Comparison of unit- and area-level small area estimators

Johannes Breidenbach and Rasmus Astrup

Norwegian Forest and Landscape Institute,
Climate Center and National Forest Inventory
P.O. Box 115, NO-1431 Ås
Tel: +47 6494 8981; JOB@SkogogLandskap.no

skog+
landskap

Contents

ALS-assited forest management inventories

- Estimates on stand level → operational forest management decisions
- Commercial providers in the Nordic countries
- Combination of field plots, ALS, (aerial images)
- Exact plot coordinates: linking field data to ALS
- Uncertainty in plot coordinates → uncertainty in estimates
- Aim: Alternative methods – no exact plot coordinates?

ALS-assited forest management inventories

- Estimates on stand level $\rightarrow$ operational forest management decisions
- Commercial providers in the Nordic countries
- Combination of field plots, ALS, (aerial images)
- Exact plot coordinates: linking field data to ALS
- Uncertainty in plot coordinates $\rightarrow$ uncertainty in estimates
- Aim: Alternative methods – no exact plot coordinates?

skog+
landskap

ALS-assited forest management inventories

- Estimates on stand level → operational forest management decisions
- Commercial providers in the Nordic countries
- Combination of field plots, ALS, (aerial images)
- Exact plot coordinates: linking field data to ALS
- Uncertainty in plot coordinates → uncertainty in estimates
- Aim: Alternative methods – no exact plot coordinates?

ALS-assited forest management inventories

- Estimates on stand level → operational forest management decisions
- Commercial providers in the Nordic countries
- Combination of field plots, ALS, (aerial images)
- Exact plot coordinates: linking field data to ALS
- Uncertainty in plot coordinates → uncertainty in estimates
- Aim: Alternative methods – no exact plot coordinates?

ALS-assited forest management inventories

- Estimates on stand level → operational forest management decisions
- Commercial providers in the Nordic countries
- Combination of field plots, ALS, (aerial images)
- Exact plot coordinates: linking field data to ALS
- Uncertainty in plot coordinates → uncertainty in estimates
- Aim: Alternative methods – no exact plot coordinates?

skog+ landskap

ALS-assited forest management inventories

- Estimates on stand level $\rightarrow$ operational forest management decisions
- Commercial providers in the Nordic countries
- Combination of field plots, ALS, (aerial images)
- Exact plot coordinates: linking field data to ALS
- Uncertainty in plot coordinates $\rightarrow$ uncertainty in estimates
- Aim: Alternative methods – no exact plot coordinates?

Study area

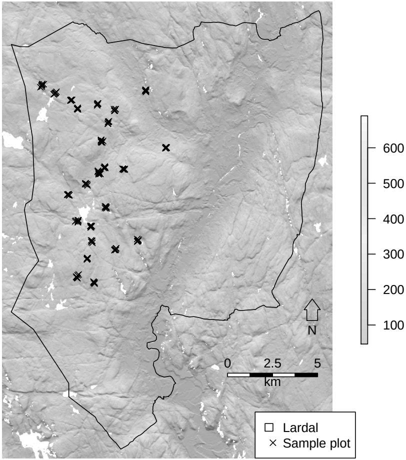
- Lardal municipality, southern Norway
- 30 stands, 5-7 plots of 250 m^2 in each
- Timber volume on plot level
- High density ALS
- CloudMetrics() and GridMetrics() in FUSION
- Hat-notation omitted for convenience

Curtesy Mr. Johannes Rahlf

skog+ landskap

Study area

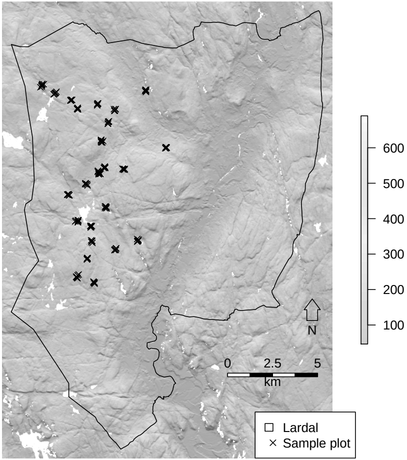
- Lardal municipality, southern Norway
- 30 stands, 5-7 plots of 250 m^2 in each
- Timber volume on plot level
- High density ALS
- CloudMetrics() and GridMetrics() in FUSION
- Hat-notation omitted for convenience



