



UNIVERSITY OF CALIFORNIA PRESS
JOURNALS • DIGITAL PUBLISHING



Using Multicriteria Analysis of Simulation Models to Understand Complex Biological Systems

Author(s): Maureen C. Kennedy and E. David Ford

Reviewed work(s):

Source: *BioScience*, Vol. 61, No. 12 (December 2011), pp. 994-1004

Published by: University of California Press on behalf of the American Institute of Biological Sciences

Stable URL: <http://www.jstor.org/stable/10.1525/bio.2011.61.12.9>

Accessed: 14/03/2012 19:26

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



University of California Press and American Institute of Biological Sciences are collaborating with JSTOR to digitize, preserve and extend access to *BioScience*.

Using Multicriteria Analysis of Simulation Models to Understand Complex Biological Systems

MAUREEN C. KENNEDY AND E. DAVID FORD

Scientists frequently use computer-simulation models to help solve complex biological problems. Typically, such models are highly integrated, they produce multiple outputs, and standard methods of model analysis are ill suited for evaluating them. We show how multicriteria optimization with Pareto optimality allows for model outputs to be compared to multiple system components simultaneously and improves three areas in which models are used for biological problems. In the study of optimal biological structures, Pareto optimality allows for the identification of multiple solutions possible for organism survival and reproduction, which thereby explains variability in optimal behavior. For model assessment, multicriteria optimization helps to illuminate and describe model deficiencies and uncertainties in model structure. In environmental management and decisionmaking, Pareto optimality enables a description of the trade-offs among multiple conflicting criteria considered in environmental management, which facilitates better-informed decisionmaking.

Keywords: Pareto, environmental management, optimal biological structures, model assessment, multiple criteria

Complex biological systems have many interacting components, system constraints, and trade-offs in system functions. We follow Gallagher and Appenzeller (1999), who wrote, "We have taken a *complex system* to be one whose properties are not fully explained by an understanding of its component parts" (p. 79). For example, the life of a grasshopper may seem simple; in order to survive, the grasshopper must forage effectively for energy-rich food, and it may be reasonable to assume that it forages so as to maximize its daily energy intake. Suppose that a research group establishes an experiment in which the foraging behavior of a grasshopper is observed, and to their surprise, they find that the grasshopper does not forage in a manner that would be considered optimal. They discover that they had failed to take into account the spiders that have entered the experimental unit, which pose a risk of predation to the grasshopper. The spiders are removed, and the foraging behavior of the grasshopper changes (Rothley et al. 1997). If we want to understand grasshopper behavior, we cannot evaluate just their energy intake; we also have to consider constraints on their behavior, such as avoiding predation. In many biological systems, the individual components do not interact in a straightforward modular fashion, and only limited information can be obtained by studying components in isolation (Mitchell 2009, Beardsley 2010).

We are dependent on models to help us understand the consequences of integrating the components important to

explaining how a complex system functions. The process of model development is complicated by the trade-offs inherent in these complex systems (e.g., should a grasshopper forage for food or remain vigilant to predation by spiders?) and in the often nonlinear and possibly irregular relationships among system components that cannot be reduced to a set of equations and that require process-based representations. It is also common in process-based models for the final model output to depend on intermediate calculations, and these calculations tend to be hidden within the model structure. These models can become complicated with many model outputs and intermediate model results.

For example, an optimal foraging model that maximizes foraging behavior for energy intake could be proposed to explain grasshopper behavior, and such a model would predict a single optimal foraging behavior. However, it was observed that the grasshopper exhibits variable foraging behavior (Rothley et al. 1997), so how can there be only one optimal solution? Since it was observed that grasshopper behavior changes in the presence of predation pressure (spiders), it may be that the grasshopper is balancing the trade-off between actively foraging and avoiding predation. Because both are necessary for survival, it is reasonable to modify the modeling study to optimize grasshopper behavior for both energy intake and predator avoidance.

This argument leads to the question of how the optimization of two quantities, which we will call *criteria* (box 1), is

Box 1. Key concepts for Pareto optimality.

A *criterion* is a measurement of model performance. For example, in an optimality study, a criterion might be a measure of organism performance, such as energy intake. In a model assessment, a criterion might be a goodness-of-fit statistic that compares a model output with its corresponding observed data set; in this case, the criterion is minimized. In environmental management, a criterion might be a goal of the management action, such as reducing fire risk.

For two model solutions (A and B), solution A is said to *dominate* solution B if solution A performs at least as well as solution B for all criteria and performs strictly better than solution B for at least one criterion. In table 1, there are four criteria that measure aspects of branch performance (the mean number of junction constrictions to terminal foliage, the mean path length to terminal foliage, the overall mechanical load of the branch, and the mean foliage overlap). Any branch able to minimize these criteria simultaneously is assumed to perform better than a branch with higher criterion values. Solution A dominates solution B because they perform

identically for the first and third criteria, but solution A performs better (has lower values) for the second and fourth criteria. Lower (preferred) values are marked by a footnote in the table.

If one solution does not dominate the other, the solutions are said to be *mutually codominant*. Solutions A and C are mutually codominant if improvement with respect to one criterion corresponds to a worse performance for another criterion. In table 2, solution C performs better for the first criterion, the two solutions perform identically for the second criterion, and solution A performs better for the third and fourth criteria. Lower (preferred) values are highlighted in the table. If we compare solution B (from table 1) to solution C (from table 2), we see that they are also mutually codominant.

The *Pareto optimal frontier* is the set of criteria vectors that represent mutually codominant solutions that are not dominated by any other solution (figure 1). This is defined for the *feasible model space* (the set of possible model outcomes subject to the optimization constraints). If solutions A, B, and C above constitute the feasible model space, solutions A and C represent the set of mutually codominant solutions not dominated by any other solution. Although solutions B and C are mutually codominant, solution B is not included in the Pareto optimal frontier because it is dominated by solution A.

The *Pareto optimal set* is the set of parameter values that comprise the solutions in the Pareto optimal frontier.

Table 1. Comparing two solutions for dominance with respect to four criteria to be minimized simultaneously.

	Number of junction constrictions to terminal foliage	Hydraulic path length to terminal foliage (centimeters)	Relative branch mechanical requirements (dimensionless)	Mean foliage overlap (number of shoots)
Solution A	5	200†	0.9	2.3†
Solution B	5	300	0.9	4.0

Note: Solution A dominates solution B.

†The lower (preferred) value for the criterion.

Table 2. Comparing two solutions for dominance with respect to four criteria to be minimized simultaneously.

	Number of junction constrictions to terminal foliage	Hydraulic path length to terminal foliage (centimeters)	Relative branch mechanical requirements (dimensionless)	Mean foliage overlap (number of shoots)
Solution A	5	200	0.9†	2.3†
Solution C	2†	200	1.1	3.5

Note: Solutions A and C are mutually codominant.

