Supplementary data to article *Uncertainty of upland soil carbon sink estimate for Finland* by Aleksi Lehtonen and Juha Heikkinen, Natural Resources Institute Finland (Luke)

Simulation of litter input to Yasso07

To simulate litter series with variability between series reflecting uncertainty and correlations of the estimates, we need to construct the covariance matrix of random vector Z consisting of estimators of $L_{\text{lb},T,c}$, $L_{\text{nm},Y,c}$, and $W_{\text{logg},c}$ for $T \in \{\text{NFI8}, \text{NFI9}, \text{NFI10}, \text{NFI11}\}$, $Y \in \{1990, 1998, 2003, 2008\}$, and $c \in \{\text{foliage, branches, stem+bark, stump, roots, fine roots}\}$. The included, mutually uncorrelated, sources of uncertainty were

- sampling variances $\text{Var}(V_{\text{lb},T})$ of stem volume estimators for living trees, uncorrelated between NFI rotations T,
- sampling variances $\text{Var}(V_{\text{nm},Y})$ of stem volume estimators for natural mortality, uncorrelated between years Y,
- sampling covariances $\text{Cov}_{\text{sampl}}(B_{s,\tau,c}, B_{s,\tau,c'})$ of BEF estimators, uncorrelated between sources s and time points τ, but correlated between biomass components c,
- covariances $\text{Cov}_{\text{model}}(B_{s,\tau,c}, B_{s',\tau',c'})$ due to uncertainty in the parameter estimates of biomass models, considered below in more detail, and
- variances $\text{Var}(P_{s,c})$ of the estimators of those litter production rates not equal to 1, uncorrelated between sources s and biomass components c.

The results were presented separately for southern and northern Finland, so we did not need the correlations between regions. Furthermore, errors in litter estimates were assumed to be uncorrelated between tree species groups. We can thus construct the required covariance matrices from those derived separately for each species group and region. The derivations presented here can be understood as applicable to a generic species and region, which will not be indexed for the sake of less cumbersome notation.

The covariance matrix C of Z containing contributions from all sources of uncertainty was derived as a sum of four matrices, C_{vol} (sampling uncertainty in stem volume), C_{sampl} (sampling uncertainty and correlations in BEFs), C_{model} (uncertainty in biomass models), and C_{litter} (uncertainty in litter rates). In order to specify the contents of each of these matrices, let V_i, B_i, and P_i refer to the stem volume, BEF, and litter rate estimate associated to the i'th element of Z, and let $s(i)$, $\tau(i)$, and $c(i)$ denote the associated litter source, time point, and biomass component. Then the elements of the covariance matrices are as follows:

$$C_{\text{vol},ij} = \begin{cases} B_i B_j P_i P_j \,\text{Var}(V_i), & \text{if } s(j) = s(i) \neq \text{logg and } \tau(j) = \tau(i) \ (\implies V_j = V_i), \\ 0, & \text{otherwise,} \end{cases}$$

$$C_{\text{sampl},ij} = \begin{cases} V_i V_j P_i P_j \,\text{Cov}_{\text{sampl}}(B_i, B_j), & \text{if } s(j) = s(i) \text{ and either } s(i) \neq \text{lb or } \tau(j) = \tau(i) \\ & (\implies B_{s(j),\tau(j),c} = B_{s(i),\tau(i),c}), \\ 0, & \text{otherwise,} \end{cases}$$

$$C_{\text{model},ij} = V_i V_j P_i P_j \,\text{Cov}_{\text{model}}(B_i, B_j), \text{ and}$$

$$C_{\text{litter},ij} = \begin{cases} V_i V_j B_i B_j \,\text{Var}(P_i), & \text{if } s(j) = s(i) \text{ and } c(j) = c(i), \\ 0, & \text{otherwise.} \end{cases}$$

Each of these four matrices is singular, consisting of blocks of equal covariances, but their sum is a proper covariance matrix.