Curtesy Mr. Johannes Rahlf

Study area

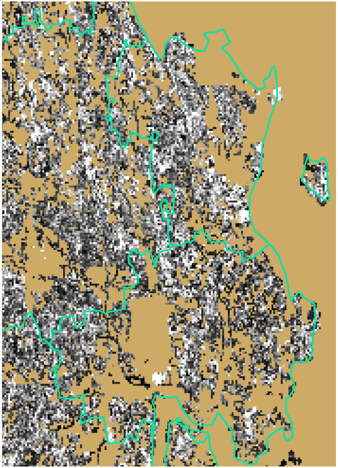
- Lardal municipality, southern Norway
- 30 stands, 5-7 plots of 250 m^2 in each
- Timber volume on plot level
- High density ALS
- CloudMetrics() and GridMetrics() in FUSION
- Hat-notation omitted for convenience



Curtesy Mr. Johannes Rahlf

Study area

- Lardal municipality, southern Norway
- 30 stands, 5-7 plots of 250 m^2 in each
- Timber volume on plot level
- High density ALS
- CloudMetrics() and GridMetrics() in FUSION
- Hat-notation omitted for convenience

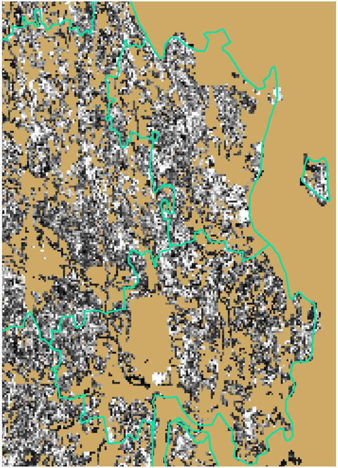


Curtesy Mr. Johannes Rahlf

skog+ landskap

Study area

- Lardal municipality, southern Norway
- 30 stands, 5-7 plots of 250 m^2 in each
- Timber volume on plot level
- High density ALS
- CloudMetrics() and GridMetrics() in FUSION
- Hat-notation omitted for convenience

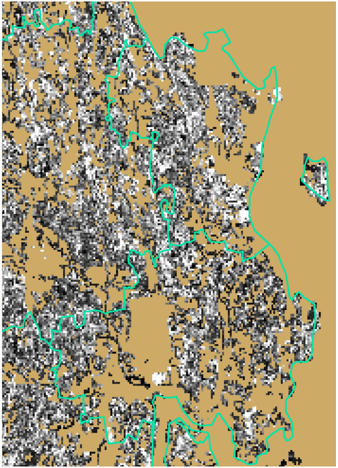


Curtesy Mr. Johannes Rahlf

skog+ landskap

Study area

- Lardal municipality, southern Norway
- 30 stands, 5-7 plots of 250 m^2 in each
- Timber volume on plot level
- High density ALS
- CloudMetrics() and GridMetrics() in FUSION
- Hat-notation omitted for convenience

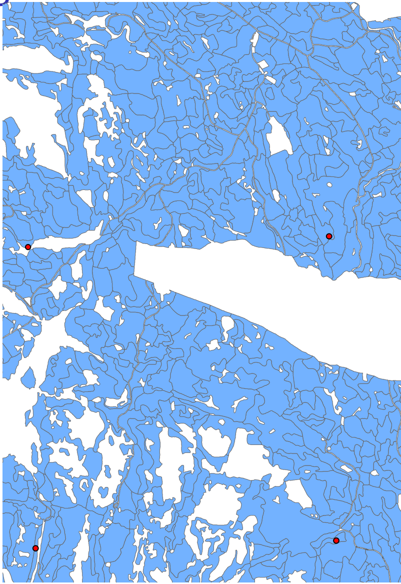


Curtesy Mr. Johannes Rahlf

skog+ landskap

Contents

The (unit-level) synthetic regression estimator



- Current approach: Næsset method
- Systematical sampling (GPS needed)
- Linear model
- Mean of the model predictions for the small area *i*

$$\Theta_i^S = \frac{1}{N_i} \sum \mathbf{X}_i^T \beta = \bar{\mathbf{X}}_i^T \beta \qquad (1)$$

$$MSE_i = \bar{\mathbf{X}}_i^T Cov(\beta) \bar{\mathbf{X}}_i$$

e.g., Prasad and Rao (1990), Madallaz (2013)

