Improving spatial predictions of animal resource selection to guide conservation decision making

Brian D. Gerber[1], Joseph M. Northrup[2]

[1]Department of Natural Resources Science, University of Rhode Island, Kingston, RI 02881-2018, USA. Corresponding Author: bgerber@uri.edu.

[2]Wildlife Research and Monitoring Section, Ontario Ministry of Natural Resources and Forestry and Environmental and Life Sciences Graduate Program, Trent University, Peterborough, ON, K9L 1Z8, Canada.

Running Head: Optimizing RSF predictions.

This is the unedited and accepted draft of a manuscript in press at Ecology:
https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2953

Abstract

Resource selection is often studied by ecologists interested in the environmental drivers of animal space use and movement. These studies commonly produce spatial predictions, which are of considerable utility to resource managers making habitat and population management decisions. It is thus paramount that predictions from resource selection studies are accurate. We evaluated model building and fitting strategies for optimizing resource selection function predictions in a use-availability framework. We did so by simulating low- and high-intensity spatial sampling data that respectively predicted study area and movement-based resource selection. We compared one of the most commonly used forms of statistical regularization, Akaike's Information Criterion (AIC), with the lesser used least absolute shrinkage and selection operator (LASSO). LASSO predictions were less variable and more accurate than AIC and were often best when considering additive and interacting variables. We explicitly demonstrate the predictive equivalence using the logistic and Poisson likelihoods and how it is lost when the available sample is too small. Regardless of modeling approach, interpreting the sign of coefficients as a measure of selection can be misleading when optimizing for prediction.

Key words: AIC; habitat selection; LASSO; movement ecology; optimal; prediction; regularization; resource selection function; RSF; spatial ecology.

Introduction

An understanding of habitat selection is integral to the study of animal ecology and evolution. By tracking individual animal movements, we can understand the behavioral processes by which animals choose locations to maximize fitness (McLoughlin et al. 2010). This selection process subsequently provides important insights into population and community dynamics (Morris 2002). Advances in animal tracking data (e.g., global positioning system radio collars) have revolutionized our ability to assess habitat selection patterns. The most common method for examining habitat selection from animal tracking data is the resource selection function (RSF), fit in a use-availability framework (Manly et al. 2002; Hooten et al. 2017). Under this

framework, animal locations (the used sample) and their underlying environmental covariates (e.g., land cover type) are contrasted with random locations and their underlying environmental covariates that were considered available to the animal (the available sample).

The available sample can be defined as the spatial region an animal could have accessed from each used location. In low-intensity tracking studies (e.g., a few locations per day), it is reasonable to assume highly mobile species (e.g., large mammal or bird) could traverse their home range or larger between used locations. The key word being *could*, rather than did or likely. These studies are common, as researchers favoring long tracking periods, perhaps for estimating demographic processes, can extend battery life of telemetry devices by acquiring fewer locations per day. In contrast, high-intensity tracking studies (e.g., 1 location/30 minutes) acquire many temporally correlated used locations thereby limiting the available sample to the area along the path of used locations; the available sample may be estimated based on a movement process that accounts for this correlation (Johnson et al. 2008; Hooten et al. 2017). Importantly, how the available sample is defined dictates the inference on the scale of resource selection (Northrup et al. 2013; Hooten et al. 2017; Gerber et al. 2018).

The RSF can be understood as a spatial point process (Hooten et al. 2017), in which the ith used location ($\boldsymbol{\mu}_i$; consisting of x-y coordinates in space) is a realization from a weighted probability distribution, $\boldsymbol{\mu}_i \sim [\boldsymbol{\mu}_i|\boldsymbol{\beta}, \boldsymbol{\theta}]$, such that

$$[\boldsymbol{\mu}_i|\boldsymbol{\beta}, \boldsymbol{\theta}] \equiv \frac{g(\mathbf{x}(\boldsymbol{\mu}_i, \boldsymbol{\beta}))f(\boldsymbol{\mu}_i, \boldsymbol{\theta})}{\int g(\mathbf{x}(\boldsymbol{\mu}_i, \boldsymbol{\beta}))f(\boldsymbol{\mu}_i, \boldsymbol{\theta})d\boldsymbol{\mu}}, \tag{1}$$

