(p,S) = \frac{\sum_{i=1}^{n} w_i(p,P_i)y(P_i)G_i}{\sum_{i=1}^{n} w_i(p,P_i)G_i}$$

where G_i are not-negative random variables equal to 0
if point i has not been selected in the second-phase sample S

$$\boxed{\text{different choices of } G_i\text{s give rise to different estimators}}$$

Two-phase ratio and direct interpolators

- Two-phase ratio interpolator: $G_i = 1/\pi_i \quad i \in S$

$$\hat{y}_R(p, S) = \frac{\sum_{i \in S} w_i(p, P_i) y(P_i) \pi_i^{-1}}{\sum_{i \in S} w_i(p, P_i)\, \pi_i^{-1}}$$

Two-phase ratio and direct interpolators

- Two-phase ratio interpolator: $G_i = 1/\pi_i \quad i \in S$

$$\hat{y}_R(p, S) = \frac{\sum_{i \in S} w_i(p, P_i) y(P_i) \pi_i^{-1}}{\sum_{i \in S} w_i(p, P_i)\, \pi_i^{-1}}$$

- Two-phase direct interpolator : $G_i = 1 \quad i \in S$

$$\hat{y}_D(p, S) = \frac{\sum_{i \in S} w_i(p, P_i) y(P_i)}{\sum_{i \in S} w_i(p, P_i)}$$

Two-phase ratio and direct interpolators

- Two-phase ratio interpolator: $G_i = 1/\pi_i \quad i \in S$

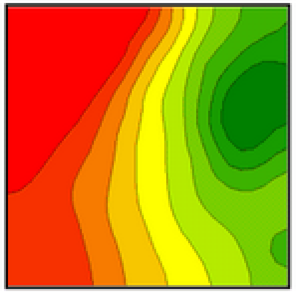
$$\hat{y}_R(p,S) = \frac{\sum_{i\in S} w_i(p,P_i)\, y(P_i)\, \pi_i^{-1}}{\sum_{i\in S} w_i(p,P_i)\, \pi_i^{-1}}$$

- Two-phase direct interpolator : $G_i = 1 \quad i \in S$

$$\hat{y}_D(p,S) = \frac{\sum_{i\in S} w_i(p,P_i)\, y(P_i)}{\sum_{i\in S} w_i(p,P_i)}$$

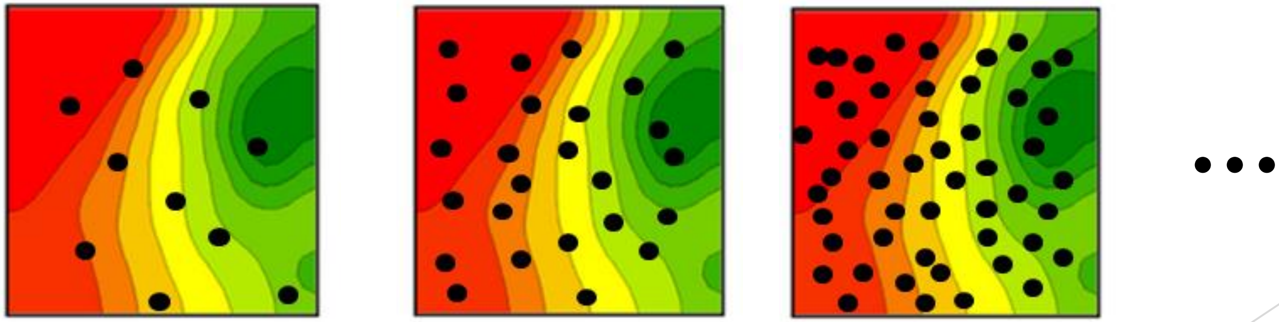
When $\pi_i = \pi_o$ for any $i \in S$ $\quad \hat{y}_R(p,S) = \hat{y}_D(p,S)$

Asymptotic framework

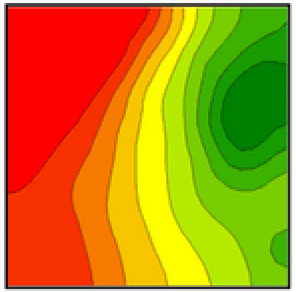


The surface $y(p)$ is fixed

Suppose a sequence of designs $\{D_k\}$ to select a second phase sample of locations $S_k = (P_1^{(k)}, \ldots, P_{m_k}^{(k)})$ of size m_k from $\mathcal{A}$, $n_k \to \infty$, $m_k \to \infty$ as k increases

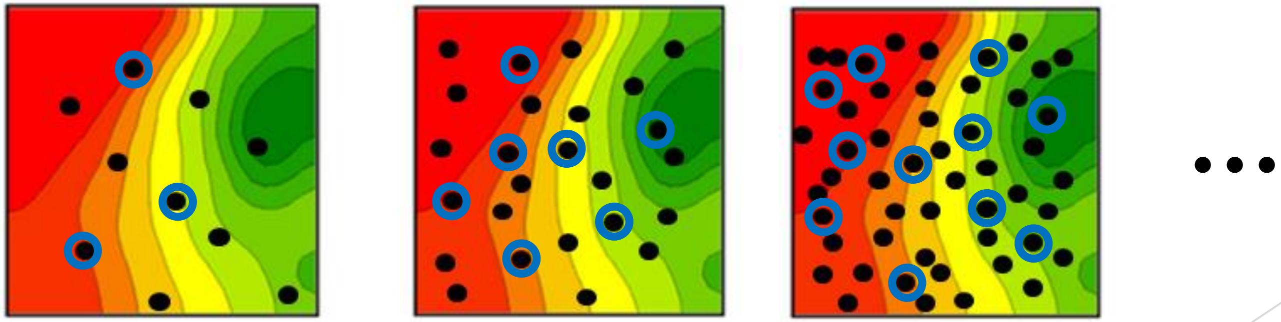


Asymptotic framework



The surface $y(p)$ is fixed

Suppose a sequence of designs $\{D_k\}$ to select a second phase sample of locations $S_k = (P_1^{(k)}, \ldots, P_{m_k}^{(k)})$ of size m_k from $\mathcal{A}$, $n_k \to \infty$, $m_k \to \infty$ as k increases