†The lower (preferred) value for the criterion.

accomplished. One idea is to combine the two criteria into a single value, which is then optimized. This requires the expression of the two criteria as a single currency in such a way that is reasonable to combine them. It is not obvious how to express the benefit of avoiding a predator in the same units one uses to express energy intake. In this sense, the two requirements are incommensurable, as is often the case in biological systems. There is no clear way to quantify the relative importance of these competing requirements, and we need to consider both of them, measured separately.

Pareto optimality (box 1) is a multicriteria optimization technique that allows one to characterize the trade-offs in multiple criteria that represent incommensurable quantities (figure 1). Pareto optimality is a concept attributed to the economist Vilfredo Pareto (Cirillo 1979) that has been recently used for the multicriteria evaluation of complex problems in ecology and biology (Rothley et al. 1997,

Schmitz et al. 1998, Reynolds and Ford 1999, Komuro et al. 2006, Vrugt et al. 2007, Kennedy et al. 2008, 2010, Kennedy and Ford 2009, Tóth et al. 2009, Turley and Ford 2009, Efstratiadis and Koutsoyiannis 2010, Rabotyagov et al. 2010, Thompson et al. 2010). In the grasshopper example, Rothley and colleagues (1997) used Pareto optimality to optimize both energy intake and predator avoidance (figure 1). The trade-off between the two is obvious: The more time a grasshopper spends avoiding predation, the less energy it can take in per day. When Rothley and colleagues (1997) conducted an experiment to record grasshopper behavior with varying predation pressure, they found that the grasshoppers' behavior followed a set of what is called *Pareto optimal solutions*. In the absence of spiders, the grasshoppers in the experiment achieved energy intake near the predicted maximum. In the presence of two spiders, the grasshoppers in the experiment achieved lower levels of energy intake and spent more time on predator avoidance. The values of both of these solutions

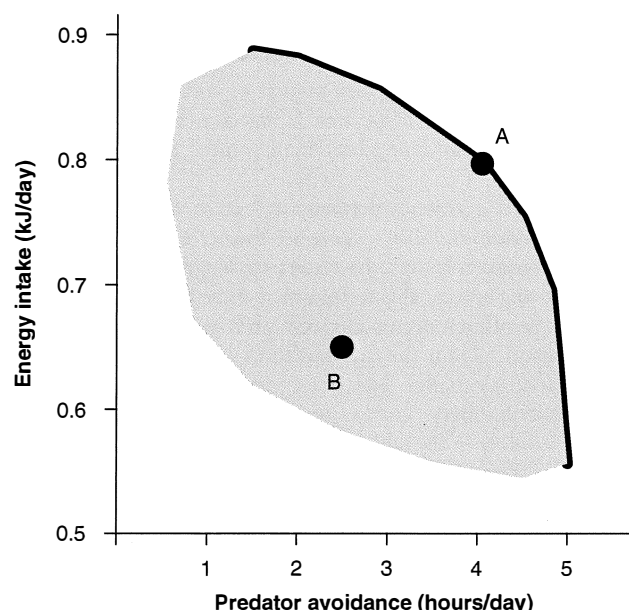


Figure 1. Hypothetical Pareto optimal frontier for the maximization of two criteria (energy intake and predator avoidance for a grasshopper foraging). Higher values are preferred for both criteria. The grey area denotes the feasible model space; the thick black line denotes the Pareto optimal frontier. Along the Pareto optimal frontier, there is no possible further improvement in any combination of energy intake and predator avoidance. Point B is not included in the Pareto optimal frontier because point A has higher values for both energy intake and predator avoidance. Abbreviation: kJ, kilojoules.

were predicted by the multicriteria model analysis using Pareto optimality. In this example, there is no single optimal solution for grasshopper behavior, and optimizing a single criterion would have been inadequate to explain this complex biological system.

Rothley and colleagues (1997) provided an example that uses Pareto optimality to optimize multiple incommensurable criteria that represent trade-offs in organism functions, in a kind of model analysis that we refer to as an *optimality study*. Pareto optimality is also used in model assessment: For a model to be deemed an adequate abstraction of a biological system, it should be able to replicate multiple system components simultaneously. The optimization of multiple criteria is also important in a decisionmaking context, in which there are competing stakeholders with differing priorities for a given environmental-management action and no clear way to combine those priorities into a single measure. In this article, we describe how to perform a multicriteria analysis for complex biological problems in three categories of model analysis. Then we give the context for and examples of multicriteria optimizations for these three kinds of model analysis.

Pareto optimality

In box 1, a four-criteria example of branch development in old-growth trees is used to provide descriptions and definitions of the concepts important to Pareto optimality. Here, we use a simpler, hypothetical two-criteria grasshopper optimal foraging example (adapted from Rothley et al. 1997) to illustrate the underlying concepts. In this hypothetical example, an optimal foraging model maximizes grasshopper behavior for two criteria—energy intake and predator avoidance (figure 1)—in which higher values are preferred to lower ones for both criteria. These two criteria are maximized under optimization constraints of digestive capacity, daily feeding time and minimum energy requirements (Rothley et al. 1997), which define feasible foraging strategies. In a hypothetical example for such a model, a maximum value of 0.87 kilojoules per day could be achieved for energy intake, but a grasshopper that achieves that level of energy intake may only be able to spend 1.5 hours per day being vigilant to predation (predator avoidance). This possibly leaves the grasshopper vulnerable to predation. A grasshopper could spend five hours per day being vigilant to predation, but the model estimates that such a grasshopper would only be able to achieve 0.56 kilojoules per day in energy intake, which is possibly suboptimal for grasshopper survival. For this hypothetical example, the solid line in figure 1 represents the points for which no further improvement in either energy intake or predator avoidance is possible (this is the *Pareto optimal frontier*), and the gray space represents all possible model outcomes (the *feasible space*). Along the solid line, any improvement with respect to energy intake corresponds to a worsened performance with respect to predator avoidance. If we compare points A and B in figure 1, we see that point A has a higher value than point B for both energy intake and predator avoidance, so point B is not considered to be Pareto optimal.

How to use Pareto optimality

There are typically four steps to using Pareto optimality to help understand complex biological systems, the details of which depend on what is being modeled. Recall that we are describing Pareto optimality for three main kinds of model analysis: (1) optimality studies (such as the grasshopper example), (2) model assessment (simulated outputs are compared with different empirically observed patterns), and (3) environmental management and decisionmaking (a management action is optimized for multiple competing priorities). The following four steps are adapted from Kennedy and colleagues (2008).

Step 1: Define the criteria. The choice of criteria drives the optimization results. The modeler should choose a set of criteria that capture the essential features of the biological system in question. In an optimality study, the criteria should be a set of requirements necessary for survival and reproduction (e.g., energy intake and predator avoidance). For model assessment, these criteria would be the set by

which, if all criteria are satisfied simultaneously, the model can be judged adequate for the stated model objective. The criteria for model assessment are usually determined by comparing the model with empirical observations by calculating an error measure. For environmental management and decisionmaking, the criteria should adequately cover the range of interests for competing stakeholders. In general, it is best to choose criteria that are unrelated (or as close to unrelated as possible). If two criteria are positively correlated, the same set of parameter values will be able to optimize both, and having a second criterion does not further elucidate model performance.

