

Smooth Height/Age Curves from Stem Analysis with Linear Programming

Juha Lappi

Lappi, J. 2006. Smooth height/age curves from stem analysis with linear programming. *Silva Fennica* 40(2): 291–301.

Stem analysis data defines a range of possible heights for each age. A smooth stem/age curve is obtained with linear programming (LP) when the sum of the absolute second differences of heights is minimized subject to constraints obtained from the stem analysis. The method is analogous to cubic splines. A LP problem can include additional constraints that are based on the assumption that the crosscut is randomly located within the annual height increment. The method produces smoother height curves than *Issa* method which is utilizing second order differences of ring counts. It was found using simulated data that the method provides better results than earlier methods for short bolts if height growth is sufficiently regular.

Keywords spline, ring counts, second order difference, Issa, height growth

Author's address Finnish Forest Research Institute, Suonenjoki Research Unit, Juntintie 154, FI-77600 Suonenjoki, Finland **E-mail** juha.lappi@metla.fi

Received 11 May 2005 **Revised** 2 December 2005 **Accepted** 20 March 2006

Available at <http://www.metla.fi/silvafennica/full/sf40/sf402291.pdf>

1 Introduction

In a stem analysis, stems are cut at different positions and rings are counted from the crosscuts. Ring widths are often also measured. Each ring disappearing in a bolt corresponds to a hidden tip in the bolt. The unknown locations of the hidden tips determine the height/age curve of the stem. Several authors have presented methods for estimating the height/age curves from stem analysis data, e.g., Carmean (1972) and Lenhart (1972). A review in a unified formalism and test results were

given by Dyer and Bailey (1987). More recently Fabbio et al. (1994) presented an *Issa* method that used second differences of ring counts to get smoother height/age curves. The previous methods used stem analysis data only locally, and hence the estimated heights and height increments were not smooth over crosscuts. All methods have assumed that all height increments that are totally hidden in a bolt are of equal length. The methods are based on different assumptions how crosscuts occur within an increment that is divided into two bolts.

The method proposed in this paper will produce smoother height/age curves than the *Issa* method using simultaneously data from all crosscuts. The motivation for striving for smoothness is the same as why smooth regression functions are used. Even if it is known that there are fluctuations in height/age curves it is better to use smooth curves if the measurements do not indicate how the curves fluctuate. The smoothness is not a goal as such but a means for removing artificial fluctuations from the estimated curves. An estimated curve can also be too smooth if it removes nonlinear trends in addition to random fluctuations.

The definition of smoothness is taken from the theory of interpolating cubic splines (see, e.g., DeBoor 1978). Interpolating cubic splines are curves that pass through known points and minimize $\int_a^b |f''(x)| dx$. In this paper smooth curves are obtained by minimizing the sum of absolute values of second differences of heights (note that in the *Issa* method second differences of ring counts are used). Because a height/age curve is not assumed to describe the within-year growth pattern, the second order difference is an appropriate measure of smoothness and not just a numerical approximation to the second derivative. Constraints for the height/age curve, corresponding to known points in splines, are obtained from the stem analysis. The resulting problem is a standard linear programming problem.

As the resulting curve is the smoothest curve consistent with the data, it may be too smooth. Additional constraints that make the curve less smooth are obtained by assuming that the crosscut occurs randomly within the annual height increment. This sounds similar to but is not equivalent to the assumption behind the method of Carmean (1972).

The proposed method is illustrated using several bolt lengths in simulated data. The results are compared with the *Issa* method of Fabbio et al. (1994) and with Carmean's method that Dyer and Bailey (1987) found to produce most consistent results before *Issa*.

The basic ideas of the paper are simple. The implementation of the ideas in the linear programming framework is also straightforward but rather tedious. To serve readers not interested in technical details, the symbols are presented in

Appendix 1, and linear programming details in other appendices. Carmean's and *Issa* methods are also presented in Appendix 1.

2 The Smoothest Possible Curve

The second order difference is a natural measure of the smoothness of a function defined on integer values. Let D_t denote the second (forward) difference of height at age t , $t=0, \dots, n-2$ (n is the total age, see Appendix 1):

$$D_t = (H_{t+2} - H_{t+1}) - (H_{t+1} - H_t), \text{ or} \\ D_t = H_{t+2} - 2H_{t+1} + H_t \tag{1}$$

where H_t is the height at age t (see e.g. Zwilliger 1995 for the difference calculus needed). The smoothest possible curve can be obtained by

finding H_1, \dots, H_{n-1} , that minimize $\sum_{t=0}^{n-2} |D_t|$ subject

to the constraints implied by the stem analysis data. This problem can be solved with linear programming (linear programming theory used in this paper can be found in any text book, e.g., Luenberger 1973). Using the standard trick in linear programming, the absolute value $|D_t|$ is treated by expressing D_t and $|D_t|$ as follows:

$$D_t = D_t^+ - D_t^- \tag{2}$$

$$|D_t| = D_t^+ + D_t^- \tag{3}$$

where D_t^+ and D_t^- are nonnegative variables. When the sum of (3) is minimized in LP, then in the solution either $D_t^+ = D_t$ and $D_t^- = 0$ or $D_t^- = -D_t$, and $D_t^+ = 0$. The complete linear programming problem producing the smoothest possible curve is shown in Appendix 2.

