Ecological Modelling 404 (2019) 21–26

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Selecting ecological models using multi-objective optimization

Perry J. Williams[a,*], William L. Kendall[b], Mevin B. Hooten[b,c]

[a] Department of Natural Resources and Environmental Science, University of Nevada, Reno, 1664 N Virginia St, Reno, NV, 89557, United States
[b] U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, United States
[c] Department of Statistics, Colorado State University, Fort Collins, CO 80523, United States

ARTICLE INFO

Keywords:
Competing models
Decision theory
Model selection
Multiple objectives
Pareto frontier
Optimal solution

ABSTRACT

Choices in ecological research and natural resource management require balancing multiple, often competing objectives. Examples include maximizing species persistence in a wildlife conservation context, while minimizing cost, or balancing opposing stakeholder objectives when managing wildlife populations. *Multiple-objective optimization* (MOO) provides a unifying framework for solving multiple objective problems. Model selection is a critical component of ecological inference and prediction and requires balancing the competing objectives of model fit and model complexity. The tradeoff between model fit and model complexity provides a basis for describing the model-selection problem within the MOO framework. We discuss MOO and two strategies for solving the MOO problem; modeling preferences pre-optimization and post-optimization. Most conventional model selection methods can be formulated as solutions of MOO problems via specification of pre-optimization preferences. We reconcile model selection within the MOO framework. We also consider model selection using post-optimization specification of preferences. That is, by first identifying Pareto optimal solutions, and then selecting among them. We demonstrate concepts with an ecological application of model selection using avian species richness data in the continental United States.

1. Introduction

The goal of modeling ecological processes varies from identifying the important factors driving a system, to robust prediction into the future or across space. Multiple competing models are considered in most cases, each model based on hypotheses of spatial or temporal structure in parameters, heterogeneity among individuals within a population, or candidates for covariates influencing the process of interest. Ultimately, one model from the candidate models, or a composition of candidate models, is selected for inference or prediction. Statistical model selection is one of the most common problems in scientific research, and numerous model-selection methods are available (e.g., Akaike, 1973; Mallows, 1973; Schwarz et al., 1978; Gelfand and Ghosh, 1998; Burnham and Anderson, 2002; Hooten and Hobbs, 2015). Each model-selection method represents an approach to balancing the bias due to missing important factors (model fit) with imprecision due to over-fitting the data (model complexity). Each method represents a different *a priori* weighting of the relative importance of model fit and model complexity. While guidelines exist, there is no consensus among statisticians on best methods for this model selection process (Hooten and Hobbs, 2015).

In the field of decision theory, multiple-objective problems are often addressed using the methods of multi-objective optimization (MOO; Marler and Arora, 2004; Williams, 2016; Williams and Kendall, 2017). MOO is common in engineering, economics, and other fields for which decisions must balance trade-offs between ≥ 2 competing objectives (Marler and Arora, 2004). When a decision maker has competing objectives, a solution that is optimal for one objective might not be optimal for the other objective and a single solution that optimizes multiple objectives does not exist. With competing objectives there exist possibly infinitely many solutions that might be considered "optimal" (i.e., Pareto optimal; Williams and Kendall, 2017). However, in most decision contexts, a decision maker can only make one choice (e.g., which model to use to predict into the future?). To choose among solutions, a decision maker must include their preferences among objectives to identify a final solution. MOO provides a mathematical framework for quantifying preferences for examining multi-objective problems.

Our objective is to frame and examine the model selection problem through the lens of the MOO framework. By doing so, we can gain new insight into the model selection problem. Specifically, (1) we bridge a gap and reconcile the literature between methodologies from two

* Corresponding author.
E-mail address: perryw@unr.edu (P.J. Williams).

https://doi.org/10.1016/j.ecolmodel.2019.04.012
Received 29 November 2018; Received in revised form 11 April 2019; Accepted 17 April 2019
Available online 14 May 2019

separate fields of study: the decision-theoretic field of MOO, and the statistical field of model selection; (2) we reveal that several methods commonly used for model selection are specific cases of the MOO problem solved using *a priori* specification of preferences; (3) we examine concepts of the MOO framework, specifically Pareto optimality, and show that common model selection methods meet the criteria for Pareto optimality; and (4) we consider model selection *via* the MOO approach of post-specification of a decision maker's preferences. We demonstrate the concepts presented using an example from the field of ecology involving variable selection in a generalized linear regression model for avian species richness data.

