Smoothed Prediction of the Onset of Tree Stem Radius Increase Based on Temperature Patterns

Mikko Korpela[1], Harri Mäkinen[2], Mika Sulkava[1],
Pekka Nöjd[2], and Jaakko Hollmén[1]

[1] Helsinki Institute for Information Technology,
Helsinki University of Technology, Department of Information and
Computer Science, P.O. Box 5400, FI-02015 TKK, Espoo, Finland
http://ics.tkk.fi
[2] Finnish Forest Research Institute (Metla), Vantaa Research Unit,
P.O. Box 18, FI-01301 Vantaa, Finland
http://www.metla.fi

Abstract. Possible changes of the growing season of trees would have significant consequences on forest production. Predicting the onset of tree growth on the basis of climate records can be used for estimating the magnitude of such changes. Conventional methods for estimating the onset of tree growth use cumulative temperature sums. These estimates, however, are quite coarse, and raise questions about making better use of the weather information available. We approach the problem of predicting the onset of tree growth with a predictor based on a combination of a k-nearest neighbor regressor and a linear regressor. The inputs are weighted sums of daily temperatures, where the weights are determined by a subset of Bernstein polynomials chosen with a variable selection methodology. The predictions are smoothed for consecutive days to give more accurate results. We compare our proposed solution to the more conventional approach. The proposed solution is found to be better.

1 Introduction

The amount and properties of wood produced are related to the timing and rate of wood formation during the growing season. Wood formation depends on genetical signaling, availability of resources for growth, temperature, tree water and nutrient status, and the stage of ontogenic development (e.g. [1,2,3]).

Despite the basic nature of the wood formation process, our present knowledge concerning the timing and rate of the phases of tree growth is surprisingly fragmentary. An important factor behind the gaps of knowledge is the difficulty in measuring xylem formation at short intervals. Point and band dendrometers have been used traditionally to monitor cambial dynamics (e.g. [4,5]). Changes in stem dimensions are not solely a result of wood formation; they are often caused by other processes, especially changes in stem hydration [6,7,8,9]. Since dendrometers measure stem radius or circumference changes rather than wood formation, it is difficult to distinguish between true wood formation and hydrological swelling and shrinking.

In our previous study, we evaluated the use of cumulative sum (CUSUM) charts for automatically determining the onset and cessation dates of radial increase based on dendrometer data [10]. The CUSUM chart is a tool for automatically detecting small changes in the mean of a signal solely based on the data. In order to produce reliable results, one has to choose suitable parameter values for the chart and onset and cessation levels for stem radius. Once configured properly, the method produced results similar to those determined by an expert. This study is a sequel to [10].

Our aim in this paper was to provide new insights into statistical methods that could be used for predicting the onset of radial increase based on weather data. During dormancy release and growth initiation in the boreal zone, the role of temperature is more important than the role of photoperiod or water availability [11,12,13]. Therefore, we compared different statistical methods to evaluate whether the expressed signal between temperature and the onset of radial increase could be strengthened by selection of appropriate method.

In this paper, we are dealing with two stochastic processes: one concerning the environmental factors, specifically temperature, and the other concerning the radial growth of trees. Our interest lies in modeling the relationship between these processes. We control the complexity of our research problem by reducing the growth process to a single number. Thus, our goal is to predict the onset date. Similarly, the complexity in the temperature data is controlled by computing features from a reasonably long sequence of past data. A subset of the features is selected based on their performance in the prediction task.

The rest of the paper is organized as follows: Section 2 introduces the methods used, from traditional and new ways of summarizing temperatures, through variable selection, to the prediction machinery. Section 3 describes our results. Section 4 concludes the paper with a summary, some comments, and future directions.

2   Material and Methods

2.1   Traditional Temperature Sums

One of the oldest ecological concepts is the "degree-day" or "physiological time" unit which mainly stems from the relationship between development rate of many organisms and temperature (e.g. [14]). Phenological phases of plants have been found to be related to accumulation of air temperatures above some threshold below which the development stops.