3 Crosscuts in the Middle of Annual Increment

The solution for the LP problem in Appendix 2 may be too smooth (too linear) so we may try

to append more constraints that would make the curve more realistic and less smooth. Assume for the moment that all bolts contain hidden tips. If there are no systematic interaction between the locations of crosscuts and height growth, then the location of a crosscut within a height increment behaves as a uniform random number. This implies that the expected location is in the middle of the height increment. Thus we might add to the basic LP problem additional constraints requiring that each crosscut is in the middle of height increment, or crosscuts are on the average in the middle of the height increments for any stem section.

The smoothness of the height curve becomes more interesting when the bolts are short. With short bolts there can be bolts without hidden tips, or equivalently an annual height increment can include whole bolts. In that case the constraint that each crosscut occurs in the middle of the height increment is not logical. The proper condition is then that the middle point of a bolt segment occurring within a height increment is in the middle of the height increment. Appendix 3 shows how to implement these constraints. Note that Carmean’s method is also based on the concept that crosscut is in the middle of height increment. But this is implemented so that the crosscut occurs at a point where the distance from the previous hidden tip is half of the constant increment occurring in the previous bolt, and the distance from the next hidden tip is the half of the increment in the next bolt. In Carmean’s method ‘in the middle’ refers to time, and in the linear programming method ‘in the middle’ refers to distance. If crosscuts are independent of the height growth, the interpretation given in this paper is theoretically sounder.

4 A Closer Look at the Solutions

The proposed problems can be directly solved with any linear programming software. If there are more than two hidden tips in any bolt, a further analysis will provide us smaller LP problems, alternative solutions and further insight.

Let us assume, using a case illustrated in Fig. 1, that there are at least three hidden tips in bolt with

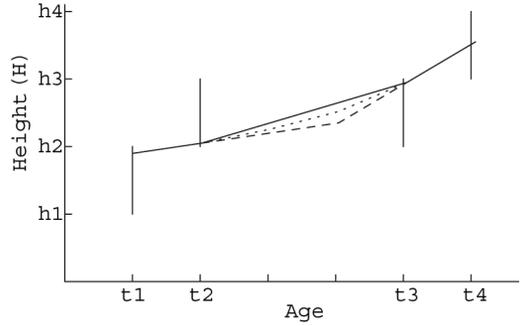


Fig. 1. Illustration of possible stem analysis data. The last hidden tip in bolt $[h_1, h_2]$ corresponds to the unknown height at age t_1 , $H(t_1)$, the height of the first hidden tip in bolt $[h_2, h_3]$ is $H(t_2)$, the height of the last hidden tip is $H(t_3)$ and the first hidden tip in bolt $[h_3, h_4]$ is $H(t_4)$. The stem analysis thus implies that $H(t_1) < h_2 \leq H(t_2) < \dots < H(t_3) < h_3 \leq H(t_4)$. The three curves shown are examples of possible estimated height/age curves consistent with these constraints. For ages between t_2 and t_3 all such estimated heights that the overall curve between t_1 and t_4 is convex yield height/age curves having equal sum of second order differences. The solid line shows the highest such curve, provided by standard linear programming solution. The lowest curve (dashed line) is obtained by continuing the lines in $[t_1, t_2]$ and in $[t_3, t_4]$ to the point where they cross. The middle line (dotted line) is obtained by minimizing the sum of absolute third differences.

endpoints at h_2 and h_3 . Let t_2 , and t_3 be the ages of the first and last hidden tips in the bolt ($t_3 > t_2 + 1$). For given $H(t_2)$ and $H(t_3)$, all nondecreasing heights $H(t)$, $t_2 < t < t_3$ satisfy the constraints of the LP problems proposed above. It can be shown that an optimal solution is such that $H(t)$ for $t_2 < t < t_3$ is on the line connecting $H(t_2)$ and, (the second differences are zero for intermediate points). That solution will be called the standard solution, and it is provided by standard LP algorithms which will keep variables zero (nonbasic) as long as the solution cannot be improved by making them nonzero (basic). The standard solution is convex in $[t_1, t_4]$ if

$$\begin{aligned}
 H(t_2) - H(t_1) &\leq [H(t_3) - H(t_2)] / (t_3 - t_2) \\
 &\leq H(t_4) - H(t_3)
 \end{aligned}
 \tag{4}$$

and the solution is concave if \leq in (4) is replaced by \geq .

If the standard solution is neither convex nor concave in $[t_1, t_4]$ then the standard solution is unique, that is for other feasible solutions the objective function would increase. If the standard solution is convex or concave, then the standard solution is not unique. More specifically, if the standard solution is convex, then the properties of difference operators imply that for any general convex solution such that

$$H(t_2) - H(t_1) \leq H(t_2 + 1) - H(t_2) \tag{5}$$

$$\leq \dots \leq H(t_4) - H(t_3)$$

the value of objective function is the same, and the contribution of this section is equal to the total change of the first difference, i.e.,

$$\sum_{t=1}^{t_3-1} |D_t| = H(t_4) - H(t_3) - [H(t_2) - H(t_1)] \tag{6}$$

Similarly, if the standard solution is concave, the objective function has the same value for any general concave solution and the contribution to the objective function is the negative of Eq. (6).

There are two possibilities to utilize the above properties. First, instead of using the standard solution we can use other solutions producing the same sum of the absolute values of the second differences. As indicated in Fig. 1, one solution is to continue the line through $H(t_1)$ and $H(t_2)$, and the line through $H(t_3)$ and $H(t_4)$ to the point where these lines intersect. For convex sections, this curve is the smallest feasible curve having the same value for the objective function, and for concave sections this is the largest curve. Note that the standard solution provides the other extreme feasible solution. The average between these extreme solutions can be taken to be a compromise solution. Another compromise solution can be obtained by putting a small penalty to the sum of the absolute values of third differences. This will force the second differences to change in smaller step than in the standard solution or in the solution obtained by continuing the lines. Appendix 4 describes how this can be implemented.

