

# Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing

I. CHUINE,<sup>1</sup> P. COUR<sup>1</sup> & D. D. ROUSSEAU<sup>1,2</sup>

<sup>1</sup>UMR CNRS 5554, Institut des Sciences de l'Evolution, Université Montpellier II, place Eugène Bataillon, 34095 Montpellier Cedex 5, France and <sup>2</sup>Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA

## ABSTRACT

The aim of the present study was to test the four commonly used models to predict the dates of flowering of temperate-zone trees, the spring warming, sequential, parallel and alternating models. Previous studies concerning the performance of these models have shown that they were unable to make accurate predictions based on external data. One of the reasons for such inaccuracy may be wrong estimations of the parameters of each model due to the non-convergence of the optimization algorithm towards their maximum likelihood. We proposed to fit these four models using a simulated annealing method which is known to avoid local extrema of any kind of function, and thus is particularly well adapted to fit budburst models, as their likelihood function presents many local maxima. We tested this method using a phenological dataset deduced from aeropalynological data. Annual pollen spectra were used to estimate the dates of flowering of the populations around the sampling station. The results show that simulated annealing provides a better fit than traditional methods. Despite this improvement, classical models still failed to predict external data. We expect the simulated annealing method to allow reliable comparisons among models, leading to a selection of biologically relevant ones.

*Key-words:* convergence; flowering time; optimization algorithm; phenology; simulated annealing method; tree budburst.

## INTRODUCTION

During the last 20 years, the causes and consequences of variation in phenology have received much interest from ecologists concerned with problems related to global warming (Lechowicz & Koike 1995). Initial interest in phenology appeared in the first International Biological Program (1969–1974), as phenology was recognized to play a prominent role in the modelling of ecosystem productivity (Lieth 1971). Some models predicting the date of budburst of temperate-zone trees have already been developed (Cannell & Smith 1983; Hunter & Lechowicz 1992a; Kramer 1994a) and are now widely used to predict the consequences of

particular global warming scenarios on tree phenology (Cannell & Smith 1986; Murray, Cannell & Smith 1989; Hänninen 1991, 1995, 1996; Hänninen *et al.* 1993, 1996; Kramer 1994b, 1995) and also on terrestrial carbon productivity (Lieth 1971; Kramer & Mohren 1996).

Temperature is recognized to be the main determinant of the timing of budburst. Some models consider only the action of forcing temperatures ['Thermal Time model' (Cannell & Smith 1983) or 'spring warming model' (Hunter & Lechowicz 1992a)], whereas others also consider the action of chilling temperatures ['parallel model' (Landsberg 1974; Hänninen 1987, 1990; Kramer 1994a); 'sequential model' (Sarvas 1974; Hänninen 1987, 1990b; Kramer 1994a) 'alternating model' (Cannell & Smith 1983; Murray *et al.* 1989; Kramer 1994a, 1994b)].

Previous studies of models predicting tree phenology have usually concerned a single species [*Fagus sylvatica* L. (Kramer 1994a), *Pinus palustris* Mill. (Boyer 1973), *Picea sitchensis* (Bong.) Car. (Cannell & Smith 1983)], and rarely several species (Hunter & Lechowicz 1992). In addition, because of a lack of data, the accuracy of predictions on external data (i.e. not used to fit the parameters), has been seldom estimated (Kramer 1994a).

However, estimates of these models have no external and even no internal validity (Kramer 1994a). We define here external validity as the accuracy of prediction of external data. In contrast, internal validity is the accuracy of predictions of data used to fit the models. The use of external validity to estimate the accuracy of a model is called cross-validation (Lebreton, Burhnam & Clobert 1992), and is especially useful when the functional relationships of model parameters have been specified. A second critical point concerning phenology modelling is the non-convergence of the optimization algorithms of the likelihood function (Kramer 1994a). This implies that best-fit parameters cannot be properly estimated and, therefore, cannot be used for accurate prediction and hypothesis testing. To make the best of any given model, new algorithms are therefore needed. The use of convergent algorithms will allow the identification of models with the best internal validity, the performances of which then have to be further tested on external data, especially if they are used in futurist climatic scenarios.

In this study, our purpose was to fit the spring warming, sequential, parallel and alternating models using a new method of optimization. The aim was to find reliable

Correspondence: I. Chuine. Tel: 33 4 67 14 39 25; fax: 33 4 67 04 20 32; e-mail: chuine@isem.univ-montp2.fr

estimates of the models in improving the convergence of the optimization algorithm of their likelihood function. The dataset used to fit the models is based on the dates of flowering of seven European tree species over 19 years. Although most of the models available in the literature were developed in order to predict leaf budburst, flower budburst may respond to the same kind of model (Boyer 1973; Cannell 1989).

## MATERIALS AND METHODS

### Data

#### Pollen data

The dates of flowering were deduced from peaks of pollen shedding in the atmosphere. These data consist of pollen concentrations in the atmosphere as weekly averages from Montpellier, France (43.3°N, 3.6°E) over 19 years (1974–1992). Pollen was trapped on filters fixed vertically on a wind cock which continuously oriented the filters to the wind. Filters were exposed all year long and were changed every week. The density of the pollen intercepted on the filters was estimated for each identified taxon according to Cour (1974). Given the amount of wind passed through the filters and measured by an anemometer, the concentration of pollen per m<sup>3</sup> air, for each taxon, was calculated. This method allowed us to determine the temporal variation in pollen concentrations of each taxon present in the atmosphere and, thus, to determine the week during which the pollen was released.

Pollen identification was made to the genus level as pollen morphologies of different species within a genus are usually indistinguishable. However, only genera represented by one species in the area of Montpellier were considered. Seven such genera were found from regional flora and vegetation maps (1/200000). The species studied were: *Aesculus hippocastanum* L., *Alnus glutinosa* Gaertn., *Buxus sempervirens* L., *Olea europaea* L., *Platanus acerifolia* Willd., *Taxus baccata* L. and *Ulmus minor* Mill.

#### Dates of flowering

The middle day of the week of pollination was assumed to be the mean date of anthesis of the populations around the pollen sampling station (≈ 50 km). The models were used to predict the dates of anthesis which corresponded to the dates of flowering of the male flowers (*A. glutinosa*, *P. acerifolia*, *T. baccata*) or to the dates of the male maturation stage (*A. hippocastanum*, *B. sempervirens*, *O. europaea*, *U. minor*).

#### Meteorological data

Pollen traps were placed in the meteorological station near the airport of Frejorgues (6 km south of Montpellier). Daily minimum and maximum temperatures, wind speed, and weekly precipitation were recorded. The average temperature of each day was estimated to be the mean of the daily minimum and maximum temperatures. The speed of the wind was used to calculate the weekly pollen concentration in the atmosphere and the weekly precipitation was used to control if the pollination was disturbed by unfavourable meteorological conditions which would bias the dates of flowering.

