

Additional file 1: Appendix

1 A new approach to derive the generalized nonlinear hierarchical model-based estimation method

In the HMB estimation method we define two models:

$$\text{Model } F: \mathbf{y} = \mathbf{f}(\mathbf{x}; \boldsymbol{\beta}) + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega}) \quad (1)$$

and

$$\text{Model } G: \mathbf{y} = \mathbf{g}(\mathbf{z}; \boldsymbol{\alpha}) + \mathbf{v}, \mathbf{v} \sim N(\mathbf{0}, \boldsymbol{\Sigma}). \quad (2)$$

In our study, the dataset S was used to estimate model parameters $\boldsymbol{\beta}$ by means of field measurements, and dataset Sa to estimate $\boldsymbol{\alpha}$ using LiDAR data.

There are several methods to estimate model parameters in nonlinear models. In our example we employed the REML estimation method available in R package “nlme”. For deriving the approximated covariances of the estimated model parameters we employed Gauss-Newton Regression (GNR) associated with the generalized nonlinear least squares estimation (GNLS) method. However, other methods can be employed (e.g., Davidson and MacKinnon (1993)).

The GNR method is based on the linearization by Taylor-series to obtain approximate or asymptotic results. An outline is as follows. We start with the model (1) and define the residual column vector $\boldsymbol{\epsilon}_S$ as

$$\boldsymbol{\epsilon}_S = \mathbf{y}_S - \mathbf{f}_S(\mathbf{x}, \boldsymbol{\beta}). \quad (3)$$

The estimator $\hat{\boldsymbol{\beta}}_S$ is obtained by minimizing the weighted sum of squared residuals,

$$\Gamma = \boldsymbol{\epsilon}_S^T \boldsymbol{\Omega}_S^{-1} \boldsymbol{\epsilon}_S. \quad (4)$$

To obtain $\hat{\boldsymbol{\beta}}_S$, the partial derivatives of Γ with respect to $\boldsymbol{\beta}$ should be set to 0. First we have the

partial derivatives $\frac{\partial \epsilon_S}{\partial \beta} = 0 - \frac{\partial \mathbf{f}_S(\mathbf{x}, \beta)}{\partial \beta} = -\tilde{\mathbf{X}}_S$, i. e. the matrix of partial derivatives over data set S . This gives us a column vector of size $(p+1)$ of partial derivatives of Γ , $\frac{\partial \Gamma}{\partial \beta} = -2\tilde{\mathbf{X}}_S^T \Omega_S^{-1} \epsilon_S$ and the $(p+1)$ equations

$$\tilde{\mathbf{X}}_S^T \Omega_S^{-1} \epsilon_S = \mathbf{0}. \quad (5)$$

Now, there is no closed form for a solution to Eq. (5) and we have to rely on an approximation based on the linearization of the function f and iterations, starting with an initial value $\beta^{(1)}$. The following $\beta^{(k)}$ are derived by using the approximation

$$\mathbf{f}_S(\mathbf{x}, \beta) \approx \mathbf{f}_S(\mathbf{x}, \beta^{(k)}) + \tilde{\mathbf{X}}_S(\beta - \beta^{(k)}). \quad (6)$$

Let us first define

$$\Delta \mathbf{y}_S^{(k)} = \mathbf{y}_S - \mathbf{f}_S(\mathbf{x}, \beta^{(k)}). \quad (7)$$

Then we have, by (3) and (7), with $\Delta \beta^{(k)} = \beta - \beta^{(k)}$, and approximately

$$\begin{aligned} \epsilon_S &= \mathbf{y}_S - \mathbf{f}_S(\mathbf{x}, \beta^{(k)}) + \mathbf{f}_S(\mathbf{x}, \beta^{(k)}) - \mathbf{f}_S(\mathbf{x}, \beta) \\ &= \Delta \mathbf{y}_S^{(k)} - \tilde{\mathbf{X}}_S \Delta \beta^{(k)}. \end{aligned} \quad (8)$$

By inserting $(\Delta \mathbf{y}_S^{(k)} - \tilde{\mathbf{X}}_S \Delta \beta^{(k)})$ into expression (5) we have

$$\tilde{\mathbf{X}}_S^T \Omega_S^{-1} (\Delta \mathbf{y}_S^{(k)} - \tilde{\mathbf{X}}_S \Delta \beta^{(k)}) = \mathbf{0}. \quad (9)$$

That is

$$\tilde{\mathbf{X}}_S^T \Omega_S^{-1} \tilde{\mathbf{X}}_S \Delta \beta^{(k)} = \tilde{\mathbf{X}}_S^T \Omega_S^{-1} \Delta \mathbf{y}_S^{(k)}. \quad (10)$$

From (10) we obtain

$$\Delta \beta^{(k)} = (\tilde{\mathbf{X}}_S^T \Omega_S^{-1} \tilde{\mathbf{X}}_S)^{-1} \tilde{\mathbf{X}}_S^T \Omega_S^{-1} \Delta \mathbf{y}_S^{(k)}, \quad (11)$$

and $\beta^{(k+1)} = \beta^{(k)} + \Delta \beta^{(k)}$. Expressions (6)–(11) are repeated until the vector $\Delta \beta^{(k)}$ is small enough.