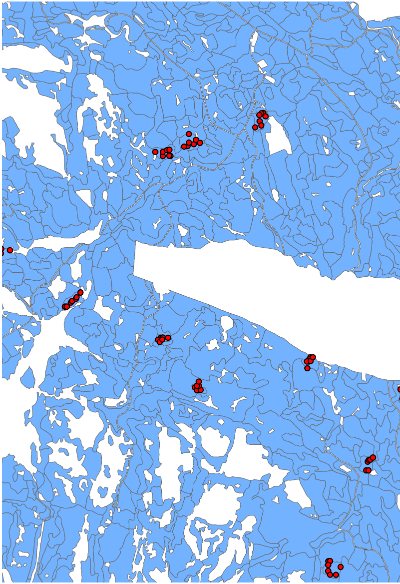
The unit-level EBLUP estimator

- Sample plots clustered in stands (GPS needed)
- Mixed-effects model
- Sum of synthetic estimator and realization of a random effect on the small area level *i*

$$\Theta_i^B = \bar{\mathbf{X}}_i^T \beta + \upsilon_i. \qquad (2)$$

$$MSE_i = g_1 + g_2 + 2g_3$$

$g_1 = f(\sigma_\upsilon)$
$g_{2,i} = f(\bar{\mathbf{X}}_i^T Cov(\beta)\bar{\mathbf{X}}_i)$
$g_3 = f(Var(\sigma_\upsilon))$



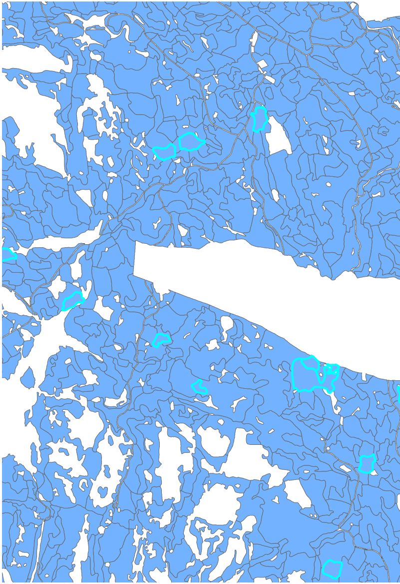
The area-level EBLUP estimator

- Direct stand-level (area-level) estimate, e.g., sample plots clustered in stands (no GPS needed)
- Decomposition: A mixed-effects model is fit to domain level estimates

$$\Theta_i^D = \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + \upsilon_i + e_i$$

No repeated observations on stand level:

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}(\sigma_\upsilon) = f(\Theta_i^D, Var(\Theta_i^D), \bar{\mathbf{X}}_i^T)$$



The area-level EBLUP estimator

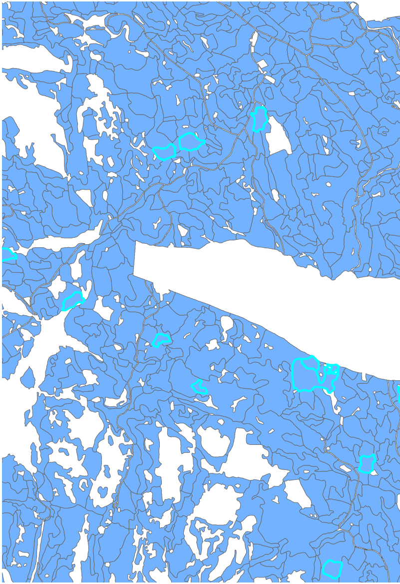
- Area-level estimates using the Fay-Harriot model
- Composite estimator

$$\Theta_i^{FH} = \gamma_i \Theta_i^D + (1 - \gamma_i) \bar{\mathbf{X}}_i^T \tilde{\beta} \qquad (3)$$