Second, we can get a smaller LP problem utilizing the fact that the standard solution is a solution to the original problem. We can reformulate the

problem by considering only ages t such that there is a crosscut just below or just above t . This formulation is given in Appendix 5. It is possible to implement the constraints that force the crosscuts to occur on the average in the middle of height increments also in this formulation.

5 Test Results

The Carmean, Issa, and LP methods were compared using simulated data as no real data with accurate height measurements were available. Using simulated data it can be demonstrated how the relations between different methods are dependent on the smoothness of height growth and on the bolt length. Height curves are smooth if there is little random variation in height growth or the autocorrelation is large. High autocorrelation and large variance lead to curves that are locally smooth but have irregular fluctuations.

It was assumed that the expected height curve is the following form of the Chapman-Richards function:

$$E(H_t) = \mu(t) = a(1 - e^{-bt})^c \tag{7}$$

where the asymptote parameter a was assumed to be 20 m, b was 0.1 and c was 2.

It was further assumed that height increments have a multiplicative error term:

$$H_t - H_{t-1} = \varepsilon_t (\mu(t) - \mu(t-1)) \tag{8}$$

where ε has log-normal distribution with expected value 1, i.e.

$$\varepsilon_t = \exp\left(e_t - \frac{1}{2}\sigma^2\right) \tag{9}$$

where e_t is normally distributed with mean zero and variance σ^2 (recall that if e is normally distributed with mean μ and variance σ^2 , then expected value of $\exp(e)$ is $\exp\left(\mu + \frac{1}{2}\sigma^2\right)$ (Johnson and Kotz 1970). Parameter σ can be interpreted as the relative standard deviation.

LP problems were solved using bqpdp-software of Prof. R. Fletcher, University of Dundee.

The behavior of different methods for short bolts

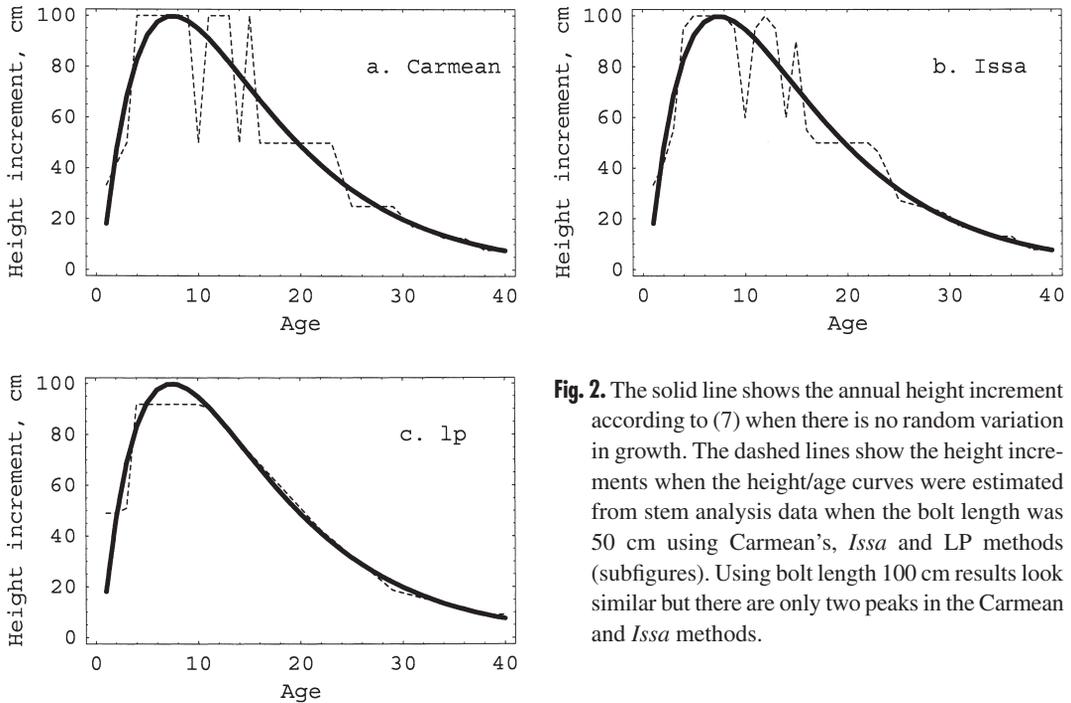


Fig. 2. The solid line shows the annual height increment according to (7) when there is no random variation in growth. The dashed lines show the height increments when the height/age curves were estimated from stem analysis data when the bolt length was 50 cm using Carmean's, *Issa* and LP methods (subfigures). Using bolt length 100 cm results look similar but there are only two peaks in the Carmean and *Issa* methods.

and smooth growth is illustrated in Fig. 2. Table 1 shows the test results when σ gets values 0, 0.1 and 0.3, and the autocorrelation r between e_t and e_{t-1} is 0.2 or 0.4. The results are averages over 6000 trees generated using the assumed random fluctuations (except only one curve was used when $\sigma=0$). The LP method was applied without constraints for the crosscuts, and with one overall constraint, and having three constraints for different (overlapping) parts of stem. For each case, the results were computed both with and without the additional penalty for third differences. Tables 2 and 3 collect information from Table 1 so that differences between methods are easier to notice. The principal differences between methods were:

Second Differences

The true second differences are larger the larger is the error deviation σ . For given σ , the larger is the autocorrelation, the smaller are the second differences. For the estimation methods, the sum of absolute second differences is naturally smallest for the LP method and largest for Carmean's

method. For each method the second differences are smaller for the long bolts.

