The use of multi-criteria assessment in developing a process model

Rié Komuro [a], E. David Ford [b,*], Joel H. Reynolds [c]

[a] Department of Applied Mathematics, Box 352420, University of Washington, Seattle, WA 98195, USA
[b] College of Forest Resources, Box 352100, University of Washington, WA 98195, USA
[c] U.S. Fish and Wildlife Service, 1011 East Tudor, Anchorage, AK 99503, USA

ABSTRACT

Ecological data frequently contain multiple patterns. A process model of the system producing the data should be able to recreate those patterns. We describe a method and associated software where components of the pattern are used as multiple criteria to assess a model during its construction. Successive improvements are made to the model so that it describes more components of the pattern effectively.

The software, Pareto_Evolve is an evolutionary algorithm. Initially it creates many individual sets of model parameters, each is used in the model to produce results that are then compared to patterns in a data set. Different individuals may achieve different components of the pattern and Pareto_Evolve calculates those that are most effective and uses them to produce new individuals. This is done by changing parameter values of individuals, called mutation, or exchanging parameter values between individuals, called crossover. This process is repeated over many generations so that a most effective parameterization evolves.

We illustrate this method with a model for hourly increments of extension of the leading shoot of a conifer tree. The particular task here is to model water uptake by the plant in response to water loss due to transpiration, calculate a water deficit between uptake and loss, and calculate contraction and re-expansion of shoot tissue due to diurnal changes in tissue water deficit. We choose criteria corresponding to different phases in the diurnal pattern of expansion and model for up to six consecutive days. A value is set marking the limit within which the model must achieve the criteria for it to be judged as success.

This value is called a binary discrepancy measure. Pareto_Evolve is used to make multiple searches with successively smaller binary discrepancy measures until all criteria are no longer achieved. At this point different parameter sets achieve different groups of criteria and we use these as indicators of how the structure of the model must be improved to achieve an overall better fit. In our example we find that contraction is more rapid than re-expansion, which is a hysteresis effect, and that re-expansion of tissue continues after water deficit within the tissue is estimated to have been removed.

We discuss how this method can be used in model development and particularly how multiple criteria assessment can be used in model development.

© 2006 Published by Elsevier B.V.

* Corresponding author. Tel.: +1 11 206 685 9995; fax: +1 11 206 543 3254.
  E-mail address: edford@u.washington.edu (E.D. Ford).

1. Introduction

Ecological process models attempt to represent the workings of complex systems (Wu and Marceau, 2002). During development of a process model choices must be made about what components of the system to include and the detail with which they should be represented. These are the problems of defining model bounds and deciding about aggregating or disaggregating model components (e.g., Auger et al., 2000) and such decisions may determine difficulties in defining model uncertainty (Oreskes, 2003). Ideally model assessment against data should inform about the effectiveness of these decisions. However, standard goodness-of-fit approaches to assessment may not adequately answer questions about model structure, particularly if based upon a single measure, e.g., R^2 between data and model output (Håkanson, 2003). Typically ecological process models have many parameters often with undefined limits for their values so that deficiencies in model structure may be overcome by variation in parameter values. It may not be possible to place deterministic bounds on the values parameters take that would restrict compensation for structural deficiencies (Norton, 1996) because variation in parameter values has not been studied when the model is first constructed.

Wiegand et al. (2003) suggest that greater use should be made of patterns that occur in data, particularly time series and spatial data. They suggest models be constructed that include the minimal set of processes necessary for reproducing patterns in such data and propose systematic comparison of observed patterns with those produced by different modifications of a model. The strength of this approach, as illustrated by their examples, is that the effects of different component processes in a system may be apparent in the time series or spatial structure produced. Model development should progress from obtaining an overall fit towards describing patterns in dependent data. However, this objective presents a technical challenge in analyzing the multiple signals that may occur in such data and in how to assess a model's effectiveness in replicating them.

An answer to this technical challenge is to use Pareto optimality to summarize the model's simultaneous performance in reproducing the suite of patterns (Reynolds and Ford, 1999). This summary can be derived by multi-criteria optimization where a model fitting procedure based on evolutionary computation searches out arrays of parameter values for the model that achieve as many criteria as possible (Deb, 2001). We apply such a method to develop a model simulating fluctuation in shoot extension in a conifer tree in response to changes in the environment, particularly the processes of contraction and re-expansion that occurs during the day–night cycle. This is the first stage of assessment—finding whether a proposed model structure is adequate for calibration. The method involves calculating the Pareto set consisting of all groups of assessment criteria that can be achieved by different arrays of parameter values. The critical technical difficulty is to find the Pareto set, particularly where large numbers (say > 5) of assessment criteria are used. We use an evolutionary algorithm; Pareto Evolve (Reynolds, 1997) developed and tested for use with large numbers of criteria (Komuro, 2005), and characterize its important

features. We show how the Pareto optimization approach can be used to produce incremental improvements in model structure by comparing patterns produced with different model structures.

2. The modeling problem

Shoot extension in temperate zone conifers such as Sitka spruce (*Picea sitchensis* (Bong.) Carr.) takes place rapidly in spring and early summer through extension of terminal buds formed in the previous summer. In a first investigation, daily measurements of shoot length were made through the growing season in the canopy of an 11 years plantation (Ford et al., 1987a,b) but analysis of growth response was restricted by accuracy of measurement. An electro-mechanical shoot growth sensor was constructed (Milne et al., 1977) that enabled continuous automatic recording of shoot length. The sensor is supported on previous year's wood below the growing shoot and measures extension using a pulsed light emitting diode that does not touch the growing tissue. Hourly values of shoot extension over a sample 7-day period are shown in the upper section of Fig. 1.