where we interpret $g(\mathbf{x}(\boldsymbol{\mu}_i, \boldsymbol{\beta}))$ (i.e., the RSF) as how animals preferentially choose resources based on selection coefficients ($\boldsymbol{\beta}$) and what is considered available to them ($f(\boldsymbol{\mu}_i, \boldsymbol{\theta})$) based on availability coefficients ($\boldsymbol{\theta}$). Note, the denominator integrates over the spatial region that is available to the animal for all used locations and when the availability is uniform over this region, $\boldsymbol{\theta}$ drops out of this equation (Hooten et al. 2017). Commonly, the RSF is defined using the exponential form as, $g(\mathbf{x}(\boldsymbol{\mu}_i, \boldsymbol{\beta})) \equiv e^{\mathbf{x}'(\boldsymbol{\mu}_i)\boldsymbol{\beta}}$, but also sometimes using the logistic form (Lele and Keim 2006) as, $g(\mathbf{x}(\boldsymbol{\mu}_i, \boldsymbol{\beta})) \equiv \frac{e^{\mathbf{x}'(\boldsymbol{\mu}_i)\boldsymbol{\beta}}}{1+e^{\mathbf{x}'(\boldsymbol{\mu}_i)\boldsymbol{\beta}}}$. Resources chosen in greater or lesser proportion to their availability are considered *selected* and *avoided*, respectively. The main

difference is that the logistic form (also called the resource selection probability function) makes inference to the probability of selection and relies on strict parameteric assumptions that are not robust (Hastie and and Fithian 2013), while the exponential form makes inference to the relative density of used points, interpreted as the relative intensity of selection, which is proportional to the probability of selection. The latter is a relative intensity because the number of possible locations is unknown or realistically infinite under a continuous process, such that the intercept reflects the observed sample size (Warton and Sheperd 2010).

Most researchers do not fit the weighted distribution directly (but see, Lele and Keim 2006; Hooten et al. 2017; Gerber et al. 2018). Rather, it is more common to make inference on the exponential form of the RSF via approximation using generalized linear modeling (i.e., logistic or Poisson regression) with familiar and available software (Northrup et al. 2013), such as via the glm() function in R. Both logistic and Poisson regression can provide equivalent inference on the selection coefficients ($\boldsymbol{\beta}$) when certain conditions are met (Aarts et al. 2012; Fithian and Hastie 2013); namely, the number of grid cells in the Poisson regression and the number of locations in the available sample of the logistic regression needs to be very large (Northrup et al. 2013) to ensure the integral in the denominator of Eq. 1 is approximated well (Warton and Sheperd 2010). Furthermore, the logistic regression is improved when the available sample is infinitely weighted (Fithian and Hastie 2013), which in practice means weighting these data by a large number (e.g., 1000) and weighting the used sample by one.

The objectives of resource selection studies are typically focused on evaluating ecological and conservation driven hypotheses (e.g., Chetkiewicz and Boyce 2009; McLoughlin et al. 2010) to infer how spatial factors influence habitat selection. However, the practical utility of an RSF for many resource managers and conservationists is the spatially mapped predictions produced from these models (Morris et al. 2016), which can influence on the ground management decisions. Resource selection predictions are used for land-use planning (Coates et al. 2016), managing populations (Hebblewhite et al. 2011; Northrup et al. 2016), and more (Morris et al. 2016). Because RSF predictions are widely relied upon in conservation and

management decision making, it is of paramount importance to obtain accurate predictions.

Resource selection studies typically adopt an explanatory modeling process (Shmueli 2010; Gerber et al. 2015) aimed at inferential model building and estimation based on a relatively small set of hypotheses and associated covariates (Burnham and Anderson 2002). Specifically, models are often a limited combination of potential covariates (typically only assumed to affect selection in an additive manner) compared using Akaike's Information Criterion (AIC; Boyce et al. 2002). Since many covariates are expected to have small effects or are collinear with other covariates, there are many potential variables that are never synthetically considered within a model comparison framework. Multicollinearity is a well known estimation issue in ecology (Graham 2003), which is prevalent in resource selection studies that often rely on remotely sensed data to produce many covariates from the same source products. Ecologists commonly ameliorate multicollinearity by excluding variables from a model or model set. Maintaining a small set of models fits into the hypothetico-deductive scientific framework, as it focuses on inference to specific and hypothesized factors driving resource selection. As such, variables are often not considered so as to maintain a single model or a small set of models, which is encouraged when using information criterion (Burnham and Anderson 2002). Prediction within a resource selection study is usually considered apart from model building and estimation to evaluate a final selected model (Boyce et al. 2002).

