Modelling the Size of Forest Trees Using Statistical Distributions

Lauri Mehtätalo

University of Helsinki

Guest lecture at SLU, February 19, 2009

Outline of the presentation

Tree size

- Let X be a random variable characterizing the size of a tree in a forest stand.
- Randomness may be due to that
  - a tree has been selected randomly from the stand, or
  - the tree is regarded as a realization of an underlying, stochastic model of the stand.
- Most commonly diameter is used as tree size.
- Other alternatives are tree height, crown diameter, crown area, basal area, tree volume, etc.

Distribution function

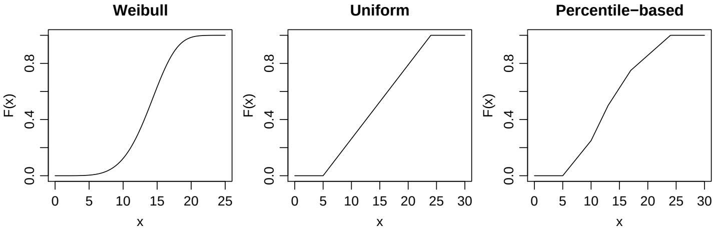
- The variability in tree size within the stand is accounted for through tree size distribution.
- The size distribution is defined as

$$F(x) = P(X \leq x)$$

- Two alternative interpretations for the distribution are
  - The probability for the size of a randomly selected tree to be below x
  - The proportion of trees with size below x

Distribution function

- For a function $F(x)$ to be a cdf, the following conditions need to hold:
  1. $F(x)$ is defined for $-\infty < x < \infty$, has minimum of 0 and maximum of 1 (i.e., $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$).
  2. $F(x)$ is a nondecreasing function of x.
  3. $F(x)$ is (right) continuous, i.e., for any x_0, $\lim_{x\to x_0} F(x) = F(x_0)$
- Examples of a distribution function

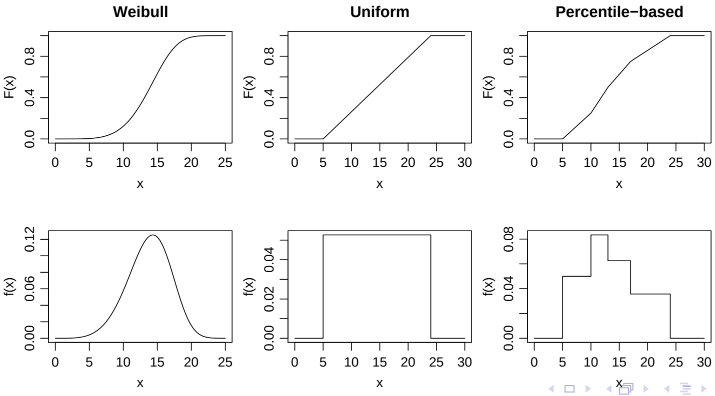


The density

The density corresponding to distribution function $F(x)$ is

$$f(x) = F'(x)$$

Examples



Distribution or density?

- The density is most commonly used
  - for illustration, as it is related to the commonly used histogram or stand table,

Distribution or density?

- The density is most commonly used
  - for illustration, as it is related to the commonly used histogram or stand table,
  - for computing expected values $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$,

Distribution or density?

- The density is most commonly used
  - for illustration, as it is related to the commonly used histogram or stand table,
  - for computing expected values $E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$,
  - for computing the likelihood $L(x) = \prod_{i=1}^{n} f(x)$ in ML-estimation,

Distribution or density?

- The density is most commonly used
  - for illustration, as it is related to the commonly used histogram or stand table,
  - for computing expected values $E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$,
  - for computing the likelihood $L(x) = \prod_{i=1}^{n} f(x)$ in ML-estimation,
  - for computing weighted distributions,

Distribution or density?