The shoot extension sensor produced more accurate data than manual daily measurements but introduced a further problem in understanding the dynamics of growth. Each morning the hourly shoot extension rate declines or becomes negative and later, particularly during the early night, shoot extension increases. The 7 days of Fig. 1 are selected from a larger set because they represent the onset of a warm and sunny period of weather when substantial diurnal contraction and re-expansion cycles first appeared. Kanninen (1985) measured shoot length of 5 years *Pinus sylvestris* and developed a time series model of growth rate as a function of 3 h lagged air temperature but his measurements show no actual contractions.
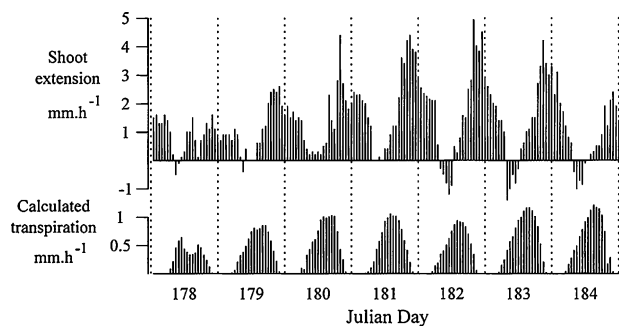

Fig. 1 – Top line: hourly measured shoot growth rate of the terminal shoot of a 14 years plantation growth tree of Sitka spruce in southwest Scotland from Julian days 178 through 184 in 1976. Dashed vertical lines mark midnight. The shoot has substantial contractions during the middle of the day and most rapid extension during the early part of the night. Note that the amplitude between maximum and minimum values generally increases over the period. Bottom line: hourly calculated transpiration using the Penman–Monteith equation. Note that maxima generally occur after the time of greatest shoot contraction.

Contractions, the shrinking of newly formed tissue, are considered to result from a decline in tissue water potential with re-expansion due to a gain in water potential. Milne et al. (1983) made a time series analysis of the relationship between transpiration, shoot water potential and the contraction and re-expansion of stem diameter, measured with linear variable differential transformers (LVDT) recorded electronically, on trees in the same forest. They found transpiration rate and shoot water potential were in phase but changes in stem radius lagged up to 3 h behind them. Calculation of hourly transpiration (Fig. 1) was made using the Penman–Monteith equation (Monteith, 1965; see Milne et al., 1983; Milne, 1989 for details of calculation) and this method has been tested against direct measurements of evapo-transpiration (Milne, 1979). In the shoot extension data contraction appears to be rapid and, in apparent contrast to the changes in stem diameter, minimum extension values are reached a number of hours before maximum transpiration. Re-expansion appears more gradual than contraction but it is not immediately clear if a distinction can be made between re-expansion and the growth of new tissue. Furthermore the amplitude of the daily change is not constant.

Important questions include: is contraction more rapid than re-expansion? Is the full amount of a daytime contraction regained in the following night period? Why does the amount of contraction vary between days with similar calculated transpiration (compare days 180, 181 and 182; Fig. 1)? A process-based model should be able to simulate different characteristics of the time series; hence we should use different criteria to assess each characteristic.

The time series model of daily shoot extension, y at day d, solved for measured daily increment of a leading shoot (Ford et al., 1987b) was

$$y_d = 0.133T_{d-1} - 0.042T_{d-2} + 0.0107R^D_{d-2} + 0.0150R^D_{d-3}, \quad (1)$$

where T is the mean daily temperature (°C) and R^D is the daily radiation (MJ m^{-2}). We start with this overall structure of lagged temperature and radiation effects, so, for each hour daily average temperature and radiation over the lagged values were calculated. This starting point provides a mean hourly shoot growth rate that varies little over the limited periods we examine and our task is to investigate model structure and parameter values for contraction and expansion that explain the diurnal variation around that mean.

Our initial model for hourly extension is

$$S_t = x_1 \left(\frac{1}{24} \sum_{k=1}^{24} T_{t-k} \right) - x_2 \left(\frac{1}{24} \sum_{k=25}^{48} T_{t-k} \right)$$
$$+ x_3 \sum_{k=25}^{48} R^H_{t-k} + x_4 \sum_{k=49}^{72} R^H_{t-k} + x_5 \Delta_t D \sum_{k=1}^{24} S^*_k \quad (2)$$

with the last term describing contraction and re-expansion. The four parameters to be estimated for temperature and radiation effects, x_1–x_4, correspond to the coefficients given in Eq. (1). Note that the temperature, T (°C), and radiation, R^H (MJ m^{-2}), contributions to growth are re-calculated for each hour and the contraction/expansion component is added to that—contraction does not inhibit growth and re-expansion does not accelerate it. For the last term we calculate two components: a water deficit and a response in shoot extension to a change in that deficit (Table 1a). Deficit, D, is the balance between loss of water from the plant through evapo-transpiration and replenishment by an estimated value for uptake from the soil. We assume that water loss is equivalent to evapo-transpiration as calculated from the Penman–Monteith equation (Monteith, 1965). That calculation requires a measure for canopy resistance (Milne et al., 1983; Milne, 1989). We estimate a parameter, x_6, for the rate of water uptake per unit of deficit (mm uptake/mm deficit/h). So

---

Table 1 – Developmental stages in modeling water deficit and its effect of contraction and re-expansion of shoot extension

(a) Calculation of the last term of Eq. (2)
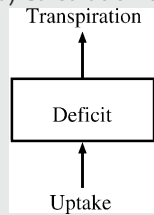
Uptake = deficit × x_6; contraction and re-expansion = $\Delta D \times x_5$ Fig. 5a and b

(b) The replacement of a single contraction and expansion parameter in that last term
Uptake = deficit × x_7; contraction = $\Delta D \times x_5$; re-expansion = $\Delta D \times x_6$ Fig. 5c

(c) The addition of a further expansion component to that shown in (b)
Uptake = deficit × x_7; contraction = $\Delta D \times x_5$; re-expansion = $\Delta D \times x_6$; when deficit = 0 and   Figs. 5 and 7
$\sum$ contraction > $\sum$ re-expansion then re-expansion = $(\sum$ contraction $- \sum$ re-expansion$) \times x_8$