### Models

The spring warming model for the timing of budburst was introduced by Reaumur (1735), used by Robertson (1968) and later by Cannell & Smith (1983). This model assumes that there is a linear relationship between energy (sum of temperature degrees above a given threshold) and the growth state of buds. Budburst thus appears after a certain amount of accumulated heat units. This model is the simplest model of budburst prediction, as it involves only three parameters: the base temperature ( $T_b$ ) (thermal threshold above which the degrees are counted), the sum of degree-days ( $F^*$ ) and the starting date of the heat sum ( $t_0$ ) (Eqn 1, Table 1).

$y$	date of flowering
$x_t$	daily mean temperature (°C)
$R_f(x_t)$	forcing rate function
$R_c(x_t)$	chilling rate function
$S_f$	state of forcing
$S_c$	state of chilling
$K_m$	minimum potential of unchilled buds to respond to forcing temperature
$C^*$	critical value of state of chilling for the transition from rest to quiescence
$F^*$	critical value of state of forcing for the transition from quiescence to flowering
$t_0$	starting day of the heat sum calculation
$t_1$	date of onset of rest
$t_2$	date of onset of quiescence
$T_b$	base temperature
$T_o$	optimal temperature of the rate of chilling
$a, b$	constants

#### Spring warming model

$$y \text{ such as } S_f = \sum_{t_0}^y R_f(x_t) = F^*$$

**Table 1.** Description of the models [from Sarvas (1974); Cannell & Smith (1983); Murray *et al.* (1989); Hänninen (1990); Kramer (1994a)]

Table 1. Continued

$$R_f(x_t) = \begin{cases} 0 & x_t \leq T_b \\ x_t - T_b & x_t > T_b \end{cases} \quad (1)$$

**Alternating model**

$$y \text{ such as } S_f = \sum_{t_2}^y R_f(x_t) = F^* = a \exp(bSc) \quad (2)$$

$$t_2 \text{ such as } \sum_{t_1}^{t_2} R_c(x_t) = C^*$$

$$S_c = \sum_{t_1}^y R_c(x_t)$$

$$R_c(x_t) = \begin{cases} 0 & x_t > T_b \\ 1 & x_t \leq T_b \end{cases} \quad (3)$$

$$R_f(x_t) = \begin{cases} 0 & x_t \leq T_b \\ x_t - T_b & x_t > T_b \end{cases}$$

**Sequential model**

$$y \text{ such as } S_f = \sum_{t_2}^y R_f(x_t) = F^*$$

$$R_f(x_t) = \begin{cases} 0 & x_t \leq T_b \\ \frac{28.4}{1 + e^{-0.185(x_t - 18.4)}} & x_t > T_b \end{cases} \quad (4)$$

$$t_2 \text{ such as } \sum_{t_1}^{t_2} R_c(x_t) = C^*$$

$$R_c(x_t) = \begin{cases} 0 & x_t \leq -3.4 \text{ or } x_t \geq 10.4 \\ \frac{x_t + 3.4}{T_o + 3.4} & -3.4 < x_t \leq T_o \\ \frac{x_t - 10.4}{T_o - 10.4} & T_o < x_t < 10.4 \end{cases} \quad (5)$$

**Parallel model**

$$y \text{ such as } S_f = \sum_{t_1}^y R_f(x_t) = F^*$$

$$R_f(x_t) = \begin{cases} 0 & \\ (K_m + \frac{1 - K_m}{C^*} S_c) f(x_t) & \begin{matrix} x_t \leq 0 \text{ and } S_c < C^* \\ x_t > 0 \text{ and } S_c < C^* \\ x_t > 0 \text{ and } S_c \geq C^* \end{matrix} \\ f(x_t) & \end{cases} \quad (6)$$

$$f(x_t) = \frac{28.4}{1 + e^{-0.185(x_t - 18.4)}}$$

$$S_c = \sum_{t_1}^y R_c(x_t)$$

$$R_c(x_t) = \begin{cases} 0 & x_t \leq -3.4 \text{ or } x_t \geq 10.4 \\ \frac{x_t + 3.4}{T_o + 3.4} & -3.4 < x_t \leq T_o \\ \frac{x_t - 10.4}{T_o - 10.4} & T_o < x_t < 10.4 \end{cases}$$

The three other models (parallel, sequential and alternating) consider, in addition, the effect of cool temperatures or chilling degrees which are involved in the break of the bud dormancy. It is now commonly assumed that a period with chilling temperatures ( $-3.4$  to  $10.4$  °C) [Sarvas (1974) in Hänninen (1990a)] followed by a period with forcing temperatures ( $> 0$  °C) induces budburst (Kramer 1994a).

The sequential model (Sarvas 1974; Hänninen 1987, 1990b; Kramer 1994a) assumes that the effect of forcing temperatures cannot be effective unless chilling requirements have already been fulfilled. On the contrary, the parallel model (Landsberg 1974; Hänninen 1987, 1990b; Kramer 1994a) assumes that forcing temperatures can be active concomitant with the time spent for chilling conditions and they are not fully active as long as full chilling is not reached. Both models assume that the state of chilling and forcing is the summation of the rates of chilling and forcing. The rate of chilling was assumed to show an optimum between the minimum and the maximum thermal thresholds (Sarvas 1974; Hänninen 1990b; 1991; Kramer 1994a) (Table 1, Eqn 5). The rate of forcing was assumed to be logistically related to temperature (Table 1, Eqns 4 and 6). Because we intended to estimate the external validity of the models, half the data were needed to fit the models (9 years), and at least half the data were needed to test them (10 years). Thus, parameters of the forcing and the chilling rate have been fixed and not released in this study, as in Hänninen (1990b, 1991), because the number of parameters would have been too high compared with the number of data available to fit the models. For the same reason, the base temperature of the parallel model was fixed to  $0$  °C in this study.

Finally, the alternating model (Murray *et al.* 1989; Cannell & Smith 1983; Kramer 1994a) assumes that the state of chilling is defined as the number of days when the temperature is under a certain thermal threshold (Table 1, Eqn 3). The state of forcing is defined as a heat sum of degree-days above this thermal threshold (Table 1, Eqn 1). This model assumes in addition the existence of a negative relationship (Table 1, Eqn 2) between the heat sum required to flower and the number of chilling days received during the autumn and/or the winter. The starting date of the accumulation of degree-days was not fixed as in Kramer (1994a) but defined by the day when the sum of chilling days attained a minimum ( $C^*$ ).