- The density is most commonly used
  - for illustration, as it is related to the commonly used histogram or stand table,
  - for computing expected values $E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$,
  - for computing the likelihood $L(x) = \prod_{i=1}^{n} f(x)$ in ML-estimation,
  - for computing weighted distributions,
- The distribution can be used e.g.,
  - for computing the proportion of trees between x_1 and x_2 as $F(x_2) - F(x_1)$,

Distribution or density?

- The density is most commonly used
  - for illustration, as it is related to the commonly used histogram or stand table,
  - for computing expected values $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$,
  - for computing the likelihood $L(x) = \prod_{i=1}^{n} f(x)$ in ML-estimation,
  - for computing weighted distributions,
- The distribution can be used e.g.,
  - for computing the proportion of trees between x_1 and x_2 as $F(x_2) - F(x_1)$,
  - for computing distributions of random variables that are related to X, e.g., that of transformed random variables.

Transformation

- A transformed random variable is random variable Y that is obtained from X using transformation $Y = g(X)$.
- Useful examples are volume, height, basal area, or crown area (Y) as a function of tree diameter (X).

Transformation

- A transformed random variable is random variable Y that is obtained from X using transformation $Y = g(X)$.
- Useful examples are volume, height, basal area, or crown area (Y) as a function of tree diameter (X).
- The distribution of $Y = g(X)$ is

$$F_Y(y) = F_X(g^{-1}(y)) \quad \text{if } g \text{ is increasing}$$
$$F_Y(y) = 1 - F_X(g^{-1}(y)) \quad \text{if } g \text{ is decreasing} \ ,$$

  where $g^{-1}(y)$ is the inverse transformation
- If X is tree diameter, then $g^{-1}(y)$ is tree diameter as a function of volume, height, basal area or crown area.

Transformation

- A transformed random variable is random variable Y that is obtained from X using transformation $Y = g(X)$.
- Useful examples are volume, height, basal area, or crown area (Y) as a function of tree diameter (X).
- The distribution of $Y = g(X)$ is

$$\begin{aligned} F_Y(y) &= F_X(g^{-1}(y)) \quad \text{if } g \text{ is increasing} \\ F_Y(y) &= 1 - F_X(g^{-1}(y)) \quad \text{if } g \text{ is decreasing} \ , \end{aligned}$$

  where $g^{-1}(y)$ is the inverse transformation
- If X is tree diameter, then $g^{-1}(y)$ is tree diameter as a function of volume, height, basal area or crown area.
- Applications
  - Formulating different distributions based on allometric relationships of trees

Weighted distribution

- Frequencies are often proportional to the number of stems.
- It may be more convenient to have them proportional to some other characteristics, such as basal area or volume
- The density of a weighted diameter distribution is

$$f_X^w(x) = \frac{w(x)f_X(x)}{\int_0^\infty w(u)f_X(u)\mathrm{d}u}.$$

- The nominator is the mean of $w(x)$, e.g., mean basal area (G/N), or mean volume (V/N).

Weighted distribution

- Frequencies are often proportional to the number of stems.
- It may be more convenient to have them proportional to some other characteristics, such as basal area or volume
- The density of a weighted diameter distribution is

$$f_X^w(x) = \frac{w(x)f_X(x)}{\int_0^\infty w(u)f_X(u)\mathrm{d}u}.$$

- The nominator is the mean of $w(x)$, e.g., mean basal area (G/N), or mean volume (V/N).
- The weighted distribution function is $F_X^w(x) = \int_0^x f_X^w(u)\mathrm{d}u$