Bias in Predicted Heights

The bias in the estimated heights (estimated-true) is generally very small. Linear programming gave the largest bias (almost 10 cm) when the height growth is completely regular ($\sigma=0$) and the bolt is 200 cm. This is probably because results were computed just for one single curve. There are no differences that could be easily interpreted. If the bias would be computed separately for the convex and concave parts of the height/age curves, linear programming methods might have some patterns in the bias for heights.

Standard Deviation of the Prediction Errors for the Height

For short bolts and regular growth, linear programming provides the best results with respect to predicting heights. When the bolt size and irregularity of height growth start to increase,

Table 1. The height/age curves were predicted with the Carmean’s method (Carm), with the *Issa* method and with the linear programming (LP) method using different bolt lengths. Six variants of the LP method were used. Variant ‘LP1’ is the basic method described in Appendix 2 (LP1). Variant ‘LP1b’ the basic method together with the additional penalty for third differences (see Appendix 4). Variant ‘LP2’ contained additional constraint forcing the average of crosscuts to be the middle the height increments (see Appendix 3). Variant ‘LP2b’ included the penalty for third differences. In variant ‘LP3’ the constraint for average location of crosscuts was done for ranges [0,0.4*H*] , [0.3*H*,0.7*H*] and [0.6*H*,*H*] where *H* is the total height of the stem. Variant ‘LP3b’ contained again the penalty for the third differences. Six thousand ‘true’ curves were simulated using Eqs. (7)–(9) and with different values for standard deviation σ and for the autocorrelation *r* (when $\sigma=0$, only one curve was used). \bar{D} denotes the average absolute second difference both for the ‘true’ curve and for the predicted curves, ‘hbias’ denotes the average value for predicted height–true height, ‘hsd’ denotes the standard deviation of predicted height–true height, and ‘isd’ denoted the standard deviation of predicted annual height increment–true increment (as total height is known, mean of increment errors is always zero).

	Bolt=25 cm				Bolt=50 cm				Bolt=100 cm				Bolt=200 cm			
	\bar{D}	hbias cm	hsd cm	isd cm	\bar{D}	hbias cm	hsd cm	isd cm	\bar{D}	hbias cm	hsd cm	isd cm	\bar{D}	hbias cm	hsd cm	isd cm
$\sigma=0, \bar{D}=4.45$																
Carm	8.70	0.0	5.2	7.2	9.20	0.7	11.6	12.2	6.47	2.2	19.2	11.6	5.06	-0.4	22.2	12.1
<i>Issa</i>	7.99	-0.1	4.9	6.6	7.66	0.1	10.6	9.3	5.90	1.1	18.8	9.7	3.76	-3.6	23.4	10.1
LP1	4.03	0.3	3.2	2.0	3.22	3.1	9.6	6.3	2.90	6.5	17.6	8.2	2.25	8.4	26.1	10.8
LP1b	4.03	0.2	3.1	2.0	3.22	3.7	9.8	6.3	2.90	6.6	17.0	7.4	2.25	6.1	25.0	10.5
LP2	4.03	0.0	3.2	2.0	3.22	1.5	10.6	6.3	2.90	4.6	18.5	8.2	2.25	3.8	25.8	10.8
LP2b	4.03	0.1	3.2	2.0	3.22	0.2	10.6	6.3	2.90	2.6	18.1	7.5	2.25	2.0	25.3	10.4
LP3	4.03	0.3	3.2	2.0	3.22	2.9	9.7	6.3	2.90	6.5	17.6	8.2	2.27	9.2	25.1	10.8
LP3b	4.03	0.3	3.2	2.0	3.22	3.1	9.6	6.3	2.90	6.5	17.6	8.2	2.27	9.9	25.4	10.7
$\sigma=0.1, r=0, \bar{D}=6.18$																
Carm	10.87	-0.2	6.0	8.2	12.60	0.0	10.6	13.5	8.05	0.0	15.7	12.0	4.46	-2.5	20.3	11.0
<i>Issa</i>	9.93	-0.4	5.7	7.5	9.77	-0.6	9.2	10.2	6.12	-1.4	15.1	9.4	3.65	-5.7	21.6	10.0
LP1	4.18	-0.6	6.3	4.6	3.33	0.7	12.8	8.1	3.09	0.8	18.9	8.9	2.50	1.0	27.7	11.8
LP1b	4.18	-0.4	5.1	4.4	3.33	1.5	11.0	7.7	3.09	2.3	16.9	8.4	2.50	1.8	25.6	11.1
LP2	4.18	-0.4	5.7	4.5	3.33	0.0	11.9	8.0	3.09	0.7	17.8	8.9	2.50	1.2	26.4	11.7
LP2b	4.18	-0.3	5.3	4.3	3.33	-0.1	11.5	7.7	3.09	0.8	17.2	8.5	2.50	1.3	25.8	11.2
LP3	4.18	-0.2	5.5	4.6	3.35	1.4	11.9	8.3	3.13	2.5	18.0	9.3	2.58	3.8	25.9	12.1
LP3b	4.18	-0.2	5.3	4.4	3.35	1.4	11.8	8.1	3.13	2.6	17.9	9.1	2.58	4.4	26.0	12.0
$\sigma=0.1, r=0.2, \bar{D}=6.77$																
Carm.	11.14	-0.2	6.0	8.2	12.78	0.0	10.6	13.5	8.29	0.1	15.9	12.3	4.51	-2.4	20.5	11.3
<i>Issa</i>	10.19	-0.4	5.7	7.5	9.91	-0.6	9.3	10.3	6.27	-1.3	15.3	9.6	3.67	-5.6	21.8	10.2
LP1	4.21	-0.6	6.4	5.0	3.35	0.7	12.8	8.4	3.10	0.9	18.9	9.2	2.50	1.2	27.7	12.0
LP1b	4.21	-0.4	5.3	4.9	3.35	1.5	11.1	8.0	3.10	2.3	16.9	8.7	2.50	1.9	25.7	11.3
LP2	4.21	-0.4	5.8	5.0	3.35	0.0	11.9	8.3	3.11	0.8	17.8	9.1	2.50	1.3	26.5	11.9
LP2b	4.21	-0.3	5.5	4.8	3.35	-0.1	11.6	8.0	3.11	0.9	17.2	8.7	2.50	1.4	25.8	11.4
LP3	4.21	-0.2	5.6	5.1	3.37	1.4	12.0	8.5	3.15	2.6	18.0	9.5	2.58	4.0	26.0	12.3
LP3b	4.21	-0.2	5.5	4.9	3.37	1.4	11.8	8.4	3.15	2.7	17.9	9.3	2.58	4.6	26.0	12.2