Most attempts to improve the concept have dealt with either different calculation procedures used to obtain degree day estimates [15,16,17], or including additional climatic parameters, such as soil water availability (e.g. [18]). We will limit ourselves to the modification of the method [19] commonly used in Finland, which computes degree-day sum as the cumulative sum of the differences between the daily mean temperature and the threshold of $+5\,^{\circ}\mathrm{C}$ (negative values

are set to zero). At least five consecutive days meeting or exceeding the threshold are needed in order for the degree-days to accumulate from a given day.[1]

2.2   Bernstein Polynomials as Basis Functions on the Past

Instead of reducing the sequence of previous daily temperatures into just one value, we can compute several temperature features. Thus by making more effective use of temperature data, we hope to improve our ability to predict the onset. For this purpose, we use Bernstein polynomials [20]:

$$B_i^d(x) = \binom{d}{i} x^i (1-x)^{d-i} , \quad i = 0, \ldots, d. \tag{1}$$

Bernstein polynomials have a special property: for any $d \in \{0, 1, 2, \ldots\}$, the polynomials of degree d sum to unity:

$$\sum_{i=0}^d B_i^d(x) = 1 \quad \forall \, x \in \mathbb{R}. \tag{2}$$

When limited to the range $x \in [0, 1]$, the Bernstein polynomials are non-negative. The polynomials are symmetric, $B_i^d(x) = B_{d-i}^d(1-x)$, and have roots at both 0 and 1, except that $B_0^d(0) = B_d^d(1) = 1$. They also have a unique maximum value in the range, at $x = i/d$ for $B_i^d(x)$. Some examples of the polynomials in the limited range are shown in Fig. 1.
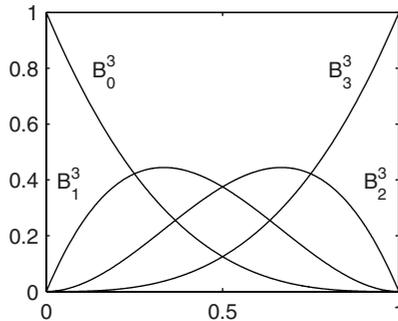
Fig. 1. The Bernstein polynomials of degree 3

The polynomials have an application related to Bézier curves [20]. We use them as weight functions in temperature features. Instead of summing temperatures above $+5\,^\circ$C as in the traditional temperature sum, we use the daily temperature values even if they are lower than the threshold. Thus, when we

---

[1] When computing the temperature sum every day in an online setting, this means a delay of up to four days, i.e. the system is not causal.

consider daily mean temperatures T_i and a history window of N previous days, the Bernstein-weighted temperature features of degree d on day i are

$$s_m(i) = \sum_{j=1}^{N} B_m^d \left(\frac{j-1}{N-1} \right) T_{i-j} \ , \quad m = 0, \dots, d \ . \tag{3}$$

Since the Bernstein polynomials sum to one, each day in the temperature sequence gets equal weight, when we consider the combination of all temperature features (3). It is clear that each temperature sum s_m in (3) has a quite strong correlation with some s_n, at least for m and n that are close to each other. In the next section, we discuss selecting a subset of the temperature features. As a result of the selection, some days – or positions in the sequence of temperatures of N previous days – will have more weight than others.

2.3   Selecting Basis Functions

We prefer relatively sparse or parsimonious models, because they are typically easier to interpret and apply than more complex ones [21]. For selecting a suitable subset of temperature features defined in (3), we employed Best First Search (BFS) [22]. Simply testing all feature sets is infeasible due to the size of the search space.

BFS is a wrapper method, i.e. it uses the machine learning algorithm as part of the feature set search. In other words, the goodness of a feature set is defined as its performance on the machine learning task at hand. The basic idea of the algorithm is to advance in the state space, where a state represents a feature set. New states are created by adding or removing one variable (simple operator) or combining several additions or removals (compound operator) (Fig. 2). The next state chosen is the one which has the lowest cost among the candidate states. The search may start from any state. Typical choices are the empty set or the full set of variables. The search proceeds until no major improvements to the cost have been observed for a while. The exact behavior of the algorithm is controlled with two user-definable parameters: one for setting the threshold for considering a new state an improvement, the other for adjusting the stopping criterion.