Weighted distribution

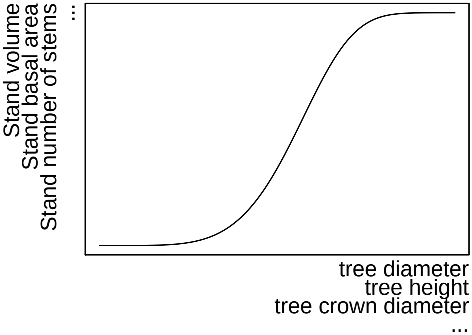
- Frequencies are often proportional to the number of stems.
- It may be more convenient to have them proportional to some other characteristics, such as basal area or volume
- The density of a weighted diameter distribution is

$$f_X^w(x) = \frac{w(x)f_X(x)}{\int_0^\infty w(u)f_X(u)\mathrm{d}u}.$$

- The nominator is the mean of $w(x)$, e.g., mean basal area (G/N), or mean volume (V/N).
- The weighted distribution function is $F_X^w(x) = \int_0^x f_X^w(u)\mathrm{d}u$
- Applications
  - weighted sampling schemes, such as angle count sampling or overlapping crowns in a aerial forest inventory, or varying sample plot radius according to tree diameter.
  - computing total basal area, number of stems, volume, or other aggregate characteristics between sizes x_1 and x_2,
  - scaling diameter distributions with other characteristics than the number of stems, e.g., the total volume or basal area.

Transformation and weighting

- Transformation changes the x-axis
- Weighting changes the y-axis



Estimating the distribution

- If individual trees have been measured for size x

- If only stand characteristics (G, N, D, etc.) are known

Estimating the distribution

- If individual trees have been measured for size x
  - Use the measured sample as such
- If only stand characteristics (G, N, D, etc.) are known

Estimating the distribution

- If individual trees have been measured for size x
  - Use the measured sample as such
  - Smooth the sample using, e.g., kernel smoothing

- If only stand characteristics (G, N, D, etc.) are known

Estimating the distribution

- If individual trees have been measured for size x
  - Use the measured sample as such
  - Smooth the sample using, e.g., kernel smoothing
  - Fit an assumed distribution function to the data using e.g., the method of Maximum likelihood. Alternative distributions are e.g. Weibull(2 or 3 parameters), logit-logistic (4 parms), SB (4 parms), percentile-based (2,...,n parms)...
- If only stand characteristics (G, N, D, etc.) are known

Estimating the distribution

- If individual trees have been measured for size x
  - Use the measured sample as such
  - Smooth the sample using, e.g., kernel smoothing
  - Fit an assumed distribution function to the data using e.g., the method of Maximum likelihood. Alternative distributions are e.g. Weibull(2 or 3 parameters), logit-logistic (4 parms), SB (4 parms), percentile-based (2,...,n parms)...
- If only stand characteristics (G, N, D, etc.) are known
  - Use existing models (regresion, kNN..) to predict the parameters of a distribution function to the data (Parameter Prediction Method, PPM)

Estimating the distribution

- If individual trees have been measured for size x
  - Use the measured sample as such
  - Smooth the sample using, e.g., kernel smoothing
  - Fit an assumed distribution function to the data using e.g., the method of Maximum likelihood. Alternative distributions are e.g. Weibull(2 or 3 parameters), logit-logistic (4 parms), SB (4 parms), percentile-based (2,...,n parms)...
- If only stand characteristics (G, N, D, etc.) are known
  - Use existing models (regresion, kNN..) to predict the parameters of a distribution function to the data (Parameter Prediction Method, PPM)
  - Use mathematical relationships between stand characteristics and distribution parameters to solve the parameters analytically (Parameter Recovery Method, PRM). Part of the stand cahracteristics may be predicted using regression models.

Estimating the distribution

- If individual trees have been measured for size x
  - Use the measured sample as such
  - Smooth the sample using, e.g., kernel smoothing
  - Fit an assumed distribution function to the data using e.g., the method of Maximum likelihood. Alternative distributions are e.g. Weibull(2 or 3 parameters), logit-logistic (4 parms), SB (4 parms), percentile-based (2,...,n parms)...
- If only stand characteristics (G, N, D, etc.) are known
  - Use existing models (regresion, kNN..) to predict the parameters of a distribution function to the data (Parameter Prediction Method, PPM)
  - Use mathematical relationships between stand characteristics and distribution parameters to solve the parameters analytically (Parameter Recovery Method, PRM). Part of the stand cahracteristics may be predicted using regression models.
  - Also combinations of these are possible

Note on computations

- Analytical solutions for several of the presented results do not exist or are hard to obtain,

Note on computations

- Analytical solutions for several of the presented results do not exist or are hard to obtain,
- but this is not a problem, because numerical algorithms exist, and they are usually fast and stable.