2. Materials and methods

2.1. Model selection as a MOO problem

The MOO framework is described generally as

$$\min_{\theta\in\Theta}(f_1(\theta), f_2(\theta), \ldots, f_k(\theta))' \;\; = \theta_i^* \tag{1}$$

where $(f_1(\boldsymbol{\theta}),\ f_2(\boldsymbol{\theta}),\ \ldots,\ f_k(\boldsymbol{\theta}))$ are k different, potentially competing, objective functions (e.g., Eqs. (4) and (5)), and $\boldsymbol{\theta} \equiv (\theta_1,\ \ldots,\ \theta_p)'$ are design variables of the objective function (e.g., potential choices an investigator has). The objective function $f_i(\boldsymbol{\theta})$, $i = 1, \ldots, k$ maps a choice $\boldsymbol{\theta}$ to the real line. Regarding model selection, the design variables are the potential parameters (and their specific value) of a model. The design variables of the objective function must be *feasible* to the investigator. That is, make sense (logically or mathematically) for the problem at hand, and can be constrained to the feasible set using inequality and equality constraints. Generally, these constraints are represented as functions (g and h) of the design variables $g_j(\boldsymbol{\theta}) \leq c_j$ and $h_m(\boldsymbol{\theta}) = d_m$ where $j = 1, \ldots, J$ are the J inequality constraints that must be less than or equal to some arbitrary constant c_j, and where $m = 1, \ldots, M$ are the M equality constraints that must be equal to some arbitrary constant d_j (Marler and Arora, 2004; Cohon, 2013). The point θ_i^* minimizes the objective functions $(f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \ldots, f_k(\boldsymbol{\theta}))$.

Pareto optimality is a concept of optimality used for Eq. (1) when no value of $\boldsymbol{\theta}$ simultaneously optimizes all functions $f_i\ \forall i \in 1, \ldots, k$. A Pareto optimal solution for a minimization problem is a solution $\boldsymbol{\theta}^* \in \Theta$ for which there is no other solution $\boldsymbol{\theta} \in \Theta$ such that both $f(\boldsymbol{\theta}) \leq f(\boldsymbol{\theta}^*)$, and $f_i(\boldsymbol{\theta}) < f_i(\boldsymbol{\theta}^*)$ for at least one function i (Deb, 2001; Marler and Arora, 2004). For decision problems with competing objectives, there are many (potentially infinite), Pareto optimal solutions. The set of solutions that are Pareto optimal is known as the Pareto set (or Pareto frontier or efficiency frontier). Each solution in a Pareto set has an implied set of preferences among the objective functions f_i (Deb, 2001; Williams and Hooten, 2016). Thus, choosing among a set of Pareto optimal solutions assumes (either implicitly or explicitly) preferences among the objective functions f_i. Such preferences can be specified pre- or post-optimization, representing two separate strategies to solving Eq. (1) (Williams and Kendall, 2017). When specifying preferences pre-optimization, decision makers explicitly describe preferences of objective functions and select the Pareto optimal solution associated with their choice of preferences. When specifying preferences post-optimization, decision makers first examine the set of Pareto optimal solutions. Then the decision maker chooses the final Pareto optimal solution based on the trade-offs observed among the set. The choice implies decision-maker preferences.

2.2. Model selection using pre-optimization selection of weights

One of the most common methods for incorporating preferences for f_i in Eq. (1) into a decision problem pre-optimization, is the weighted-sum method (Athan and Papalambros, 1996; Das and Dennis, 1997; Cohon, 2013; Williams and Kendall, 2017). The weighted-sum method is described by

$$f(\theta) = \sum_{i=1}^{k} w_i f_i(\theta), \tag{2}$$

for which the optimal solution is

$$f(\theta^*) = \min_\theta \sum_{i=1}^{k} w_i f_i(\theta). \tag{3}$$

The weights w_i are chosen by the decision maker to reflect the importance of each objective function f_i. The weighted-sum method is a composition that results in a single objective function over which to optimize. When optimizing an objective function under one fixed set of weight values, an unequivocal optimal choice can be made.

Model selection methods typically consist of minimizing a weighted sum of two functions, often described heuristically as a function for model fit and a function for model complexity (e.g., Gelfand and Ghosh, 1998; Burnham and Anderson, 2002, p. 87). This can be described as a specific case of Eq. (3), where $\boldsymbol{\theta}^*$ represents the optimal solutions from the set of design variables $\boldsymbol{\theta}$ (i.e., model parameters), describing fit and complexity of any model, w_i are weights for the importance of the objectives associated with model fit and complexity, and f_i are functions that quantify the value of model fit and complexity. The statistical literature is rich in excellent work developing the theoretical justification for choices of objective functions $f_i(\boldsymbol{\theta})$ and their corresponding weights w_i (Akaike, 1973; Mallows, 1973; Schwarz et al., 1978; Gelfand and Ghosh, 1998; Burnham and Anderson, 2002; Link and Barker, 2006; Hooten and Hobbs, 2015). Although there is no consensus among statisticians on specific model selection methods, most of the theoretical development related to model selection can be described by two general functions for f_i. Differences in model selection criteria are often the result of different theoretical motivation for the selection of weights. The most common objective function for model fit is the negative log-likelihood of the parameters, given the data (i.e., the deviance). That is, if f_1 is the objective function associated with model fit, it is described as

