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

In the HMB estimation method we define two models:

Model $F$: \( y = f(x; \beta) + \epsilon, \epsilon \sim N(0, \Omega) \) \hspace{1cm} (1)

and

Model $G$: \( y = g(z; \alpha) + \nu, \nu \sim N(0, \Sigma) \). \hspace{1cm} (2)

In our study, the dataset $S$ was used to estimate model parameters $\beta$ by means of field measurements, and dataset $S_a$ to estimate $\alpha$ 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 a follows. We start with the model (1) and define the residual column vector $\epsilon_S$ as

\[ \epsilon_S = y_S - f_S(x, \beta). \] \hspace{1cm} (3)

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

\[ \Gamma = \epsilon_S^T \Omega_S^{-1} \epsilon_S. \] \hspace{1cm} (4)

To obtain $\hat{\beta}_S$, the partial derivatives of $\Gamma$ with respect to $\beta$ should be set to 0. First we have the
partial derivatives \( \frac{\partial \epsilon_S}{\partial \beta} = 0 - \frac{\partial f_S(x, \beta)}{\partial \beta} = -\tilde{X}_S \), 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 \( \Gamma \),

\[
\frac{\partial \Gamma}{\partial \beta} = -2 \tilde{X}_S^\top \Omega_S^{-1} \epsilon_S
\]

and the \((p + 1)\) equations

\[
\tilde{X}_S^\top \Omega_S^{-1} \epsilon_S = 0. \tag{5}
\]

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

\[
f_S(x, \beta) \approx f_S(x, \beta^{(k)}) + \tilde{X}_S(\beta - \beta^{(k)}). \tag{6}
\]

Let us first define

\[
\Delta y_S^{(k)} = y_S - f_S(x, \beta^{(k)}). \tag{7}
\]

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

\[
\epsilon_S = y_S - f_S(x, \beta) = \Delta y_S^{(k)} - \tilde{X}_S \Delta \beta^{(k)}. \tag{8}
\]

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

\[
\tilde{X}_S^\top \Omega_S^{-1} (\Delta y_S^{(k)} - \tilde{X}_S \Delta \beta^{(k)}) = 0. \tag{9}
\]

That is

\[
\tilde{X}_S^\top \Omega_S^{-1} \tilde{X}_S \Delta \beta^{(k)} = \tilde{X}_S^\top \Omega_S^{-1} \Delta y_S^{(k)}. \tag{10}
\]

From (10) we obtain

\[
\Delta \beta^{(k)} = (\tilde{X}_S^\top \Omega_S^{-1} \tilde{X}_S)^{-1} \tilde{X}_S^\top \Omega_S^{-1} \Delta y_S^{(k)}, \tag{11}
\]

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 \( \tilde{\beta}_S = \beta^{(k)} \) and \( \Delta y_S^{(k)} = y_S - f_S(x, \tilde{\beta}_S) \approx y_S - f_S(x, \beta) = \epsilon_S \). Since the covariance matrix of \( \Delta \beta^{(k)} \) and \( \beta^{(k)} \) are identical (\( \beta \) is constant) and by inserting \( \Delta y_S^{(k)} = \epsilon_S \) into (11)
we obtain

$$\text{Cov} \left( \hat{\beta}_S \right) = (\tilde{X}_S^T \Omega_S^{-1} \tilde{X}_S)^{-1}. \quad (12)$$

It is both an approximated theoretical expression and an estimator, assuming that $\Omega_S$ 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 $\alpha$. The generalized nonlinear least squares estimator is approximated through (e.g., Davidson and MacKinnon (1993))

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

Since $\hat{\alpha}$ is a function of $\tilde{y}_{F_{Sa}}$, 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)

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

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

$$E \left[ \text{Cov} (\hat{\alpha} \mid \tilde{y}_{F_{Sa}}) \right] = (\tilde{Z}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{Z}_{Sa})^{-1}, \quad (15)$$

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

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) as

$$\text{Cov} \left( E \left[ \hat{\alpha} \mid \tilde{y}_{Sa} \right] \right) = (\tilde{Z}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{Z}_{Sa})^{-1} \tilde{Z}_{Sa} \Sigma_{Sa}^{-1} \text{Cov} (\tilde{y}_{F_{Sa}}) \Sigma_{Sa}^{-1} \tilde{Z}_{Sa} (\tilde{Z}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{Z}_{Sa})^{-1}, \quad (16)$$

using that

$$E \left[ \hat{\alpha} \mid \tilde{y}_{F_{Sa}} \right] = E \left[ (\tilde{Z}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{Z}_{Sa})^{-1} \tilde{Z}_{Sa} \Sigma_{Sa}^{-1} \tilde{y}_{F_{Sa}} \mid \tilde{y}_{F_{Sa}} \right]$$

$$= (\tilde{Z}_{Sa}^T \Sigma_{Sa}^{-1} \tilde{Z}_{Sa})^{-1} \tilde{Z}_{Sa} \Sigma_{Sa}^{-1} \tilde{y}_{F_{Sa}}. \quad (17)$$

\footnote{Here we used the following rule: if $f = A x$, then the covariance matrix of $f$ is expressed as $\text{Cov} (f) = A \text{Cov} (x) A^T$, where $\text{Cov} (x)$ is a covariance matrix of $x$.}
The covariance of \( \hat{\alpha} \) is then

\[
\text{Cov}(\hat{\alpha}_{S_a}) = \left( \mathbf{Z}_{S_a}^\top \Sigma_{S_a}^{-1} \mathbf{Z}_{S_a} \right)^{-1} + \left( \mathbf{Z}_{S_a}^\top \Sigma_{S_a}^{-1} \mathbf{Z}_{S_a} \right)^{-1} \mathbf{Z}_{S_a}^\top \Sigma_{S_a}^{-1} \text{Cov}(\hat{y}_{F_{S_a}}) \Sigma_{S_a}^{-1} \mathbf{Z}_{S_a} \left( \mathbf{Z}_{S_a}^\top \Sigma_{S_a}^{-1} \mathbf{Z}_{S_a} \right)^{-1} \cdot \Sigma_{S_a}^{-1} \mathbf{Z}_{S_a}.
\]

(18)

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

\[
\hat{\text{Cov}}(\hat{y}_{F_{S_a}}) = \hat{\mathbf{X}}_{S_a} \hat{\text{Cov}}(\hat{\beta}_S) \hat{\mathbf{X}}_{S_a}.
\]

(19)

References