with weight

$$\gamma_i = \sigma_v^2 / (Var(\Theta_i^D) + \sigma_v^2).$$

$$MSE_i = g_1 + g_{2,i} + 2g_3$$

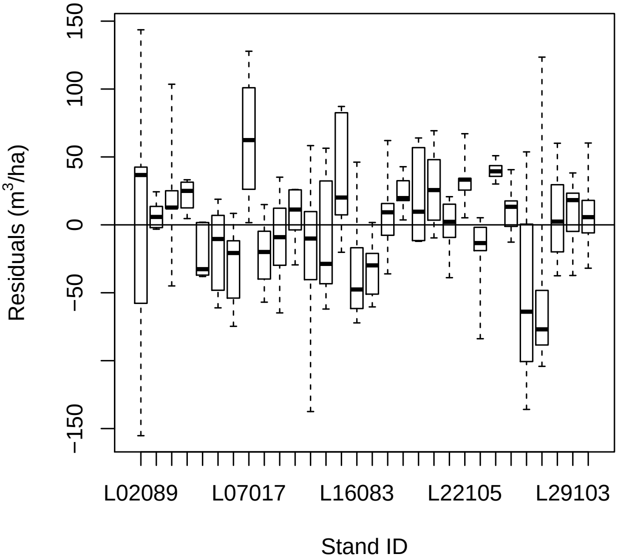


Contents

skog+ landskap

Significance of random effect

x = mean.als.ht; $R^2 = 0.88$



Significance of random effect

- Random effect barely significant in unit-level model (ΔBIC=2)
- No significant random effect in area-level model
- Area-level EBLUP is applicable without direct estimate: γ-weight = 0

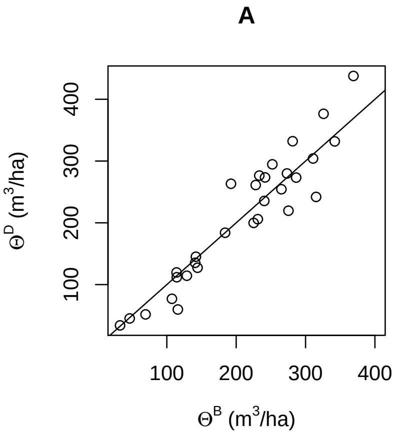
$$\Theta_i^B = \bar{\mathbf{X}}_i^T \boldsymbol{\beta} \quad (4)$$

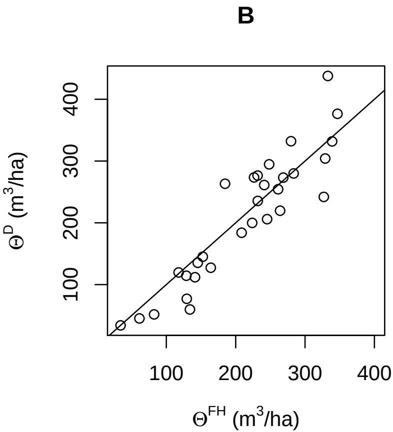
$$\Theta_i^{FH} = \bar{\mathbf{X}}_i^T \tilde{\boldsymbol{\beta}} \quad (5)$$