Note on computations

- Analytical solutions for several of the presented results do not exist or are hard to obtain,
- but this is not a problem, because numerical algorithms exist, and they are usually fast and stable.
  - It is quite easy to numericalle compute integrals, perform maximization or minimization, solve inverse functions numerically, or solve systems of equations.
  - Mathematicians call these results as approximations, but from foresters' point of view, these results are usually as accurate as analytical results, even though some care is needed.

Note on computations

- Analytical solutions for several of the presented results do not exist or are hard to obtain,
- but this is not a problem, because numerical algorithms exist, and they are usually fast and stable.
  - It is quite easy to numericalle compute integrals, perform maximization or minimization, solve inverse functions numerically, or solve systems of equations.
  - Mathematicians call these results as approximations, but from foresters' point of view, these results are usually as accurate as analytical results, even though some care is needed.
  - The results of this presentation utilized R-functions `integrate()`, `optim()`, and my own implementations of the Newton-Raphson and simple up-and-down algorithms.

Scaling

- Usually, the distribution is scaled with the weighting variable
  - Unweighted distribution is scaled with the number of stems
  - basal area weighted is scaled using basal area
  - volume weighted is scaled using volume

Scaling

- Usually, the distribution is scaled with the weighting variable
  - Unweighted distribution is scaled with the number of stems
  - basal area weighted is scaled using basal area
  - volume weighted is scaled using volume
- This is not always the case, e.g., the unweighted distribution may need to be scaled using basal area or volume.
- Solution: compute the ratio of the scaling variable and the number of stems, and solve it for a computational N

Scaling

- Usually, the distribution is scaled with the weighting variable
  - Unweighted distribution is scaled with the number of stems
  - basal area weighted is scaled using basal area
  - volume weighted is scaled using volume
- This is not always the case, e.g., the unweighted distribution may need to be scaled using basal area or volume.
- Solution: compute the ratio of the scaling variable and the number of stems, and solve it for a computational N
- Examples
  - The mean basal area is $\int_0^\infty \pi/4u^2 f_X^N(u)\mathrm{d}u$ which is the ratio of basal area and number of stems, G/N. The number of stems corresponding to G is $N = G/\int_0^\infty \pi/4u^2 f_X^N(u)\mathrm{d}u$
  - If V is known but N or G is not known, the diameter distribution should be scaled using $N = V/\int_0^\infty v(u) f_X^N(u)\mathrm{d}u$, where $v(x)$ is a known volume model as a function of diameter.
  - If a height-diameter curve $h(x)$ and volume function $v(x, h)$ are available, use $N = V/\int_0^\infty v(u, h(u)) f_X^N(u)\mathrm{d}u$.

Basal area between given diameters

- Let X be tree diameter. The density of a basal area weighted diameter distribution is

$$f_X^G(x) = \frac{\pi/4x^2 f_X^N(x)}{\int_0^\infty \pi/4u^2 f_X^N(u)\mathrm{d}u} \quad (1)$$

- The distribution function is $F_X^G(x) = \int_0^x f_X^G(u)\mathrm{d}u$

Basal area between given diameters

- Let X be tree diameter. The density of a basal area weighted diameter distribution is

$$f_X^G(x) = \frac{\pi/4x^2 f_X^N(x)}{\int_0^\infty \pi/4u^2 f_X^N(u)\mathrm{d}u} \quad (1)$$