Table 1 continued.

	Bolt=25 cm				Bolt=50 cm				Bolt=100 cm				Bolt=200 cm			
	\bar{D}	hbias cm	hsd cm	isd cm	\bar{D}	hbias cm	hsd cm	isd cm	\bar{D}	hbias cm	hsd cm	isd cm	\bar{D}	hbias cm	hsd cm	isd cm
$\sigma=0.3, r=0.4, \bar{D}=13.28$																
Carm	15.50	-0.1	6.1	8.3	16.41	-0.2	10.6	13.8	10.99	-1.1	15.0	15.2	5.85	-4.1	20.6	15.1
Issa	14.38	-0.3	5.8	7.6	13.20	-0.7	9.5	11.2	8.24	-2.3	14.3	12.8	4.45	-6.7	22.0	14.4
LP1	5.99	-0.8	7.9	9.1	4.37	-0.1	13.9	12.3	3.50	-0.7	20.7	14.4	2.89	-2.2	28.8	16.2
LP1b	5.99	-0.5	7.6	8.9	4.37	0.6	13.2	12.2	3.50	0.7	19.7	14.3	2.89	0.6	26.8	16.0
LP2	6.00	-0.2	7.8	9.0	4.39	-0.2	13.4	12.2	3.55	-0.4	19.6	14.3	2.92	-0.8	27.4	16.2
LP2b	6.00	-0.2	7.7	8.9	4.39	-0.1	13.3	12.2	3.55	-0.2	19.5	14.2	2.92	-0.1	27.0	16.1
LP3	6.07	-0.1	7.5	9.0	4.50	0.7	12.9	12.3	3.70	1.0	18.7	14.4	3.04	1.4	25.9	16.4
LP3b	6.07	0.0	7.5	8.9	4.50	0.7	12.9	12.3	3.70	1.1	18.7	14.4	3.04	2.4	26.2	16.4
$\sigma=0.3, r=0.2, \bar{D}=15.47$																
Carm	17.14	-0.1	6.1	8.3	17.48	-0.2	10.7	13.9	11.57	-1.1	15.2	15.8	5.95	-4.1	21.1	16.5
Issa	15.95	-0.3	5.8	7.7	14.10	-0.7	9.7	11.5	8.62	-2.3	14.6	13.7	4.51	-6.7	22.4	15.9
LP1	6.66	-0.8	8.2	10.0	4.65	-0.1	14.3	13.7	3.59	-0.7	21.1	15.9	2.93	-2.2	29.2	17.6
LP1b	6.66	-0.6	8.0	9.9	4.65	0.6	13.7	13.6	3.59	0.7	20.1	15.8	2.93	0.6	27.2	17.5
LP2	6.68	-0.2	8.1	9.9	4.67	-0.2	13.8	13.5	3.65	-0.4	20.0	15.8	2.95	-0.9	27.8	17.6
LP2b	6.68	-0.1	8.1	9.8	4.67	-0.1	13.8	13.5	3.65	-0.2	19.9	15.7	2.95	-0.2	27.4	17.5
LP3	6.77	0.0	7.9	9.9	4.80	0.6	13.3	13.6	3.81	0.9	19.1	15.9	3.08	1.4	26.4	17.8
LP3b	6.77	0.0	7.9	9.9	4.80	0.6	13.3	13.6	3.81	1.0	19.1	15.8	3.08	2.4	26.8	17.8

Table 2. The relative MSE (hbias² + hsd²) of predicted heights for different bolt lengths, estimation methods and height curve parameters, results obtained from Table 1. For each bolt length and height curve set, the method with smallest MSE is scaled into one. The LP method is the method ‘LP1b’ in Table 1. Best method is written in bold.