Step 2: Define the feasible space. The feasible space is subject to the constraints of the optimization. For example, in optimality studies, the feasible space should be defined by the structures available to the organism (e.g., digestive capacity, daily feeding time, and minimum energy requirements define the feasible space for the grasshopper foraging example). For model assessment, the modeler sets the feasible space by the range of parameter values that is explored in the optimization. This is often subject to theoretical constraints (e.g., a reproductive rate cannot be negative). In environmental management and decisionmaking, the feasible space should be defined by the actions available to the manager.

Step 3: Choose a model. There is no standard recommendation for how to choose a model for a given modeling exercise, because model selection is driven by the objectives of the analysis. In optimality studies and for environmental management and decisionmaking, a model is chosen that has been determined to be adequate for the question at hand. Presumably, the chosen model has already undergone an extensive assessment. During model assessment, a model structure is proposed, the parameter values are optimized for the multiple criteria, and a judgment is made of whether the model is adequate when compared with the set of criteria. If the model cannot satisfy all criteria simultaneously, the underlying structure is investigated for model deficiencies, the model is modified, and the optimization is repeated. This procedure is repeated until the model is found to be adequate for its set of criteria.

Step 4. Evaluate the Pareto optimal solutions. This step requires an optimization algorithm to search the feasible space and to return an approximation of the Pareto optimal frontier. An approximation must be used because it is usually computationally prohibitive to evaluate the entire feasible space. There is a large field of study surrounding the best algorithms for multicriteria optimization, usually variations of evolutionary algorithms (Deb 2001). These algorithms use evolutionary principles such as selection and breeding to produce generations of parameter values that are improvements on the previous generations and, in general, converge to reliable approximations of the Pareto optimal frontier.

Evaluation of the set of Pareto optimal solutions is necessary to understand the complex biological problem of interest. For optimality studies, the Pareto optimal solutions are compared with observed organism structure and behavior. For model assessment, the Pareto optimal solutions are evaluated for whether the model performs adequately for the set of criteria. For environmental management and decisionmaking, the Pareto optimal solutions are presented to the stakeholders to facilitate understanding of the trade-offs inherent in choosing one action over other feasible actions and a discussion of the merits of alternative actions.

Examples of multicriteria analysis

In optimality studies, the observed structure or behavior of an organism is compared with some state theoretically thought to optimize a measure of performance, such as fitness. This comparison requires that a model calculate the proposed optimal state. It is presumed in such studies that evolution works toward optimization (Smith 1978, Gould and Lewontin 1979, Parker and Smith 1990), although it is not assumed that organisms themselves are optimal (Parker and Smith 1990). In fact, organisms are rarely, if ever, optimal for a single feature (Honda and Fisher 1978, 1979, Farnsworth and Niklas 1995, Rothley et al. 1997, Vrugt et al. 2007), and organisms with different forms are able to survive and reproduce under similar constraints. That organisms are rarely optimal for a single feature does not preclude the use of optimality models to study organism form and function. Schmitz and colleagues (1998) recommended multicriteria optimization to reconcile variability in optimal behavior. Farnsworth and Niklas (1995) stated, "We have already indicated our belief that most, if not all, of the Pareto optimal solutions will be accepted in natural selection, given a range of environmental circumstances and strategy options" (p. 359).

Box 2 gives an example of an optimality study in which the persistence of old-growth trees is accomplished, despite constraints on net growth. It is well established that large, old trees exhibit a growth constraint whereby net growth is limited at a characteristic height in a given environment. In the old-growth forests of the Pacific Northwest, centuries-old Douglas fir (*Pseudotsuga menziesii*) trees have negligible annual growth with respect to height and crown expansion, yet the trees are projected to persist for centuries more. There are other species that also manage to persist in the old-growth canopy that exhibit clearly distinctive morphologies. There have been two main schools of thought to explain the growth limitation. The first is that increased maintenance respiration for large, old structures tips the carbon balance away from growth and toward maintenance (i.e., toward balanced carbon; Hunt et al. 1999). The second is that hydraulic resistance increases with increasing height, which causes stomata to close and which reduces photosynthesis (i.e., hydraulic limitation; Ryan and Yoder 1997). Hydraulic constrictions are also observed at branch junctions, which potentially reduce water potential at the

Box 2. Multicriteria optimization of branch development in old trees.

Kennedy and colleagues (2010) used multicriteria optimization of a process model of branch development in old-growth Douglas fir to optimize branch development for four design criteria simultaneously. The first two criteria represented components of hydraulic limitation (minimizing the number of junction constrictions from the base of the branch to terminal shoots and the hydraulic path length from the base of the branch to terminal shoots), and the second two represented components of carbon balance (minimizing the relative mechanical requirements of a branch and, for each foliated shoot, the average number of other foliated shoots that overlap it). We found two distinct branch forms along the resulting Pareto optimal frontier (figure 2), corresponding to the unique domains of the Pareto optimal solutions. One of the branch forms resembled the observed branching pattern of Douglas fir, and this pattern tended to minimize path length at the expense of increased hydraulic constrictions (table 3). The second branch form resembled the observed branching pattern of true firs (such as grand fir), which tended to minimize the number of hydraulic constrictions at the expense of increased path length (table 3). Within each of these model domains, there were trade-offs among the remaining criteria, such that with decreasing hydraulic constrictions, the mean foliage overlap increased.

Table 3. Two example realizations from the Pareto optimal frontier for branch development (see figure 2).

	Number of junction constrictions to terminal foliage	Hydraulic path length to terminal foliage (centimeters)	Relative branch mechanical requirements (dimensionless)	Mean foliage overlap (number of shoots)
Solution A	4.78	195.60 [†]	0.692	7.64
Solution C	2.49 [†]	276.63	0.634 [†]	5.05 [†]

Note: The simulated branch that resembles Douglas fir performs better for path length but worse for the remaining criteria (therefore, these two solutions are mutually codominant).

[†]The lower (preferred) value for the criterion.

By using multicriteria analysis with Pareto optimality, Kennedy and colleagues (2010) were able to identify two main solutions to the problem of branch development in old, large trees and to use their relative performance for the different criteria to explain why those different morphologies might occur. If the model analysis had focused on one or the other of the criteria (for example, minimizing just the number of branch junctions), only one optimal solution would have been identified, and the importance of the remaining criteria would have been overlooked. If a weighted sum of the criteria had been calculated, the optimal solution would depend greatly on the chosen weights, which again would mask the variability that we observed in the Pareto optimal solution space.