$$MSE_i = g_{2,i}$$

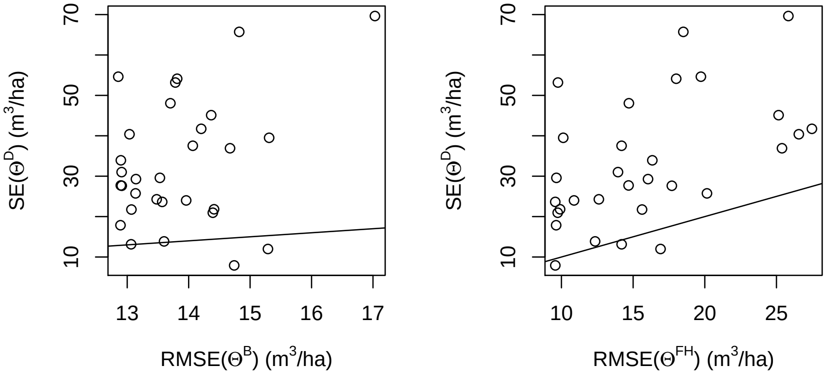
skog+ landskap

Comparison of estimates





Comparison of uncertainties



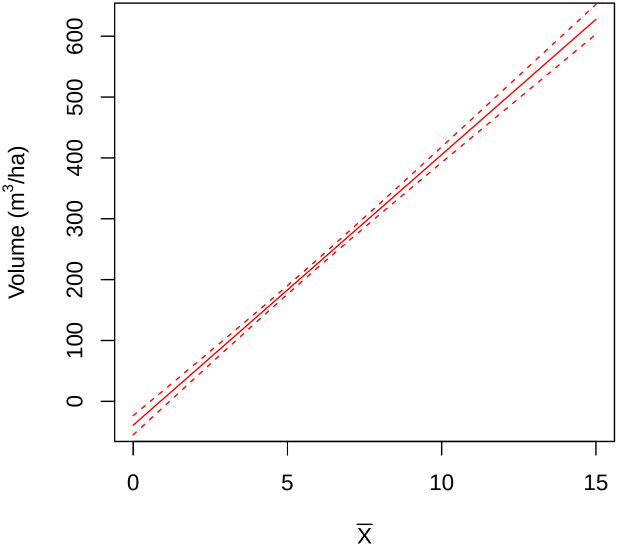
skog+ landskap

Comparison of unit- and area-level MSE

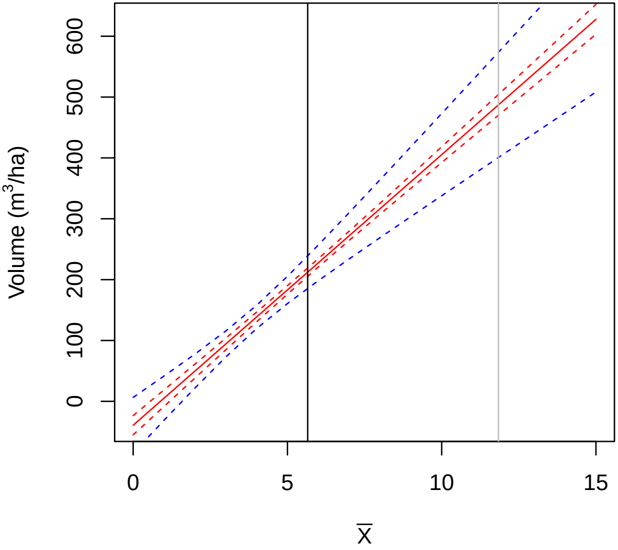
$$
\begin{aligned}
CI(\Theta_{x_0}) &= \Theta_{x_0} \pm \Phi_q \sqrt{MSE_{MIX-SYN}(x_0)} \\
&= \Theta_{x_0} \pm \Phi_q \sqrt{x_0^T Cov(\beta) x_0}
\end{aligned}
$$