### Parameter estimation

Models were fitted using the least squares method. The function  $f(x) =$

$$\sum_i [r_i(x)]^2$$

is minimized in the parameter space  $x$ , where  $r_i(x)$  is the residual,  $r_i(x) = d_i(x) - d_{i\text{obs}}$ ,  $d_i(x)$  is the predicted date and  $d_{i\text{obs}}$  the observed date of the year  $i$ . This is the so-called non-linear least squares problem. A Shapiro-Wilk's  $W$ -test for normality (Shapiro, Wilk & Chen 1968) was performed

on the residuals for each model and each species. Test results were combined using Fisher's method (Manly 1985).

We chose a simulated annealing method to fit the models. Simulated annealing refers to an analogy with a thermodynamic principle on the way metals cool and anneal or liquids freeze and crystallize. This principle is the Boltzmann probability distribution,  $\text{Prob}(E) \approx \exp(-E/kT)$ , a system in thermal equilibrium at temperature  $T$  has its energy probabilistically distributed among all different energy states  $E$ . In 1953, Metropolis *et al.* first used this principle for numerical calculations. The algorithm of optimization issued from this principle, named the Metropolis algorithm was used to fit the models. The principle is as follows. Considering that a set of parameter values  $x$  is analogous to a thermodynamic state  $S$  of a system, and the sum of squares  $f(x)$  is analogous to the energy  $E$  of that system, the range of possible states ( $S$ ) is explored by randomly changing from  $S_1$  to  $S_2$  according to the following rule. The probability of transition  $S_1$  to  $S_2$  is 1 if  $S_2 < S_1$ , and  $\exp(-(S_2 - S_1)/kT)$  if  $S_2 > S_1$ , where  $T$  is the temperature of the system. A slow enough decrease of  $T$  leads to a quasi certain convergence of the system towards its state of minimum energy (Press *et al.* 1989). The algorithm needs a generator of random changes. We chose the Marsaglia, Zaman & Tsang (1990) procedure. The first set of parameters is arbitrary, then all the parameters are changed by a random amount. If the change decreases the residual sum of squares, it is accepted; if it increases the residual sum of squares (from  $S_1$  to  $S_2$ ), then it is accepted with probability  $P = \exp(-(S_2 - S_1)/kT)$ . This generates a random walk which tends to decrease the residual sum of squares. The whole parameter space is explored roughly a first time, then in detail around the highest relative minimum or the absolute minimum found. The parameters of the Metropolis algorithm must be carefully adjusted to ensure the convergence of the algorithm towards the absolute least square minimum. In particular,  $T$  normally decreases from the start to the end of the search of the minimum. The more  $T$  decreases, the higher the probability of rejecting a change that increased  $S$ . In our case, because there exist an especially large number of local minima,  $T$  has to decrease very slowly so that the algorithm does not fall into one of the local minima. Another important condition to improve the search of the absolute minimum was to allow  $r_i(x)$  to take a real value (not only integer ones). This condition completely smooths the distribution of  $f$ , which makes the convergence toward the absolute minimum much easier.

The accuracy of the estimates found by the Metropolis algorithm was checked by replication. The quality of the convergence was measured by the standard error on each parameter for the repetitions.

### Tests

#### *Internal and external validity*

First, parameters were fitted using 9 years of the Montpellier dataset (odd years from 1975 to 1991). The internal validity of each model was measured by the

percentage variance explained ( $R^2$ ), and tested for each species separately using  $F$ -tests. The internal validity measures the goodness of fit for the years used to fit the parameters only. The external validity (Chatfield 1988) was measured by the percentage of variance explained of external observed dates (10 even years from 1974 to 1992). External validity was tested as follows. The percentage variance explained was assimilated to a coefficient of determination, the associated correlation coefficient of which was compared with critical values for correlation coefficients (Rohlf & Sokal 1969).

**RESULTS**

**Use of the residual sum of squares**

Normality of residuals is verified for each model (Fisher's combined tests over independent Shapiro-Wilk's test results:  $\chi^2_{14} = 2.3; 7.7; 11.1; 8.7$  for the spring warming, alternating, sequential, parallel models, respectively, all non-significant). Residuals and observed dates of flowering were not correlated except for the parallel model (Fisher's combined tests over independent Shapiro-Wilk's test results:  $\chi^2_{14} = 16.7; 14.5; 9.6; 23.9$  for the spring warming, alternating, sequential, parallel models, respectively,  $\chi^2_{14(5\%)} = 23.7$ ). Correlation between residuals and the dates of flowering for the parallel model are the result of low internal validity. However, we can say that the measurement errors of the dates of flowering were independent and normally distributed. The use of the minimum sum of squares of the residuals as a criterion to identify the optimal set of parameter values is therefore justified.

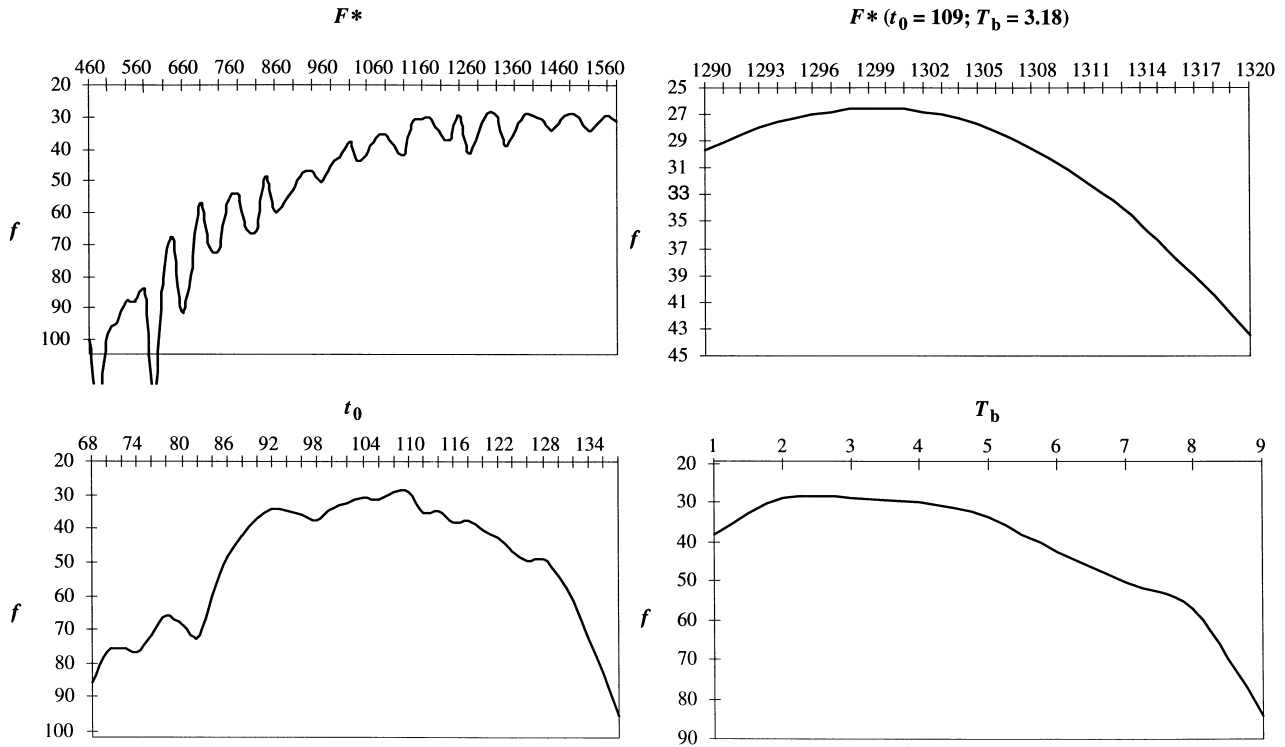
**Convergence of the Metropolis algorithm**