- The distribution function is $F_X^G(x) = \int_0^x f_X^G(u)\mathrm{d}u$
- The basal area between diameters x_1 and x_2 is $G\left[F_X^G(x_2) - F_X^G(x_1)\right]$

Basal area between given diameters

- Let X be tree diameter. The density of a basal area weighted diameter distribution is

$$f_X^G(x) = \frac{\pi/4x^2 f_X^N(x)}{\int_0^\infty \pi/4u^2 f_X^N(u)\mathrm{d}u} \tag{1}$$

- The distribution function is $F_X^G(x) = \int_0^x f_X^G(u)\mathrm{d}u$
- The basal area between diameters x_1 and x_2 is $G\left[F_X^G(x_2) - F_X^G(x_1)\right]$
- If G is not known, it can be solved from $G/N = \int_0^\infty \pi/4u^2 f_X^N(u)\mathrm{d}u$

The distribution of height

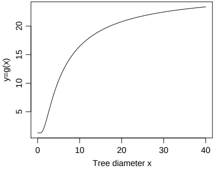
Let X be tree diameter and Y tree height. The diameter distribution is two-parameter Weibull

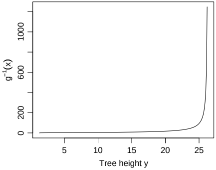
$$F_X(x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right].$$

and height-diameter curve is

$$y = g(x) = 1.3 + a\exp\left(\frac{b}{x}\right), \tag{2}$$

where parameters are $\alpha = 4$, $\beta = 15$, $a = 25$ and $b = -5$. Solving (2) for x gives $g^{-1}(y) = \frac{b}{\ln\left(\frac{y-1.3}{a}\right)}$.





Height distribution continued

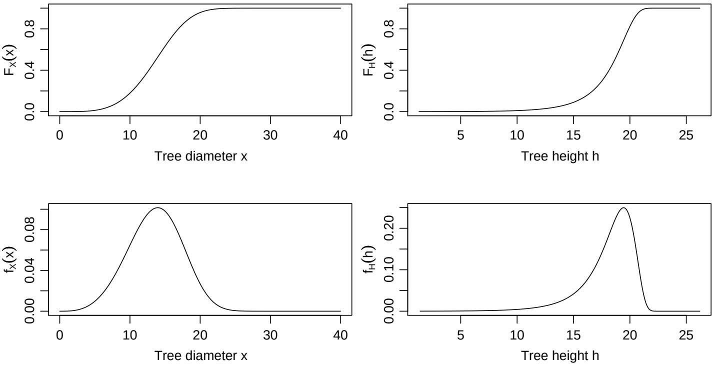
The distribution of tree height becomes

$$F_Y(y) = 1 - exp\left[-\left(\frac{b}{\beta \ln\left(\frac{y-1.3}{a}\right)}\right)^{\alpha}\right].$$

Note that this is no more of the Weibull form.
The density is obtained through differentiation as

$$f_Y(y) = \frac{\alpha}{\beta}\left(\frac{b}{\beta \ln\left(\frac{y-1.3}{a}\right)}\right)^{\alpha-1} \exp\left[-\left(\frac{b}{\beta \ln\left(\frac{y-1.3}{a}\right)}\right)^{\alpha}\right] \frac{-b}{(y-1.3)\left[\ln\left(\frac{y-1.3}{a}\right)\right]^2}$$

Plots of the example



Dominant height

- The height distribution of dominant trees is obtained by truncating the height distribution and rescaling it to unity. Dominant height is the expected value of that distribution
- There are N=500 stems/ha in the stand. The limit of dominant trees is obtained from the height distribution as $F_H^{-1}((N-100)/N) = 19.89572$
- The dominant height is

$$H_{dom} = \frac{N}{100} \int_{F_H^{-1}((N-100)/N)}^{H_{max}} u f_H(u) \mathrm{d}u = 20.44752$$