We contend that viewing model fitting and selection in a statistical regularization framework has benefits when seeking to optimize resource selection models for prediction. Regularization is a statistical technique that seeks to optimize the generalizabiltiy of a model by trading off bias and variance (Bickel et al. 2006). Regularization encompasses most forms of model selection commonly used in ecology, which in resource selection studies is the use of information criterion and specifically, AIC (Boyce et al. 2002). Information criterion is used to evaluate discrete model sets and relies on asymptotic assumptions to justify predictive performance (Stone 1977). In contrast, alternative regularization techniques use a continuous model selection process from a global to an intercept-only model by constraining estimated

coefficients via a shrinkage parameter that can be optimally chosen via cross-validation, often leading to improved prediction performance (Hastie et al. 2009; Gerber et al. 2015). Continuous model selection can also be computationally more efficient than evaluating all possible subsets of discrete models, which can be a prohibitively large number of models.

We highlight one of the more common continuous regularization techniques in applied statistics, the least absolute shrinkage and selection operator (LASSO; Tibshirani 1996). Notably, LASSO has variable selection properties and can remove effects of variables by constraining them to be zero, which gives the optimal model an additional amount of interpretability over other techniques (e.g., ridge regression; Hastie et al. 2009). Further, LASSO can accommodate the numerical issues of moderate multicollinearity, maintaining good predictive performance (Dormann et al. 2013), and thus does not necessitate removing partially collinear variables from models or model sets. Unless variables are completely correlated, there is potential information that could be useful to improve predictions; such information is lost when only one set of collinear variables is considered. Simply, LASSO is an integrated model-selection and estimation technique that leverages the power of cross-validation to identify a set of coefficients that optimizes predictive performance. We focus on LASSO because it identifies sparse models that may be useful for inference on resource selection, as well as optimal prediction.

We can compare LASSO and AIC by their optimization routines, in which we estimate model parameters (e.g., $\boldsymbol{\beta}$) by minimizing {model lack of fit + $\lambda \times$ model complexity}, where λ is a penalization or shrinkage factor. Model lack of fit for both is the deviance ($-2 \times \log(\mathcal{L}(\boldsymbol{\beta}))$). While AIC defines $\lambda = 2$ based on theory, LASSO allows this value to be chosen, typically using cross-validation. Further, AIC considers model complexity as the number of parameters (q = 0) and includes the intercept (a = 1), while LASSO measures the number and magnitude of the absolute value of parameters (q = 1) and does not penalize the intercept (a = 2), such that the optimization argument for estimating K total parameters is $\arg\min_{\boldsymbol{\beta} \in \mathbb{R}^K} \{-2 \times \log(\mathcal{L}(\boldsymbol{\beta})) + \lambda \times \sum_{k=a}^{K} |\beta_k|^q\}$. Note, q = 2 and a = 2 defines ridge

regression. Both LASSO and ridge have natural Bayesian interpretations (Hastie et al. 2009; Gerber et al. 2015). For more specifics, see Bickel et al. 2006 and Hastie et al. 2009.

We considered two types of common animal telemetry data for predicting resource selection, low- and high-intensity individual sampling. Low-intensity sampling data represent individuals that are tracked infrequently (relative to their potential rate of movement), such that we assume no temporal correlation in sequential used locations, and inference to selection is over a large spatial region (i.e., home-range) that is considered available. High-intensity sampling data represent individuals tracked frequently where used locations are realizations from an animal movement process with temporal correlation between sequential used locations and the availability is defined by estimated step-lengths and turning angles. We focus on individual-level analyses as they are the fundamental unit of interest in resource selection studies and selection is expected to vary by individual (Montgomery et al. 2018). We compared LASSO and AIC using two model building strategies, only additive combinations of variables, and additive and pairwise interactions of all variables. Further, while statistical theory has clarified the equivalence between the logistic and Poisson approximation of the weighted distribution (Warton and Sheperd 2010; Fithian and Hastie 2013), there has yet to be a simple comparison of models with equivalent covariates fit with both likelihoods that is approachable for practitioners; as such, using the low-intensity data we compared all model building and fitting strategies using Poisson and logistic linear models. Last, we compared empirical results from individual movement-based RSF analyses optimized by AIC or LASSO, using location data from 44 mule deer (*Odocoileus hemionus*) in Colorado (Northrup et al. 2015).