Coefficients of variation of the residual sum of squares of the different repetitions of the fit were low in most cases (between 0 and 10% in 82% of cases) (Table 2). Coefficients of variation greater than 10% are due to frequent convergences towards a second local minimum. The coefficients of variation of the estimates were inferior to 15% in 67% of cases. Coefficients greater than 15% are due to one or two convergences towards a local minimum far away from the absolute, or to a parameter of very low variance and mean, such as  $T_o$ ,  $T_b$ ,  $b$  or  $C^*$ . This indicates that the algorithm almost always converged, or at least more than half the time, towards the same region defining a minimum. We can expect that the lowest minima found with the highest frequency is the absolute minima.

To verify this hypothesis,  $f$  should be plotted for all possible parameter values, for all models and all species. This was carried out for the spring warming model for *O. europaea* (Fig. 1). The support curve shows that the absolute minimum at  $f = -29$  was defined by the (1320, 110, 3) parameter set of the model. The Metropolis algorithm found a minimum at  $f = -27$  for the parameter set (1297, 109, 3.2). A zoom on the  $F^*$  parameter (step size: 1 °C) with the values 109 for  $t_0$  and 3.2 for  $T_b$ , shows that the

**Table 2.** Coefficients of variation (first line) and standard errors (second line) of  $f$  and each parameter of each model across 10 replicate runs of the algorithm. *Italic characters correspond to coefficients associated with a frequent convergence towards a second local minimum*

	Spring warming					Sequential					Parallel					Alternating				
	$T_b$	$F^*$	$t_0$	$f$	$T_b$	$F^*$	$C^*$	$T_o$	$f$	$T_b$	$F^*$	$C^*$	$K_m$	$f$	$T_b$	$C^*$	$a$	$b$	$f$	
<i>A. hippocastanum</i>	1.6	15.8	0.5	3.1	2.4	24.7	78.7	280	5.0	8.9	0.1	20.7	-	0.2	2.0	2.5	51.7	17.7	2.1	
	2.2	33.1	0.9	1.9	0.4	24.7	21.6	4.6	3.5	0.7	0.5	0.3	0.0	1.5	2.7	2.9	1349	0.0	0.9	
<i>A. glutinosa</i>	111	1.4	0.2	0.3	13.0	4.2	7.7	13.7	8.0	275.1	3.2	15.0	62.5	2.2	1.0	26.6	32.8	16.7	29.1	
	0.8	48.1	0.2	2.4	0.6	6.1	1.8	0.7	11.0	1.6	9.4	3.0	0.0	11.8	0.7	2.5	1005	0.0	103	
<i>B. sempervirens</i>	4.8	8.7	4.3	0.8	0.5	0.2	2.0	8.0	0.6	15.0	4.0	11.1	62.7	4.0	13.2	65.5	28.3	87.9	6.1	
	3.5	14.6	5.6	4.8	0.0	0.4	0.8	0.4	2.5	0.9	16.8	4.5	0.0	37.7	10.4	19.4	81.6	0.0	31.8	
<i>O. europaea</i>	11	4.5	1.1	4.2	28.3	7.7	23.2	76.6	2.3	1329	0.6	29.1	-	0.2	2.6	8.0	4.6	0.0	1.4	
	3.6	57.4	1.2	1.1	1.9	78.5	2.1	2.0	0.6	2.2	6.5	6.1	0.0	0.3	2.3	1.2	247	0.0	0.8	
<i>P. acerifolia</i>	18.2	3.2	0.1	1.5	0.6	8.5	18.3	10.3	2.9	7.2	1.1	5.5	-	3.6	1.1	1.1	29.8	37.6	2.8	
	2.7	140	0.2	0.6	0.1	13.4	2.9	0.6	2.4	0.5	4.9	2.5	0.0	4.8	19.5	11.7	1175	0.0	90.0	
<i>T. baccata</i>	22.3	38.8	8.6	4.9	0.3	0.0	8.6	51.0	0.0	80.5	4.5	12.5	107	1.6	9.9	32.6	64.7	51.5	12.2	
	1.7	47.4	0.2	0.6	0.0	0.0	1.1	1.1	0.0	1.1	13.8	2.7	0.0	5.7	13.2	15.8	434	0.0	27.5	
<i>U. minor</i>	2.4	4.2	0.3	0.2	16.4	20.9	27.7	458	2.7	13.3	0.1	2.8	33.4	0.1	11.8	24.3	20.0	68.3	21.1	
	1.4	46.2	0.3	0.5	1.3	20.4	3.9	2.4	4.2	0.4	0.2	0.4	0.0	0.2	4.8	6.0	134	0.0	52.9	



**Figure 1.** Least square function of the spring warming model for *Olea europaea*. For each parameter value, only the maximum of  $f$  is represented among all possible values. Step size of the parameters:  $F^*$ : 10;  $t_0$ : 2;  $T_b$ : 1. A zoom of the range 1290–1320 of the  $F^*$  parameter is shown for  $t_0 = 110$  and  $T_b = 3.18$ .  $t_0$  = number of days from 1 September.

absolute minimum is defined for  $F^* = 1299$  with a minimum of  $-26.5$ . Thus, considering the step size that was used to plot the  $f$  function (10; 2; 1), it appears that the Metropolis algorithm found the global minimum of this function.

### Internal validity of the models

Each model had internal validity whatever the species (Table 3). The explained variance is from 14% for the worst performance to 91% for the highest performance, with a mean of 70% over all species and models.

### External validity of the models

Based on the number of species for which the model shows significant external validity, the sequential and the parallel models are the worst predictor, followed by the alternating model and the spring warming model (Table 3). The explained variance (when external validity exists) varied from 33% for the worst performance to 69% for the best performance.