$$f_1(\theta) = -\log(L(\theta; \mathbf{y})), \tag{4}$$

where $L(\boldsymbol{\theta}; \boldsymbol{y})$ represents the likelihood of the parameters ($\boldsymbol{\theta}$), given the data ($\boldsymbol{y}$). Although the deviance is a commonly used objective function for model fit, others have been used. For example in Mallows' C_p, $f_1(\theta) = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_{\text{sub}})^2}{\sum_{i=1}^{n}(y_i - \hat{\mu}_{\text{full}})^2} - n$, where $\hat{\mu}_{\text{sub}}$ equals the estimated mean of a sub-model parameter in consideration, $\hat{\mu}_{\text{full}}$ equals the estimated mean of the full model parameter in consideration, and n equals the sample size (Mallows, 1973).

Tibshirani (1996), Tikhonov et al. (2013), and Hooten and Hobbs (2015) describe several objective functions for model complexity using a function proportional to

$$f_2(\theta) = \sum_{j=1}^{p} |\theta_j - \mu_j|^\gamma, \tag{5}$$

known as the *regulator*, *regularizer*, or *penalty*. In Eq. (5), p represents the number of parameters in the model, γ is the degree of the norm; a user-defined parameter that controls the relative penalty of the distance between θ_j and μ_j, θ_j are parameter estimates for centered and scaled covariates, and μ_j is a location parameter, often set to 0. Substituting the choices of $f_1(\boldsymbol{\theta})$ and $f_2(\boldsymbol{\theta})$ from Eqs. (4) and (5) into Eq. (3), we obtain the following multi-objective optimization problem

$$\begin{aligned} f(\theta) &= w_1 f_1(\theta) + w_2 f_2(\theta), \\ &= w_1(-\log(L(\theta; \mathbf{y}))) + w_2 \sum_{j=1}^{p} |\theta_j - \mu_j|^\gamma, \end{aligned} \tag{6}$$

with the objective of $\min_{\boldsymbol{\theta}}(f(\boldsymbol{\theta}))$.

Table 1
Values of weights (w_i) and γ for the multi-objective optimization problem of model selection described in Eq. (6) for various model selection methods. These methods are typically used for modeling preferences pre-optimization. The objective function for model fit is $-log(L(\boldsymbol{\theta}; \boldsymbol{y}))$, where $\boldsymbol{\theta} \equiv \boldsymbol{\beta}$; the objective function for model complexity is $\sum_{j=1}^{p} |\beta_j - \mu_j|^{\gamma}$, $j = 1, \ldots, p$. AIC = Akaike's information criterion; AIC_c = Second-order information criterion; QAIC = quasi-AIC; BIC = Schwartz information criterion; n = sample size; p = no. parameters in model. A * indicates objective function for model fit defined by: $\sum_{i=1}^{n} (y_i - \beta_0 - \boldsymbol{x}\,\boldsymbol{\beta})^2$. See Burnham and Anderson (2002) and Hooten and Hobbs (2015) for additional details.

Model selection method	w_1	w_2	γ	Note
AIC	2	2	0	
AIC_c	2	$2\left(\frac{n}{n-p-1}\right)$	0	
QAIC	$\frac{2}{\hat{c}}$	2	0	$\hat{c} = \chi^2/\text{df}$
$QAIC_c$	$\frac{2}{\hat{c}}$	$2\left(\frac{n}{n-p-1}\right)$	0	$\hat{c} = \chi^2/\text{df}$
BIC	2	$\log(n)$	0	
*Ridge regression	1	User defined or estimated	2	Larger values of w_2 shrink $\boldsymbol{\beta}$ to 0.
*LASSO	1	User defined or estimated	1	Larger values of w_2 shrink $\boldsymbol{\beta}$ to 0.

2.3. Model selection using post-optimization selection of weights

Solving a MOO problem with competing objectives using post-optimization specification of weights requires first identifying as many Pareto optimal solutions as possible, then choosing among the Pareto optimal solutions (Williams and Kendall, 2017). Pareto optimal solutions for the objective functions in Eqs. (4) and (5) are models for which increasing the value of Eq. (4) requires a decrease in the values in Eq. (5), and *vice versa*. Values for Eqs. (4) and (5) for any candidate model can be obtained from standard software (see Appendix for R code to obtain these values). To identify Pareto optimal solutions, consider a 2-dimensional plot of candidate models with values of Eq. (4) on the y-axis and Eq. (5) on the x-axis (e.g., Fig. 2). After the Pareto frontier is identified, the decision maker can select the model based on the trade-offs observed in the Pareto frontier. This is analogous to *best subset selection*, an active area of statistical research (e.g., Hastie et al., 2009). Thus, the selection of the final model is made without explicitly choosing weights $\boldsymbol{w}$ associated with the model selection criteria listed in Table 1. However, if a choice from the Pareto frontier is also optimal with respect to specific model-selection criterion, the weights of that selection criterion are implied.