For each stage hourly deficit, mm, is calculated as the sum of transpiration up to that point in the day minus an hourly uptake and the change in deficit, ΔD, for each hour is used to calculate contraction, if ΔD is positive, and re-expansion if ΔD is negative. Parameter values for these versions of the model are given in Table 3. The figures referenced in the right hand column indicate where the defined model is graphed.
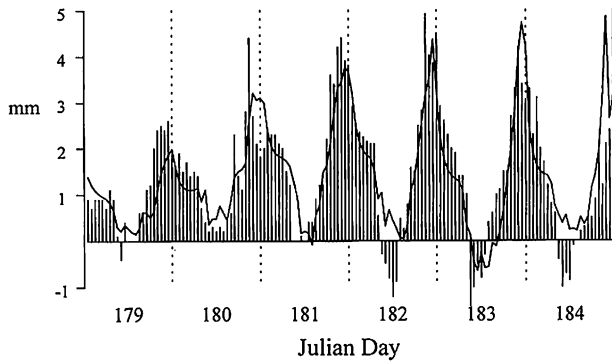
Fig. 2 – The initial model, Eq. (2) (solid line) as fit to data (histogram lines) using the simplex method to minimize residual sum of squares between model and data. Parameter values for this model are given in Table 3.

the change in deficit for any hour is

$$\Delta D_t = W_t - x_6 D_{t-1}, \tag{3}$$

where W_t is the calculated evapo-transpiration for that hour. The last term of Eq. (2) calculates change in contraction or re-expansion dependent upon the change in deficit and with x_5 an estimated parameter. As water deficit increases, i.e., when hourly transpiration values exceed water uptake, then we estimate the shrinkage in that hour as $x_5 \Delta D \sum S^*$, where $\sum S^*$ is the previous day's net increase in shoot length and used as an initial estimate of the amount of tissue that can shrink, i.e., has not undergone sufficient cell wall thickening to prevent shrinkage. Subsequently we investigate the time period over which $\sum S^*$ should best be calculated. As the deficit starts to decrease then re-expansion is calculated as $x_5 \Delta D \sum S^*$. Initially, Eq. (2), just one parameter, x_5, is used for both the contraction and re-expansion phases but a question of interest is whether the magnitude of contraction and re-expansion is identical in response to ΔD.

To illustrate the difficulty in assessing a model using a single criterion we use the residual sum of squares (RSS) between model and data (Fig. 2) and the downhill simplex method of optimization (Nelder and Mead, 1965) with the *Numerical Recipes in C* implementation of Press et al. (2002). Parameter

values minimizing RSS for days 179–184 were sought. During day 178 the calculation of water deficit equilibrates so we do not use values from that day in any of the optimizations described. The RSS from a repeated simplex fit was 14.97% so one approach is to say that the model explains 85% of the variation in the data. However, problems with this attained fit are apparent from Fig. 2. The largest deviations between model and data occur around the minima and maxima, i.e., when the shoot shows maximum contraction and expansion, which are features of particular interest in the development of the contraction/expansion component for the model. The period of minimum values is underestimated for each day. The data-model series for Fig. 2 has a partial residual autocorrelation term at lag 1, $\alpha(1) = 0.56 \pm 0.17$ ($p < 0.05$) reflecting this pattern.

In multi-criteria assessment one can focus on the attainment of minima and maxima by selecting criteria specifically defined on them. Our objective is not to achieve a lower residual sum of squares, although we do, but to model the process of contraction and expansion so that patterns in the data are explained (Wiegand et al., 2003).

3. The Pareto optimal set of model solutions

Assessing a model using multiple criteria reveals that some arrays of parameter values satisfy some assessment criteria while other arrays of parameter values satisfy others (Reynolds and Ford, 1999). This can be useful information. It suggests the model has not completely failed but is only partially successful and how different parameter values can satisfy different groups of criteria may indicate how model structure can be improved.

We use four concepts: binary discrepancy measures, the Pareto set, the Pareto frontier, and parameter bounds (Table 2). A binary discrepancy measure (Hornberger and Spear, 1981) defines a range of assessment criterion values considered as successful achievement of the criterion. If we chose, say, four assessment criteria, and set binary discrepancy measures within which they each must be achieved, then we may obtain a solution such as illustrated in Fig. 3. The one, or more, individuals in Pareto group 1 are arrays of parameter values that produce outputs within the specified discrepancy bounds for each of assessment criteria A, C and D. These

Table 2 – Definition of terms used in multi-criteria assessment	
Term	Definition and explanation
Individual	An array of values, one for each parameter of a model, which produces a unique model output. A number of individuals form a population
Assessment criterion	An objective function identified with a specific assessment characteristic. The vector of assessment criteria is used to define the multi-criteria optimization problem
Binary discrepancy measure	A binary discrepancy measure is where maximum and minimum values are placed on what will be considered a successful achievement of an assessment criterion. This enables classification of an individual's performance relative to a specific assessment criterion as: 1 if divergence between model and data is within the discrepancy and 0 otherwise
Pareto group	A set of assessment criteria that are achieved by one, or a number of, non-dominated individuals. Requires use of binary discrepancy measures
Pareto frontier	The set of all Pareto groups
Pareto set	The set of all non-dominated individuals

Criteria

Fig. 3 – Illustration of a set of assessment vectors where four criteria are not satisfied by any one group. Satisfied criteria are represented by black squares, unsatisfied criteria by white squares. Individuals in groups 1 and 2 both achieve three criteria and are non-dominated but those in group 3 and group 4 are both dominated by group 1.

individuals are superior to those in group 3 because they achieve assessment criterion D and superior to those in group 4, because they achieve criterion A. Individuals (i.e., arrays of parameter values) in groups 3 and 4 are dominated by Pareto group 1. However, Pareto group 2 achieves an assessment criterion that individuals in Pareto group 1 do not achieve, and vice versa, so the individuals in Pareto groups 1 and 2 are both members of the Pareto set, i.e., the collection of non-dominated individuals.

In this work, we use an example where structural components are added to a model to explain particular features in a data set. It is important to note that *all* versions of the model can fit the data to some extent. We need to know how a specific change in model structure may produce an improvement by accounting for some pattern in the data. Of course the ability of a model to satisfy all criteria simultaneously is directly related to the widths of the binary discrepancy measure chosen for each criterion. In our application we chose to use the same width for all criteria although this is not essential and widths could be allowed to vary if there is different confidence in particular criteria. Our approach is to find the width of the binary discrepancy measures at which the model is able to satisfy all of the criteria, then decrease that width, find which criteria are failed, and consider how a structural change may enable the model to achieve all criteria at this reduced binary discrepancy width. Comparing the contrasting parameter values obtained for different Pareto groups in a Pareto frontier can indicate what type of improvement needs to be made.