Asymptotic framework

Sequence of IDW interpolators

$$\hat{y}_k(p,S) = \frac{\sum_{i=1}^{n} w_i\left(p, P_i^{(k)}\right) y(P_i^{(k)}) G_i}{\sum_{i=1}^{n} w_i\left(p, P_i^{(k)}\right) G_i}$$

The IDW interpolator is defined to be **point-wise design-consistent** at $p \in \mathcal{A}$ if

$$p\lim_{k\to\infty} \left|\hat{y}_k(p) - y(p)\right| = 0$$

Properties of the two-phase interpolators

Properties of the two-phase interpolators

Under conditions

C1. $y(p)$ is a piecewise continuous function 🙂

C2. weights should decrease as a power of distance 🙂

Properties of the two-phase interpolators

Under conditions

C1. $y(p)$ is a piecewise continuous function

C2. weights should decrease as a power of distance

and

D3. for any $p \in \mathcal{A}$ and any $\epsilon > 0$ there exists $T > 0$ and an integer k_0 such that

$$\Pr\left\{Z_k(p, T/\sqrt{m_k}, \mathcal{S}_k) = 0\right\} < \varepsilon \quad \forall\, k > k_0$$

D4. $|G_i| \leq const$

Properties of the two-phase interpolators

Under conditions

C1. $y(p)$ is a piecewise continuous function

C2. weights should decrease as a power of distance

and

D3. for large m the sampling scheme is able to evenly spread the sample points in such a way that any unsampled location is likely to have neighbouring second-phase points sampled

D4. $|G_i| \leq const$

Properties of the two-phase interpolators

Under conditions

C1. $y(p)$ is a piecewise continuous function

C2. weights should decrease as a power of distance

and

D3. for large m the sampling scheme is able to evenly spread the sample points in such a way that any unsampled location is likely to have neighbouring second-phase points sampled

D4. $|G_i| \leq const$

Properties of the two-phase interpolators

Under conditions

C1. $y(p)$ is a piecewise continuous function

C2. weights should decrease as a power of distance

and

D3. for large m the sampling scheme is able to evenly spread the sample points in such a way that any unsampled location is likely to have neighbouring second-phase points sampled

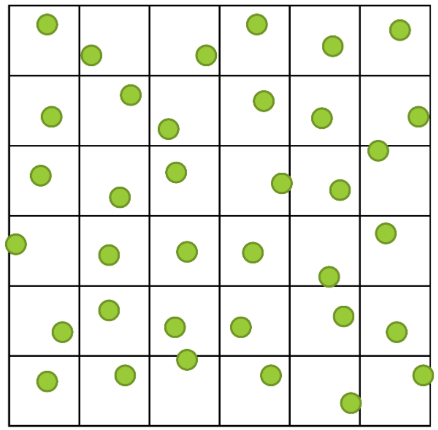
D4. $|G_i| \leq const$

the two-phase IDW interpolators
are almost everywhere **DAU&C**

Some two-phase sampling schemes ensuring DAU&C of the two-phase IDW interpolators

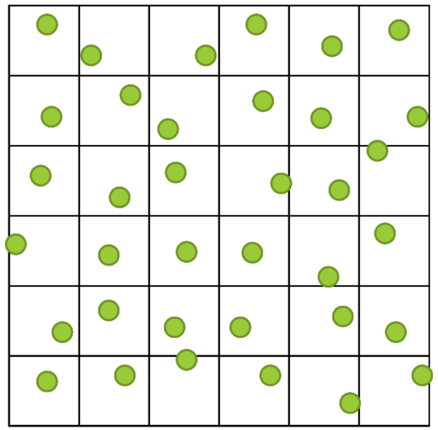
Some two-phase sampling schemes ensuring DAU&C of the two-phase IDW interpolators

First phase:
Tesselation stratified sampling

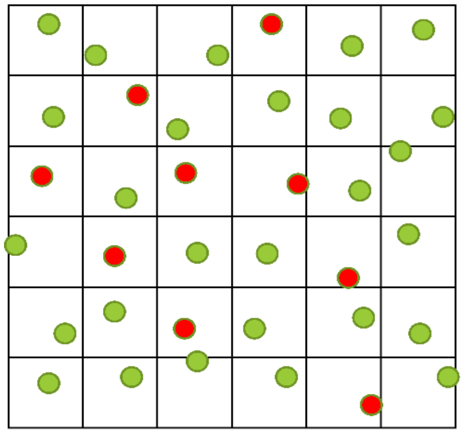


Some two-phase sampling schemes ensuring DAU&C of the two-phase IDW interpolators

First phase:
Tesselation stratified sampling

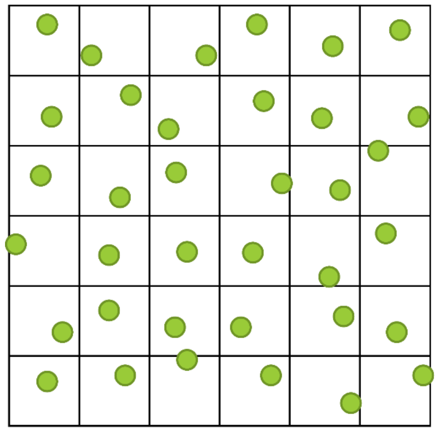


Second phase:
Simple random sampling without replacement

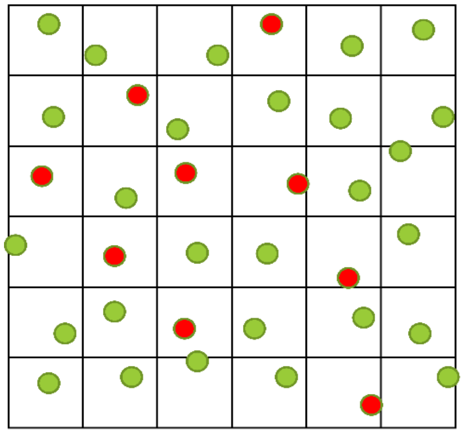


Some two-phase sampling schemes ensuring DAU&C of the two-phase IDW interpolators