Another way to compute the dominant height

- The diameter limit for dominant trees is $d_{dom} = F_D^{-1}((N-100)/N) = 16.89506$ cm.
- The diameter distribution of dominant trees is obtained by truncating the diameter distribution at d_{dom} and rescaling with $(N-100)/N$
- Using the expected value of $g(X)$ we get

$$H_{dom} = \frac{N}{100} \int_{F_X^{-1}((N-100)/N)}^{\infty} f_X^N(u) h(u) \mathrm{d}u = 20.44752 \,,$$

Overlapping crowns in a Poisson stand

- Crown radius Z is assumed to follow Weibull distribution.
- Assume that a tree remains unobserved if the tip is within a crown of a larger tree.
- The probability for a tree being observed depends on crown radius according to

$$w(z) = \pi^{\frac{1}{\int_0^\infty t^2 f_Z(t|\alpha,\beta)dt} \int_z^\infty t^2 f_Z(t|\alpha,\beta)dt},$$

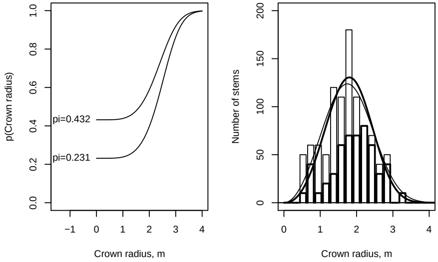
  where π is the expected canopy closure, replaced with its observed value in applications.
- The distribution of observed crown areas is

$$f_Z^w(z|\alpha,\beta) = \frac{w(z|\alpha,\beta) f_Z(z|\alpha,\beta)}{\int_0^\infty w(u|\alpha,\beta) f_Z(u|\alpha,\beta) \mathrm{d}u}.$$

- Parameters α and β can be estimated by fitting the weighted distribution to the observed sample of crown radii.
- Stand density can then be estimated as $\lambda = -\frac{\ln(\pi)}{E(Z)}$.
- The same principle can be applied by weighting the actual observations.

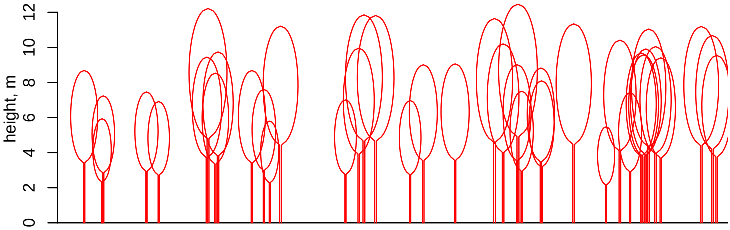
Overlapping crowns in a Poisson stand (continued)

Figure: The plot on the left shows the probability of a tree being observed. The plot on the right shows an example of an estimated distribution using simulated data. The Histogram with thick lines is the observed sample, the histogram with thin lines shows all trees of the plot. the thin line shows the true underlying Weibull distribution and the Thick line the estimated distribution.



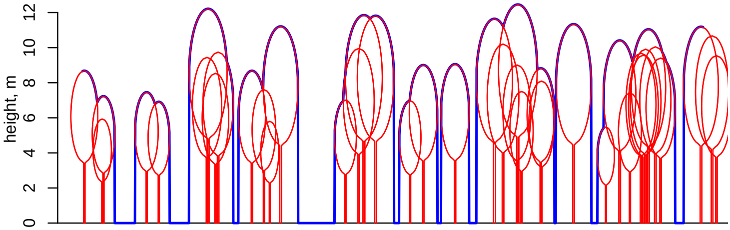
From Mehtätalo 2006, CJFR, Further developments to 3D (laser) data in Mehtätalo and Nyblom 2009 (Manuscript, For Sci).