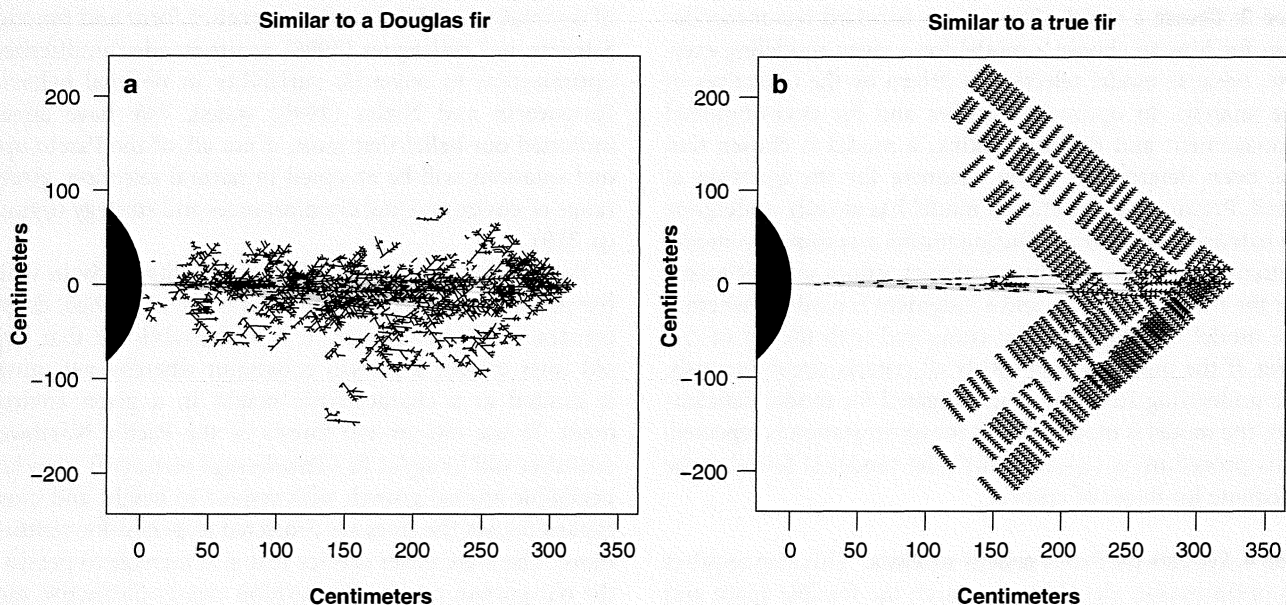


Figure 2. Maps of simulated branches from the Pareto optimal frontier (viewed from above). One group (a) in the frontier resembles observed Douglas fir branch patterns, and another group (b) resembles the branch morphology of true firs (e.g., grand fir). The Douglas fir-like branch utilizes the existing architecture within the interior crown of the branch, whereas the true fir depends on extending lateral axes to produce new foliage.

terminus of complex branch structures. Many studies have been conducted to evaluate which of these hypotheses explains the height limitation, but evidence partially supports both hypotheses, and there is no clear reason to choose

one over the other. Kennedy and colleagues (2010) showed that two major growth patterns that appear in old-growth trees are simulated at different points along a Pareto optimal frontier calculated for four design criteria (box 2), and the

Pareto optimal frontier is able to explain the variability in optimal behavior in the old-growth canopy.

Recently, the US Environmental Protection Agency published guidance for environmental models that emphasized the importance of model evaluation (USEPA 2009), as opposed to validation or verification (Oreskes et al. 1994). They recognized that model assessment is essential to determine whether a model is adequate for its stated objective. In biology and ecology, process-based models are often used. The goal of a process model of a biological system is to represent the functioning of many system components, with fidelity to the processes that govern the components' interactions. For example, a population model could be expressed as a system of ordinary differential equations, with each population summarized as a single-state variable (e.g., the population size). In a process-based model, individual organisms can be represented, the behavior of each defined by a set of rules and parameters intended to mimic actual behavior, and the modeled population dynamics is the culmination of the individual behaviors (as it is in the real world). It is important to evaluate the underlying structure of a process-based model in order to assess whether it adequately replicates system processes.

For a process-based model, even if there is a single characteristic of interest (e.g., population size over time) for the model to replicate, there are many components of the model that should be accurately represented in order for the model to be considered adequate (e.g., reproductive rate, death rate, maturation). A common method of model assessment is to overlay simulated data on an empirically observed pattern (such as a time series) and to visually evaluate whether the model follows the pattern in the data. A statistic such as model likelihood can be calculated, and the model is calibrated to maximize the likelihood relative to the data. This method is great for finding model parameter values that optimize that single measure. However, this method does not determine whether the underlying model structure is an adequate representation of the target system. For that determination, we can define the limits of acceptability for model outputs (a target range) such that the model is deemed adequate if it produces outputs within a target range and is deemed inadequate if it produces outputs outside of the target range (box 3).

With many model components and kinds of model outputs, it is essential to increase the demands on model structure by requiring the model to replicate not just one data pattern but multiple data patterns. For over a decade, the hydrological modeling community has been developing techniques for multicriteria model calibration with Pareto optimality (Yapo et al. 1998, Efstratiadis and Koutsoyannis 2010). It is unlikely that a single model can replicate all system characteristics, and multicriteria analysis with Pareto optimality allows a modeler to describe the contexts under which different combinations of criteria can be satisfied. If parameter values shift significantly when they are calibrated to a second data pattern, there is likely a deficiency in the

underlying model structure (box 4). If the model is able to match two relevant data patterns with a single set of parameter values, confidence in the model is increased. If the model is able to match more than two relevant data patterns simultaneously, confidence in the model is further increased.

Box 4 provides an example model assessment for a stochastic model of fire spread intended to integrate fire as a disturbance into projections of watershed processes in a changing climate. The model must be able to replicate multiple features that characterize a fire regime, including the fire size distribution (e.g., many small fires or few large fires) and the spatial pattern of fire spread (compact fires or more-dispersed fires with irregular shapes). Fire history reconstructions spanning 200 years were calculated for several watersheds throughout the Pacific Northwest in order to provide estimates of the fire size distribution and the spatial structure of fires to be compared with model outputs using multicriteria assessment. Model development begins with a simple stochastic model with three parameters (see Kennedy and McKenzie 2010 for a description of the model). In this model, a simple Monte Carlo sampling is used to search the parameter space, which results in tens of thousands of model runs conducted with different combinations of parameter values. It was found that the model could not simultaneously replicate the fire size distribution and the spatial pattern of spread, which led to a modification of the model processes that determine eventual fire size.