For the final iteration we have $\hat{\beta}_S = \beta^{(k)}$ and $\Delta \mathbf{y}_S^{(k)} = \mathbf{y}_S - \mathbf{f}_S(\mathbf{x}, \hat{\beta}_S) \approx \mathbf{y}_S - \mathbf{f}_S(\mathbf{x}, \beta) = \epsilon_S$. Since the covariance matrix of $\Delta \beta^{(k)}$ and $\beta^{(k)}$ are identical (β is constant) and by inserting $\Delta \mathbf{y}_S^{(k)} = \epsilon_S$ into (11)

we obtain

$$\text{Cov}(\hat{\beta}_S) = (\tilde{\mathbf{X}}_S^\top \mathbf{\Omega}_S^{-1} \tilde{\mathbf{X}}_S)^{-1}. \quad (12)$$

It is both an approximated theoretical expression and an estimator, assuming that $\mathbf{\Omega}_S$ is known (e.g., Davidson and MacKinnon (1993)).

Now we derive the approximated covariance matrix estimator of $\hat{\alpha}_{Sa}$. First we present the estimator used to estimate α . The generalized nonlinear least squares estimator is approximated through (e.g., Davidson and MacKinnon (1993))

$$\hat{\alpha}_{Sa} \approx (\tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1} \tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \hat{\mathbf{y}}_{F_{Sa}}. \quad (13)$$

Since $\hat{\alpha}$ is a function of $\hat{\mathbf{y}}_{F_{Sa}}$, the vector of predicted AGB using the F model over the dataset Sa (the field estimated AGB); the covariance of $\hat{\alpha}$ can be written using the law of total covariance Rudary (2009)

$$\text{Cov}(\hat{\alpha}_{Sa}) = \text{E}[\text{Cov}(\hat{\alpha} \mid \hat{\mathbf{y}}_{F_{Sa}})] + \text{Cov}(\text{E}[\hat{\alpha} \mid \hat{\mathbf{y}}_{F_{Sa}}]). \quad (14)$$

The first term on the right-side of expression (14) is

$$\text{E}[\text{Cov}(\hat{\alpha} \mid \hat{\mathbf{y}}_{F_{Sa}})] = (\tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1}, \quad (15)$$

following generalized nonlinear least squares using the GNR method described above, conditionally on $\hat{\mathbf{y}}_{F_{Sa}}$.

The second term on the right-side in expression (14) can be developed using the most general expression of the error propagation theory Ku (1966) asⁱ

$$\text{Cov}(\text{E}[\hat{\alpha} \mid \hat{\mathbf{y}}_{Sa}]) = (\tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1} \tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \text{Cov}(\hat{\mathbf{y}}_{F_{Sa}}) \mathbf{\Sigma}_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa} (\tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1}, \quad (16)$$

using that

$$\begin{aligned} \text{E}[\hat{\alpha} \mid \hat{\mathbf{y}}_{F_{Sa}}] &= \text{E} \left[(\tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1} \tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \hat{\mathbf{y}}_{F_{Sa}} \mid \hat{\mathbf{y}}_{F_{Sa}} \right] \\ &= (\tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1} \tilde{\mathbf{Z}}_{Sa}^\top \mathbf{\Sigma}_{Sa}^{-1} \hat{\mathbf{y}}_{F_{Sa}}. \end{aligned} \quad (17)$$

ⁱHere we used the following rule: if $\mathbf{f} = \mathbf{A}\mathbf{x}$, then the covariance matrix of $\mathbf{f}$ is expressed as $\text{Cov}(\mathbf{f}) = \mathbf{A} \text{Cov}(\mathbf{x}) \mathbf{A}^\top$, where $\text{Cov}(\mathbf{x})$ is a covariance matrix of $\mathbf{x}$.

The covariance of $\hat{\alpha}$ is then

$$\begin{aligned} \text{Cov}(\hat{\alpha}_{Sa}) &= (\tilde{\mathbf{Z}}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1} \\ &+ (\tilde{\mathbf{Z}}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1} \tilde{\mathbf{Z}}_{Sa}^T \Sigma_{Sa}^{-1} \text{Cov}(\hat{\mathbf{y}}_{F_{Sa}}) \Sigma_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa} (\tilde{\mathbf{Z}}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{\mathbf{Z}}_{Sa})^{-1}. \end{aligned} \quad (18)$$

By replacing $\text{Cov}(\hat{\mathbf{y}}_{F_{Sa}})$ with the corresponding estimator, we obtain the approximated estimator $\widehat{\text{Cov}}(\hat{\alpha}_{Sa})$. The $\text{Cov}(\hat{\mathbf{y}}_{F_{Sa}})$ is estimated using model-based inference (e.g., Ståhl et al. (2016)) as

$$\widehat{\text{Cov}}(\hat{\mathbf{y}}_{F_{Sa}}) = \tilde{\mathbf{X}}_{Sa} \widehat{\text{Cov}}(\hat{\beta}_S) \tilde{\mathbf{X}}_{Sa}. \quad (19)$$

References

- Davidson, R. and MacKinnon, J. G. (1993). Estimation and inference in econometrics, Oxford University Press.
- Rudary, M. R. (2009). On Predictive Linear Gaussian Models, PhD thesis, Ann Arbor, MI, USA. AAI3354094.
- Ku, H. H. (1966). Notes on the use of propagation of error formulas, J Res Natl Bur Stand **70**(4):263–273.
- Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S. P., Patterson, P. L., Magnussen, S., Næsset, E., McRoberts, R. E. and Gregoire, T. G. (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation, Forest Ecosyst **3**(5): 1–11.