### Estimates

Table 4 shows the estimates of each model for each species in order to permit comparison with further studies. Estimates have not been published in the previous studies

for these species so that no comparison was possible. Budburst models pertained to mechanist model type. One of the aims and advantages of such models is the possibility of giving a biological meaning to the estimates. Our knowledge of tree phenology does not permit such an interpretation, all the more so because estimates have, for example, never been tested on experimental data. However, it can be remarked that first, the  $K_m$  parameter of the parallel model is always estimated to be 0, a result that indicates that buds have no potential to respond to forcing temperature when they are unchilled, which is the main hypothesis of this model.

## DISCUSSION

### Data

The aeropalynological dataset, which covers relatively long periods, has a large advantage compared with direct phenological observations in natural populations. Pollen trapped in the filters came from a wide region so that the pollen peaks corresponded to the timing of flowering of a whole population. Although such data do not take into account intrapopulation variability due to genetic differences or local variation in thermal conditions (solar exposure), they represent phenological trends of populational/regional significance. Thus, the dataset avoids one of the most important problems associated with the historical phenological

**Table 3.** Internal and external validities: percentage variance explained ( $R^2$ ) on internal and external data

	Internal validity				External validity			
	Spring warming	Sequential	Parallel	Alternating	Spring warming	Sequential	Parallel	Alternating
<i>A. hippocastanum</i>	0.84ns	0.82ns	0.48ns	0.88*	–	–	–	–
<i>A. glutinosa</i>	0.41ns	0.90*	0.27ns	0.79ns	–	–	–	–
<i>B. sempervirens</i>	0.58ns	0.72ns	0.45ns	0.66ns	0.58**	–	–	0.69***
<i>O. europaea</i>	0.90*	0.91*	0.74ns	0.78ns	0.64***	0.33*	0.64***	0.67***
<i>P. acerifolia</i>	0.90*	0.79ns	0.81ns	0.73ns	0.65***	–	–	–
<i>T. baccata</i>	0.80ns	0.74ns	0.14ns	0.59ns	–	–	–	–
<i>U. minor</i>	0.75ns	0.83ns	0.57ns	0.76ns	–	–	–	–

Significance level: ns,  $P > 0.05$ ; \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . –, residual sum of squares superior to the total sum of squares.

observations made *in natura*, for which there is often no certainty that the same trees were observed each year, nor do we know how representative of the population these trees were (Hunter & Lechowicz 1992; Kramer 1994b; Sparks & Carey 1995).

It is important to point out here the exact geographical coincidence between the vegetation samples and the meteorological station. This is of major importance for two reasons. Firstly, local adaptation of phenology to climate is still possible although it has not yet been proved. Secondly, the geographical separation of meteorological and phenological data make predictions and fits imprecise. Coincidence has not always been verified, as phenological observations in phenological gardens or in natural populations usually come from isolated places where no meteorological stations exist (Kramer 1994a,b).

The pollen dataset also suffers disadvantages. First, pollen can only be identified to the genus level and a large number of taxa cannot be used in the search study. Second, the pollen data indicate the time of anthesis and not the time of burst of the flower buds. Meteorological conditions for anthesis are relatively high temperatures or increasing temperatures and dry conditions. If these conditions are not met, anthesis can be delayed or reduced although flowers were potentially ready for anthesis. The pollen signal can be in such cases biased from 1 to 7 d. Nevertheless, the data used to fit the models have been controlled in order to avoid biases due to meteorological conditions as previously explained. Third, pollen data are weekly. This precision would be low for a study on an individual tree, but is sufficient for a population study because the magnitude of flowering from an individual to another is more than 1 month. In this condition, it is useless to try to obtain sums of squares lower than  $9 \text{ (years)} \times 3 \text{ (minimum measurement error)}^2 = 81$ , i.e. for one-third of the cases in our study.

### The Metropolis algorithm

As  $f(x)$  is a non-linear function of many parameters, finding its minimum is not trivial; numerical methods to evaluate the maximum likelihood estimate by iteration

are necessary. Furthermore, the support curve of  $-f$  shows a large number of local minima (Fig. 1), so that the traditional Downhill Simplex and Newton methods will converge to the global maximum with a very low probability. Virtually nothing is known about finding global extrema in general. Traditionally, two kind of methods can be distinguished: (i) find all local extrema, starting from widely varying starting values of the parameters and keep the extreme one, (ii) perturb a local extremum by taking a finite amplitude step away from it, and see if the algorithm returns a better parameter set or always the same parameter set (Press *et al.* 1989). Among the other possible methods, Barnett (1966) has shown that the only method to ensure convergence when the likelihood equation has multiple roots was the method of false position. This method, however, does not guarantee that the absolute extremum is found. All the relative extrema should be found to determine the absolute extremum.

The problem of convergence of the algorithm of optimization used for budburst model fits has already been underlined by Kramer (1994a) using the GENSTAT directive Fitnonlinear and the subroutines EO4FCF of the Fortran library (Newton method); and Amoeba of the Pascal library transposed in Fortran (Downhill Simplex method) (Kramer, personal communication).

The method of simulated annealing is a technique that has received a lot of attention for its suitability for all kinds of optimization problems, as any function can be optimized. It was seen 20 years ago as a very attractive method of optimization, especially for anomalous cases (Press *et al.* 1989). It solved the famous travelling salesman problem and was also used successfully for designing complex integrated circuits. One of its most attractive features is its accuracy for not falling into unfavourable local maxima. The Metropolis algorithm had already been used successfully for genetic problems (Szymura & Barton 1986). The results obtained in the present study, that is the convergence towards the absolute minimum despite many local minima, show that the algorithm also seems to be well adapted to budburst models. Even if models have no external validity, we are now sure that it

**Table 4.** Estimates of the parameters for each species.  $T_b$ ,  $F^*$ ,  $C^*$ ,  $t_0$ ,  $T_o$ ,  $K_m$ ,  $a$ ,  $b$ : defined as in Table 1

	Spring warming			Sequential			Parallel			Alternating					
	$T_b$	$F^*$	$t_0$	$T_o$	$T_b$	$F^*$	$C^*$	$T_o$	$F^*$	$C^*$	$K_m$	$T_b$	$C^*$	$a$	$b$
<i>A. hippocastanum</i>	13.9	20.1	3/3	10.1	15.5	50.6	69.5	8.0	899.5	1.9	0.000	13.5	116	170	-0.0110
<i>A. glutinosa</i>	0.0	344.4	1/1	5.8	4.1	141.9	26.2	2.7	292.3	24.4	0.015	7.6	11.7	224	-0.0215
<i>B. sempervirens</i>	7.1	178.1	1/12	5.1	6.8	207.4	38.4	6.2	410.8	41.8	0.008	9.2	55.3	352	-0.0211
<i>O. europaea</i>	3.2	1296.7	12/20	3.7	5.8	1050.7	10.2	4.8	1078.1	37.3	0.000	8.9	15.1	525	-0.0002
<i>P. acerifolia</i>	1.7	424.5	2/15	6.0	12.0	173.9	13.1	7.0	443.7	48.6	0.000	10.3	36.2	206	-0.0099
<i>T. baccata</i>	8.5	58.4	1/1	-0.7	10.8	65.6	13.6	2.6	301.3	24.9	0.006	10.5	61.4	340	-0.0346
<i>U. minor</i>	5.8	111.4	1/1	2.9	6.4	131.9	18.1	-2.5	276.2	14.9	0.000	8.0	24.6	61.6	-0.0024

cannot be the consequence of wrong estimations, as the convergence ensures correct estimates.