Decisionmaking in environmental management is complicated by the interests of competing stakeholders and the intricacies of complex environmental systems. Furthermore, the currently observed dynamic behavior of an ecological system may differ from its long-term behavior, which may introduce uncertainty to the decisionmaking process. Models are used to evaluate the consequences of management alternatives for the values identified as important to various stakeholders, and these models should incorporate potentially transient dynamics (Hastings 2010). Decision-support systems are expected to identify the management alternatives that are considered to be the best, which requires balancing the priorities of the competing stakeholders. A common decisionmaking method when faced with multiple criteria is to solicit preferences for each criterion and to use these preferences to estimate the relative weight of each (Kim and Smith 2005). The weights are used to calculate a scalar-valued function, which allows for the use of single-criterion optimization tools to find the optimal decision. This decision method thus relies heavily on the choice of weighting scheme, which might not accurately reflect the values of the competing stakeholders. The optimal decision may be sensitive to even small changes in the weighting values (Hyde et al. 2005), and the decisions required to quantify the weighting scheme introduce uncertainty in the optimal decision (Schoemaker and Waid 1982). In contrast, Pareto optimality allows stakeholders to evaluate the possible range of trade-offs among Pareto optimal decisions before relative preferences are specified (Thompson et al. 2010).

Box 3. Error structures for the application of Pareto optimality.

In model assessment, the criterion used to evaluate model performance is usually some error measure, and the deviations between the model predictions and the observed data are minimized. Another common measure is the model likelihood: A probability structure is assigned to the errors, and the likelihood of the parameter set given the data is maximized. In the context of multiple criteria, we can choose multiple model outputs and calculate error measures in order to compare the model outputs with the appropriate data sets. These criteria can be calculated using many kinds of statistics, including multiple likelihoods, goodness-of-fit statistics, residual sums of squares, or some mixture of these, as is appropriate for each of the model–data comparisons. In contrast to the grasshopper example in figure 1, in which the two criteria were maximized, the Pareto optimal frontier can also be calculated for the minimization of all criteria (e.g., model errors) simultaneously (figure 3a).

When we approximate the Pareto optimal frontier to minimize multiple error measures simultaneously, we find the solutions that perform best for the multicriteria optimization, but the approximation does not establish whether those solutions perform adequately. The judgment of model adequacy is important when the goal is model assessment. We can establish limits of acceptability for each criterion (a target range), such that the model is deemed adequate for that criterion if it achieves a model result within the target range. Each criterion is assigned a score of 1 if the model result is within the target range and a score of 0 otherwise. These are called *binary error measures*, and the optimization goal is to achieve a score of 1 for all of the criteria (figure 3). The concept of binary errors has a history in hydrological modeling, in which a model is defined as *behavioral* (binary error = 1) or *not behavioral* (binary error = 0) for one or more criteria (Hornberger and Cosby 1985). We can express this result as a vector. For example, in the two-criteria case, if the model achieves both criteria, the binary errors are expressed as {1,1}. An improvement with respect to performance for a given criterion occurs when one parameterization is able to satisfy that criterion and another is not.

Binary errors are useful if the true value for a given criterion is not known precisely, and binary errors are robust to the possibility of target-value misspecification (Hornberger and Spear 1981, Hornberger and Cosby 1985, Jaffe et al. 1987). There are several strategies for choosing target ranges for binary errors. If data are available for which confidence intervals can be calculated, those can be used to construct the error ranges. Expert opinion can be used to determine the range of values for a criterion over which a model can be considered adequate for that criterion. Reynolds and Ford (1999) used both confidence intervals and expert opinion in determining the error ranges for 10 criteria for the assessment of a spatially explicit model of competition among trees. Kennedy and Ford (2009) chose binary error intervals based on the observed range of data for a limited sample of shoot demography (the number of foliated shoots and the number of foliage clusters) for a functional–structural model of branch development in old-growth Douglas fir trees.

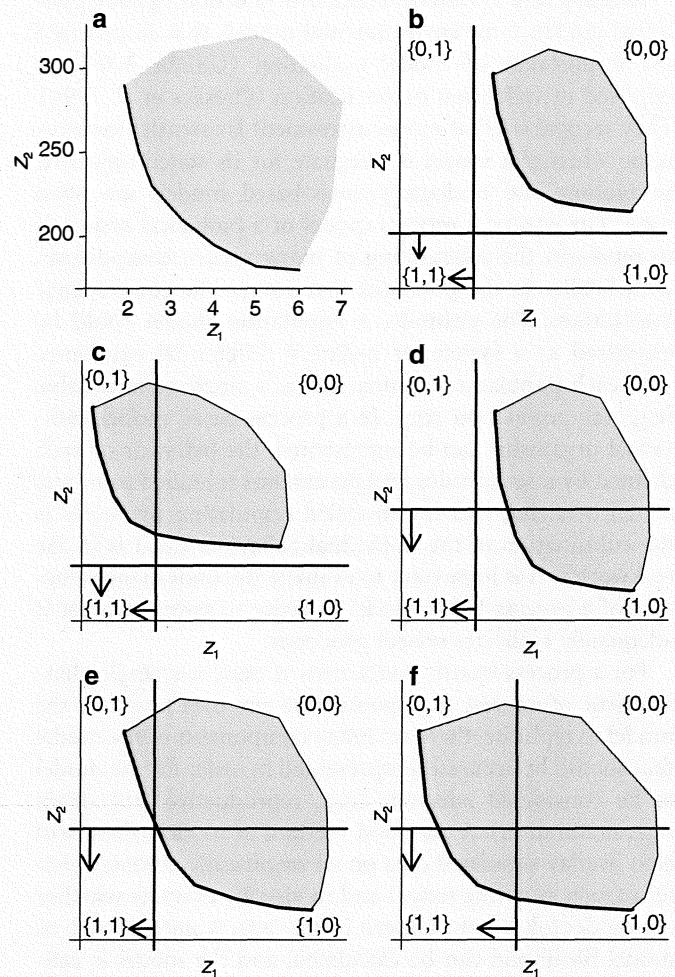


Figure 3. Hypothetical Pareto optimal frontier for two criteria (Z_1 and Z_2). In this case, the optimization problem is to minimize both criteria simultaneously (lower values are preferred for both criteria). The gray area denotes the feasible model space; the thick black line denotes the Pareto optimal frontier. (a) If target ranges of acceptability are set for each criterion (vertical and horizontal lines) and the frontier is expressed as binary errors, there are five possible outcomes in the two-criteria case. (b) The model may fail to achieve either of the criteria {0,0}; the feasible space does not overlap either target range. (c) The feasible space of the model could overlap with the target region of the first criterion but not that of the second {1,0}, or (d) the model could achieve the second criterion but not the first {0,1}. (e) The model may achieve both criteria, but not with a single parameterization. The Pareto optimal frontier then comprises two groups of solutions: In the first group, some parameterizations achieve {1,0}, and in the second group some parameterizations achieve the vector {0,1}. (f) In the best case, the model satisfies both criteria simultaneously and with a single set of parameter values, achieving a binary vector {1,1}.

Box 4. Multicriteria assessment of a model of fire spread.

Two parameters in a stochastic fire-spread model have particular importance to the model output: the spread probability from a burning pixel to each of its neighbors (independently) and the mean fire size. A fire size value is drawn from a probability distribution with the mean fire size. Fires that spread in the simulation can either burn out stochastically or be forced to stop by the randomly drawn fire size. If the spread probability is low, the fire is likely to burn out stochastically, but if the spread probability is high, the fire spreads until it reaches the randomly drawn fire size or until it spans the raster grid.