2.4. Example: avian species richness in the U.S.

Model selection is commonly used for selecting variables to include in linear and generalized linear regression models. We examine the variable selection problem within a MOO framework by considering avian species richness in the contiguous U.S. as a function of state-level covariates. These data were originally used to demonstrate model selection techniques in Hooten et al. (2019). We seek to model the number of avian species y_i counted in each of the contiguous states in the U.S. and Washington D.C. (i.e., $i = 1, \ldots, n$, where sample size $n = 49$), based on covariate information $\mathbf{x}_i$ collected in each state. Covariates include the area of the state, the mean annual temperature, and mean annual precipitation. Response data and covariates are shown in Fig. 1. We modeled count data using a Poisson distribution, $y_i \sim \text{Poisson}(\lambda_i)$, where λ_i represents the mean species richness in each state. We linked mean species richness to the covariate data using the log link function, $\log(\lambda_i) = \mathbf{x}_i'\boldsymbol{\beta}$. We considered a total of 24 different models for $\mathbf{x}_i'\boldsymbol{\beta}$, representing different linear and quadratic combinations of the covariates. The 24 candidate models are provided in Table 2. We fit each model using the `glm` function in R statistical software (R Core Team, 2013, see Appendix).

To conduct model selection for the avian species richness data, we used the objective function in Eq. (6) with values of $\boldsymbol{w} \equiv 2$, and $\gamma = 0$ (i.e., AIC). That is, for a Poisson likelihood, the weighted objective function was

$$f(\boldsymbol{\beta}_m) = 2\left(\sum_{i=1}^{n} (\lambda_{i,m} - y_i \log(\lambda_{i,m}) + \log(y_i!))\right) + 2\sum_{j=1}^{p_m} |\beta_{j,m}|^0, \quad (7)$$

where $\boldsymbol{\beta}_m$ is the subset of parameters for model $m = 1, \ldots, 24$, n is the sample size, and the term $\log(y_i!)$ can be omitted because it does not affect the optimization of $\boldsymbol{\beta}_m$, provided the likelihood is not changed.

3. Results and discussion

3.1. Pre-optimization selection of weights and relationship among model selection methods

Eq. (6) is the general function used in many model selection methods including Akaike's information criterion (AIC), AIC for small samples (AIC_c), quasi-AIC (QAIC), QAIC for small samples ($QAIC_c$), Schwartz's information criterion (BIC), ridge regression, LASSO (least absolute shrinkage and selection operator), natural Bayesian shrinkage, and some forms of posterior predictive loss (Table 1; Gelfand and Ghosh, 1998; Hooten and Hobbs, 2015). Each of the listed model selection methods result from specific choices of $\boldsymbol{w}$ and γ, which we report in Table 1. For example, let the weights be: $w_1 = 2$, $w_2 = 2$, and set γ to zero. With these weights, Eq. (6) simplifies to $-2\log(L(\boldsymbol{\theta}; \boldsymbol{y})) + 2p$, or AIC (Table 1).

Additionally, expressing model selection methods in terms of Eq. (3) has an important result that links model selection to Pareto optimality. For positive weights $\boldsymbol{w}$, any solution to Eq. (3) is a Pareto optimal solution (Marler and Arora, 2010). Thus, any model selection method that can be expressed in terms of Eq. (6) (i.e., the methods in Table 1) results in a solution that is Pareto optimal with respect to the objectives of maximizing model fit and minimizing model complexity.

The explicit application of multi-objective optimization to model selection using the objective functions defined in Eqs. (4) and (5) ties several important properties of MOO to common methods used in ecological research to select a model. First, many different model selection methods are special cases of the weighted-sum method; each method representing different objective weights. This provides a unifying framework to quantitatively and visually compare model-selection methods based on different theoretical foundations. Practitioners of multi-objective optimization in operations research or other decision-theoretic fields usually recommend evaluating sensitivity of the resulting decisions, given the choice of objective weights (Keeney and Raiffa, 1976; Williams and Kendall, 2017). A sensitivity analysis for the model selection problem consists of evaluating multiple model selection criteria (representing different objective weights) to examine the robustness of the solution to the choice of criterion. Many practitioners argue against this approach, suggesting that a criterion should be selected based on its theoretical motivation (e.g., AIC is asymptotically efficient; BIC is consistent, Aho et al., 2014). Others view a specific information criterion as one line of evidence to assist in a decision and report different criteria side-by-side (e.g., Araújo and Luoto, 2007; Parviainen et al., 2008). Preferences vary by field; the former tends to be the dominant paradigm in ecological research, whereas the latter is common in other fields. Second, many model selection methods result in Pareto optimal solutions because they are specific formulations of Eq. (2), which is sufficient for Pareto optimality. Thus, there is a decision-theoretic basis for model selection methods that can be expressed in the form of Eq. (6) in terms of optimality criteria.