The sampling variances and covariances $\mathrm{Var}(V_i)$, $\mathrm{Var}(P_i)$, and $\mathrm{Cov}_{\mathrm{sampl}}(B_i, B_j)$ were estimated in the usual NFI manner (Tomppo et al. 2011, sec. 3.5), and the model covariances $\mathrm{Cov}_{\mathrm{model}}(B_i, B_j)$ through approximations similar to those of Ståhl et al. (2014): Since the applied biomass models (Repola 2008, 2009) are of the general form

$$y_{c,m} = \exp\left(\sum_{k=1}^{p(c)} \alpha_{c,k} x_{m,c,k}\right),$$

where $y_{c,m}$ is the predicted biomass of component c of tree m, $\alpha_{c,k}$'s are the $p(c)$ parameters of the model for component c, and $x_{m,c,k}$'s some tree measurements, we can express BEF estimators as

$$B_{s,\tau,c} = \frac{\sum_{m \in S(s,\tau)} w_{s,\tau,m} y_{c,m}}{\sum_{m \in S(s,\tau)} w_{s,\tau,m} v_m} = \sum_{m \in S(s,\tau)} W_{s,\tau,m} y_{c,m},$$

where $S(s,\tau)$ is the sample of trees representing litter source s at time τ, $w_{s,\tau,m}$ the weight assigned to tree m in that sample (inversely proportional to the inclusion probability), v_m the stem volume of tree m, and

$$W_{s,\tau,m} = w_{s,\tau,m} \bigg/ \sum_{m' \in S(s,\tau)} w_{s,\tau,m'} v_{m'}.$$

Following Ståhl et al. (2014), covariances due to uncertainty in model parameters can then be approximated by

$$\mathrm{Cov}_{\mathrm{model}}(B_i, B_j) = \sum_{k=1}^{p(c(i))} \sum_{k'=1}^{p(c(j))} \frac{\partial B_i}{\partial \alpha_{c(i),k}} \frac{\partial B_j}{\partial \alpha_{c(j),k'}} \mathrm{Cov}(\alpha_{c(i),k}, \alpha_{c(j),k'}), \qquad (1)$$

where

$$\frac{\partial B_i}{\partial \alpha_{c,k}} = \sum_{m \in S(s(i),\tau(i))} x_{m,c,k} W_{s(i),\tau(i),m} y_{c,m}$$

and covariances $\mathrm{Cov}(\alpha_{c,k}, \alpha_{c',k'})$ are available in the appendix tables of Ståhl et al. (2011).

Example. To illustrate the computations described above as well as our simulations, let us consider a small example restricted to litter from the above-ground biomass components of living pine trees in southern Finland (Table S1). R code and input data for reproducing this example are included in the zip-file given as additional Supplementary data.

Table S1: Stem volumes V, BEFs B, litter rates P and litter Z of living pines in southern Finland.

i	$c(i)$	$\tau(i)$	V_i	B_i	P_i	$Z_i = V_i B_i P_i$
1	foliage	NFI8	408.36	29.52	0.245	2953.57
2	branches	NFI8	408.36	73.21	0.020	597.94
3	stem+bark	NFI8	408.36	391.29	0.005	830.89
4	foliage	NFI9	450.01	27.92	0.245	3077.87
5	branches	NFI9	450.01	70.84	0.020	637.60
6	stem+bark	NFI9	450.01	389.60	0.005	911.68
7	foliage	NFI10	493.77	27.23	0.245	3294.68
8	branches	NFI10	493.77	67.71	0.020	668.68
9	stem+bark	NFI10	493.77	388.20	0.005	996.74
10	foliage	NFI11	528.16	25.45	0.245	3293.79
11	branches	NFI11	528.16	64.48	0.020	681.08
12	stem+bark	NFI11	528.16	387.71	0.005	1064.82

Each of the four stem volume estimators V_i (one from each NFI, Table A1.1) contributes to those three litter estimators Z_j, which are based on the same NFI, i.e., $\tau(j) = \tau(i)$. Similarly, each of the three litter rate estimators P_i (one for each biomass component, Table A1.5) contributes to each NFI. Same biomass models are used in each NFI to compute the BEFs B_i (Table A1.2), but they are slightly different between NFIs, because the models are applied to different sets of trees.