The assessment criteria are the p -values calculated using a Monte Carlo goodness-of-fit test, which compares simulated data with observed fire history patterns: the spatial pattern of fire spread and the fire size distribution. We use binary criteria, and the model is deemed adequate for a criterion if the results of a goodness-of-fit test (which tests whether the observed pattern is a typical model realization; Waller et al. 2003) are nonsignificant when $p > \alpha$, so that the limits of acceptability for each criterion are $(\alpha, 1)$. Figure 4 gives the results of the analysis for one of the watersheds. The same model is able to replicate the fire size distribution and the spatial pattern but not simultaneously with a single set of parameter values (figure 4). There is clearly a deficiency in the model structure that prevents the model from being able to replicate both patterns simultaneously, and this deficiency would have been masked if only one or the other of the patterns had been compared with the model output.

An analysis of the parameter distributions that correspond to each group led us to the conclusion that the method by which eventual fire size is determined is inadequate if fire size is to be replicated, which led to an improvement of the representation of the model processes that influence fire size.

This example is based on results found during the development of an ongoing project at the US Forest Service's Pacific Wildland Fire Sciences Laboratory in Seattle. Although the results are simulated, they follow the general pattern of what was found through the course of model development.

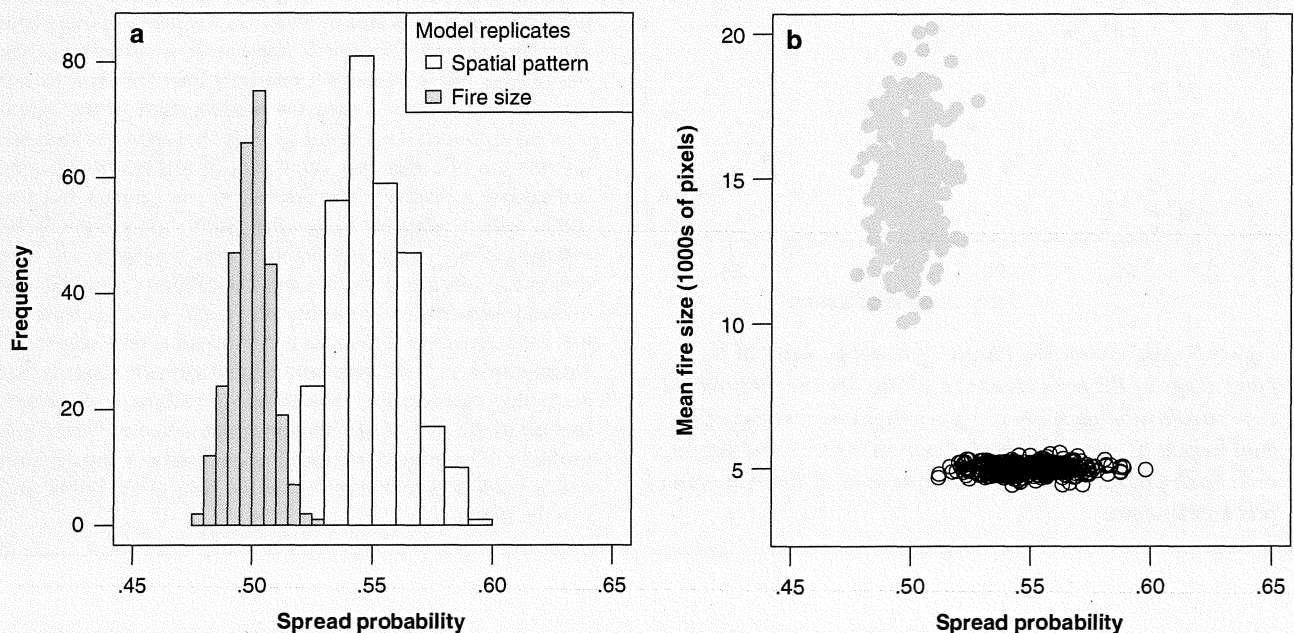


Figure 4. Example parameter distributions in the two-criteria assessment of a fire-spread model. (a) The spread probabilities clearly differ between solutions able to replicate the spatial pattern of fire spread and those able to replicate the fire size distribution, and (b) there is an obvious relationship between the two parameters. In order to satisfy the spatial pattern the spread probability is higher whereas the mean fire size is lower. In order to satisfy the fire size distribution the spread probability is lower and the mean fire size is higher.

Box 5 provides an example of a multicriteria optimization conducted for a fire- and fuels-management problem. The land-management policies of the past century have increased the risk of catastrophic fire in regions that historically experienced frequent low-severity fires, particularly in the Pacific Northwest. It has been shown that fuel-reduction treatments can mitigate this fire risk (Safford et al. 2009,

Johnson et al. 2011), yet the question of how to implement such treatments is not without controversy. The implementation of a fuel-treatment strategy should balance the need to reduce fire risk with the need to protect the valuable ecological landscape features, such as wildlife habitat and old-growth forest reserves. Kennedy and colleagues (2008) performed a multicriteria analysis of a fuel-treatment

Box 5. Multicriteria analysis of fire and fuels management.

For a multicriteria analysis of fire and fuels management, the decisionmaking goal is to find an efficient spatial distribution of fuel treatments that minimizes the fire risk (quantified as a burn probability) with as few units treated as possible in the watershed. In the Pacific Northwest, fire risk to owl habitat and old-growth reserves is of particular concern, so the multiple criteria to judge management action are the percentage of the area that has been treated, the fire risk to owl habitat, and the fire risk to old-growth reserves (all to be minimized). In the Pareto optimal frontier calculated by Kennedy and colleagues (2008), there is an obvious trade-off between the percentage of the area that was treated and the fire risk; with an increasing area treated, the fire risk lessens (figure 5). There is a steep decline in fire risk at low percentages of area treated, which implies that a small increase in area treated would have

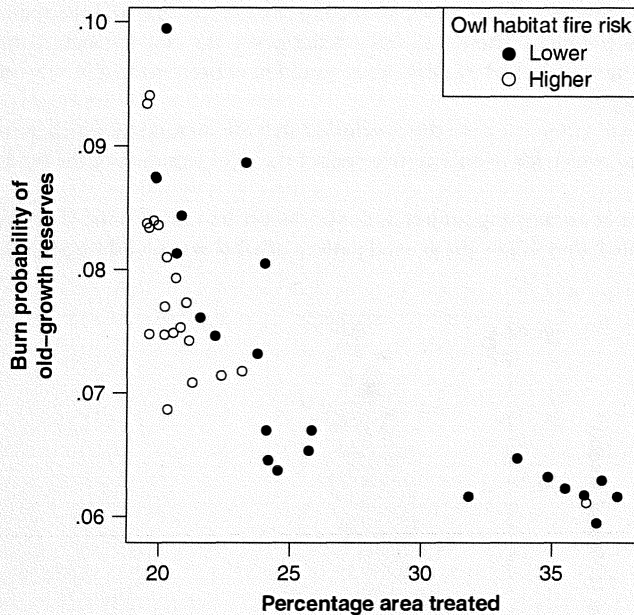


Figure 5. Approximated Pareto optimal frontier of the percentage of the area treated and the fire risk (expressed as a burn probability) to old-growth forest reserves. The goal here is to minimize both the area treated and the fire risk. Each point represents a unique spatial distribution of fuel treatments.

a greater impact on fire risk to old-growth forest reserves. Fire risk levels off at higher percentages of area treated, which implies diminishing returns at higher percentages of area treated. Given limited resources, it is unlikely that the high value for the percentage of the area treated that minimizes fire risk can be implemented, and at lower (and more likely to be implemented) levels of percentage of area treated, there are trade-offs between the fire risk to owl habitat and the fire risk to old-growth reserves.