First phase:
Tesselation stratified sampling

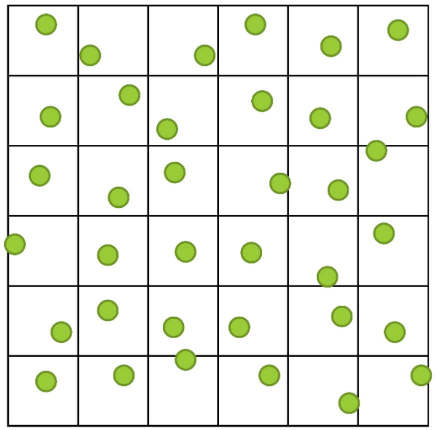


Second phase:
Simple random sampling without replacement

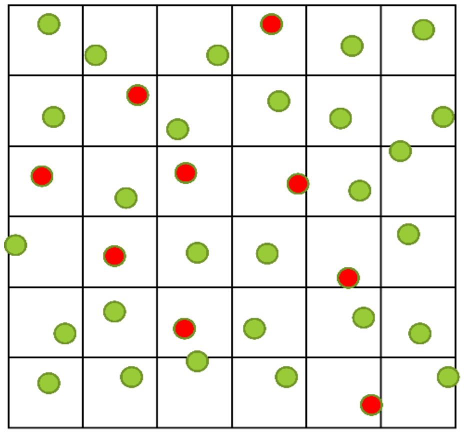


Some two-phase sampling schemes
ensuring DAU&C of the two-phase IDW interpolators

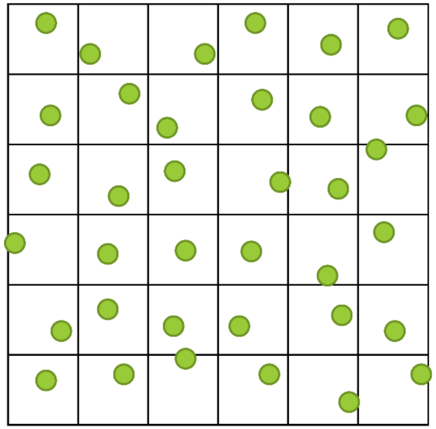
First phase:
Tesselation stratified sampling



Second phase:
Simple random sampling without replacement

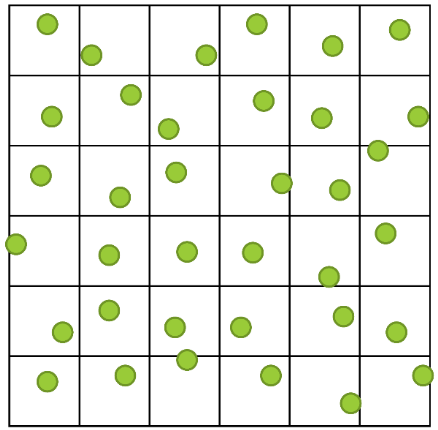


First phase:
Tesselation stratified sampling

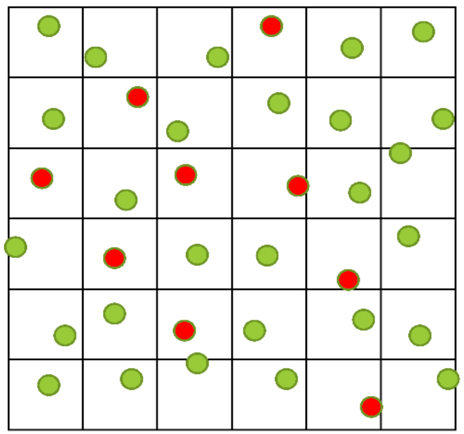


Some two-phase sampling schemes ensuring DAU&C of the two-phase IDW interpolators

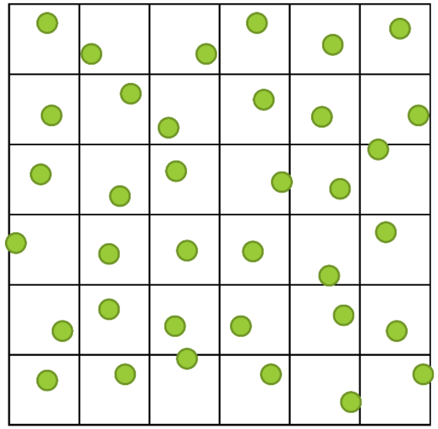
First phase:
Tesselation stratified sampling



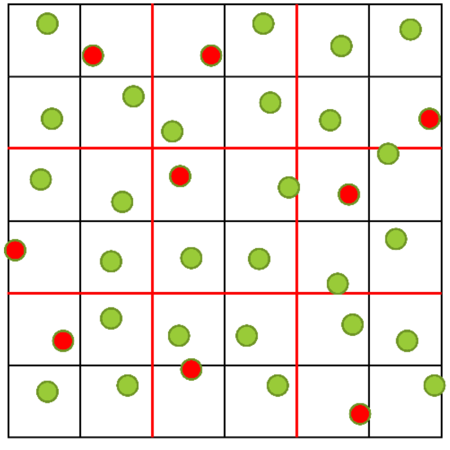
Second phase:
Simple random sampling without replacement