4. Evolutionary algorithm solution for the Pareto set

Pareto_Evolve (Reynolds, 1997; Komuro, 2005) is an EA based on the non-dominated sorting genetic algorithm (NSGA) (Srinivas and Deb, 1995). The search for a Pareto set begins with an initial population, for our problems typically 100 individuals, with their parameter values selected randomly from within the specified bounds and a search will continue for 500 or more generations. At each generation model outputs and values for the assessment criteria are calculated for each individual in the population. Each individual X is assigned a *fitness* which measures how good X is as a solution to the multi-objective function $F(X)$. Based on the fitness values there is then *selection* of some individuals to be parents for the next generation. Where a number of individuals have similar fitness then random selection is made between them. The selected parents produce offspring (*reproduction*) either by exchanging some parameter values between two parents, called *crossover* (Bäck et al., 1997), or changing some parent parameter values in small amounts, called *mutation* (Fig. 4). The cycle of assigning *fitness*, *selection* and *reproduction* is repeated until a maximum allowed number of generations is reached.

Many EAs use a procedure called *elitism* (Deb et al., 2000) to preserve individuals that have emerged as more successful. This can make a search faster and ensure the random changes produced by crossover and mutations do result in progress towards a solution. Elite individuals are chosen from non-dominated groups and are preserved for a number of generations. We found that a Pareto frontier can degenerate if elitism is not used, i.e., favorable individuals can be lost (Komuro, 2005) which happens because both crossover and mutation change the individuals they are applied to. In Pareto_Evolve, elites are preserved for a number of generations.

After extensive testing (Komuro, 2005) the following rules for selecting and maintaining elites were constructed:

(a) If a Pareto group achieves the most criteria in the Pareto frontier of the current generation then some of its individuals are stored as elites in an *external pool*.
(b) If a Pareto group achieves more criteria than at least one of the Pareto groups in the Pareto frontier defined at the previous generation then some of its individuals are stored as elites in the external pool.
(c) The external pool lets an elite survive until a better one appears. All elites stored in the external pool are compared with the Pareto frontier at the beginning of the next generation and are removed if better ones have been produced.
(d) Where a criterion remains unachieved by any of the Pareto groups chosen under conditions a–c, and if there exist Pareto groups satisfying this criterion in the current generation, the one achieving the most criteria is chosen among them.

The external pool of elites should not contain many individuals numerically close to each other and having the same assessment vectors. The number of elites corresponding to each Pareto group selected under one of the above four conditions is determined by

l = {population size}/{number of Pareto groups in

the current generation Pareto frontier}.

If the number of individuals in a Pareto group is smaller than or equal to the limit l, all of the individuals are accepted. Otherwise, only l individuals are chosen based on a measure of
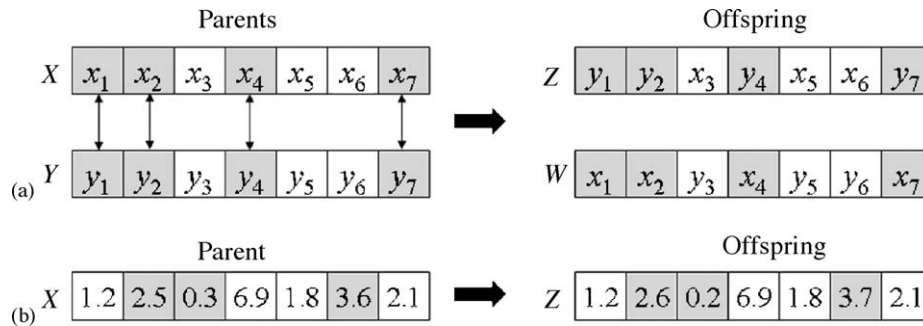
Fig. 4 – (a) Crossover randomly selects parameters from two randomly selected parents and exchanges their values to generate two offspring. In this example each individual has seven parameters. For two parents $X = (x_1, x_2, \ldots, x_7)$ and $Y = (y_1, y_2, \ldots, y_7)$, the first, second, fourth and seventh parameters are randomly chosen to be crossed over, producing offspring Z and W. (b) Mutation randomly selects parameters from a randomly selected parent and changes the parameter values by small, randomly generated amounts relative to the search range width. For example, each individual has seven parameters, and, for parent $X = (x_1, x_2, \ldots, x_7)$, the second, third, and sixth parameters are randomly chosen to be mutated to generate offspring Z. The random change is 0.1 with parameter values in the range [0, 10]. Whether the value is added or subtracted from the original parameter value is also determined randomly. Mutation is valuable for local searching in the objective space.

crowdedness in the objective space (Deb et al., 2000; Komuro, 2005) are chosen. If many Pareto groups are obtained l does not become large. In selecting elites a balance is maintained between the members of the Pareto set and this in turn helps to maintain the breadth of the search process. The number of generations for which an elite is maintained can be varied.

Using the described rules for elitism we found that the search to be most effective if crossover and mutation probabilities are dynamic (Komuro, 2005). If an elite individual is chosen to be a parent, then it is inefficient to apply mutation, i.e., to change parameter values by small amounts because the created offspring stays numerically close to the parent in the search space. When using binary discrepancy measures the task is to achieve additional criteria, rather than to improve the accuracy of presently achieved criteria so it is more effective to apply crossover to the parent so that the search becomes broad. If a parent has not been selected as elite it is mutated.

Pareto_Evolve maintains an archive of all non-dominated solutions encountered during the search. This archive is updated each generation so that even if Pareto optimal solutions are lost from a specific generation, they are retained in the archive. The archive is the source of the final Pareto frontier estimate, not the generation-specific Pareto frontier.

Searching for solutions with Pareto_Evolve uses random procedures in crossover, mutation, and the selection of elites. Consequently it is important to repeat searches and to examine the variability of solutions that may be obtained.