2.4   Smoothed Prediction with a Combination of a Parametric and a Non-parametric Model

The k-nearest neighbor method (k-NN) is a traditional method used in classification [23], but it can also be used for regression in an analogous way. We employed a combination of a k-NN regression model and a linear regression model to predict the number of days until onset of radial increase. The aggregate prediction is simply a weighted mean of the two parts. Temperature features (3) were used as inputs and the number of days until the onset of radial increase as the target variable. Figure 3 shows an example of the output of the predictor machine, when predictions are made as a sequence, on consecutive days.

The k-NN part of the predictor finds k days of the training set (reference database) that are closest to the query day with respect to the temperature features. The prediction of the k-NN part is effectively the mean of the target variable on the neighboring days. When s is the vector of selected temperature features (3) on the query day, and s_i contains the corresponding values on reference day i, the k closest days are found by computing the Euclidean distances (4) (or equivalently squared distances) over all the reference data, sorting
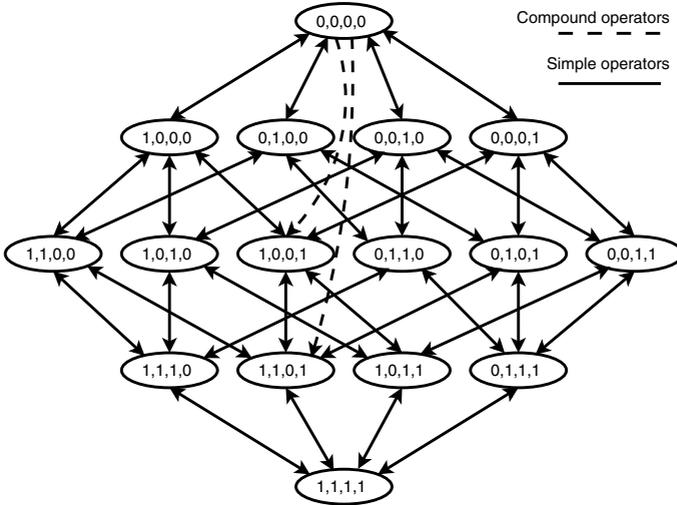
Fig. 2. Structure of state space in BFS, when selecting from four possible variables and creating new nodes using both simple and compound operators. Each state is a set of input variables (1 = "variable present", 0 = "variable absent"). Arrow from node a to node b means that b is a child of a.
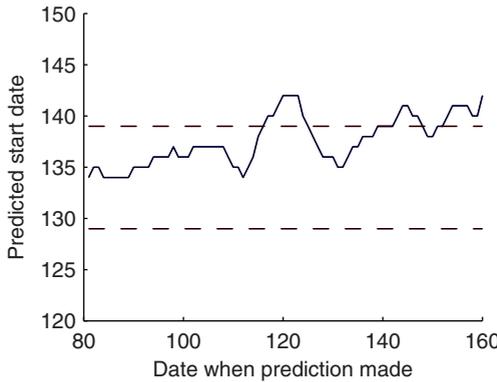
Fig. 3. Example output of the prediction machine. The measured onset region, determined by the minimum and maximum onset dates of the individual trees, is marked by the dashed lines.

the distances, and choosing the indices i that correspond to the k smallest distances. Due to the non-parametric nature of the k-NN part, the training data set becomes a part of the model, and is always needed when new predictions are made.

$$d_i = \sqrt{(\boldsymbol{s} - \boldsymbol{s}_i)^T(\boldsymbol{s} - \boldsymbol{s}_i)} \tag{4}$$

The other part of the composite predictor is a simple linear regression on selected temperature features (3), with time until onset of radial increase as the target variable. When $\boldsymbol{s}$ and $\boldsymbol{\beta}$ are the chosen temperature features (accompanied by a constant term) and regression coefficients, respectively (both column vectors), the regression model is as in (5). The regression result y_r is rounded to the closest integer, which may be considered superfluous, because the final output of the combined model is also rounded.