First phase:
Tesselation stratified sampling

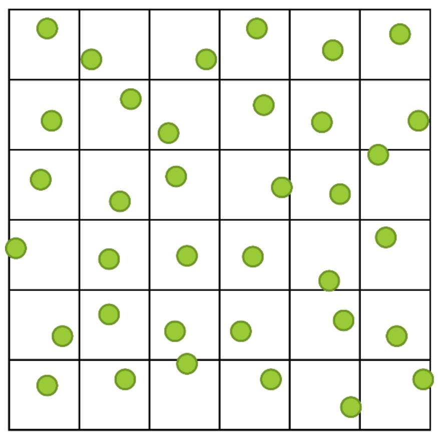


Second phase:
One-per-stratum-stratified sampling

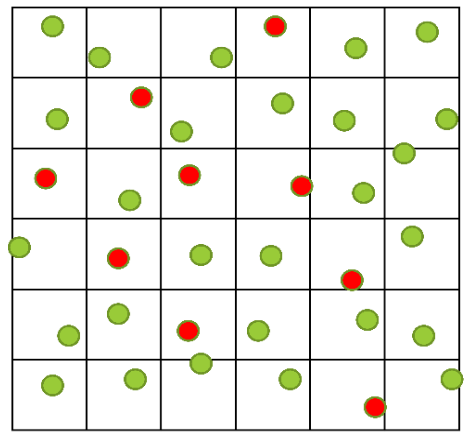


Some two-phase sampling schemes ensuring DAU&C of the two-phase IDW interpolators

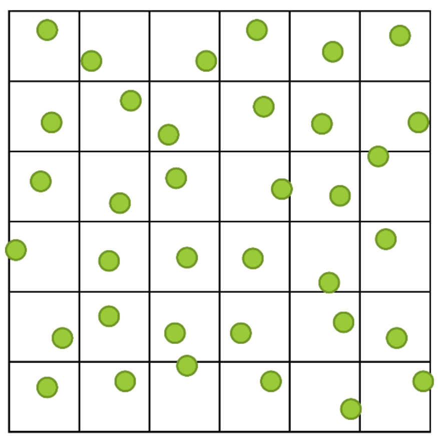
First phase:
Tesselation stratified sampling



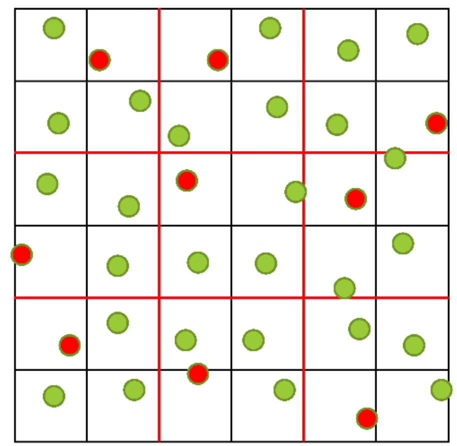
Second phase:
Simple random sampling without replacement



First phase:
Tesselation stratified sampling



Second phase:
One-per-stratum-stratified sampling



Mean square estimator

The pseudo-population bootstrap method is used

Mean square estimator

The pseudo-population bootstrap method is used

- the pseudo-population is the map obtained by means of the two-phase IDW interpolator

Mean square estimator

The pseudo-population bootstrap method is used

- the pseudo-population is the map obtained by means of the two-phase IDW interpolator
- B bootstrap samples S_b (b=1,...,B) are selected from the map following the original sampling scheme

Mean square estimator

The pseudo-population bootstrap method is used

- the pseudo-population is the map obtained by means of the two-phase IDW interpolator
- B bootstrap samples S_b (b=1,...,B) are selected from the map following the original sampling scheme
- $\hat{y}_b^*(p, S_b)$ is the two-phase interpolator with the b-th bootstrap sample

Mean square estimator

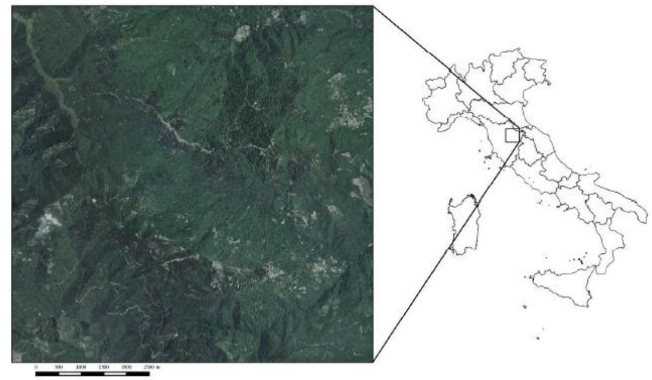
The pseudo-population bootstrap method is used

- the pseudo-population is the map obtained by means of the two-phase IDW interpolator
- B bootstrap samples S_b (b=1,...,B) are selected from the map following the original sampling scheme
- $\hat{y}_b^*(p, S_b)$ is the two-phase interpolator with the b-th bootstrap sample
- the mean square error of $\hat{y}(p, S)$ can be estimated by

$$\frac{1}{B}\sum_{b=1}^{B}(\hat{y}_b^*(p, S_b) - \hat{y}(p, S))^2$$

Case study

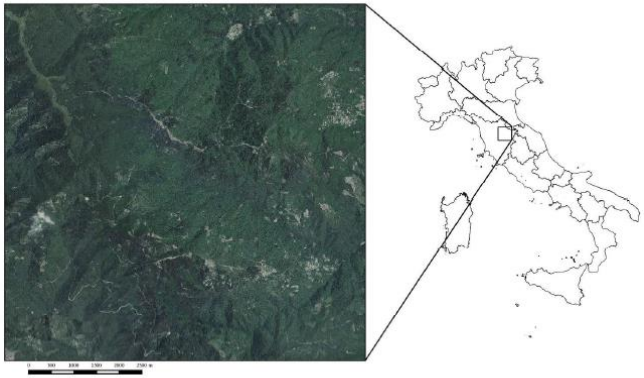
Case study



Study area

- size 4,900 ha
- located in the Casentino Valley, in Central Italy
- prevailing tree species are beech, European black pine, Turkey oak, Douglas fir, and silver fir.

Case study



Study area

- size 4,900 ha

- located in the Casentino Valley, in Central Italy

- prevailing tree species are beech, European black pine, Turkey oak, Douglas fir, and silver fir.