The elements of covariance matrix C_{vol}, describing variation in Z due to sampling errors in V_i's, are $C_{\text{vol},ij} = B_i B_j P_i P_j C'_{\text{vol},ij}$, where

$$C'_{\text{vol}} = \begin{bmatrix} 22.53 & 22.53 & 22.53 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 22.53 & 22.53 & 22.53 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 22.53 & 22.53 & 22.53 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 21.04 & 21.04 & 21.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 21.04 & 21.04 & 21.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 21.04 & 21.04 & 21.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 30.17 & 30.17 & 30.17 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 30.17 & 30.17 & 30.17 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 30.17 & 30.17 & 30.17 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 24.93 & 24.93 & 24.93 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 24.93 & 24.93 & 24.93 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 24.93 & 24.93 & 24.93 \end{bmatrix}$$

Dividing the square roots of the four distinct values in C'_{vol} by the stem volume estimates V_i results in the first four rse-values of Table A1.1.

Similarly, $\text{Cov}_{\text{sampl}}(B_i, B_j) = C'_{\text{sampl},ij}$, where

$$C'_{\text{sampl}} = \begin{bmatrix} 0.11 & 0.15 & -0.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.15 & 0.27 & -0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.04 & -0.06 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.04 & 0.06 & -0.02 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.06 & 0.11 & -0.03 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -0.02 & -0.03 & 0.03 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.03 & 0.05 & -0.02 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.05 & 0.09 & -0.03 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.02 & -0.03 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.02 & 0.03 & -0.01 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.03 & 0.06 & -0.02 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.01 & -0.02 & 0.03 \end{bmatrix}$$

the diagonal values of C'_{sampl} corresponding to rse,s -values in Table A1.2 and off-diagonal values to correlations in Table A1.3, and $C_{\text{litter},ij} = V_i V_j B_i B_j C'_{\text{litter},ij}$, where the non-zero values of C'_{litter}, $\text{Var}(P_i)$, correspond to CV:s of Table A1.5. Finally, matrix

$$C'_{\text{model}} = \begin{bmatrix} 2.04 & 2.14 & -0.74 & 1.95 & 2.05 & -0.74 & 1.88 & 1.96 & -0.73 & 1.76 & 1.87 & -0.73 \\ 2.14 & 5.63 & -0.43 & 2.05 & 5.37 & -0.46 & 2.00 & 5.18 & -0.48 & 1.88 & 4.98 & -0.49 \\ -0.74 & -0.43 & 15.29 & -0.72 & -0.42 & 14.93 & -0.72 & -0.40 & 15.03 & -0.68 & -0.37 & 15.25 \\ 1.95 & 2.05 & -0.72 & 1.86 & 1.97 & -0.72 & 1.80 & 1.88 & -0.72 & 1.69 & 1.80 & -0.72 \\ 2.05 & 5.37 & -0.42 & 1.97 & 5.14 & -0.45 & 1.92 & 4.96 & -0.47 & 1.81 & 4.78 & -0.48 \\ -0.74 & -0.46 & 14.93 & -0.72 & -0.45 & 14.65 & -0.72 & -0.42 & 14.80 & -0.68 & -0.39 & 15.04 \\ 1.88 & 2.00 & -0.72 & 1.80 & 1.92 & -0.72 & 1.75 & 1.85 & -0.72 & 1.65 & 1.77 & -0.72 \\ 1.96 & 5.18 & -0.40 & 1.88 & 4.96 & -0.42 & 1.85 & 4.84 & -0.44 & 1.74 & 4.69 & -0.44 \\ -0.73 & -0.48 & 15.03 & -0.72 & -0.47 & 14.80 & -0.72 & -0.44 & 15.03 & -0.69 & -0.41 & 15.31 \\ 1.76 & 1.88 & -0.68 & 1.69 & 1.81 & -0.68 & 1.65 & 1.74 & -0.69 & 1.56 & 1.67 & -0.69 \\ 1.87 & 4.98 & -0.37 & 1.80 & 4.78 & -0.39 & 1.77 & 4.69 & -0.41 & 1.67 & 4.55 & -0.41 \\ -0.73 & -0.49 & 15.25 & -0.72 & -0.48 & 15.04 & -0.72 & -0.44 & 15.31 & -0.69 & -0.41 & 15.62 \end{bmatrix}$$

such that $\mathrm{Cov}_{\mathrm{model}}(B_i, B_j) = C'_{\mathrm{model},ij}$, is obtained using equation Eq. 1 with partial derivatives $\partial B_i/\partial\alpha_{c(i),k}$ listed in Table S2 and covariance matrix of model parameters (Table S3) derived from Table A1.2 of Ståhl et al. (2014). The within-component correlations in C'_{model} between the NFI's are close to 1, as expected, and also the model correlations between foliage and branch BEFs are quite high, as seen more clearly in Table A1.4.