The frontier appears irregular because of the third criterion, the fire risk to owl habitat. At this point, there is a relatively high fire risk to old-growth forest reserves, and a higher percentage of area treated than the nearest points with lower fire risk to old-growth reserves. It is included in the frontier because it achieves a relatively lower fire risk to owl habitat (defined as less than the median value in the Pareto optimal frontier). This result implies that through strategic distribution of a relatively small area of fuel treatments, one can achieve a relatively low fire risk to owl habitats, but this comes with a relatively higher fire risk to old-growth forest reserves. Using Pareto optimality, forest managers can take these trade-offs into account when they allocate a limited area of fuel treatments and to potentially justify small increases in fuel treatments when the area to be treated is relatively small. Decisionmakers must consider the fuel-treatment strategy that acceptably balances fire risks, within a realistic level of percentage of the area of the landscape that is treated. These risk profiles can be used to enhance the credibility of any decision that is made and to communicate the risks to the public and various stakeholders.

strategy for a watershed in the Cascade Mountains of Washington State. They searched spatial distributions of fuel treatments that simultaneously minimized three criteria: (1) the percentage of the area that was treated, (2) the fire risk to spotted owl habitat, and (3) the fire risk to old-growth reserve sites. They found a trade-off between the area treated and fire risk, such that with increasing area treated, the fire risk decreased. At lower percentages of area treated, there was a trade-off in the fire risk to owl habitat and old-growth reserves. This result implies that if resources are limited for fuel-treatment application, the spatial distribution of those treatments must be chosen carefully, and a decision must be made about how to balance protection from catastrophic fire of owl habitats with that of old-growth reserves. Rather than the modeler providing the forest manager with a single answer that the modeler claims optimizes the manager's

interest, by using Pareto optimality, the modeler gives the manager insight into the possible solutions that balance the priorities in different ways. This allows the manager to choose among the alternatives with a better understanding of the trade-offs associated with alternative actions.

Comparison of Pareto optimality with other forms of model analysis

Pareto optimality for model analysis is often compared with Bayesian analysis, but there are some important differences between them (well-articulated in Engeland et al. 2006). In Bayesian analysis, prior information is combined with known (or assumed) error structures to calculate model likelihoods and the posterior probability of a given model or set of parameter values (Robert 2001). Although we do not propose that Pareto optimality supplants Bayesian analysis,

a Bayesian analysis accommodates only a single criterion (usually the model likelihood). That single function could be an aggregation of multiple measures, but the solution would collapse the full range of trade-offs into a single solution that is a compromise among the multiple measures. In contrast, Pareto optimality-based multicriteria techniques identify the full trade-off space among the criteria, which is then evaluated in order to determine an acceptable compromise (if one exists). Pareto optimality-based multicriteria analysis can accommodate a likelihood measure as a criterion, as Turley and Ford (2009) demonstrated, but it can also include measures that are not necessarily amenable to likelihood analysis (for example, an error structure that does not conform well to any known probability distribution from which a likelihood can be calculated).

Another common method to contend with multiple criteria is to combine the criteria in a single scalar function. This is usually accomplished with a weighted sum, whereby each criterion is assigned a weight that represents its relative importance. The criterion is multiplied by its assigned weight, and those are summed to produce a single value. That value is then optimized with the usual single-criterion techniques. Bayesians have used this method to combine likelihoods from multiple sources into a single likelihood measure (Beven and Binley 1992). This method depends on the ability to convert the criteria into a single currency for which it is reasonable to perform a summation, and the choice of weights becomes an essential and important component of the optimization exercise. This technique is less computationally challenging than approximating the Pareto optimal frontier, but it presents an incomplete picture of the multicriteria surface. It presumes a single optimal solution that can always be agreed on, except the optimal solution depends on the weighting scheme, and the choice of weights might vary depending on whom one asks. The Pareto optimal frontier contains all optimal combinations of weights for an optimization problem, such that for any choice of weightings, the optimal solution can be found along the Pareto optimal frontier. Approximation of the Pareto optimal frontier is superior to combining the multiple criteria in a scalar function if the weightings are not known and if the criteria are not easily converted to a single currency.

Conclusions

Pareto optimality should be applied where systems are complex and defined by many interacting and often conflicting features. Attempts to describe these features with a single value often fail and mask important relationships that contribute to our understanding of and to an explanation of the system's function, or they decrease the information available for decisionmaking in environmental management.

Beardsley (2010) called for alternative scientific methods that integrate the components of complex systems and allow for multiple solutions to complex problems. He cited Mitchell's (2009) analysis of knowledge development for

complex systems, in which she said, "Instead of eternally true universal laws applicable to all space and time, an epistemology tuned to complexity will produce a host of contingent, domain-restricted generalizations that describe more or less stable causal structures" (p. 18). Pareto optimization used as we describe it here allows for the description of those multiple solutions and provides a framework through which their integration can be accomplished; if unity in explanations exists, Pareto optimality can help us find it. Each Pareto optimal solution may be considered one of those "contingent, domain-restricted generalizations," in which the domain is described by the relative performance for the criteria for each of the solutions and in which the model structure and parameter values that correspond to the Pareto optimal solutions help to describe the causal structure of the biological system.

In optimality studies, one is able to integrate multiple system features to better describe whole-organism observations and multiple optimal solutions for reproduction and survival. In model assessment, one can use Pareto optimality to evaluate a model against multiple system components simultaneously, thereby first describing and then reducing model uncertainty. In a decisionmaking context, rather than subsuming the uncertainty in the complex system to yield a single supposedly optimal solution, by applying Pareto optimality, one can balance the competing priorities of multiple solutions.

Acknowledgments

The authors thank Miranda Fix, Shawn Behling, and Bert Loosmore for valuable comments to improve a prior draft of the manuscript. Don McKenzie and three anonymous reviewers provided feedback that helped improve the clarity of our presentation. Much of the research that enabled us to develop the concepts presented here was conducted at the canopy dynamics laboratory at the University of Washington, in Seattle, particularly by Joel Reynolds, Marianne Turley, and Rie Komuro. Funding during the production of this manuscript was provided by the Pacific Northwest Research Station, US Forest Service, and the University of Washington.