Attribute of interest

$y(p)$ total volume of all trees - taller than 1.30 *m* and with stem diameter at breast height greater than 5 *cm* - lying within the circular plot of radius 20 *m* centred at p

Case study

SAMPLING: two-phase sampling

Case study

SAMPLING: two-phase sampling

- ✓ **First phase**
  the area was covered by a grid of 1,225 quadrats of side 200 *m* and a point was randomly selected within each quadrat

Case study

SAMPLING: two-phase sampling

- ✓ **First phase**

  the area was covered by a grid of 1,225 quadrats of side 200 *m* and a point was randomly selected within each quadrat

- ✓ **Second phase**
  - on the basis of aerial imagery, first phase points were partitioned into two strata: 1,151 forest points and 74 points located outside forests

Case study

SAMPLING: two-phase sampling

✓ **First phase**
the area was covered by a grid of 1,225 quadrats of side 200 *m* and a point was randomly selected within each quadrat

✓ **Second phase**

- on the basis of aerial imagery, first phase points were partitioned into two strata: 1,151 forest points and 74 points located outside forests
- by means of simple random sampling without replacement a sample of 95 forest points was selected

Case study

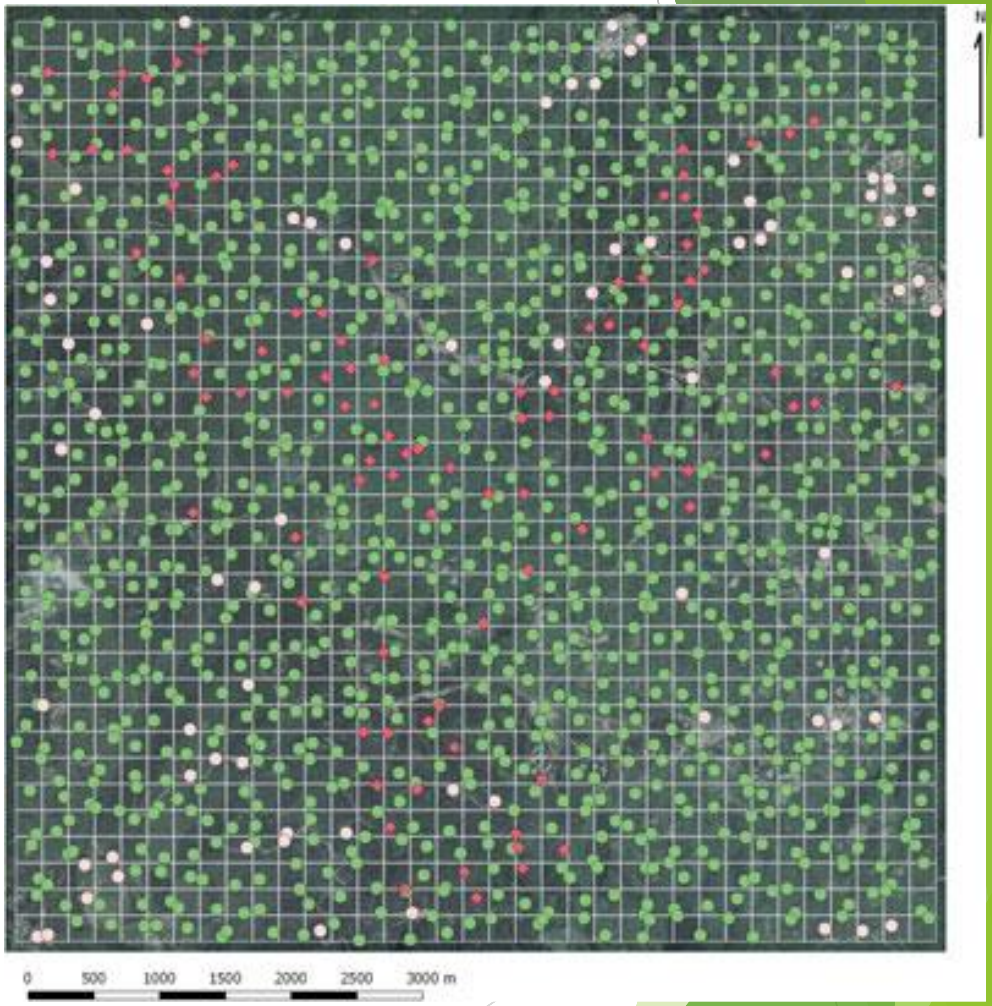
SAMPLING: two-phase sampling

- ✓ **First phase**
  the area was covered by a grid of 1,225 quadrats of side 200 *m* and a point was randomly selected within each quadrat

- ✓ **Second phase**
  - on the basis of aerial imagery, first phase points were partitioned into two strata: 1,151 forest points and 74 points located outside forests
  - by means of simple random sampling without replacement a sample of 95 forest points was selected
  - each second phase points was visited, recording the total volume of all the trees lying within the circular plot of radius 20 *m* centered at the point

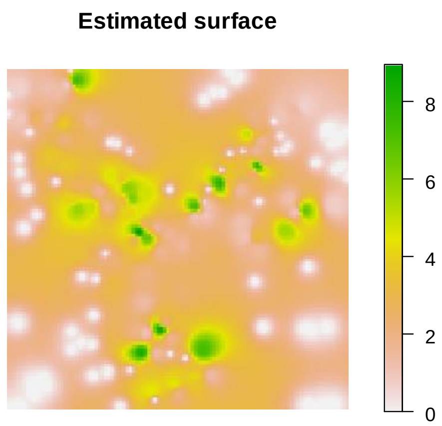
Case study

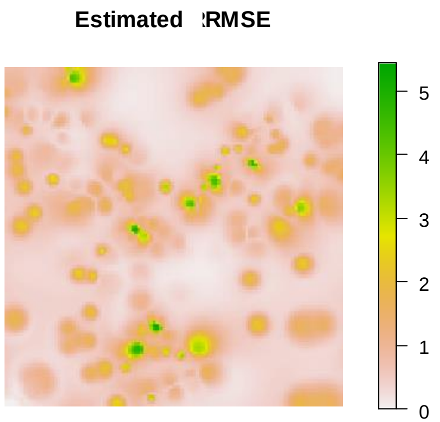
- 74 first-phase plots located outside forest
- 1056 first-phase plots located within forest and not visited on the ground
- 95 second-phase plots visited on the ground