5. Model development using Pareto_Evolve

The failure of simplex minimization of RSS to obtain a model capturing daily minima could be due to an incomplete or incorrect model and/or to the choice of criteria—minimization of RSS for data of this type will seek a solution where minima and maxima of model and data are in phase with the size of extreme points being secondary. For multi-criteria assess-

ment we initially use four criteria for each day: deviations from the measured mean shoot extension for hours 3 through 6, 9 through 12, 13 through 16, and 21 through 24. These represent, respectively: the pre-dawn expansion period when we might expect that water deficit would be at its lowest; the period of maximum contraction; the period where recovery from contraction appears to start; the period of maximum expansion which occurs in the hours before midnight. All criteria use the same binary discrepancy measure defining the acceptable deviation, + or −, from measured values.

The model is first applied to days 179–181 and then days 182–184 separately. There is a difference between these two parts of the sequence with lower minima and greater amplitude of variation over days 182–184. Deans (1979) shows that there is an increase in soil moisture tension and a decrease in fine root length over this time period in this forest so there is the possibility that additional processes may become important over this time.

An initially large binary discrepancy measure of $\pm 1.5\,\text{mm}\,\text{h}^{-1}$ produced a Pareto frontier of one group satisfying all criteria for days 179–181. This discrepancy value was decreased for successive runs of Pareto_Evolve until the smallest binary discrepancy measure achieving a Pareto frontier of one group, i.e., all 12 criteria, was $\pm 0.45\,\text{mm}\,\text{h}^{-1}$ (Fig. 5a). This value is indicated on a scale at the top of Fig. 5a along with short horizontal lines showing actual values that were attained for each of the 12 criteria that indicate whether the model over- or underestimated that data for that period. The dot–dash line across the middle of Fig. 5a is the overall trend defined by the lagged temperature and radiation terms. Visually (Fig. 5a) the model fit is similar to that obtained using the simplex procedure but now we can quantify the direction of the failures. The criteria closest to the $\pm 0.45\,\text{mm}\,\text{h}^{-1}$ limit, and where the model failed when binary discrepancy measures were decreased, are the contraction periods of days 180 and 181. A plot (not shown) of the pattern of the parameter search for the
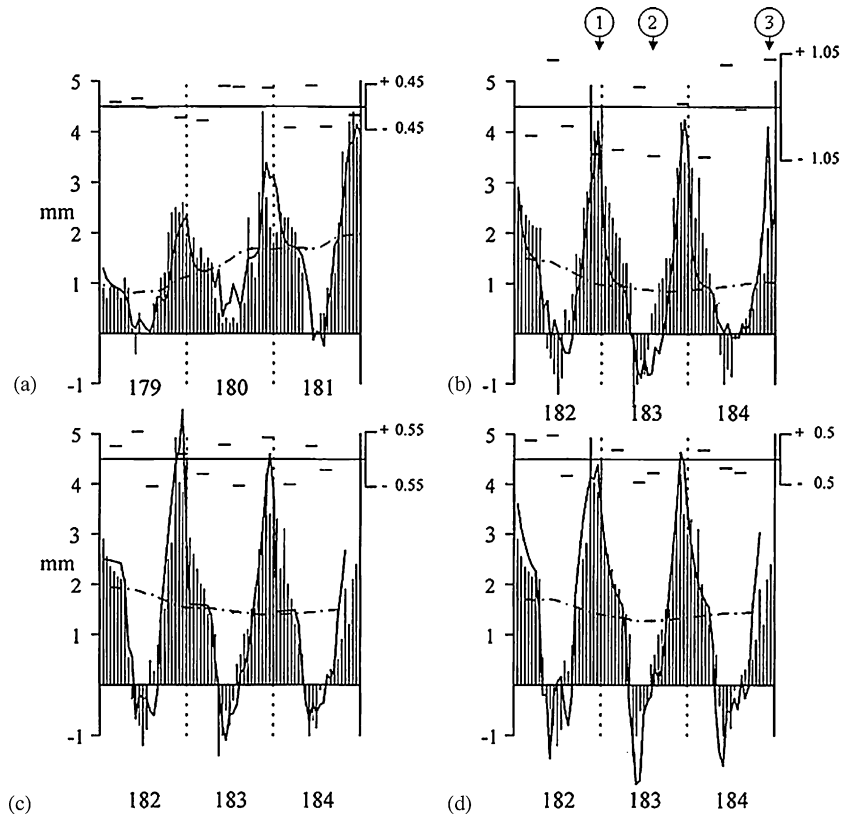
Fig. 5 – Each graph consists of measured data (vertical bars), and model values (continuous line), and trend contribution to the model (dash–dot line) determined by lagged temperature and radiation. Above each is a continuous horizontal line that provides a zero reference for the assessment criteria used in the model fitting procedure. At the right end of that line is the scale defining the smallest binary discrepancy measure which the model satisfied for all criteria graph and above and below the reference are lines indicating the particular value achieved for each criteria. In (b) circled figures indicate the criteria failed in each of three Pareto groups that form the Pareto frontier when binary discrepancy measures are decreased. Parameter values for these models are given in Table 3.

model in Fig. 5a showed that stable parameter values had developed.

Pareto_Evolve was applied to the model defined in Eq. (2) for days 182–184 (Fig. 5b) and compared to the fit obtained for days 179–181. A single Pareto group achieving all criteria was achieved at minimum binary discrepancy measure of no less than $\pm 1.05\,\mathrm{mm\,h^{-1}}$ over the 4 h periods—more than twice the minimum binary discrepancy measure achieved for days 179–181 with the model over-estimating extension during the period of pre-noon shrinkage and under-estimating it during the post-midnight contraction. This model (Fig. 5b) has a greater rate of contraction and expansion per unit change in water deficit (parameter x_5) than for days 179–181 and a lower rate of water uptake per unit deficit (x_6) (Table 3). We conclude that the model is inadequate, particularly in its inability to achieve minima effectively. The differences in parameter values between the two periods indicate how the fitting procedure attempts to accommodate the model to the data and in particular to achieve the larger changes in amplitude through changes in parameter values—when really a more effective model structure is required. Increase in x_5 causes more rapid change in extension per unit change of deficit but decreasing the rate of uptake prolongs the duration of the deficit enabling

the model to achieve the maximum extension rate in the late evening hours. Differences in parameters x_1–x_4 (Table 3) cause change in the overall trend of growth but do not contribute directly to the diurnal cycles of contraction and expansion.