$$y_r = \boldsymbol{s}^T\boldsymbol{\beta} \tag{5}$$

In an ideal situation, the predictor would always point to the same onset date, independent of the day on which the prediction is made. When the predictions are made as a sequence, the predicted onset date will, however, fluctuate. We did a simple smoothing of the predictor's output by taking a weighted mean of the new prediction and the previous one. Then the effect of old, unsmoothed predictions, undergoes exponential decay as the prediction sequence is extended. The smoothing operates with unrounded values, but the final output is rounded to a precision of one day. Since the predictor actually gives the time until onset, the number of days separating the previous and current predictions is subtracted from the previous result before it is used as part of the new prediction. The smoothing process (without integer rounding) can be written as

$$\begin{aligned}\tilde{y}_i &= \alpha y_i + (1 - \alpha)\tilde{y}_{i-1}\\ &= \alpha[y_i + (1 - \alpha)y_{i-1} + \ldots + (1 - \alpha)^{i-2}y_2] + (1 - \alpha)^{i-1}y_1 ,\end{aligned} \tag{6}$$

where $\alpha \in (0, 1]$ adjusts the level of smoothing, $\tilde{y}_i$ is the i:th smoothed prediction in the prediction sequence, and y_i is the corresponding "novelty", the new part in the prediction. Note that the first prediction in the sequence is a special case, where no smoothing is applied.

2.5   Dendrometer Data

Sample trees were selected at two sites located 300 m from each other in Tuusula, southern Finland. In the first stand, the Norway spruce trees were growing in a pure spruce stand on fertile mineral soil classified as *Oxalis-Myrtillus* forest type [24]. Mean stem diameter of the sample trees at breast height was 27 cm and relative crown length was 68 %. The sample trees were monitored during the growing seasons of 2001–2005. In the second stand, four Norway spruce and four Scots pine trees were monitored during the growing seasons of 2002–2003, and another four spruce and four pine trees during the growing seasons of 2004–2005. They were growing in a mixed spruce-pine stand on a relatively fertile mineral

soil classified as *Myrtillus* forest type [24]. The total number of observations (year × tree combinations) was, thus, 57. The stands and sample trees have been described in detail in [10].

Stainless-steel band-dendrometers were installed on each tree at a height of about 2 m. Changes in tree girth were measured at a resolution of 0.1 mm, corresponding to diameter change of about 0.03 mm. The output of the dendrometers was stored as 1-h averages. From these measurements, the daily values of stem circumference were calculated as the mean of hourly values and the circumference changes were converted to radial changes assuming a circular stem cross-section. The dendrometer has been described in detail in [5].

For each year and tree, the onset date of radial increase was determined visually and verified by the CUSUM method [10]. Depending on the year, we have the onset date of either 5 (year 2001) or 13 (2002–2005) trees. Because we wanted to evaluate the methods described above, i.e. detailed biological interpretation of the increment onset is beyond the scope of this study, we averaged the data from both sites and tree species. As mentioned above, swelling and shrinkage of the tree, caused by changes in water level, add noise to the data.

Temperature data at 3-hour intervals were obtained from a meteorological station of the Finnish Meteorological Institute located about 5 km from the study stands. From the temperature measurements, the average daily values were calculated as arithmetic means.

2.6   Test Setting

We run feature selection with BFS, selecting from a total of 40 or 20 temperature features weighted with Bernstein polynomials, corresponding to all polynomials of degree 39 or 19, respectively. The search is started from the empty set of features. Years 2001–2004 are used for the feature selection in a 4-fold cross-validation setting, where the onset date on one year is predicted based on data from the remaining three years. Year 2005 is reserved as a (very limited) test set for later use. From each year, days 81–200 are included, counting from the start of the year. The temperature features are computed from 80 previous days, i.e. the temperatures on days $\{x - 80, \ldots, x - 1\}$ are reduced to 40 or 20 numbers to be associated with day x.

The feature selection also incorporates the selection of k for k-NN. In our setting, the cost function of BFS tests the prediction machine with values of k ranging from 5 to 50 in steps of 5. The cost is the MSE of all predictions, where the predictions are made for each date in the range of 101–160 (April 11–June 9 in the case of "no leap year"). For comparison, the measured annual average onset dates range from 133 to 144. The average was taken over all the years; there are $4 \cdot (160 - 101 + 1) = 240$ error values to be averaged. In the feature selection phase, the prediction sequence is not smoothed: $\alpha = 1$ in (6). For later tests, $\alpha = 0.4$ was used, and the predictor was started at day 81, which allowed a stabilizing period of 20 days before the measurement of prediction error. The value of α was chosen by non-rigorous testing towards the goal of minimizing error. The results are not sensitive to the specific choice of this parameter.