Table S2: Partial derivatives of BEF-estimators with respect to the biomass model parameters.

i	$c(i)$	k	$\tau(i)$	$\partial B_i/\partial\alpha_{c(i),k}$
1	foliage	1	8	29.39
1	foliage	2	8	23.11
1	foliage	3	8	26.93
2	branches	1	8	73.39
2	branches	2	8	50.85
2	branches	3	8	40.02
3	stem+bark	1	8	391.31
3	stem+bark	2	8	265.69
3	stem+bark	3	8	226.45
4	foliage	1	9	28.19
4	foliage	2	9	22.12
4	foliage	3	9	25.90
5	branches	1	9	70.81
5	branches	2	9	48.68
5	branches	3	9	38.63
6	stem+bark	1	9	389.47
6	stem+bark	2	9	261.39
6	stem+bark	3	9	224.78
7	foliage	1	10	27.24
7	foliage	2	10	21.59
7	foliage	3	10	25.21
8	branches	1	10	67.72
8	branches	2	10	46.95
8	branches	3	10	37.94
9	stem+bark	1	10	388.20
9	stem+bark	2	10	260.86
9	stem+bark	3	10	226.82
10	foliage	1	11	25.49
10	foliage	2	11	20.37
10	foliage	3	11	23.69
11	branches	1	11	64.54
11	branches	2	11	45.16
11	branches	3	11	36.86
12	stem+bark	1	11	387.69
12	stem+bark	2	11	261.87
12	stem+bark	3	11	229.23

Table S3: Covariances between parameter estimates in the biomass models for the above-ground components of Scots pine (Ståhl et al. 2014, Table A1.2).

i	$c(i)$	$k(i)$	$\mathrm{Cov}(\alpha_{c(i),k(i)}, \alpha_{c(j),k(j)})$								
			j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8	j=9
1	foliage	1	0.275	0.116	-0.395	0.009	0.025	-0.047	-0.001	-0.003	0.006
2	foliage	2	0.116	0.122	-0.228	-0.004	0.025	-0.022	-0.000	-0.000	0.001
3	foliage	3	-0.395	-0.228	0.623	-0.005	-0.047	0.070	0.001	0.004	-0.008
4	branches	1	0.009	-0.004	-0.005	0.008	-0.011	-0.002	0.000	-0.001	0.000
5	branches	2	0.025	0.025	-0.047	-0.011	0.068	-0.062	-0.000	0.003	-0.002
6	branches	3	-0.047	-0.022	0.070	-0.002	-0.062	0.081	-0.000	-0.002	0.002
7	stem+bark	1	-0.001	-0.000	0.001	0.000	-0.000	-0.000	0.001	-0.002	0.000
8	stem+bark	2	-0.003	-0.000	0.004	-0.001	0.003	-0.002	-0.002	0.011	-0.010
9	stem+bark	3	0.006	0.001	-0.008	0.000	-0.002	0.002	0.000	-0.010	0.011

Covariance matrix $C = C_{\text{vol}} + C_{\text{sampl}} + C_{\text{model}} + C_{\text{litter}}$ containing all sources of uncertainty in Z implies relative standard deviations and correlations given in Table S4. Although both sampling and model errors of the BEFs were strongly correlated between foliage and branches (matrices C'_{sampl} and C'_{model}), litter estimates do not inherit these correlations. The reason is that the uncertainty in litter rates (uncorrelated between components) dominates the total uncertainty of the estimates of litter from living trees.

Table S4: Relative standard errors (rse) and mutual correlations ρ_{ij} of litter estimators Z_i of Table S1.