Idea of the generalization



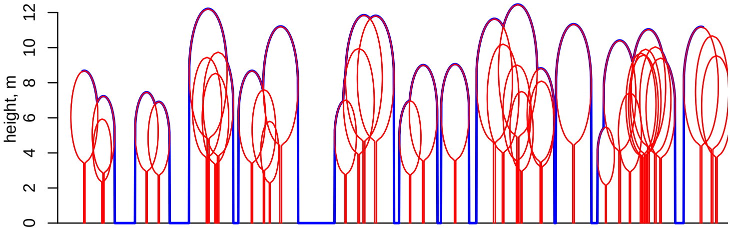
- In forest inventory, we are interested in the number, species and size of the trees.

Idea of the generalization



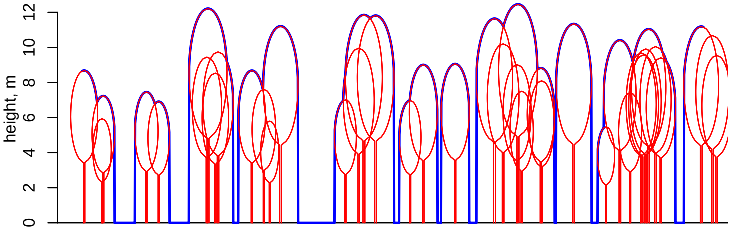
- In forest inventory, we are interested in the number, species and size of the trees.
- From above we can see only the surface of the forest stand.

Idea of the generalization



- In forest inventory, we are interested in the number, species and size of the trees.
- From above we can see only the surface of the forest stand.
- In airborne laser scanning, we essentially measure height of this surface at given points, i.e. the distribution of canopy height.

Idea of the generalization



- In forest inventory, we are interested in the number, species and size of the trees.
- From above we can see only the surface of the forest stand.
- In airborne laser scanning, we essentially measure height of this surface at given points, i.e. the distribution of canopy height.
- At each reference height z, $1 - P(Z \leq z)$ is obtained as a union of crown intersections at the reference height.

PRM for ALS-inventory

- Assume that diameter distribution is $Weibull(\alpha, \beta)$ and H-D-curve is $Korf(a)$. denote $\boldsymbol{\theta} = (\alpha, \beta, a)'$.
- Using the results of weighted distributions and transformations, we can write expressions for volume, mean height and mean diameter as $V(\boldsymbol{\theta}, \widehat{N})$, $D(\boldsymbol{\theta})$, and $H(\boldsymbol{\theta})$.
- Equating these expressions to values obtained using ALS, $\widehat{V}$, $\widehat{D}$ and $\widehat{H}$ we get

$$\begin{cases} V(\boldsymbol{\theta}, \widehat{N}) = \widehat{V} \\ D(\boldsymbol{\theta}) = \widehat{D} \\ H(\boldsymbol{\theta}) = \widehat{H} \end{cases}$$

Evaluation

- 213 almost pure Scots pine plots from Juuka, Eastern Finland.
- Regression models were used to estimate V, H, D and N

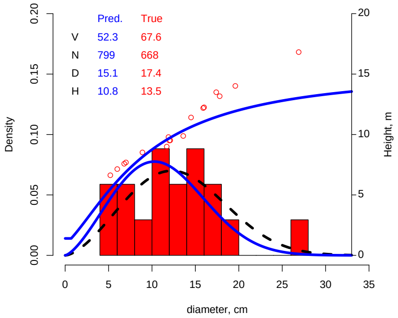
Evaluation

- 213 almost pure Scots pine plots from Juuka, Eastern Finland.
- Regression models were used to estimate V, H, D and N
- On 2 out of 213 the system was infeasible
- The RMSE of volumes of trees above 10, 15 and 20 cm in diameter are given below.

	RMSE		Bias	
	Absolute	%	Absolute	%
V_{10}, m^3 ha^{-1}	19.7	16.67	-0.79	-0.67
V_{15}, m^3 ha^{-1}	22.2	22.7	-2.2	-2.25
V_{20}, m^3 ha^{-1}	24.33	42.76	-1.72	-3.02

From Mehtätalo, Maltamo and Packalen 2007, SilviLaser.