3   Results

The traditional temperature sum alone does not provide an accurate prediction about the onset date (Fig. 4). When excluding the smallest values of temperature sum, the average onset date over different years could quite well be explained by a straight line. The variance between years, however, is large.

The performance of the Bernstein-weighted temperature features is presented in Table 1. The bottom part of the table lists comparative results, with no feature selection or using just the traditional temperature sum. In these cases, the k range tried was 10 to 110 in steps of 5. RMSE 1 means the square root of the error measure used in the feature selection tests: the mean of squared errors in the predictions sequences. For RMSE 2, each annual prediction sequence is reduced to a single error measure, where the onset date given by the predictor is set to be the first date in the sequence, where the predicted onset date and the date of prediction intersect, i.e. according to the predictor, growth has already started. The error is derived from the difference of this intersection date and the date derived from the girth band data. When only predicting the onset on one year, RMSE 2 would reduce to the absolute value of this error, but in the table, we also show the sign of the error. "Valid." is the error in the validation tests used for selecting the features and k. "Test" and "Pseudo" are errors done with the selected features and k, using data from all the year (2001–2005). In the former case, only the year 2005 is predicted with the other years serving as reference data. In the latter case, all the years are predicted one by one, with the remaining years as reference data.

With our settings for BFS, the number of states evaluated was about 8000 out of 2^{20} or 25000–35000 out of 2^{40}. The search was set to stop, when the 500 latest state expansions have not resulted in a cost smaller than 1.001 times the best known cost.
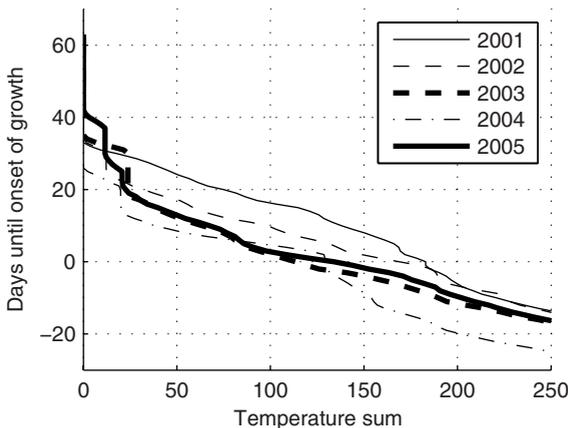
Fig. 4. The number of days until the onset of radial increase plotted against temperature sum in different years

Table 1. Prediction error. The k-NN component was always used. Also linear regression was used in some tests (see first column). The bottom part of the table shows comparative results. The column marked with $*$ holds the signed absolute error instead of RMSE.