Although we described the model selection problem heuristically in terms of maximizing model fit and minimizing model complexity, we could have replaced model fit with predictive ability as the objective of

Fig. 1. Top left: Response variable – the number of avian species in each state. Top right: Predictor variable 1 – mean annual temperature. Bottom left: Predictor variable 2 – area. Bottom right: Predictor variable 3 – annual precipitation.

interest. Predictive ability is the most commonly sought model characteristic for model selection, and many information criteria and other model selection methods were developed to optimize predictive ability (Akaike, 1973; Stone, 1977; Gelfand and Ghosh, 1998; Hoeting et al., 1999; Burnham and Anderson, 2002; Hooten and Hobbs, 2015). Many information criteria have weights and penalties that serve as bias corrections for optimization in terms of predictive ability (Konishi and Kitagawa, 1996). That is, many information criteria are based on bias-corrected log likelihoods, for which the model complexity is a correction factor to remove asymptotic bias of the log likelihood of a fitted

Table 2
Model selection results from avian species richness data. AIC is Akaike's information criterion, ΔAIC = is the difference in AIC compared to the top model. Asterisks indicate Pareto optimal models, $f_1(\boldsymbol{\beta})$ and $f_2(\boldsymbol{\beta})$ are described in Eqs. (4) and (5), respectively.

$\log(\lambda_i) =$	$f(\boldsymbol{\beta})$ (i.e., AIC)	ΔAIC	$f_1(\boldsymbol{\beta})$ (fit)	$f_2(\boldsymbol{\beta})$ (complexity)
β_0	741.1	229.8	369.6	1*
$\beta_0 + \beta_1 x_{\text{area},i}$	571.2	59.9	283.6	2*
$\beta_0 + \beta_1 x_{\text{temp},i}$	669.2	157.9	332.6	2
$\beta_0 + \beta_1 x_{\text{precip},i}$	706.1	194.8	351.0	2
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x_{\text{temp},i}$	526.7	15.4	260.3	3*
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x_{\text{precip},i}$	567.7	56.4	280.9	3
$\beta_0 + \beta_1 x_{\text{temp},i} + \beta_2 x_{\text{precip},i}$	536.7	25.5	265.4	3
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i}$	572.8	61.5	283.4	3
$\beta_0 + \beta_1 x_{\text{temp},i} + \beta_2 x^2_{\text{temp},i}$	668.0	156.7	331.0	3
$\beta_0 + \beta_1 x_{\text{precip},i} + \beta_2 x^2_{\text{precip},i}$	704.9	193.6	349.4	3
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x_{\text{temp},i} + \beta_3 x_{\text{precip},i}$	515.3	4.0	253.7	4*
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i} + \beta_3 x_{\text{temp},i}$	524.3	13.0	258.1	4
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i} + \beta_3 x_{\text{precip},i}$	565.6	54.3	278.8	4
$\beta_0 + \beta_1 x_{\text{temp},i} + \beta_2 x^2_{\text{temp},i} + \beta_3 x_{\text{area},i}$	528.2	16.9	260.1	4
$\beta_0 + \beta_1 x_{\text{temp},i} + \beta_2 x^2_{\text{temp},i} + \beta_3 x_{\text{precip},i}$	535.5	24.2	263.8	4
$\beta_0 + \beta_1 x_{\text{precip},i} + \beta_2 x^2_{\text{precip},i} + \beta_3 x_{\text{area},i}$	568.1	56.8	280.0	4
$\beta_0 + \beta_1 x_{\text{precip},i} + \beta_2 x^2_{\text{precip},i} + \beta_3 x_{\text{temp},i}$	524.2	12.9	258.1	4
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i} + \beta_3 x_{\text{temp},i} + \beta_4 x^2_{\text{temp},i}$	525.7	14.4	257.8	5
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i} + \beta_3 x_{\text{precip},i} + \beta_4 x^2_{\text{precip},i}$	567.1	55.8	278.6	5
$\beta_0 + \beta_1 x_{\text{temp},i} + \beta_2 x^2_{\text{temp},i} + \beta_3 x_{\text{precip},i} + \beta_4 x^2_{\text{precip},i}$	518.0	6.7	254.0	5
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i} + \beta_3 x_{\text{precip},i} + \beta_4 x^2_{\text{precip},i} + \beta_5 x_{\text{temp},i}$	512.2	0.9	250.1	6
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i} + \beta_3 x_{\text{precip},i} + \beta_5 x_{\text{temp},i} + \beta_6 x^2_{\text{temp},i}$	519.2	7.9	253.6	6
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x_{\text{precip},i} + \beta_3 x^2_{\text{precip},i} + \beta_4 x_{\text{temp},i} + \beta_5 x^2_{\text{temp},i}$	511.3	0	249.6	6*
$\beta_0 + \beta_1 x_{\text{area},i} + \beta_2 x^2_{\text{area},i} + \beta_3 x_{\text{precip},i} + \beta_4 x^2_{\text{precip},i} + \beta_5 x_{\text{temp},i} + \beta_6 x^2_{\text{temp},i}$	512.8	1.5	249.4	7*