Case study

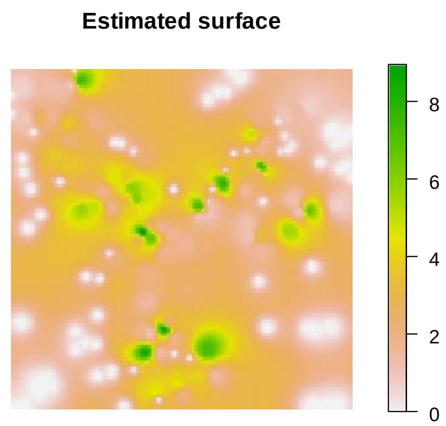
Two-phase ratio interpolator

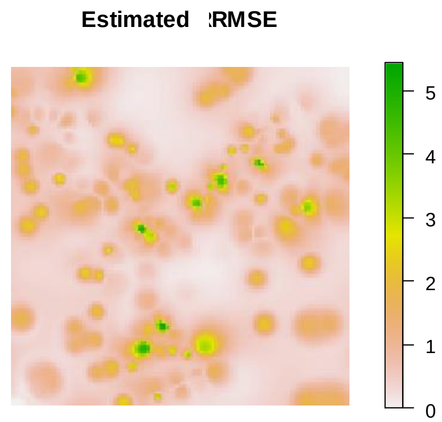




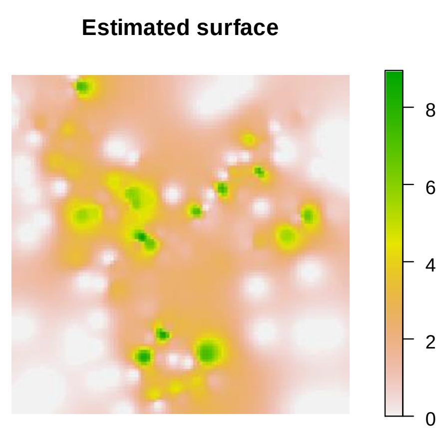
Case study

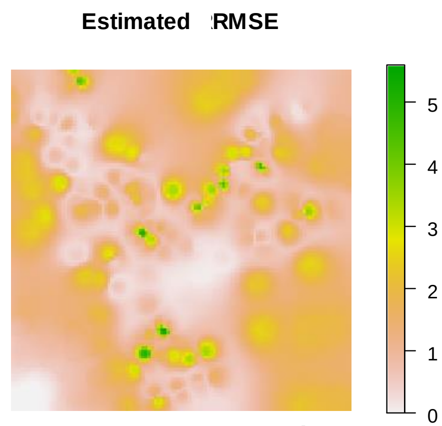
Two-phase ratio interpolator





Two-phase direct interpolator





Simulation study

Simulation study

Real surface

- Public forest estate in Val di Sella region
- Size 604.5 ha, trees 1213
- $y(p)$ basal area (m^2) within a plot of radius 13 m centred at p of trees taller than 1.30 m with stem diameter at breast height greater than *5*

Simulation study

Real surface

- Public forest estate in Val di Sella region
- Size 604.5 ha, trees 1213
- $y(p)$ basal area (m^2) within a plot of radius 13 *m* centred at p of trees taller than 1.30 *m* with stem diameter at breast height greater than *5*

Artificial surface

- Unit square
- $y\,(p) = C(3 sin p_1 sin^2 p_2)^2$, C such that the maximum value is 10

Simulation study

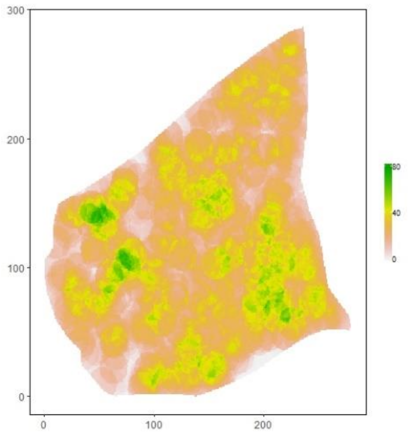
Real surface

- Public forest estate in Val di Sella region
- Size 604.5 ha, trees 1213
- $y(p)$ basal area (m^2) within a plot of radius 13 m centred at p of trees taller than 1.30 m with stem diameter at breast height greater than *5*

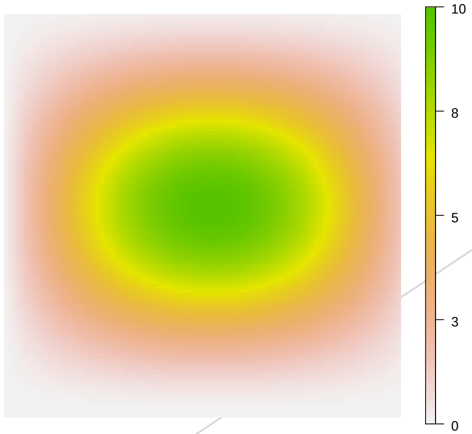
Artificial surface

- Unit square
- $y\,(p) = C(3 sin p_1 sin^2 p_2)^2$, C such that the maximum value is 10

Real surface



Artificial surface



Simulation study

Sampling

Simulation study

Sampling

First phase: tesselation stratified sampling $n = 100, 225, 400$

Simulation study

Sampling

First phase: tesselation stratified sampling $n=100, 225, 400$

- ✓ For the artificial surface the unit square was partitioned into quadrats of equal size and one point per quadrat was selected
- ✓ For the real surface, a regular tessellation was not possible and the study area was tessellated into n irregular patches of equal size