Curve	Bolt=25 cm			Bolt=50 cm			Bolt=100 cm			Bolt=200 cm		
	Car	Issa	LP	Car	Issa	LP	Car	Issa	LP	Car	Issa	LP
$\sigma=0$	2.8	2.5	1.0	1.3	1.0	1.0	1.1	1.1	1.0	1.0	1.1	1.3
$\sigma=0.1, r=0.4$	2.0	1.2	1.0	1.3	1.0	1.5	1.1	1.0	1.3	1.0	1.1	1.4
$\sigma=0.1, r=0.2$	1.3	1.2	1.0	1.3	1.0	1.4	1.1	1.0	1.2	1.0	1.2	1.6
$\sigma=0.3, r=0.4$	1.3	1.0	1.7	1.2	1.0	1.9	1.1	1.0	1.9	1.0	1.2	1.7
$\sigma=0.3, r=0.2$	1.3	1.0	1.9	1.2	1.0	2.0	1.1	1.0	1.9	1.0	1.2	1.7

the Issa method provides the smallest standard deviations for height predictions. For long bolts, Carmean’s method provides smallest standard deviations.

Standard Deviations of the Prediction Errors for Height Increment

With respect to height increment prediction, the relations between methods are similar with

respect to height prediction. But the differences are more pronounced. For instance, for the bolt length 25 cm and $\sigma=0$, the error variance of Carmean’s method for the height is 2.8 times the error variance of the LP method. But for the height increment, the error variance of Carmean’s method is 13 times the error variance of the LP method. The LP method remains also optimal for longer bolts or for more irregular height growth. The Issa method is optimal for longer bolts or for

Table 3. The relative error variance (isd^2) of predicted annual height increments for different bolt lengths, estimation methods and height curve parameters, results obtained from Table 1. For each bolt length and height curve set, the method with smallest variance is scaled into one. The LP method is the method ‘LP1b’ in Table 1. Best method is written in bold.

Curve	Bolt=25 cm			Bolt=50 cm			Bolt=100 cm			Bolt=200 cm		
	Car	Issa	LP	Car	Issa	LP	Car	Issa	LP	Car	Issa	LP
$\sigma=0$	13.0	10.9	1.0	3.8	2.2	1.0	2.5	1.7	1.0	1.4	1.0	1.1
$\sigma=0.1, r=0.4$	3.5	2.9	1.0	3.1	1.8	1.0	2.0	1.3	1.0	1.2	1.0	1.2
$\sigma=0.1, r=0.2$	2.8	2.3	1.0	2.8	1.7	1.0	2.0	1.2	1.0	1.2	1.0	1.2
$\sigma=0.3, r=0.4$	1.2	1.0	1.4	1.5	1.0	1.2	1.4	1.0	1.2	1.1	1.0	1.2
$\sigma=0.3, r=0.2$	1.2	1.0	1.7	1.5	1.0	1.4	1.3	1.0	1.3	1.1	1.0	1.2

more irregular height growth than with respect to the standard deviation of height predictions.

Relations between Different Versions of the LP Method

There were two modification of the LP method that were tested. First, there were one or three additional constraints that forced the average crosscut to be in the middle of annual growth segment either for the whole stem or in three different segments, respectively. Second, there was a modification that minimized absolute values of the third differences. The differences between different versions of the LP method were small. The differences were logical in the sense that constraints decreasing the smoothness improved the results a little for long bolts and variable height growth.

6 Conclusions

The method presented in this paper extends the method of Fabbio et al. (1994) into its logical extreme by forcing the estimated height/age curve to be as smooth as possible by considering all second differences of heights simultaneously. The smoothness idea implemented in the method has certain intuitive appeal but if it is useful or not in practical estimation is dependent how stem analysis data are collected and for what purpose. If stem analysis is done accurately, i.e., using short bolts then it may be in balance with the overall measurement cost to estimate the height/age curves

also with this more complicated method. What bolt length is sufficiently short for making the LP method good is dependent on the smoothness of the height growth itself. So tests with empirical, very accurately measured data would be needed before practical recommendations can be done.

The presented method works better for annual height increments than for the accumulated height. This agrees well with the result of Fabbio et al (1994) that the *Issa* method compares well with Carmean’s method especially with respect to the annual increment. That annual height increment is accurately measured (estimated) can be important in studies where the annual increment is related to annual diameter increment or annual variation of climatic variables. From a statistical point of view, the height/age measurements obtained from stem analysis are always problematic as the measurements contain measurement errors which do not behave according to any simple stochastic model.

References

Carmean, W.H. 1972. Site index curves for upland oaks in the Central States. *Forest Science* 18(2): 109–120.
 DeBoor, C. 1978. *A practical guide to splines*. Springer, New York.
 Dyer, M.E. & Bailey, R.L. 1987. A test of six methods for estimating true heights from stem analysis data. *Forest Science* 33(1): 3–13.
 Fabbio, G., Frattegiani, M. & Manetti, M.C. 1994. Height estimation in stem analysis using second

- differences. *Forest Science* 40(2): 329-340.
- Johnson, N.L. & KOTZ, S. 1970. *Distributions in statistics. Continuous univariate distributions-1*. Wiley, New York.
- Lenhart, J.D. 1972. An alternative procedure for improving height/age data from stem analysis. *Forest Science* 18(4): 332.
- Luenberger, D.G. 1973. *Introduction to linear and nonlinear programming*. Addison-Wesley, Reading, Mass. 356 p.