model (Konishi and Kitagawa, 1996). The MOO problem in terms of maximizing predictive ability and accounting for model bias is similar in spirit to the MOO problem of maximizing model fit while minimizing model complexity.

3.1.1. Avian species richness

The model from Table 2 that minimized Eq. (7) (i.e., the AIC top model) included the intercept, a linear area effect, and a quadratic precipitation and temperature effect. All other model fitting results are shown in Table 2.

A common practice for pre-specification of weights in MOO problems in other applications includes examining the sensitivity of the optimal choice relative to the selected weights (Barron and Schmidt, 1988; Insua, 1990). An analogous procedure in the model selection framework is to examine the optimal solutions relative to different information criteria because different criteria represent different objective weights (Table 1). The AIC, AIC$_c$, and BIC criteria all resulted in the same top model suggesting the optimal solution for these data was robust to several different choices of weights.

3.2. Post-optimization model selection by examining Pareto optimal solutions

Model selection by examining trade-offs of fit and complexity post-optimization has been used in several other applications. Users of Mallows' C$_p$ often conduct similar investigations. Freitas (2004) examined Pareto optimality in the related question comparing prediction and simplicity for data mining. Viewing each model's trade-offs, in terms of objectives, provides a visual assessment of the model selection problem, a potentially useful tool for ultimately choosing a model for inference or prediction. As is the case with any multi-objective optimization problem, the additional flexibility in model choice based on post-optimization specification of preferences could be viewed as either a positive or negative trait, depending on how an investigator values the order for which preferences are specified. Specifying preferences pre-optimization for the model selection problem benefits from being objective in the sense that a decision maker chooses how to weigh their specific objective functions without being influenced by how weights will alter the outcome of optimization. Specifying preferences post-optimization has the added flexibility of choosing a Pareto optimal solution that provides the best trade-offs for context dependent decision problems.

3.2.1. Avian species richness

We examined model selection *via* specification of preferences post-optimization using avian species richness data (Fig. 2). We identified Pareto optimal solutions among the 24 models, and then considered potential methods for selecting a model. To identify Pareto optimal solutions, we plotted the values f_1 and f_2 described in Eq. (7) for each model on opposing axes to identify solutions along the Pareto frontier (Fig. 2). Identifying Pareto optimal solutions does not require specifying w_1 or w_2, and therefore does not require adhering to an information criterion. The Pareto optimal set included 6 models; one model for each number of parameters 1, …, 7, except $p = 5$, where both model fit and complexity could be simultaneously improved by using the top model containing four parameters. Each Pareto solution represented the model that minimized Eq. (7) among all models with the same number of parameters. There were 17 dominated models (i.e., models that were not Pareto optimal; Fig. 2). The AIC top model was a Pareto optimal solution; this was expected because AIC (and other information criteria) is a specific formulation of the weighted-sum method and is therefore *sufficient* for Pareto optimality (i.e., if it is a specific formulation of the weighted-sum method it is Pareto optimal; Marler and Arora, 2010). Each of the Pareto solutions correspond to a specific set of weights in Eq. (6).

Given the information on Pareto optimal solutions in Fig. 2, selecting a final model for inference can proceed in many ways, depending on the application and the nature of the parameters under consideration. A decision maker can use the information on Pareto optimal solutions to view trade-offs of fit gained by adding (or subtracting) additional parameters from the model, and choose a Pareto optimal solution with trade-offs acceptable to the decision maker. Some parameters might be associated with covariates for which annual data are difficult, expensive, or impossible to collect. The trade-offs in terms of model fit can be assessed relative to the expense of collecting additional data for these parameters. If the increase in model fit from the Pareto optimal solution that requires the additional (expensive) covariate data does not justify the additional expense, another Pareto optimal solution may be preferred.