Simulation study

Sampling

First phase: tesselation stratified sampling $n=100, 225, 400$

- ✓ For the artificial surface the unit square was partitioned into quadrats of equal size and one point per quadrat was selected
- ✓ For the real surface, a regular tessellation was not possible and the study area was tessellated into n irregular patches of equal size

Second phase: simple random sampling without replacement
sampling fraction 20%

Simulation study

Sampling

First phase: tesselation stratified sampling n=100, 225, 400

- ✓ For the artificial surface the unit square was partitioned into quadrats of equal size and one point per quadrat was selected
- ✓ For the real surface, a regular tessellation was not possible and the study area was tessellated into n irregular patches of equal size

Second phase: simple random sampling without replacement
sampling fraction 20%

Note that $\hat{y}_R(p,S) = \hat{y}_D(p,S)$ as $\pi_i = \pi_\text{o}$ for any $i \in S$

Simulation study

Sampling

First phase: tesselation stratified sampling n=100, 225, 400

✓ For the artificial surface the unit square was partitioned into quadrats of equal size and one point per quadrat was selected

✓ For the real surface, a regular tessellation was not possible and the study area was tessellated into n irregular patches of equal size

Second phase: simple random sampling without replacement
sampling fraction 20%

Note that $\hat{y}_R(p,S) = \hat{y}_D(p,S)$ as $\pi_i = \pi_o$ for any $i \in S$

For each first and second phase sampling effort, sampling was independently replicated 1000 times to empirically determine:

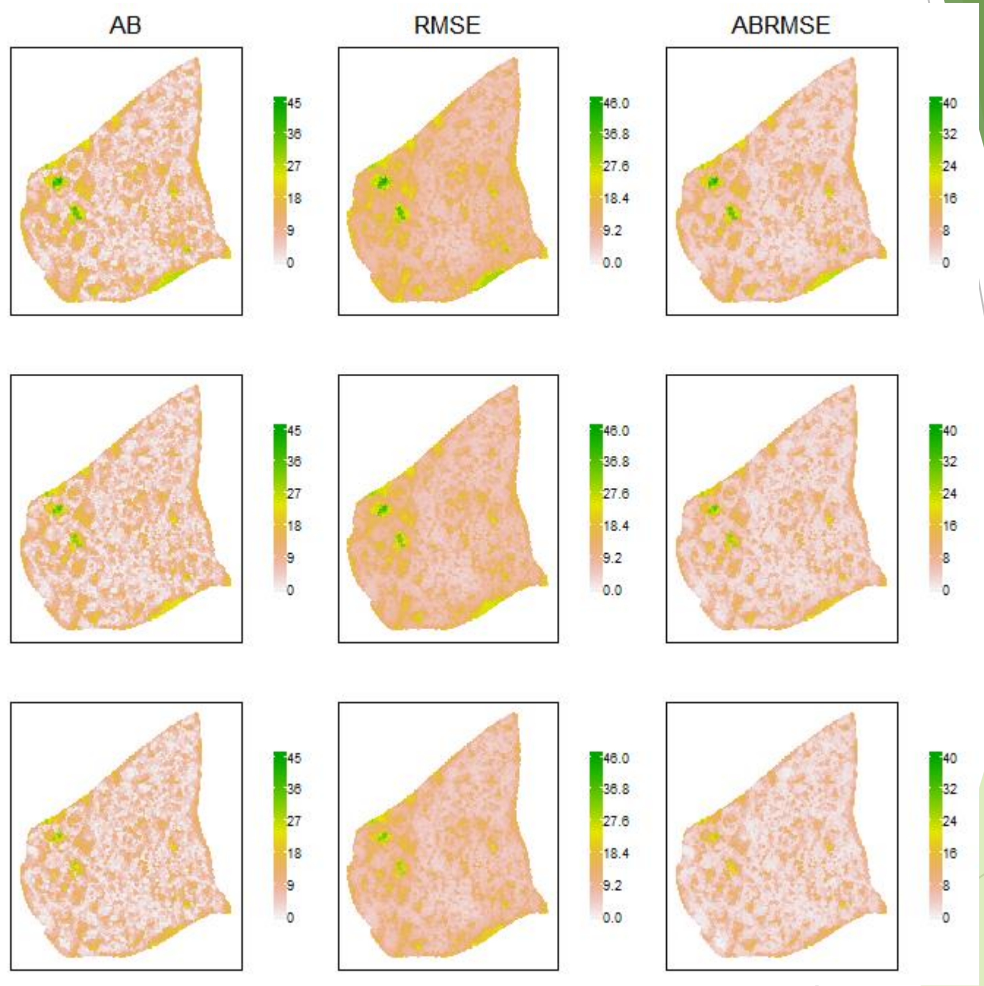
- the absolute bias (AB)
- the root mean square error (RMSE)
- the absolute bias of root mean square error (ABRMSE)