### External validity

The first important result is the existence of external validity for some species, a result that has not been previously reported for such models. The algorithm of optimization seems to play an important role in this result. However, the lack of external validity shown by the previous study of Kramer (1994a) could also have other explanations in addition to the fact that the algorithm of optimization did not converge. In his estimation of external validity, Kramer fitted the parameters of the models with the data of a Dutch dataset and tested them with a German dataset. The models fitted had no external validity except the sequential-i model. This model was similar to the sequential model except that  $T_b$  was fixed to 0 °C and that all the other parameters were not constrained into a range so that some of the estimates could take unrealistic values. Three possible reasons can be suggested for the lack of external validity. First, there may be a local adaptation of tree phenology. If the German populations of *Fagus sylvatica* are adapted to the climate of Germany, the estimates of the models should be different from the estimates of the Dutch populations. If this were the case, such a result could explain why the predictions of the dates of budburst in Germany using the Dutch estimates did not match the observations. Second, the phenological data came from observations of individual trees. Intrapopulation variation of dates of budburst is usually high (Parlange 1974; Billington & Pelham 1991), especially for early flowering species. Thus, even if there is no local adaptation of tree phenology, it can be expected that the estimates of some Dutch individual trees would be different from the estimates of German individual trees. Third, the meteorological data did not correspond to the phenological data, as the corresponding meteorology of the sampling localities was not available. One meteorological station was available in the Netherlands to fit the models and another station was available in Germany to test the estimates.

Whatever the reason for the lack of external validity, the present study has shown that such models may have external validity for some species but not others. Any one or a combination of the four reasons presented above could help to explain this result. However, external validity is still non-existent for the majority of the species. Species having their dates of flowering accurately predicted by the models are the late flowering species, and in particular, *Olea europaea* (the latest flowering species), of which the dates of flowering were accurately predicted by each model. This result may be explained by the rapid and important rise in temperature in the late spring in Montpellier when those species flower. This rise is not followed by a subsequent decrease as is often observed in winter and early spring. Such successions of cool and warm events may have two consequences. First, interruptions of a warm period by cool events in late winter and



early spring could provoke an increase in the time from quiescence to budburst. Second, interruptions of the chilling period by warmer events could provoke a less efficient impact of the cool temperatures (Farmer 1968). The consequence of this phenomenon would be a wrong estimation of the state of chilling and, thus, a wrong estimation of the starting date of the forcing phase (and a wrong estimation of the critical state of forcing in the case of the alternating model). Thus, models could fail to make accurate predictions for early flowering species. In contrast, late flowering species accumulate a lot of degree-days per day the last days before their flowering, so that the delay induced by the warming period in the winter and the cool events in early spring can be compensated.

Finally, despite the fact that the internal validity of the models is usually very high (up to 90%), this does not provide high external validity at all (Table 3). This strengthens the importance of cross-validation of these kinds of models.

### Model comparison

The comparison of the accuracy of the different models for the different species shows that there is no consensus model even if some models seem consistently more accurate than others. The results of Hunter & Lechowicz (1992) testing different models for many species also showed a similar feature. This result indicates that all the different possible models have to be tested for each species.

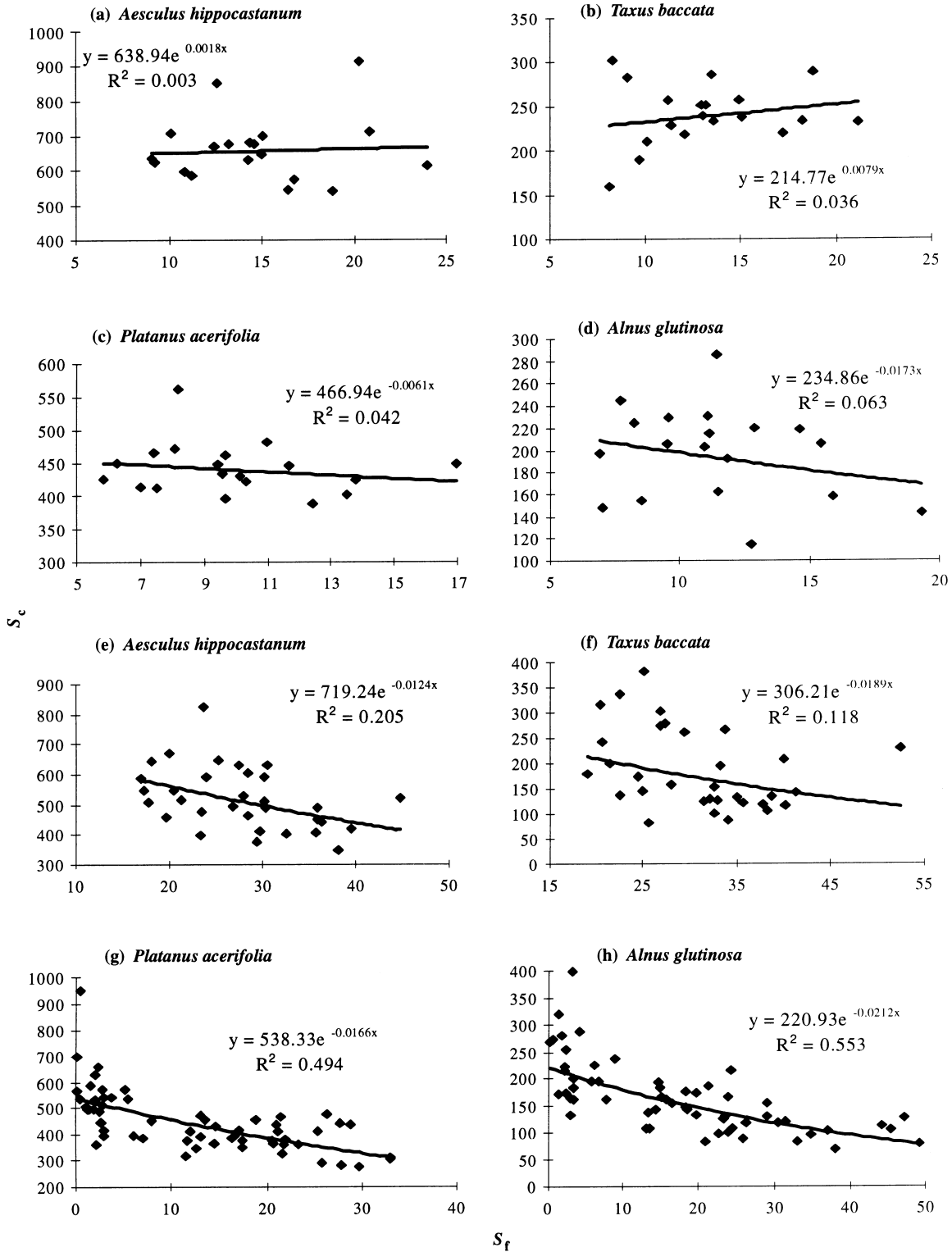
The choice of the best model requires that a certain number of conditions be checked. The aim of model selection is to identify a biologically meaningful model that explains the significant variability in the data, but excludes unnecessary parameters (Lebreton *et al.* 1992). Model selection uses mainly two kind of statistics: the likelihood ratio test and the Akaike Information Criterion (AIC) ( $I = -2\ln L - 2df$ ). AIC implies at least to be able to calculate the likelihood of each model which is impossible as the distribution of the residuals is not exactly known, even if we know it is not significantly different from a normal distribution. As we are allowed to use residual sum of squares as estimator of the likelihood, we could have used *F*-tests on the models' residual sum of squares instead of a likelihood ratio test if the models were interlocked. It is not always possible to decide upon one model that is best for a given set of data. Several similar but different models may be nearly equally applicable to the data, which is the case here. The last solution in such a case is that the models should be made on grounds external to the data used to compute the parameters of the models (Chatfield 1988).