i	rse	ρ_{ij}, for $j =$											
	%	1	2	3	4	5	6	7	8	9	10	11	12
1	12.12	1.00	0.05	0.00	0.99	0.04	-0.00	0.98	0.04	-0.00	0.98	0.04	-0.00
2	20.31	0.05	1.00	0.00	0.04	1.00	-0.00	0.04	1.00	-0.00	0.04	1.00	-0.00
3	15.08	0.00	0.00	1.00	-0.00	-0.00	0.99	-0.00	-0.00	0.99	-0.00	-0.00	0.99
4	12.10	0.99	0.04	-0.00	1.00	0.05	0.00	0.99	0.04	-0.00	0.99	0.04	-0.00
5	20.29	0.04	1.00	-0.00	0.05	1.00	0.00	0.04	1.00	-0.00	0.04	1.00	-0.00
6	15.07	-0.00	-0.00	0.99	0.00	0.00	1.00	-0.00	-0.00	0.99	-0.00	-0.00	1.00
7	12.10	0.98	0.04	-0.00	0.99	0.04	-0.00	1.00	0.05	0.00	0.99	0.04	-0.00
8	20.30	0.04	1.00	-0.00	0.04	1.00	-0.00	0.05	1.00	0.00	0.04	1.00	-0.00
9	15.07	-0.00	-0.00	0.99	-0.00	-0.00	0.99	0.00	0.00	1.00	-0.00	-0.00	0.99
10	12.09	0.98	0.04	-0.00	0.99	0.04	-0.00	0.99	0.04	-0.00	1.00	0.05	0.00
11	20.30	0.04	1.00	-0.00	0.04	1.00	-0.00	0.04	1.00	-0.00	0.05	1.00	0.00
12	15.06	-0.00	-0.00	0.99	-0.00	-0.00	1.00	-0.00	-0.00	0.99	0.00	0.00	1.00

Table S5: Weights for interpolating annual litter estimates from those based on four NFI rotations

NFI	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	...
8	0.89	0.78	0.68	0.58	0.47	0.37	0.26	0.16	0.05				...
9	0.11	0.22	0.32	0.42	0.53	0.63	0.74	0.84	0.95	0.94	0.82	0.69	...
10										0.06	0.18	0.31	...
11													...

The interpolation weights for converting simulations of Z into annual time series (Table S5) are inversely proportional to the number of days from July 1 of the target year to the average of the measurement dates in the two adjacent NFIs. Fig. S1 illustrates 10 simulations. Strong correlations between NFIs lead to very few intersections between the interpolated series. On the other hand, weak correlations between biomass components are reflected by different order of series in the three panels.

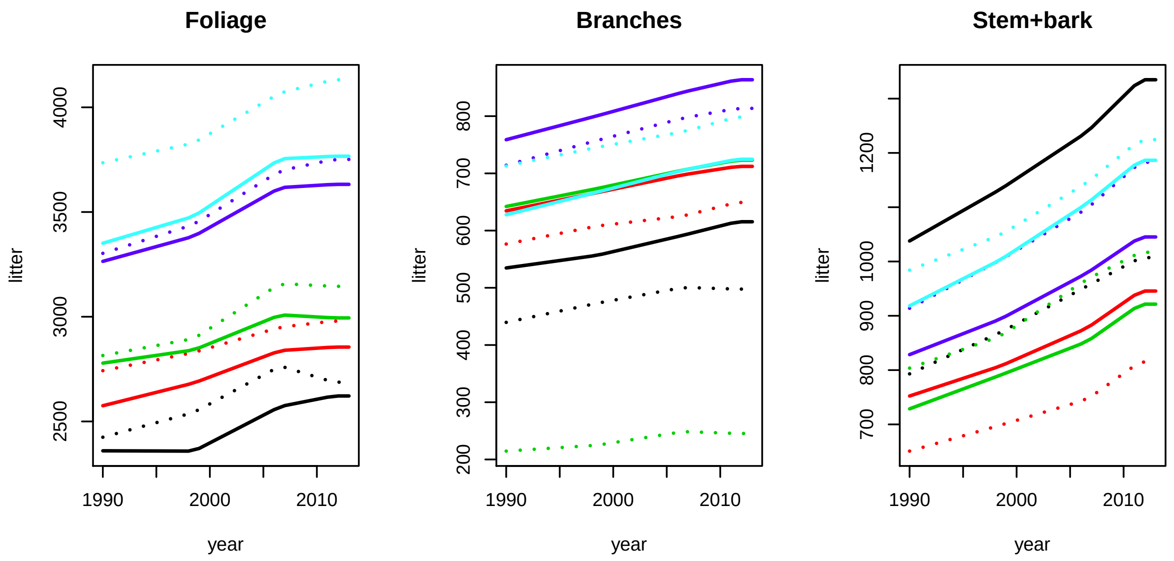

Fig. S1: Simulated litter series reflecting the uncertainty and correlations of litter estimates. The three lines with the same colour and type are from the same realization.