Appendix 1. Symbols, and Carmean's and Issa methods

The symbols used for the stem analysis data are the same as in Dyer and Bailey (1987) with some necessary additions:

- n = total tree age
- H = total tree height
- I = number of crosscuts excluding ground level (=number of bolts)
- r_i = number of rings at the i th crosscut, $i=0, \dots, I, (r_0=n)$
- h_i = height at the i th crosscut, $h_0=0$
- k_i = number of inner rings in interval $[h_i, h_{i+1})$ (inner ring is a ring that disappears in $[h_i, h_{i+1})$)
- t_{ij} = age of the tree associated with the j th ring from center at i th crosscut, i.e. $t_{ij}=n-r_i+j$, (note that the ring is not necessarily inner ring)
- t_{i0} = $t_{i1}-1$, the age of last tip disappearing before h_i .
- H_t or $H(t)$ = tree height at age t , $t=0, \dots, n, H_0=0$.
- $F(t)$ = i if crosscut at h_i is the first crosscut in interval $[H_t, H_{t+1})$
- $L(t)$ = i if crosscut at h_i is the last crosscut in interval $[H_t, H_{t+1})$
- $b(t)$ = $h_{L(t)}-h_{F(t)}$, i.e., the total length of whole bolts in $[H_t, H_{t+1})$
- H_{ij} = tree height at age t_{ij} (H_{i0} is thus the last hidden tip in the previous bolt)
- C = the set of ages t such that $t \in C$ if and only if $[H_t, H_{t+1})$ contains at least one crosscut.

Additional symbols defined in the paper:

D_t forward second difference, defined in (1)

S_t forward third difference, defined in (D1)

Newberry, J.D.1991. A note on Carmean's estimate of height from stem analysis data. *Forest Science* 37(1): 368-369.

Zwillinger, D. 1995. (ed.). *The 30th edition of the standard mathematical tables and formulas*. CRC Press.

Total of 9 references

As can be seen from the definition of k_i , it is assumed that if $H_t=h_i$ then the ring of year t is noticed at the crosscut i .

Carmean's (1972) method

In Carmean's method the locations of hidden tips are computed according to the formulation of Dyer and Bailey (1987) as follows:

$$H_{ij} = h_i + (h_{i+1} - h_i) \sqrt{\frac{2(r_i - r_{i+1})}{(j-1)(h_{i+1} - h_i) / (r_i - r_{i+1})}} \quad (A1)$$

For hidden tips located in the first bolt this must be modified as (Fabbio et al. 1994)

$$H_{ij} = jh_i / (0.5 + r_i - r_{i+1}) \quad (A2)$$

and for hidden tips located in the last bolt as (Newberry 1991)

$$H_{ij} = h_i + (H - h_i) \sqrt{\frac{2(r_i - 0.5)}{(j-1)(H - h_i) / (r_i - 0.5)}} \quad (A3)$$

Issa method

In the Issa method (Fabbio et al 1994) the second (central) difference of ring counts is first defined for i th crosscut as

$$R_i = (r_{i-1} - r_i) - (r_i - r_{i+1}) = r_{i-1} - 2r_i + r_{i+1} \quad (A4)$$

The age of i th crosscut is calculated as

$$e_i = n - r_i + 1 / (2 + R_i) \text{ if } R_i \geq 0 \tag{A5}$$

$$e_i = n - r_i + 1 - 1 / (2 - R_i) \text{ if } R_i < 0 \tag{A6}$$

Thereafter the heights of hidden tips are computed as

$$H_{ij} = h_i + (h_{i+1} - h_i) (t_{ij} - e_i) / (e_{i+1} - e_i) \tag{A7}$$

Taking into account that $t_{ij} = n - r_i + j$ we note that if $R_i = 0$ Eq. (A7) is equivalent to (A1).

Appendix 2. LP Problem for Smoothest Possible Curve

A LP formulation for the problem described in section ‘The Smoothest Possible Curve’ is:

$$\text{Minimize } \sum_{t=0}^{n-2} (D_t^+ + D_t^-) \tag{B1}$$

subject to the following constraints:

$$D_t^+ - D_t^- - H_t + 2H_{t+1} - H_{t+2} = 0, t = 0, \dots, n-2 \tag{B2}$$

$$H_{t+1} - H_t \geq 0, t = 0, \dots, n-1 \tag{B3}$$

$$H_0 = 0 \tag{B4}$$

$$H_n = H \tag{B5}$$

$$H_{i0} \leq h_i - \varepsilon, i = 1, \dots, I \tag{B6}$$

$$H_{i1} \geq h_i, i = 1, \dots, I \tag{B7}$$

$$D_t^+ \geq 0, D_t^- \geq 0, t = 0, \dots, n-2 \tag{B8}$$

Constraints of form (B2) define the second order difference variables D_t^+ and D_t^- in terms of the heights H_t . Constraints (B3) require that the height over time is nondecreasing. Constraints (B4) and (B5) express the knowledge of the initial and total heights. This information could be also substituted directly into constraints (B2) and (B3) for $t=0$, and $t=n-1$. Constraints (B6) and (B7) express the information provided by the crosscuts. Term ε in (B6) is a small positive constant used in accordance to the assumption that a ring ending exactly at the crosscut will be noticed. This way we can also directly tell whether the obtained estimate for H_t

is at the lower or upper end of the feasible range, and there will be no ambiguity if the estimated height/age curve is turned back to rings data. Note that H_{i0} and H_{i1} are for each crosscut i among variables H_0, \dots, H_n . If there are bolts without hidden tips, then some of constraints are redundant (e.g. they can include constraints $H_{22} \geq 10$ and $H_{22} \geq 11$, so that constraint $H_{22} \geq 10$ could be removed). Nonnegativity constraints (B8) guarantee, for the given objective function, that for each t either D_t^+ or D_t^- will be $|D_t|$. Nonnegativity of heights H_t is already guaranteed by (B3) and (B4). Constraints (B3) are redundant for such values of t that H_t and H_{t+1} are in different bolts (increasing heights are guaranteed by constraints B6 and B7). The problem is an ordinary linear programming problem that can be easily solved.