	#Features			RMSE 1 (sequence)			RMSE 2 (i.sect)		
Model	Chosen	Total	k	Valid.	Test	Pseudo	Valid.	Test$*$	Pseudo
lin $+$ k-NN	16	40	5	5.2	5.0	4.9	6.4	5	5.8
k-NN	16	40	35	5.3	5.2	5.0	5.5	-2	4.5
lin $+$ k-NN	8	20	5	5.4	6.4	5.3	5.1	6	5.5
k-NN	9	20	25	5.4	3.5	4.8	6.6	5	6.0
lin $+$ k-NN	40	40	90	5.7	5.9	5.2	4.9	2	4.4
lin $+$ k-NN	20	20	105	5.8	6.0	5.2	5.2	2	4.4
k-NN on temperature sum			50	8.5	2.5	7.0	8.7	1	6.9
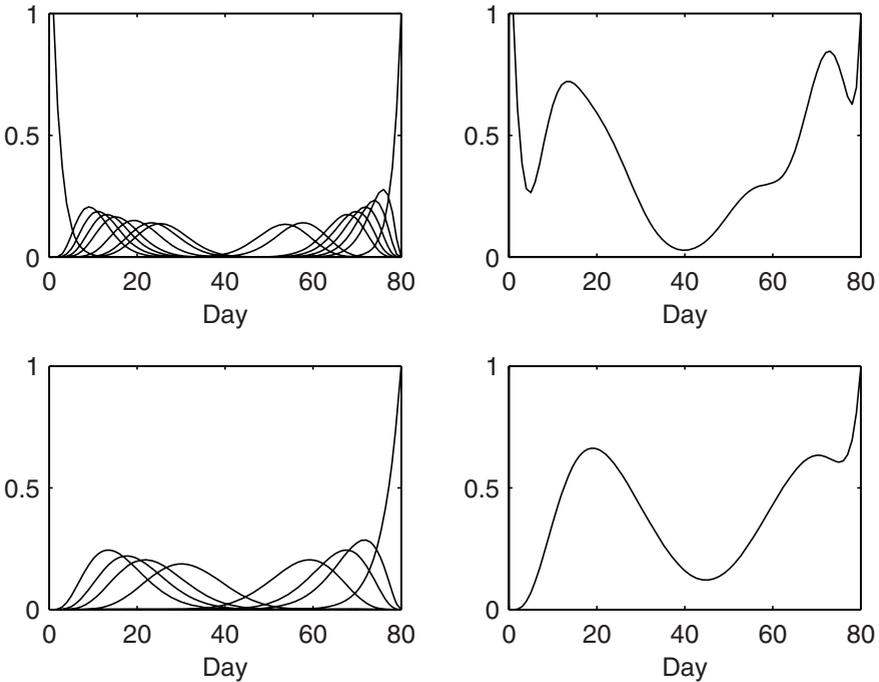
Fig. 5. Bernstein weighting functions chosen in rows 1 and 3 of Table 1. In the figures, the most distant past is at the right end of the x-axis. Top row: 16/40 features. Bottom row: 8/20 features. Left column: individual weighting functions. Right column: sum of weights on the left.

We look at the "RMSE 1" results first, because that is the optimization criterion in the validation tests. The tests, which average the error over several years, show that the Bernstein polynomial temperature features produce prediction accuracy superior to the traditional temperature sum. Adding one year

of data always improved the average performance. The results from the prediction of a single year's growth (column "Test") show that although close to each other on the average, the different predictor settings may give quite different results for any individual year. Note that the "truth" used as the basis of the error measure is quite uncertain: when considering the onset dates suggested by each girth band on a given year, the standard deviation ranges from about 3.7 to 9.0, averaging at 6.8 (days). The error measure "RMSE 2" shows generally similar results in the sense that the traditional temperature sum is outperformed by the new features. Also here the variation in a single year's results is clearly seen.

The optimal k decreases noticeably, when linear regression is also used in the predictor machine. It seems that the addition of the simple regression model, quite logically, regularizes the output of the k-NN model. A similar effect is achieved with an increased k. When feature selection is omitted (the first two rows in the bottom part of Table 1), the optimal k is very large despite the presence of linear regression.

Figure 5 shows the features chosen by BFS in two of our tests. The results look almost the same. In both cases, some highly correlated features are included, but features concentrated in the middle part of the history window are curiously missing.

4   Summary and Conclusions

In this paper, we presented a method for predicting the onset of stem radius increase. The method is based on temperature features weighted with Bernstein polynomials, and a combination of two simple regression methods operating on those features. A subset of the features was selected with BFS. The motivation for the development of the features was to extract more information from the temperature time series than what is possible with the traditional temperature sum, thus improving prediction performance.

In our cross-validation tests, the proposed temperature features outperformed the traditional temperature sum. However, prediction accuracy still shows a large year-to-year variation. There are at least two plausible reasons for the variation: the nature of the dendrometer data (swelling, shrinking), and factors left out of the model (everything but temperature).

After these initial steps with the prediction framework, we can think of applications for it. For example, we could try to find a possible trend in the past onset dates. However, the potential differences between the current and past environment would have to be considered: have the environmental factors not included in the model changed so much over the years, that a model developed using present data is not valid for the past? Another application possibility is the modeling of phenomena other than the onset of stem radius increase.