Materials and Methods

Simulation

We compared model building and fitting strategies in a simulation study where the true process that we seek to predict is known. Specifically, we simulated low- and high-intensity individual used locations using an intensity function that combines additive and pairwise

interactions of categorical (x_{i1}, x_{i2}, x_{i3}) and continuous variables $(x_{i4}, x_{i5}, x_{i6}, x_{i7}, x_{i8}, x_{i9}, x_{i10})$ with varying effect sizes ($\boldsymbol{\beta} \equiv [1, 2, 1, 1, -1, 0.5, -0.5, 0.5, -2, 0.5, 2, -2, -2, 2]'$), as

$$\begin{aligned} e^{\mathbf{x}'(\boldsymbol{\mu}_i)\boldsymbol{\beta}} \equiv \exp(&\beta_0 x_{i1} + \beta_1 x_{i2} + \beta_2 x_{i3} + \beta_3 x_{i4} + \beta_4 x_{i5} + \beta_5 x_{i6} + \beta_6 x_{i7} + \beta_7 x_{i8} + \beta_8 x_{i9} \\ &+ \beta_9 x_{i10} + \beta_{10}(x_{i4} \times x_{i7}) + \beta_{11}(x_{i8} \times x_{i10}) + \beta_{12}(x_{i2} \times x_{i7}) + \beta_{13}(x_{i3} \times x_{i7})). \end{aligned} \tag{2}$$

The categorical variable mimics land-cover type with three levels (e.g., forest, shrub and grassland), while the continuous variables mimic landscape features, such as elevation, ruggedness, etc. All spatial variables ($\mathbf{x}$) were simulated as continuous Gaussian random fields (Appendix S1: Fig. S1). We simulated a second set of variables, $\mathbf{w}$, that were considered as potential covariates hypothesized to influence resource selection, but were not directly related to the true RSF; $\mathbf{w}$ consisted of one categorical (three levels) and eight continuous variables (Appendix S1: Fig. S2). The continuous variables of $\mathbf{w}$ and $\mathbf{x}$ are minimally and maximally correlated (r) from -0.24 to 0.80 (Appendix S1: Fig. S3). As such, we are considering a common issue, in that many spatial variables are hypothesized and some or many of those are naturally or circumstantially correlated with each other.

For the low-intensity sampling simulation, we used $e^{\mathbf{x}'(\boldsymbol{\mu}_i)\boldsymbol{\beta}}$ to simulate 2000 data sets from an inhomogenous Poisson point process that ranged in the number of used locations from 150 to 350,000, such that the proportion of the landscape used at least once ranged from approximately 0-100%. For the high-intensity sampling, we simulated used locations from an equivalent intensity function using a movement-based process following the approach outlined by Muff et al. 2019 (see Appendix S2). However, we redefined the habitat variables of $\mathbf{x}$ and $\mathbf{w}$ to make them more patchy and thus were appropriately encountered when simulating animal movements (Appendix S1: Fig. S4). We varied the number of total steps (used sample) by 100, 500, 1000, 2000, 5000, and 10000. At each step, we predicted 100 random locations as the available sample for each used location. For each step size, we simulated 200 individual animal tracks; all simulations and model fitting was done in R (version 3.6.0); code can be found in Data S1.

Model Building and Fitting

We fit models to each simulated data set using model building strategies that included either a model set with all combinations of additive covariates or all combinations of additive and pairwise interactions of covariates. Model fitting strategies included either selecting an optimal predictive model via AICc (AIC with small sample-size correction; Burnham and Anderson 2002) or LASSO. For each model building strategy, we considered all $\mathbf{x}$ and $\mathbf{w}$ covariates, except for x_4, which we exclude to represent an important variable that was not hypothesized or could not be appropriately measured, and thus can not be included in a model set. We also fit each data set using the correctly specified model (i.e., exact set of covariates and their interactions used to simulate the data) as a benchmark for the best case for each strategy and data set. All continuous covariates were centered and standardized to a mean of zero and standard deviation of one. For modeling the low-intensity data using logistic regression, the available sample was the entire study area with each zero weighted by 1000. We demonstrated the predictive equivalence of the Poisson and logistic likelihoods by comparing predictions for all combinations of model building and fitting strategies and how it is lost by reducing the available sample using the logistic likelihood to 1000 random samples that are not weighted. For the high-intensity sampling data, we fit models using conditional logistic regression where each strata corresponds to a single used location that is matched with a set of corresponding available locations (Northrup et al. 2013). Mapped RSF predictions indicate the relative intensity of selection of a location conditional on all locations on the map being equally available to the animal.