Considering the number of species for which the models show external validity and the values of the coefficient of determination, the spring warming model appears to be the most accurate model followed by the alternating model, and lastly by the parallel and sequential models. The poor performance of the parallel model (in terms of internal or external validity) could be explained by the fact that its principal hypothesis, i.e. buds are potentially able to

respond to forcing temperature even if they are unchilled, is not supported by the data, as  $K_m$  (the potential of unchilled buds) is always fitted to be 0. The external validity obtained by the spring warming model for late flowering species is probably due to the relatively poor influence of chilling temperatures for these species. The case of *O. europaea* shows that chilling temperatures do not bring a lot of explained variance compared with the additional degree of freedom used. This is also shown by the high sensibility of the  $T_o$  parameter (optimal temperature for the chilling temperature action) (Table 2), which denotes a lack of information to fit it properly. However, the action of chilling temperatures on budburst has been proved in many experiments (Farmer 1968; Murray *et al.* 1989; Hänninen *et al.* 1993; Heide 1993; Myking & Heide 1995; Nelson & Lavender 1979) and the failure of the spring warming model to predict dates of flowering of early species indicates that forcing is insufficient for accurate predictions for this kind of species. The negative relationship was first shown by Cannell, Murray & Smith (1983) in a greenhouse experiment where the state of chilling was extremely variable compared with possible variation occurring in natural conditions. In the present study, models were fitted with data collected in natural conditions. In such conditions, chilling temperatures experienced by trees each year do not vary a lot, especially at the latitude of Montpellier. As a consequence, the negative relationship between the state of chilling and the state of forcing seems to be non-existent in the Montpellier area (Fig. 2a–d), and may also occur at higher latitude. If the geographical scale is expanded, the relationship becomes more obvious (Fig. 2e & f) and if it is expanded up to Europe the relationship becomes particularly clear (Fig. 2g & h).

The choice of the best model is difficult as models have been fitted for several independent data series which are the several species. Hence, the choice of the most accurate model depends on the species. According to the standard error on external data, the best models are the alternating model for *B. sempervirens* and the spring warming model for *O. europaea* and *P. acerifolia*. According to the principle of parsimony, we can say that the spring warming model is the best predictor as it has similar external validity to the others but requires less degrees of freedom. However, many studies concerning budburst models, including this one, have shown that there is no consensus model for every species. The spring warming model for example, has shown its limits for the early flowering species.

In conclusion, it appears that the choice of the algorithm of optimization used to fit budburst models is a particularly important parameter in budburst modelling. Our study indicates that the method of simulated annealing may bring a non-negligible contribution to the problem of non-convergence encountered in the former studies. Despite the improvement due to this kind of optimization method, the models tested seem to be accurate only for late flowering species. Thus, the forcing process seems apparently well formalized and sufficient to predict the flowering of these



**Figure 2.** Relationship between the state of forcing ( $S_f$  in ordinate) and the state of chilling ( $S_c$  in abscissa) for *Aesculus hippocastanum*, *Taxus baccata*, *Platanus acerifolia* and *Alnus glutinosa*: (a, b, c, d) in Montpellier and (e, f, g, h) in different localities over Europe from Abisko to Oran (dates of flowering measured with aeropalynological data, Cour, unpublished results). Each point is a given year in a given location.  $R^2$ : explained variance by the adjusted exponential function.  $S_f$  and  $S_c$  have been calculated as follows:

$$S_f = \sum_{t_1}^y R_f(x_t) \quad \text{and} \quad S_c = \sum_{t_2}^{t_1} R_c(x_t),$$

with  $t_2 = 1$  January,  $t_1 = 1$  September and  $R_c$  and  $R_f$  defined as in the sequential model.

kinds of species for which the influence of the chilling period is reduced. As proved experimentally, the chilling process is therefore necessary for early species, as the spring warming model failed to predict their dates of flowering. However, its formalization seems until now to be too different from biological reality to improve the model's accuracy. In particular, two problems can be responsible, first the unknown impact of warm periods on chilling temperature action; second, the difficulty of fitting the negative relationship between the state of forcing and the state of chilling in a single place for the alternating model. Nevertheless, convergence will now allow reliable tests and comparisons of different models that will enable the selection of accurate hypotheses involved in budburst timing prediction.

## ACKNOWLEDGMENTS

The authors are particularly grateful to Thomas Lenormand and Patrice David for constructive criticisms, and to John Thompson for revising the English language of this paper. We thank also Heikki Hänninen and anonymous referees for their constructive comments; and D. Duzer, J. Ferrier, L. Quet and G. Sare for the pollen analyses. This paper is publication number 98-037 in the ISEM series.