When the binary discrepancy measure was decreased to $\pm 1.0\,\mathrm{mm\,h^{-1}}$ for the model in Fig. 5b a Pareto frontier formed with three Pareto groups each with one failed criteria (Table 4). Interestingly the first two of these groups have similar values of x_5 to that obtained for the model of Fig. 5a, although x_6, the uptake rate is lower which, by enabling a larger deficit to develop, may enable the minima to be achieved more effectively. Group 3 individuals, which fail to meet the final criteria on day 184, have a greater uptake rate and response to changes in deficit. The measured recovery from deficit on day 184 is less rapid than on previous days and this also occurred on day 185 (not shown) and may be an indication of some change in the controlling eco-physiological processes.

The flexibility provided by x_5 and x_6 is insufficient to provide a model that can capture minima and maxima effectively—we conclude that the model structure is inadequate. Three changes were made to the model. (i) Two parameters are used to estimate the effects of changes in water deficit on extension: one when deficit increases and con-

Table 3 – Model parameters used to produce output in Figs. 2, 5 and 7

	x_1	x_2	x_3	x_4	x_5	x_6		
Fig. 2	0.1285	0.1129	0.0519	0.0013	0.1622	0.4991		
Fig. 5a	0.0955	0.0727	0.0330	0.0250	0.1865	0.5890		
Fig. 5b	0.0090	0.0765	0.1135	0.0015	0.2090	0.5420		
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
Fig. 5c	0.0005	0.0510	0.0975	0.0185	0.3650	0.2910	0.8555	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Fig. 5d	0.0925	0.0720	0.0435	0.0040	0.5700	0.1625	0.7550	0.4625
Fig. 7	0.0190	0.0490	0.0690	0.0400	0.6240	0.3910	0.7790	0.2030

In each case parameters x_1–x_4 are those defined in Eq. (2). For Figs. 2 and 5a and b, x_5 is the contraction and or expansion, mm, per unit change of calculated water deficit and x_6 is the estimated water uptake rate, mm per mm of deficit. For Fig. 5c there are separate contraction, x_5, and expansion, x_6, parameters and x_7 is the uptake parameter. Fig. 5d has the same contraction, expansion and uptake parameters as Fig. 5c and additionally parameter x_8 which is the hourly rate of re-expansion of the contraction–expansion difference existing when calculated water deficit is zero. For Figs. 5 and 7 one representative individual from the final population of the Pareto_Evolve search is shown: for Fig. 5a and b these were selected from 100 individuals after 500 generations; for Figs. 5c, d and 7 these were selected from a population of 300 after 1000 generations.

traction takes place and the other when deficit decreases and expansion occurs (Table 1b). This is the most important change because it allows for a hysteresis effect in contraction and expansion of extensible tissues under changes in water stress where contraction is greater than re-expansion per unit change in deficit as reported by Murphy and Ortega (1996) for pea seedlings. (ii) The period used to calculate the amount of shoot available to be contracted was changed from the whole of the previous day to the period between 23 h of the previous day and 6 h of the day in which the contraction takes place. This was estimated by seeking the optimal time using Pareto_Evolve. (iii) Arithmetic smoothing is made in calculation of water uptake so it depends on an average of the last 3 h. This prevents short-term oscillations during contraction and expansion. Also, during model development it was found that no model captured the final criteria effectively, i.e., pre-midnight expansion on day 184. Inspection of the longer period of measurement suggested a change in pattern of extension from that time onward so this final criterion was dropped from the assessment criteria.

In the resulting model (Fig. 5c) $x_5 > x_6$ (Table 3) giving re-expansion per unit change in deficit to be less than contraction, and estimated uptake rate, x_7, to be large compared with previous model versions. Although this model captures both maxima and minima it fails to capture the gradual decline of extension between midnight and dawn. Examination of the Pareto frontier when binary discrepancy values were decreased showed a division into groups that achieved a gradual post-midnight decline but failed to achieve the post-

noon re-expansion and vice versa. This suggested there was a longer period of re-expansion than simulated, or an increase in real growth during the period when deficit is zero, i.e., midnight to dawn. The gradual decline in measured extension from midnight suggests a discrepancy in model calculation of re-expansion. Examination of the calculated water deficit term produced by the model in Fig. 5c showed it was reduced to zero by midnight, but by that time the total calculated re-expansion was substantially less than the total contraction.

The assumption was then made that re-expansion continues after water deficit has been reduced to zero, i.e., there is a lag before full re-expansion is attained. This is, in effect, completion of the hysteresis cycle of contraction and re-expansion. The model was modified (Table 1c and Fig. 5d) so that the difference between contraction and expansion was calculated at the time of zero deficit and then distributed over subsequent hours with an estimated constant arithmetic decline rate, x_8, applied to the remaining (contraction–expansion) term at each hour. This model does capture the observed pattern of decline but it was necessary to change the assessment criteria to estimate parameter x_8. Arithmetic differences for the mean of four successive values provide information on whether the model under or overestimated measured extension but when used at the point of the maxima, large compensations could occur in the calculation of the mean hourly values for the criterion value, e.g., two large under estimates with two large over estimates. Consequently the assessment criterion for this time period was changed to a mean squared difference. Other model formulations were examined to see if they could simu-

Table 4 – Criterion failed and example parameter values for the three Pareto groups that formed when binary discrepancy measures were reduced to $\pm 1.0\,\mathrm{mm\,h^{-1}}$ for the model in Fig. 6b