Another option is to examine the curvature of the Pareto frontier. For example, the Pareto frontier in Fig. 2 makes a curved shape with a point around $p = 3$. At this point, increasing the number of parameters has diminishing marginal returns in terms of f_1, and decreasing the number of parameters has a large effect on f_1. In the avian species richness data, the largest improvement in model fit, per parameter added, was adding area to the null model (an 86 unit improvement to fit; Fig. 2). Subsequent parameter additions showed diminishing marginal returns in model fit; the second biggest improvement in model fit was adding precipitation to the area model (23 units), followed by adding temperature to the area + precipitation model (7 units). No models with five parameters occurred on the Pareto frontier. Which model the investigator should choose is context dependent. By viewing these trade-offs between model fit and complexity, the decision maker has an additional tool to help inform context-dependent decisions.

Another approach is to compare the trade-offs to scientific significance of the parameters involved and the need to make inference on those parameters. For example, if a parameter is required to inform a management decision, such as the effect of management on survival rates for conservation decisions, a decision maker would want to choose a model where the effect of decisions on survival (or any other parameter) is included. Another approach might be to choose a Pareto optimal solution such that the maximum number of parameters is constrained by the amount of data. For example, if an investigator wishes to constrain the number of parameters in the model such that $p < \frac{n}{15}$, the investigator could select the Pareto optimal solution that maximized model fit within the constrained set. In the avian species richness data, with $n = 49$, this constraint would suggest choosing the Pareto optimal model with three parameters (with log linear predictor $\beta_0 + \beta_1 x_{\text{area},t} + \beta_2 x_{\text{temp},t}$; Table 2).

Models that are optimal in terms of model selection criteria could be highlighted as reference points on the Pareto frontier to guide decisions on the final model choice. Ultimately, the use of the Pareto frontier is that it provides visual information on the trade-offs of the objectives of the decision maker; in this case, maximizing model fit and minimizing complexity. Two dimensions are trivial to visualize, and techniques exist to visualize the Pareto optimal set if the MOO problem considered has > 2 objective functions (Lotov and Miettinen, 2008).

In our reasoning as well as our example, we assumed that $\gamma = 0$, although this assumption is not necessary (Table 1). The value of γ controls the shape of the model complexity function in Eq. (5). Hooten and Hobbs suggest that the choice of γ is more a result of personal preference based on desired inference; setting $\gamma = 2$ is analogous to the penalty term used in ridge regression, while setting $\gamma = 1$ is analogous to the penalty term used in LASSO.

Acknowledgments

Funding was provided by U.S. Geological Survey–Alaska Science Center, the Colorado State University Department of Fish, Wildlife, and Conservation Biology, and the Department of Statistics. Ken Burnham, Joel Schmutz, and Yu Wei provided comments that improved an earlier version of this manuscript. Any use of trade, firm, or product names is

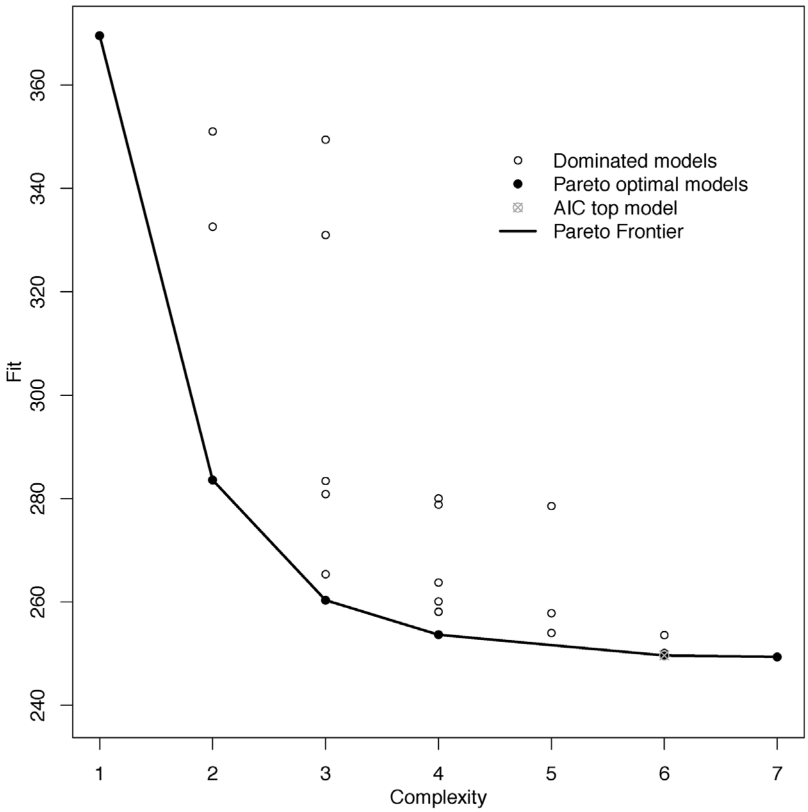

Fig. 2. Model fit ($f_1(\boldsymbol{\theta}) = -\log(L(\boldsymbol{\theta}|\mathbf{y}))$ vs. model complexity ($f_2(\boldsymbol{\theta}) =$ no. parameters) for each candidate model fit to avian species richness data. Optimal solutions minimize fit (moving towards bottom of figure) and complexity (moving to the left of figure). The top model using $f(\boldsymbol{\theta}) =$ AIC was a Pareto optimal solution. Two *candidate models* (i.e., ΔAIC < 2) were not Pareto optimal (i.e., they were dominated by another model). Note, the continuous black line is used to as a tool to help visualize the discrete Pareto optimal solutions, and is known as the Pareto frontier.