An example plot



Examples of true and recovered stand descriptions. The histogram shows the observed diameter distribution and the open circles the tree heights. The dashed line shows a Weibull-distribution fitted to the observed data using ML. The blue lines show the recovered diameter distribution and H-D curve.

Models for Weibull parameters

- Assumed that G, N, DGM and mean height are known.
- When $f_D^G(d)$ is Weibull density, the following equations hold:

$$\begin{aligned}\beta &= \frac{DGM}{\ln 2^{1/\alpha}} \\ \frac{\pi}{4\Gamma(1-2/\alpha)(\ln 2)^{2/\alpha}} &= \frac{G}{N \cdot DGM^2},\end{aligned}$$

where $\Gamma()$ is the gamma function.

Models for Weibull parameters

- Assumed that G, N, DGM and mean height are known.
- When $f_D^G(d)$ is Weibull density, the following equations hold:

$$\begin{aligned} \beta &= \frac{DGM}{\ln 2^{1/\alpha}} \\ \frac{\pi}{4\Gamma(1-2/\alpha)(\ln 2)^{2/\alpha}} &= \frac{G}{N \cdot DGM^2}, \end{aligned}$$

  where $\Gamma()$ is the gamma function.
- Two modelling approaches
  - PRM: the above system was solved for α and β.

Models for Weibull parameters

- Assumed that G, N, DGM and mean height are known.
- When $f_D^G(d)$ is Weibull density, the following equations hold:

$$\beta = \frac{DGM}{\ln 2^{1/\alpha}}$$
$$\frac{\pi}{4\Gamma(1-2/\alpha)(\ln 2)^{2/\alpha}} = \frac{G}{N \cdot DGM^2},$$

  where $\Gamma()$ is the gamma function.
- Two modelling approaches
  - PRM: the above system was solved for α and β.
  - PPM1: the following models were fitted to pine data using 2SLS

$$shtr_i = a_\alpha + b_\alpha \cdot xtr1_i[+c_\alpha \bar{H}_i] + e_{\alpha i}$$
$$\beta_i = a_\beta + b_\beta \cdot xtr2_i[+c_\beta \bar{H}_i] + e_{\beta i},$$

  where $shtr_i = \frac{\pi}{4\Gamma(1-2/\alpha_i)(\ln 2)^{2/\alpha_i}}$, $xtr1_i = \frac{G_i}{N_i \cdot DGM_i^2}$ and $xtr2_i = \frac{DGM_i}{(\ln 2)^{1/\widehat{\alpha}_i}}$.

Models for Weibull parameters

Table: Results from comparisons between different estimation methods in the modeling and test datasets.

		ML fit	PPM1	PPM2	PRM	Partial recovery
Modeling data						
Volume	RMSE	1.19	1.29	1.11	1.22	1.22
	Bias	-0.039	-0.053	0.033	-0.418	-0.416
Error index	Mean	6.10	6.43	6.30	6.56	6.29
Test data						
Volume	RMSE	0.296	0.952	0.688	0.652	1.05
	Bias	-0.013	-0.522	-0.292	-0.013	-0.84
Error index	Mean	6.64	7.68	7.62	8.33	7.59

- How does theoretically-based PPM compare to a trial-and-error PPM?
- From Mehtätalo and Nyblom, manuscript.

Models for Weibull parameters

Histogram: true distribution
Dashed: ML estimate
Thin solid: PPM2
Dotted: PRM

Volume

	True	ML-fit	PPM2	PRM
A	36.5	36.5	36.6	36.7
B	81.1	81.3	81.7	81.1
C	102.9	103	103.7	102.9

Error index

	ML-fit	PPM2	PRM
A	1.0	1.2	1.2
B	3.6	4.7	6.2
C	4.8	7.0	13.5