References cited

- Beardsley TM. 2010. Disentangling complexity in biology. *BioScience* 60: 327.
- Beven KJ, Binley AM. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* 6: 279–298.
- Cirillo R. 1979. *The Economics of Vilfredo Pareto*. Cass.
- Deb K. 2001. *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley.
- Efstratiadis A, Koutsoyiannis D. 2010. One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal* 55: 58–78.
- Engeland K, Braud I, Gottschalk L, Leblois E. 2006. Multi-objective regional modelling. *Journal of Hydrology* 327: 339–351.
- Farnsworth KD, Niklas KJ. 1995. Theories of optimization, form and function in branching architecture in plants. *Functional Ecology* 9: 355–363.

- Gallagher R, Appenzeller T. 1999. Beyond reductionism. *Science* 284: 79.
- Gould SJ, Lewontin RC. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society B* 205: 581–598.
- Hastings A. 2010. Timescales, dynamics, and ecological understanding. *Ecology* 91: 3471–3480.
- Honda H, Fisher JB. 1978. Tree branch angle: Maximizing effective leaf area. *Science* 199: 888–890.
- . 1979. Ratio of tree branch lengths: The equitable distribution of leaf clusters on branches. *Proceedings of the National Academy of Sciences* 76: 3875–3879.
- Hornberger GM, Cosby BJ. 1985. Selection of parameter values in environmental models using sparse data: A case study. *Applied Mathematics and Computation* 17: 335–355.
- Hornberger GM, Spear RC. 1981. An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management* 12: 7–18.
- Hunt ER Jr, Lavigne MB, Franklin SE. 1999. Factors controlling the decline of net primary production with stand age for balsam fir in Newfoundland assessed using an ecosystem simulation model. *Ecological Modelling* 122: 151–164.
- Hyde KM, Maier HR, Colby CB. 2005. A distance-based uncertainty analysis approach to multi-criteria decision analysis for water resources decision making. *Journal of Environmental Management* 77: 278–290.
- Jaffe PR, Paniconi C, Wood EF. 1987. Model calibration based on random environmental fluctuations. *Journal of Environmental Engineering* 114: 1136–1145.
- Johnson MC, Kennedy MC, Peterson DL. 2011. Simulating fuel treatment effects in dry forests of the western United States: testing the principles of a fire-safe forest. *Canadian Journal of Forest Research* 41: 1018–1030.
- Kennedy MC, Ford ED. 2009. Two-criteria model assessment shows that foliage maintenance in old-growth *Pseudotsuga menziesii* requires both delayed and sequential reiteration. *Trees: Structure and Function* 23: 1173–1187.
- Kennedy MC, McKenzie D. 2010. Using a stochastic model and cross-scale analysis to evaluate controls on historical low-severity fire regimes. *Landscape Ecology* 25: 1561–1573.
- Kennedy MC, Ford ED, Singleton P, Finney M, Agee JK. 2008. Informed multi-objective decision-making in environmental management using Pareto optimality. *Journal of Applied Ecology* 45: 181–192.
- Kennedy MC, Ford ED, Hinckley TM. 2010. Defining how aging *Pseudotsuga* and *Abies* compensate for multiple stresses through multi-criteria assessment of a functional–structural model. *Tree Physiology* 30: 3–22.
- Kim K-J, Smith RL. 2005. Systematic procedure for designing processes with multiple environmental objectives. *Environmental Science and Technology* 39: 2394–2405.
- Komuro R, Ford ED, Reynolds JH. 2006. The use of multi-criteria assessment in developing a process model. *Ecological Modelling* 197: 320–330.
- Mitchell SD. 2009. *Unsimple Truths: Science, Complexity, and Policy*. University of Chicago Press.
- Oreskes N, Shrader-Frechette K, Belitz K. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263: 641–649.
- Parker GA, Smith JM. 1990. Optimality theory in evolutionary biology. *Nature* 348: 27–33.
- Rabotyagov S, Campbell T, Jha M, Gassman PW, Arnold J, Kurkalova L, Secchi S, Feng H, Kling CL. 2010. Least-cost control of agricultural nutrient contributions to the Gulf of Mexico hypoxic zone. *Ecological Applications* 20: 1542–1555.
- Reynolds JH, Ford ED. 1999. Multi-criteria assessment of ecological process models. *Ecology* 80: 538–553.
- Robert CP. 2001. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer.
- Rothley KD, Schmitz OJ, Cohon JL. 1997. Foraging to balance conflicting demands: novel insights from grasshoppers under predation risk. *Behavioral Ecology* 8: 551–559.
- Ryan MG, Yoder BJ. 1997. Hydraulic limits to tree height and tree growth. *BioScience* 47: 235–242.
- Safford HD, Schmidt DA, Carlson CH. 2009. Effects of fuel treatments on fire severity in an area of wildland–urban interface, Angora Fire, Lake Tahoe Basin, California. *Forest Ecology and Management* 258: 773–787.
- Schmitz OJ, Cohon JL, Rothley KD, Beckerman AP. 1998. Reconciling variability and optimal behaviour using multiple criteria in optimization models. *Evolutionary Ecology* 12: 73–94.
- Schoemaker PJH, Waid CC. 1982. An experimental comparison of different approaches to determining weights in additive utility models. *Management Science* 28: 182–196.
- Smith JM. 1978. Optimization theory in evolution. *Annual Review of Ecology and Systematics* 9: 31–56.
- Thompson M, Session J, Boston K, Skaugset A, Tomberlin D. 2010. Forest road erosion control using multiobjective optimization. *Journal of the American Water Resources Association* 46: 712–723.
- Tóth SF, Haight RG, Snyder SA, George S, Miller JR, Gregory MS, Skibbe AM. 2009. Reserve selection with minimum contiguous area restrictions: An application to open space protection planning in suburban Chicago. *Biological Conservation* 142: 1617–1627.
- Turley MC, Ford ED. 2009. Definition and calculation of uncertainty in ecological process models. *Ecological Modelling* 220: 1968–1983.
- [USEPA] US Environmental Protection Agency. 2009. *Guidance on the Development, Evaluation, and Application of Environmental Models*. Office of the Science Advisor, Council for Regulatory Environmental Modeling, USEPA. Report no. EPA/100/K-09/003.
- Vrugt JA, van Belle J, Bouten W. 2007. Pareto front analysis of flight time and energy use in long-distance bird migration. *Journal of Avian Biology* 38: 432–442.
- Waller LA, Smith D, Childs JE, Real LA. 2003. Monte Carlo assessments of goodness-of-fit for ecological simulation models. *Ecological Modelling* 164: 49–63.
- Yapo PO, Gupta HV, Sorooshian S. 1998. Multi-objective global optimization for hydrological models. *Journal of Hydrology* 204: 83–97.

Maureen C. Kennedy (mkenn@u.washington.edu) is a research scientist and E. David Ford (edford@u.washington.edu) is a professor at the University of Washington School of Forest Resources, Seattle.