skog+ landskap

Comparison of unit- and area-level MSE



Comparison of unit- and area-level MSE

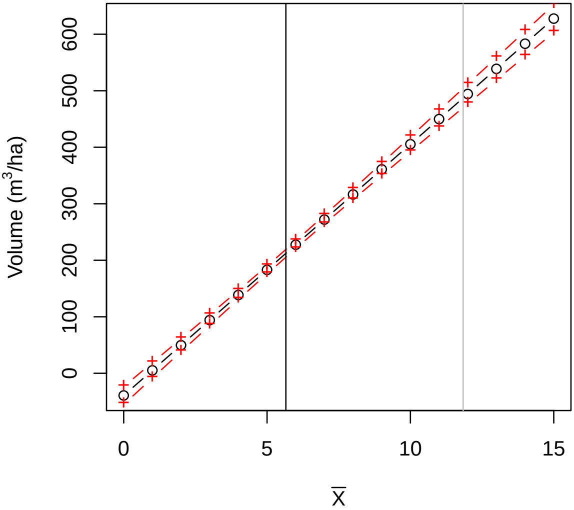


Comparison of unit-level Mixed-Synth MSE and CI of LM

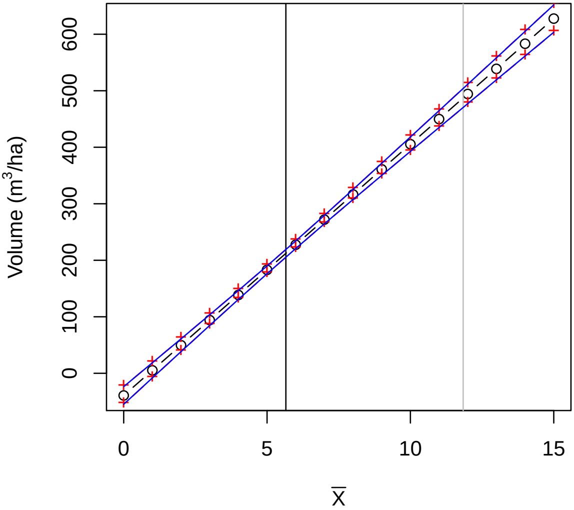
$$CI(\Theta^B_{x_0}) = \Theta^B_{x_0} \pm \Phi_q \sqrt{x_0^T Cov(\beta) x_0}$$

$$CI(E(y_{x_0})) = y_{x_0} \pm t_q \cdot \sigma_{LM} \cdot \sqrt{x_0^T (\mathbf{X}^T \mathbf{X})^{-1} x_0}$$

Comparison of unit-level Mixed-Synth MSE and CI of LM



Comparison of unit-level Mixed-Synth MSE and CI of LM



Comparison of unit-level Mixed-Synth MSE and CI of LM

$$CI(\Theta^B_{x_0}) = \Theta^B_{x_0} \pm \Phi_q \sqrt{x_0^T Cov(\beta) x_0}$$
$$= \Theta^B_{x_0} \pm \Phi_q \sqrt{x_0^T \sigma^2_\varepsilon (\mathbf{X}^T\mathbf{X})^{-1} x_0}$$

$$CI(E(y_{x_0})) = y \pm t_q \cdot \sigma_{LM} \cdot \sqrt{x_0^T (\mathbf{X}^T\mathbf{X})^{-1} x_0}$$
$$= y \pm t_q \sqrt{x_0^T \sigma^2_{LM} (\mathbf{X}^T\mathbf{X})^{-1} x_0}$$

$$t_q(df = 180) \approx \Phi_q = 1.96$$

$$CI(\Theta^B_{x_0}) \approx CI(E(y_{x_0}))$$

skog+ landskap

Contents

Summary

- Usage of ALS improved precision of estimates
- No strong/significant (random) effects on stand level (area-level)

⇒ Area-level EBLUP applicable to all stands regardless of direct estimate

- Unit-level synthetic estimates: $CI(\Theta^{B}_{x_0}) \approx CI(E(y_{x_0}))$

skog+ landskap

Summary

- Usage of ALS improved precision of estimates
- No strong/significant (random) effects on stand level (area-level)

⇒ Area-level EBLUP applicable to all stands regardless of direct estimate

- Unit-level synthetic estimates: $CI(\Theta^B_{x_0}) \approx CI(E(y_{x_0}))$

skog+ landskap

Summary

- Usage of ALS improved precision of estimates
- No strong/significant (random) effects on stand level (area-level)

⇒ Area-level EBLUP applicable to all stands regardless of direct estimate

- Unit-level synthetic estimates: $CI(\Theta^{B}_{x_0}) \approx CI(E(y_{x_0}))$

skog+ landskap

Summary

- Usage of ALS improved precision of estimates
- No strong/significant (random) effects on stand level (area-level)

⇒ Area-level EBLUP applicable to all stands regardless of direct estimate

- Unit-level synthetic estimates: $CI(\Theta^{B}_{x_0}) \approx CI(E(y_{x_0}))$

Conclusions

- Advantages of area-level EBLUP

⇒ Exact GPS coordinates not necessary

⇒ Easier application in remote areas or with cheap equipment

⇒ Sampling design does not need to 'fit' to ALS (e.g., line intersect sampling, Bitterlich)

- Disadvantages of area-level EBLUP

⇒ Larger MSE than unit-level EBLUP

Conclusions

- Advantages of area-level EBLUP

⇒ Exact GPS coordinates not necessary

⇒ Easier application in remote areas or with cheap equipment

⇒ Sampling design does not need to ’fit’ to ALS (e.g., line intersect sampling, Bitterlich)