## REFERENCES

- Barnett V.D. (1966) Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. *Biometrika* **53**, 151–165.
- Billington H.L. & Pelham J. (1991) Genetic variation in the date of budburst in Scottish birch populations: implications for climate change. *Functional Ecology* **5**, 403–409.
- Boyer W.D. (1973) Air temperature, heat sums, and pollen shedding phenology of longleaf pine. *Ecology* **54**, 421–425.
- Cannell M.G.R. (1989) Chilling, thermal time and the dates of flowering of trees. In *Manipulation of Fruiting* (ed. C.J. Wright), pp. 99–113. Butterworth, London.
- Cannell M.G.R. & Smith R.I. (1983) Thermal time, chill days and prediction of budburst in *Picea sitchensis*. *Journal of Applied Ecology* **20**, 951–963.
- Cannell M.G.R. & Smith R.I. (1986) Climatic warming, spring budburst and frost damage on trees. *Journal of Applied Ecology* **23**, 177–191.
- Chatfield C. (1988) *Problem Solving: a Statistician Guide*. Chapman and Hall, London.
- Cour P. (1974) Nouvelles techniques de détection des flux et des retombées polliniques. Etude de la sédimentation des pollens et des spores à la surface du sol. *Pollen et Spores* **16**, 103–141.
- Farmer R.E. (1968) Sweetgum dormancy release: effects of chilling, photoperiod, and genotype. *Physiologia Plantarum* **21**, 1241–1248.
- Hänninen H. (1987) Effects of temperature on dormancy release in woody plants: implications of prevailing models. *Silva Fennica* **21**, 279–299.
- Hänninen H. (1990a) Modelling bud dormancy release in trees from cool and temperate regions. *Acta Forestalia Fennica* **213**, 1–47.
- Hänninen H. (1990b) Modelling dormancy release in trees from cool and temperate regions. In *Process Modeling of Forest Growth Responses to Environmental Stress* (eds R.K. Dixon, R.S. Meldahl, G.A. Ruark & W.G. Warren), pp. 159–165. Timber Press, Portland.
- Hänninen H. (1991) Does climatic warming increase the risk of frost damage in northern trees? *Plant, Cell and Environment* **14**, 449–454.
- Hänninen H. (1995) Effects of climatic change on trees from cool and temperate regions: an ecophysiological approach to modelling of budburst phenology. *Canadian Journal of Botany* **73**, 183–199.
- Hänninen H. (1996) Effects of climatic warming on northern trees: testing the frost damage hypothesis with meteorological data from provenance transfer experiments. *Scandinavian Journal of Forestry Research* **11**, 17–25.
- Hänninen H., Kellomäki S., Laitinen K., Pajari B. & Repo T. (1993) Effect of increased winter temperature on the onset of height growth of Scots pine: a field test of a phenological model. *Silva Fennica* **27**, 251–257.
- Hänninen H., Leinonen I., Repo T. & Kellomäki S. (1996) Overwintering and productivity of Scots Pine in a changing climate. *Silva Fennica* **30**, 2–3.
- Heide O.M. (1993) Daylength and thermal time responses of budburst during dormancy release in some northern deciduous trees. *Physiologia Plantarum* **88**, 531–540.
- Hunter A.F. & Lechowicz M.J. (1992) Predicting the timing of budburst in temperate trees. *Journal of Applied Ecology* **29**, 597–604.
- Kramer K. (1994a) Selecting a model to predict the onset of growth of *Fagus sylvatica*. *Journal of Applied Ecology* **31**, 172–181.
- Kramer K. (1994b) A modelling analysis of the effects of climatic warming on the probability of spring frost damage to tree species in The Netherlands and Germany. *Plant, Cell and Environment* **17**, 367–377.
- Kramer K. (1995) Phenotypic plasticity of the phenology of seven European tree species in relation to climatic warming. *Plant, Cell and Environment* **18**, 93–104.
- Kramer K. & Mohren G.M.J. (1996) Sensitivity of FORGRO to climatic change scenarios: a case study on *Betula pubescens*, *Fagus sylvatica* and *Quercus robur* in the Netherland. *Climatic Change* **34**, 231–237.
- Landsberg J.J. (1974) Apple fruit bud development and growth: analysis and an empirical model. *Annals of Botany* **38**, 1013–1023.
- Lebreton D., Burhnam K.P. & Clobert J. (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* **62**, 67–118.
- Lechowicz M.J. & Koike T. (1995) Phenology and seasonality of woody plants: An unappreciated element in global change research. *Canadian Journal of Botany* **73**, 147–148.
- Lieth H. (1971) The phenological viewpoint in productivity studies. In *Productivity of Forest Ecosystems. Proceedings of the Brussels Symposium by UNESCO* (ed. P. Duvigneaud), pp. 71–83. UNESCO, Paris.
- Manly B.F.J. (1985) *The Statistics of Natural Selection. Population and Community Biology Series* (ed. M.B. Usher & M.L. Rosenzweig). Chapman & Hall, London.
- Marsaglia G., Zaman A. & Tsang W.W. (1990) Toward a universal random number generator. *Statistics and Probability Letters* **8**, 35–39.
- Murray M.B., Cannell M.G.R. & Smith R.I. (1989) Date of budburst of fifteen tree species in Britain following climatic warming. *Journal of Applied Ecology* **26**, 693–700.
- Myking T. & Heide O.M. (1995) Dormancy release and chilling requirements of buds of latitudinal ecotypes of *Betula pendula* and *B. pubescens*. *Tree Physiology* **15**, 697–704.
- Nelson E.A. & Lavender D.P. (1979) The chilling requirement of western hemlock seedlings. *Forest Science* **25**, 485–490.
- Parlange J.Y. (1974) Analytical solution to model passages through phenophases. In *Phenology and Seasonality Modelling*. Springer, New York.

- Press W.H., Flannery B.P., Teukolsky S.A. & Vetterling W.T. (1989) *Numerical Recipes in Pascal*, pp. 759. Cambridge University Press, Cambridge.
- Reaumur. (1735) Observations du thermomètres, faites à Paris pendant l'année 1735, comparées avec celles qui ont été faites sous la ligne, à l'isle de France, à Alger et quelques unes de nos isles de l'Amérique. *Académie des Sciences de Paris* **545**.
- Robertson G.W. (1968) A biometeorological time scale for a cereal crop involving day and night temperatures and photoperiod. *International Journal of Biometeorology* **12**, 191–223.
- Rohlf F.J. & Sokal R.R. (1969) *Statistical Tables*, pp. 350. W. H. Freeman & Co., San Francisco.
- Sarvas R. (1974) Investigations on the annual cycle of development of forest trees. II Autumn dormancy and winter dormancy. *Communicationes Instituti Forestalis Fenniae* **84**, 101p.
- Shapiro S.S., Wilk M.B. & Chen H.J. (1968) A comparative study of various tests of normality. *American Statistical Association Journal* **63**, 1343–1372.
- Sparks T.H. & Carey P.D. (1995) The responses of species to climate over two centuries: an analysis of the Marsham phenological record, 1736–1947. *Journal of Ecology* **83**, 321–329.
- Szymura J.M. & Barton N.H. (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near cracow in southern Poland. *Evolution* **40**, 1141–1159.

*Received 5 December 1997; received in revised form 5 February 1998; accepted for publication 5 February 1998*