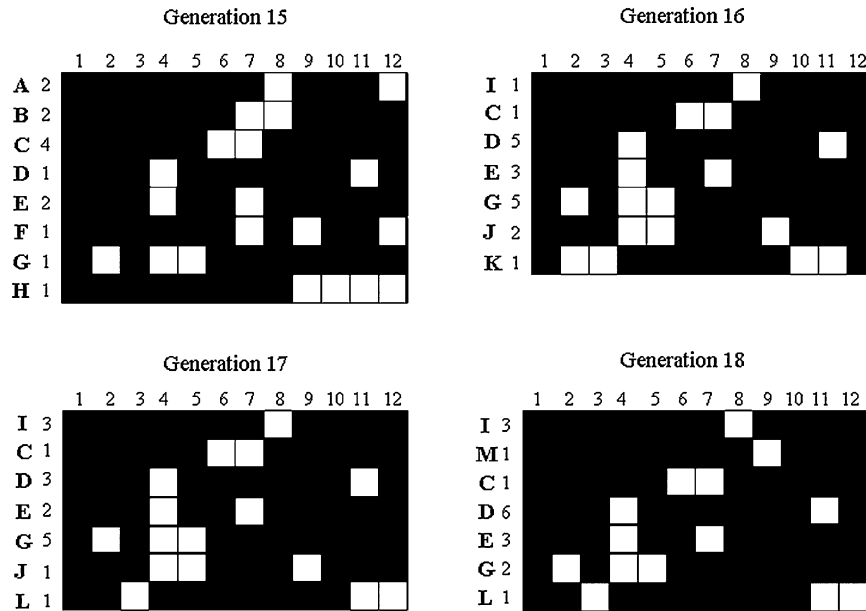| Group	Criterion failed	x_1	x_2	x_3	x_4	x_5	x_6
1	Day 182, underestimated pre-midnight expansion	0.0025	0.0550	0.0895	0.0100	0.1820	0.4680
2	Day 183, underestimated early afternoon expansion	0.0135	0.0725	0.0775	0.0270	0.1850	0.3015
3	Day 184, exceeded pre-midnight expansion	0.0335	0.1005	0.1070	0.0135	0.2805	0.7570

Fig. 6 – Pareto frontiers in the breeding population of successive generations in the search that produced the model shown in **Fig. 5**a. Criteria are numbered along the top of each diagram. Satisfied criteria are represented by black squares, unsatisfied criteria by white squares. Each group is labeled alphabetically and the number of individuals represented in the population of 100 is listed beside the label. Note that at generation 16 group I appears and dominates both A and B of generation 15 and so these groups are dropped. Groups C–E, each achieving 10 criteria, remain through these four generations.

late the midnight to dawn decline effectively. These included the use of a two compartment water model, i.e., foliage and wood with a transfer between them, and use of a root resistance. Both of those versions of the model introduced delays that substantially reduced re-expansion in late afternoon and early evening so that the criteria for those times were failed by large amounts and those changes in model structure were rejected.

The model of Fig. 5d captures minima, maxima and the gradual post-midnight decline in extension and reduces the binary discrepancy range at which all criteria are achieved. The overall residual sum of squares was 11.4% and the partial autocorrelation term of residuals at lag 1 was 0.24 ± 0.24 suggesting that where model-data discrepancies occur they do so for a number of hours, e.g., during the early morning hours of day 182 and in the middle of day 183. However, they do not occur consistently at the same points in the diurnal cycle (Fig. 6).

The model in Fig. 5d was then fit to days 179–184 using the same criteria for each day, i.e., 23 criteria in all (Fig. 7). This was the largest assessment made: using a population of 200 individuals and 1000 generations took 5 min 35 s on a 3.6 GHz computer with 4 GB of memory and running in Microsoft Visual C++. This model captures more features in the data than that shown in Fig. 2 but there are still some discrepancies between model and data. The most notable is the large modeled contraction for day 181 and the failure to capture pre-midnight expansion of days 179 and 181.

A comparison of parameter values for the model fitted for days 182–184 (Fig. 5d) and for days 179–184 (Fig. 7) shows estimated uptake rates, x_7, are within 1% of each other (Table 3) and contraction rate, x_5, for days 182–184 is 91% of that for

days 179–184. However, re-expansion rate, x_6, for days 182–184 is only 42% of days 179–184. So, when estimated for the 6 days sequence re-expansion and to a lesser extent contractions are estimated to be more rapid. Consequently, the amount and rate, x_8, of re-expansion added after the daily water deficit is reduced to zero is less, i.e., the x_8 value for days 179–184 is 44% of that for days 182–184. These differences in parameter values suggest that there may be additional processes not included in the model. For example, Génard et al. (2001) show the hysteresis effect can be greater when water potential is lower and there may be changes over the 6-day period not captured in the water deficit calculation.
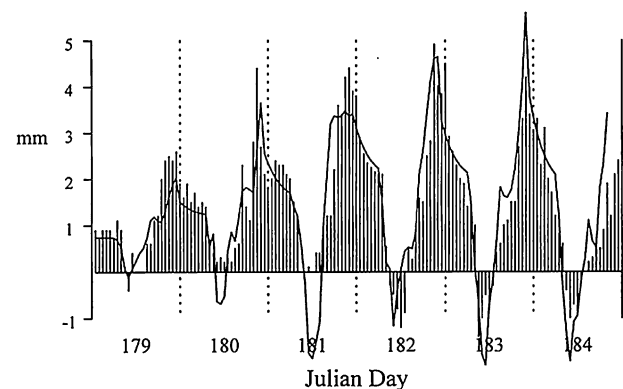

Fig. 7 – A model with different parameters for contraction and re-expansion rates per unit change in deficit, and a parameter to complete re-expansion in the early morning hours following removal of water deficit. Parameter values are given in **Table 3**.

6.   Discussion

Multi-criteria assessment enabled us to answer three questions: is contraction more rapid than re-expansion? Yes, because models with a separate parameter for contraction and expansion per unit change in deficit, and with the expansion parameter less than that for contraction, could fit multiple criteria, i.e., those defining both minima and maxima. Is the full amount of a daytime contraction regained in the following night period? Yes, in the final model. However, a component of re-expansion was required after the removal of water deficit to capture the criteria of post-midnight decline. Why does the amount of contraction vary between days with apparently similar transpiration? There are differences in the amount of growth available for contraction between different days.

An important consideration is how assessment criteria and their binary discrepancy measures should be selected. The procedure we use here has three components.

- Criteria are *chosen* to test whether the structure of the model is correct. Consequently the resulting model is not ubiquitously correct—it explains particular features of the data with a certain degree of accuracy. This means that a model and the assessment criteria used in its development must be considered together as a couple.
- Assessment criteria should be informative quantitatively. Initially we used differences between model and data so that we obtained information whether the model was under or overestimating sequences in the data. Subsequently we changed one of the criteria to least squares when accuracy of fitting the model at a particular point became important. Crucially we make decisions whether changes in model structures are improvements or not by the width of the binary discrepancy measures they require to achieve a Pareto frontier with one Pareto group achieving all criteria.
- Binary discrepancy measures are valuable when, as illustrated in this work and in Reynolds and Ford (1999), a model is being formulated and the principal objective is to develop an effective model structure. The task is to ensure that the model simulates a number of different patterns in the data and to identify which patterns are most difficult to achieve. Once a decision on model structure has been made then it may be possible to analyze its performance for particular assessment criteria by direct discrepancy analysis and without using the binary classification.