Simulation results – Real surface

first phase points: 100
second phase points: 20

first phase points: 225
second phase points: 45

first phase points: 400
second phase points: 80



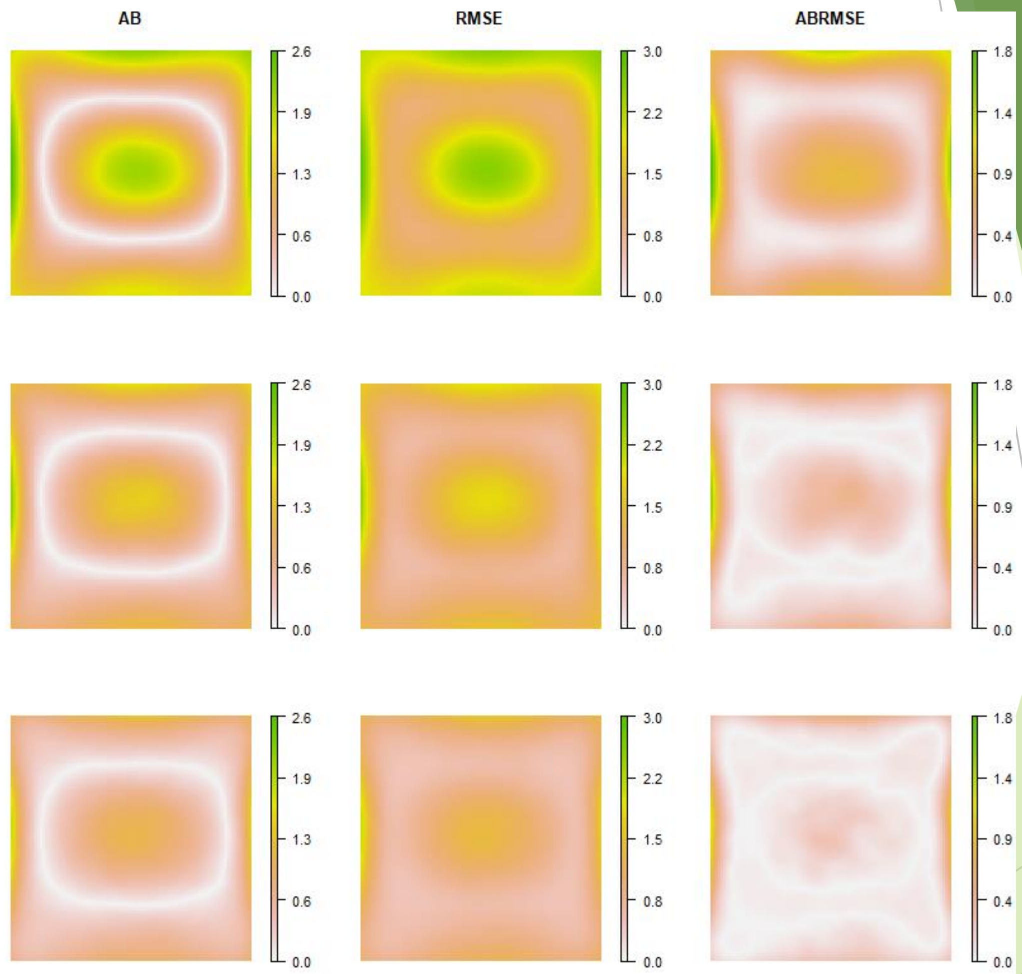
AB: absolute bias, RMSE: root mean square error
ABRMSE: absolute bias of root mean square error

Simulation results – Artificial surface

first phase points: 100
second phase points: 20

first phase points: 225
second phase points: 45

first phase points: 400
second phase points: 80



AB: absolute bias, RMSE: root mean square error
ABRMSE: absolute bias of root mean square error

Work in progress on simulation study…

Sampling

First phase: tesselation stratified sampling n=100, 225, 400

✓ For the artificial surface the unit square was partitioned into quadrats of equal size and one point per quadrat was selected

✓ For the real surface, a regular tessellation was not possible and the study area was tessellated into n irregular patches of equal size

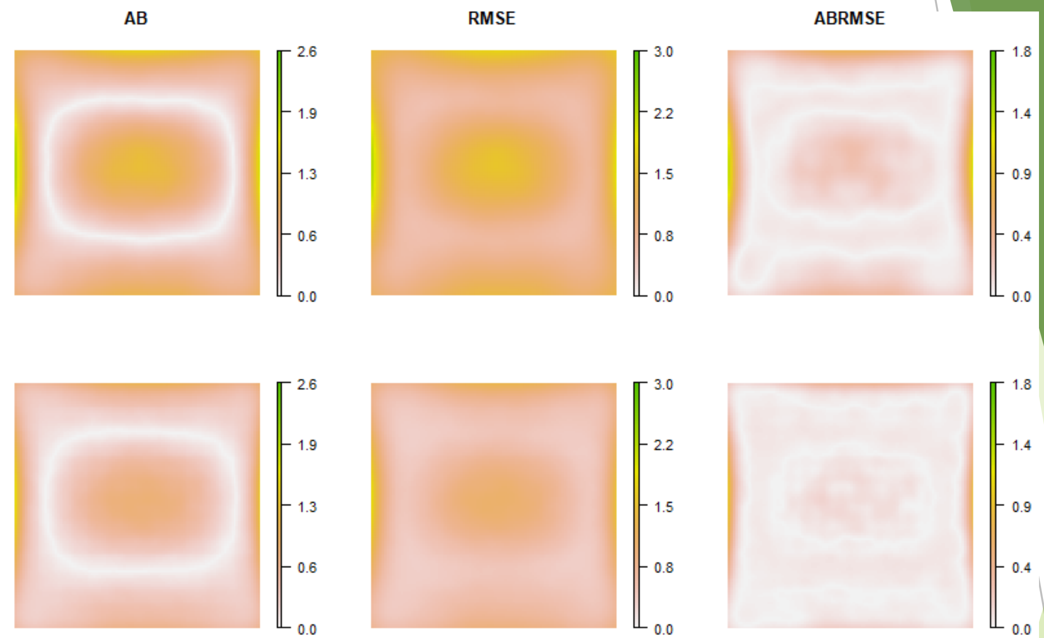
Second phase:

- simple random sampling without replacement with a sampling fraction of 50%

- One Per Stratum Sampling with sampling fraction 20% and 50%

Simulation are still running!

Simulation results – Artificial surface



first phase points: 100
second phase points: 50

first phase points: 225
second phase points: 112

first phase points: 400
second phase points: 200

Simulation are still running!

AB: absolute bias, RMSE: root mean square error
ABRMSE: absolute bias of root mean square error

Joined work with
L. Fattorini, M. Marcheselli, C. Pisani, L. Pratelli

Thanks for attention

Reference paper

L. Fattorini, M. Marcheselli, C. Pisani, L. Pratelli (2017). *Design-based asymptotics for two-phase sampling strategies in environmental surveys*, Biometrika, 104, 195-202.