References

1. Fritts, H.C.: Tree Rings and Climate. Academic Press, London (1976)
2. Savidge, R.A.: Xylogenesis, genetic and environmental regulation. IAWA Journal 17, 269–310 (1996)
3. Vaganov, E.A., Hughes, M.K., Shashkin, A.V.: Growth Dynamics of Conifer Tree Rings. Ecological Studies, vol. 183. Springer, Heidelberg (2006)
4. Yoda, K., Suzuki, M., Suzuki, H.: Development and evaluation of a new type of opto-electronic dendrometer. IAWA Journal 21(4), 425–434 (2000)
5. Pesonen, E., Mielikäinen, K., Mäkinen, H.: A new girth band for measuring stem diameter changes. Forestry 77(5), 431–439 (2004)
6. Kuroda, K., Kiyono, Y.: Seasonal rhythms of xylem growth measured by the wounding method and with a band-dendrometer: an instance of Chamaecyparis obtusa. IAWA Journal 18, 291–299 (1997)
7. Deslauriers, A., Morin, H., Urbinati, C., Carrer, M.: Daily weather response of balsam fir (*Abies balsamea* (L.) Mill.) stem radius increment from dendrometer analysis in the boreal forests of Québec (Canada). Trees 17(6), 477–484 (2003)
8. Mäkinen, H., Nöjd, P., Saranpää, P.: Seasonal changes in stem radius and production of new tracheids in norway spruce. Tree Physiology 23(14), 959–968 (2003)
9. Mäkinen, H., Seo, J.W., Nöjd, P., Schmitt, U., Jalkanen, R.: Seasonal dynamics of wood formation: a comparison between pinning, microcoring and dendrometer measurements. European Journal of Forest Research (in print, 2008)
10. Sulkava, M., Mäkinen, H., Nöjd, P., Hollmén, J.: Detecting onset and cessation of tree stem radius increase based on dendrometer data. Environmetrics (submitted, 2008)
11. Partanen, J., Leinonen, I., Repo, T.: Effect of accumulated duration of the light period on bud burst in Norway spruce (Picea abies) of varying ages. Silva Fennica 35(1), 111–117 (2001)
12. Deslauriers, A., Morin, H., Begin, Y.: Cellular phenology of annual ring formation of Abies balsamea in the Quebec boreal forest (Canada). Canadian Journal of Forest Research 33(2), 190–200 (2003)
13. Seo, J.W., Eckstein, D., Jalkanen, R., Rickebusch, S., Schmitt, U.: Estimating the onset of cambial activity in Scots pine in northern Finland by means of the heat-sum approach. Tree Physiology 28, 105–112 (2008)
14. Wang, J.Y.: A critique of the heat unit approach to plant response studies. Ecology 41(4), 785–790 (1960)
15. Baskerville, G.L., Emin, P.: Rapid estimation of heat accumulation from maximum and minimum temperatures. Ecology 50(3), 514–517 (1969)
16. Allen, J.C.: A modified sine wave method for calculating degree-days. Environmental Entomology 5, 388–396 (1976)
17. Zalom, F.G., Goodell, P.B., Wilson, L.T., Barnett, W.W., Bentley, W.J.: Degree-days: The calculation and use of heat units in pest management. Leaflet 21373, University of California, DANR (1983)
18. Idso, S.B., Jackson, R.D., Reginato, R.J.: Extending the "degree day" concept of plant phenological development to include water stress effects. Ecology 59(3), 431–433 (1978)

19. Sarvas, R.: Investigations on the annual cycle of development of forest trees. Active period. Communicationes Instituti Forestalis Fenniae 76(3), 1–110 (1972)
20. Prautzsch, H., Boehm, W., Paluszny, M.: Bézier and B-Spline Techniques. Springer, Heidelberg (2002)
21. Sulkava, M., Tikka, J., Hollmén, J.: Sparse regression for analyzing the development of foliar nutrient concentrations in coniferous trees. Ecological Modelling 191(1), 118–130 (2006)
22. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artificial Intelligence 97(1–2), 273–324 (1997)
23. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. John Wiley & Sons, Chichester (2001)
24. Cajander, A.K.: Forest types and their significance. Acta Forestalia Fennica 56(5), 1–71 (1949)