The next phase in development of this model would be to apply it to longer periods of data, and to parallel data sets for other shoots. This should be done in an exploratory rather than confirmatory framework—we should not expect to verify or reject the model of Table 1c in a straightforward way. This is because there are already indications of lack of constancy in parameter values, e.g., between the model fits used to construct Figs. 5d and 7, as well as differences in what the model may have to explain, e.g., the reduced early nighttime expansion of day 184. The exploration should be whether model structure has to change to deal with these and possibly other issues. In particular, as the duration of time modeled is extended, there has to be an investigation of factors that control the underlying growth process. In this way model development is a guide to constructing a theory for shoot extension but it may not be appropriate to anticipate that a model of best fit will encapsulates that theory.

These model development and assessment procedures support Wiegand et al. (2003) and Grimm et al. (2005) who suggest that during model development greater use be made of patterns that occur in data. However, it is important that patterns needing to be explained should be selected in a rational and quantitative way. Multi-criteria assessment using binary discrepancy measures provides an effective method for the example we use. This method recognizes that many models may be fitted to a data set, but drives model development so that the largest deviations between model and data are reduced by adding new features to the model.

The Pareto_Evolve software along with an example of how it can be used is available at http://faculty.washington.edu/edford/.

REFERENCES

Auger, P., Charles, S., Viala, M., Poggiale, J.-C., 2000. Aggregation and emergence in ecological modeling: integration of ecological levels. Ecol. Model. 127, 11–20.

Bäck, T., Hammel, U., Schwefel, H.-P., 1997. Evolutionary computation: comments on the history and current state. IEEE Trans. Evol. Comput. 1, 3–17.

Deans, J.D., 1979. Fluctuations of the soil environment and fine root growth in a young Sitka spruce plantation. Plant Soil 52, 195–208.

Deb, K., 2001. Multi-objective Optimization using Evolutionary Algorithms. John Wiley and Sons, Chichester, New York.

Deb, K., Agrawal, S., Pratap, A., Meyarivan, T., 2000. A First elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: Proceedings of the Parallel Problem Solving from Nature VI Conference. Springer, pp. 849–858.

Ford, E.D., Milne, R., Deans, 1987a. Shoots extension in *Picea sitchensis*. II. Seasonal variation within a forest canopy. Ann. Bot. 60, 531–542.

Ford, E.D., Milne, R., Deans, 1987b. Shoots extension in *Picea sitchensis*. II. Analysis of weathers influences on daily growth rate. Ann. Bot. 60, 543–552.

Génard, M., Fishman, S., Vercambre, G., Huguet, J.-G., Claude Bussi, J., 2001. A biophysical analysis of stem and root

diameter variations in woody plants. Plant Physiol. 126, 188–202.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991.

Håkanson, L., 2003. Propagation and analysis of uncertainty in ecosystem models. In: Canham, C.D., Cole, J.J., Lauenroth, W.K. (Eds.), Models in Ecosystem Science. Princeton University Press, Princeton, pp. 139–167.

Hornberger, G.M., Spear, R.C., 1981. An approach to the preliminary analysis of environmental systems. J. Environ. Manage. 12, 7–18.

Kanninen, M., 1985. Shoot elongation in Scots pine: diurnal variations and response to temperature. J. Exp. Bot. 36, 1760–1770.

Komuro, R. 2005. Multi-objective evolutionary algorithms for ecological process models. PhD Thesis. University of Washington.

Milne, R., 1979. Water loss and canopy resistance of a young Sitka spruce plantation. Boundary-layer Meteorol. 16, 67–81.

Milne, R., 1989. Diurnal water storage in the stems of *Picea sitchensis* (Bong.) Carr. plant. Cell Environ. 12, 63–72.

Milne, R., Smith, S.K., Ford, E.D., 1977. An automatic system for measuring shoot length in Sitka spruce and other plant species. J. Appl. Ecol. 14, 523–529.

Milne, R., Ford, E.D., Deans, J.D., 1983. Time lags in the water relations of Sitka spruce. For. Ecol. Manage. 5, 1–25.

Monteith, J.L., 1965. Evaporation and Environment. Symposium of the Society for Experimental Biology, vol. 19. Cambridge University Press, pp. 205–234.

Murphy, R., Ortega, J.K.E., 1996. A study of the stationary volumetric elastic modulus during dehydration and rehydration of stems of pea seedlings. Plant Physiol. 110, 1309–1316.

Nelder, J.A., Mead, R., 1965. A simplex method of function minimization. Comput. J. 7, 308–313.

Norton, J.P., 1996. Roles for deterministic bounding in environmental modeling. Ecol. Model. 86, 157–161.

Oreskes, N., 2003. The role of quantitative models in science. In: Canham, C.D., Cole, J.J., Lauenroth, W.K. (Eds.), Models in Ecosystem Science. Princeton University Press, Princeton, pp. 13–31.

Press, W.H., Teukolsky, S.A., Vettering, W.T., Flannery, B.P., 2002. Numerical Recipes in C. Cambridge University Press, Cambridge.

Reynolds, J.H., 1997. Multi-criteria assessment of ecological process models using Pareto optimization. PhD Thesis. University of Washington.

Reynolds, J.H., Ford, E.D., 1999. Multi-criteria assessment of ecological process models. Ecology 80, 538–553.

Srinivas, N., Deb, K., 1995. Multiobjective optimization using nondominated sorting in genetic algorithms. Evol. Comput. 2, 221–248.

Wiegand, T., Jeltsch, F., Hanski, I., Grimm, V., 2003. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and conservation practice. Oikos 100, 209–222.

Wu, J., Marceau, D., 2002. Modeling complex ecological systems: and introduction. Ecol. Model. 153, 1–6.