for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A

R statistical software script to fit models described in Table 2 to avian species richness data, and calculate and plot values for Eqs. (5) and (6) shown in Fig. 1.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecolmodel.2019.04.012.

References

Aho, K., Derryberry, D., Peterson, T., 2014. Model selection for ecologists: the worldviews of aic and bic. Ecology 95, 631–636.
Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csáki, F. (Eds.), 2nd International Symposium on Information Theory, pp. 267–281.
Araújo, M.B., Luoto, M., 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecol. Biogeogr. 16, 743–753.
Athan, T.W., Papalambros, P.Y., 1996. A note on weighted criteria methods for compromise solutions in multi-objective optimization. Eng. Optim. 27, 155–176.
Barron, H., Schmidt, C.P., 1988. Sensitivity analysis of additive multiattribute value models. Oper. Res. 36, 122–127.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media.
Cohon, J.L., 2013. Multiobjective Programming and Planning. Courier Dover Publications.
Das, I., Dennis, J.E., 1997. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. Struct. Optim. 14, 63–69.
Deb, K., 2001. Multi-Objective Optimization Using Evolutionary Algorithms, vol. 16 John Wiley & Sons.
Freitas, A.A., 2004. A critical review of multi-objective optimization in data mining: a position paper. ACM SIGKDD Explor. Newsl. 6, 77–86.
Gelfand, A.E., Ghosh, S.K., 1998. Model choice: a minimum posterior predictive loss approach. Biometrika 85, 1–11.
Hastie, T., Friedman, J., Tibshirani, R., 2009. The Elements of Statistical Learning, vol. 2 Springer.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14, 382–401.
Hooten, M.B., Cooch, E.G, 2019. Comparing ecological models. In: Marcot, B.G., Tri, A., Brennan, L.A. (Eds.), Quantitative Analysis in Wildlife Science. Johns Hopkins University Press, pp. 373–394.
Hooten, M.B., Hobbs, N.T., 2015. A guide to Bayesian model selection for ecologists. Ecol. Monogr. 85, 3–28.
Insua, D.R., 1990. Sensitivity Analysis in Multi-Objective Decision Making. Springer-Verlag.
Keeney, R.L., Raiffa, H., 1976. Decisions With Multiple Objectives: Preferences and Value Trade-Offs. Cambridge University Press.
Konishi, S., Kitagawa, G., 1996. Generalised information criteria in model selection. Biometrika 83, 875–890.
Link, W.A., Barker, R.J., 2006. Model weights and the foundations of multimodel inference. Ecology 87, 2626–2635.
Lotov, A.V., Miettinen, K., 2008. Visualizing the pareto frontier. Multiobjective Optimization. Springer, pp. 213–243.
Mallows, C.L., 1973. Some comments on c_p. Technometrics 15, 661–675.
Marler, R.T., Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. Optim. 26, 369–395.
Marler, R.T., Arora, J.S., 2010. The weighted sum method for multi-objective optimization: new insights. Struct. Multidiscip. Optim. 41, 853–862.
Parviainen, M., Luoto, M., Ryttäri, T., Heikkinen, R.K., 2008. Modelling the occurrence of threatened plant species in taiga landscapes: methodological and ecological perspectives. J. Biogeogr. 35, 1888–1905.
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Schwarz, G., et al., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
Stone, M., 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. J. R. Stat. Soc. Ser. B (Methodol.) 39, 44–47.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) 267–288.
Tikhonov, A.N., Goncharsky, A., Stepanov, V., Yagola, A.G., 2013. Numerical Methods for the Solution of Ill-Posed Problems, vol. 328 Springer Science & Business Media.
Williams, P.J., 2016. Methods for Incorporating Population Dynamics and Decision Theory in Cackling Goose Management. Colorado State University Ph.D. thesis.
Williams, P.J., Hooten, M.B., 2016. Combining statistical inference and decisions in ecology. Ecol. Appl. 26, 1930–1942.
Williams, P.J., Kendall, W.L., 2017. A guide to multi-objective optimization for ecological problems with an application to cackling goose management. Ecol. Model. 343, 54–67.