- Disadvantages of area-level EBLUP

⇒ Larger MSE than unit-level EBLUP

Conclusions

- Advantages of area-level EBLUP

⇒ Exact GPS coordinates not necessary

⇒ Easier application in remote areas or with cheap equipment

⇒ Sampling design does not need to 'fit' to ALS (e.g., line intersect sampling, Bitterlich)

- Disadvantages of area-level EBLUP

⇒ Larger MSE than unit-level EBLUP

Conclusions

- Advantages of area-level EBLUP

⇒ Exact GPS coordinates not necessary

⇒ Easier application in remote areas or with cheap equipment

⇒ Sampling design does not need to 'fit' to ALS (e.g., line intersect sampling, Bitterlich)

- Disadvantages of area-level EBLUP

⇒ Larger MSE than unit-level EBLUP

skog+ landskap

Conclusions

- Advantages of area-level EBLUP

⇒ Exact GPS coordinates not necessary

⇒ Easier application in remote areas or with cheap equipment

⇒ Sampling design does not need to 'fit' to ALS (e.g., line intersect sampling, Bitterlich)

- Disadvantages of area-level EBLUP

⇒ Larger MSE than unit-level EBLUP

skog+ landskap

Conclusions

- Advantages of area-level EBLUP

⇒ Exact GPS coordinates not necessary

⇒ Easier application in remote areas or with cheap equipment

⇒ Sampling design does not need to 'fit' to ALS (e.g., line intersect sampling, Bitterlich)

- Disadvantages of area-level EBLUP

⇒ Larger MSE than unit-level EBLUP

References

J.N.K. Rao
Small Area Estimation
Wiley Series in Survey Methodology, 2003.

N. Prasad & J.N.K. Rao
The estimation of the mean squared error of small-area estimators
JASA (1990) 85:163:171.

D. Mandallaz
Design-based properties of some small-area estimators [...]
Can. J. For. Res. (2013) 43:441-449.

J. Breidenbach and R. Astrup
Small area estimation of forest attributes [...]
Eur. J. For. Res. (2012) 131:1255-1267.

J. Breidenbach
The R-package JoSAE: Functions for unit-level small area estimators
CRAN (2011), http://cran.r-project.org/web/packages/JoSAE/index.html.

skog+ landskap

Thank you!

EBLUP and GREG

$$\bar{Y}_{E,i} = \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + \hat{\gamma}_i \left(\frac{1}{n_i} \sum_j \varepsilon_{ij} \right) \quad (6)$$

$$\hat{\gamma}_i = \frac{\hat{\sigma}_\upsilon^2}{\hat{\sigma}_\upsilon^2 + \hat{\sigma}_\varepsilon^2 / n_i} \quad (7)$$

skog+
landskap

The Mixed-effects model

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \upsilon_i + \varepsilon_{ij}, \quad j = 1, ..., N_i, \quad i = 1, ..., m \tag{8}$$

$$\upsilon \sim \mathrm{N}(0, \sigma_\upsilon^2), \qquad \varepsilon \sim \mathrm{N}(0, \sigma_\varepsilon^2) \tag{9}$$

skog+
landskap

EBLUP MSE

$$\text{MSE}_{\bar{Y}_{B,i}} \approx C_{1,i} + C_{2,i} + 2C_{3,i} \tag{10}$$

$$C_{1,i} = \gamma_i(\hat{\sigma}_\varepsilon^2/n_i).$$

$$C_{2,i} = (\bar{\mathbf{X}}_i - \gamma_i\bar{\mathbf{x}}_i)^T \left(\sum_i \mathbf{X}^T\mathbf{U}\mathbf{X}\right)^{-1} (\bar{\mathbf{X}}_i - \gamma_i\bar{\mathbf{x}}_i)$$

$$C_{3,i} = n_i^{-2}(\hat{\sigma}_\upsilon^2 + \hat{\sigma}_\varepsilon^2/n_i)^{-3} C_{31}$$

$$C_{31} = \hat{\sigma}_\upsilon^4 \bar{V}_{\upsilon\upsilon} + \hat{\sigma}_\varepsilon^4 \bar{V}_{\varepsilon\varepsilon} + 2\hat{\sigma}_\varepsilon^2\hat{\sigma}_\upsilon^2 \bar{V}_{\varepsilon\upsilon}.$$

skog+
landskap