Appendix 3. Crosscuts in the Middle of Annual Increment

The crosscut at h_i occurs in the middle of the annual increment if

$$\frac{1}{2} (H_{i0} + H_{i1}) = h_i$$

or

$$H_{i0} + H_{i1} = 2h_i \tag{C1}$$

If both the bolt below h_i and the bolt above h_i contains at least one hidden tip, we could force crosscut i to be in the middle of the annual increment by appending constraint (C1) to the constraints (B2)-(B8). The crosscut can be forced to occur, in the average, in the middle of annual increment for a stem segment, if we take the average of (C1), or equivalently, by summing up both sides:

$$\sum_{i=L_s}^{U_s} (H_{i0} + H_{i1}) = 2 \sum_{i=L_s}^{U_s} h_i \tag{C2}$$

where L_s is the first crosscut in the segment s and U_s is the last. Different segments can overlap.

The constraint (C2) is well defined also in case some bolts do not contain any hidden tips. However, in that case $H_t + H_{t+1}$ for the same t appears several times in (C2), and such t may get too much weight (when computing coefficient of H_t , all occurrences are counted).

The constraint (C2) can then be modified as follows. If there are several crosscuts within the same interval $[H_t, H_{t+1})$ (or equivalently, interval $[H_{i0}, H_{i1})$ is the same for several crosscuts i) we can not really assume that each h_i would occur randomly within the interval. Rather we may expect that interval $[h_{F(t)}, h_{L(t)}]$ between the first and last crosscut occurs randomly within $[H_t, H_{t+1})$, or equivalently, that $h_{F(t)}$ is randomly located within $[H_t, H_{t+1} - b_t)$, where $b_t = h_{L(t)} - h_{F(t)}$. Height $h_{F(t)}$ occurs in the middle of $[H_t, H_{t+1} - b_t)$ if

$$\frac{1}{2}(H_t + H_{t+1} - b_t) = h_{F(t)}, \text{ or}$$

$$H_t + H_{t+1} = 2h_{F(t)} - b_t \quad (\text{C3})$$

The constraint saying that the crosscuts occur, on the average, in the middle of their possible ranges can be stated:

$$\sum_{t \in C} (H_t + H_{t+1}) = \sum_{t \in C} (2h_{F(t)} - b_t) \quad (\text{C4})$$

where the summation is over all t such that there is a crosscut (belonging to a desired stem segment) in $[H_t, H_{t+1})$. The same H_u can still appear in (C4) twice, first as H_{t+1} for $t = u - 1$ and then as H_t for $t = u$. When applying the above formulas, the crosscut at h_1 is counted as the first crosscut of the stem.

Appendix 4. Additional Smoothness Using Third Differences

Third forward difference for heights is defined by

$$S_t = H_{t+3} - 3H_{t+2} + 3H_{t+1} - H_t \quad (\text{D1})$$

Using different variables for positive S and negative S we get LP constraints:

$$(D2)$$

$$S_t^+ - S_t^- - H_{t+3} + 3H_{t+2} - 3H_{t+1} + H_t = 0, t = 0, \dots, n-3$$

where both S_t^+ and S_t^- variables are required to be nonnegative.

Adding constraints (D2) to the problem given in the Appendix 2 does not change the solution at all because S_t^+ and S_t^- will just get such values that $S_t^+ - S_t^-$ is the value of the third difference in the original solution. We must change the objective row (C1) into:

$$\text{minimize } \sum_{t=0}^{n-2} (D_t^+ + D_t^-) + \lambda \sum_{t=0}^{n-3} (S_t^+ + S_t^-) \quad (\text{D3})$$

If λ is a small positive constant then the heights around crosscuts and the sum of absolute second differences do not change but the heights of the inner hidden tips do change so that the height/age curve is between the extreme possible curves (see Fig. 1). By giving λ greater values one can also get smooth, and reasonable, height/age curves but the smoothness criterion does not correspond to the standard interpretation.

Appendix 5. LP Formulation without Inner Hidden Tips

Because a solution that is linear between the first and last hidden tip in a bolt has the minimum sum of absolute second differences, we can formulate the LP problems so that the height of the inner hidden tips are not considered. The heights at the inner tips are included in the definitions of the second differences, so we cannot just drop constraint rows of type (B2) but we must express the heights of innermost tips using the heights of first and last tips and the linearity condition. If t is the age of an inner tip disappearing in bolt i , then

$$H_t = \frac{l(t) - t}{l(t) - f(t)} H_{f(t)} + \frac{t - f(t)}{l(t) - f(t)} H_{l(t)} \quad (\text{E1})$$

where $f(t)$ is the age of the first tip disappearing in the same bolt as H_t , and $l(t)$ is the age of the last tip disappearing in the same bolt as H_t .

Then the LP problem given in Appendix 2 can be modified so that:

- Remove all such constraint (B2) for which $t+l$ is an inner tip, and remove the corresponding D terms also from the objective row (B1).
- Reformulate the constraints requiring that heights are increasing by constraints that require that the height of last tip in a bolt is greater than the height of the first tip.
- For such D terms that are utilizing heights of the inner tips use (E1) to replace them in (B2) with the height at first or last tip in the same bolt.