For strategies using AICc, we randomly removed collinear variables with a correlation $\geq$ 0.6 to determine the global model before evaluating all possible subsets via an automatic model selection routine in the R package 'glmulti' (Calcagno and Calcagno 2010) for logistic and Poisson analyses and in the package MuMIn for conditional logistic regression analyses. Predictions were model averaged using Akaike weights (Burnham and Anderson 2002). For strategies using LASSO, we did not remove collinear variables and regularized coefficients via

a complete set of shrinkage parameters (λ); we evaluated each shrinkage parameter via 10-fold cross-validation using the average deviance (-2 $\times\log(\mathcal{L}(\boldsymbol{\beta}))$) of the left out data across all folds. Note, for conditional logistic regression the left out folds occurred by strata. Logistic and Poisson modeling with LASSO and cross-validation was done using the R package 'glmnet' (Friedman et al. 2010) and conditional logistic modeling was done using 'clogitL1' (Reid and Tibshirani 2014). See Appendix S2 for additional details on cross-validation.

We evaluated RSF predictions against the true RSF in three ways. First, we computed Kendall's rank correlation coefficient (τ), which measures the similarity of the ordering of continuous quantities by comparing concordant and discordant pairs. A high value of τ indicates that two continuous quantities have a similar ranking order. It does not guarantee that the relative difference between similarly ranked predictions and the true values are the same. Second, we computed the coefficient of determination (R^2), which measures the proportion of the variance in the true values that is predictable from the RSF predictions. Third, we computed the mean absolute error (MAE) between the true and predicted RSF values after standardizing them to be between zero and one. A good model fitting and selection strategy should have a high τ and R^2, a low MAE, and is consistent within a sample size, such that these measures vary little. We plot results by sample size when considering all RSF predictions together and for the low-intensity results we also binned the true RSF values into quartile groups of low to high selection (0-25%, 25-50%, 50-75%, 75-100%) and calculated τ, R^2, and MAE with their corresponding RSF predictions. Binning predictions is commonly done when creating resource selection maps for managers (Morris et al. 2016) and clarifies which values are most difficult to predict. Lastly, we investigated the inferential reliability in interpreting estimated coefficients as selection and avoidance by evaluating the proportion of coefficients with the correct sign (+, 0, -) across simulations within each strategy.

Empirical Case Study

We used location data from 44 mule deer in the piceance basin of Colorado to fit movement-based RSF models optimized using AICc or LASSO. Used locations by individual

ranged from 240 to 330 and each used location was matched with 300 available locations in a temporally dynamic manner following the process outlined by Northrup et al. 2015. We specifically compared predictions using AICc with additive variables and LASSO with additive and pairwise interactions. We evaluated predictive differences by measuring τ, R^2, and the mean standard deviation of the difference between predictions. Further, we evaluated within-sample predictive performance using the ratio in the deviance explained by the LASSO strategy relative to the AICc strategy; values >1 indicate improved prediction using LASSO. Lastly, we evaluated out of sample predictive performance by withholding 10% of each individuals data and fitting the remaining data with the LASSO and AICc strategies. Specifically, we measured the mean individual proportional change in deviance; values >1 indicate improved prediction using LASSO.

Results

We found that model building (Additive or Additive & Interactions) and fitting strategies (LASSO or AICc) led to important differences in predicting resource selection for both low- and high-intensity modeling approaches (Figs. 1-3). First, preliminary investigations determined that AICc with all possible additive and pairwise interactions led to inconsistent results (Appendix S1: Figs. S5-S7) that were rarely more accurate than using LASSO with pairwise interactions and often less accurate than using AICc or LASSO with only additive combinations of variables. Thus, due to the computational issues of fitting >1 billion models per data set we removed this approach from further consideration. Across all strategies, we found that using LASSO always led to more accurate and consistent results (i.e., low variation in τ, R^2 and MAE for a given sample size; Figs. 2, 3) than using AICc. The combination of randomly removing collinear variables, the instability of comparing many models using AICc, and the lack of explicit predictive evaluation via cross-validation led to the observed high variability in prediction agreement for similar sample sizes. Considering pairwise interactions with LASSO generally improved τ, R^2, and MAE compared to only additive models, except at the smaller sample sizes. We also found that optimizing for prediction can lead to poor

inference on the selection and avoidance of resources when interpreting the sign of estimated coefficients when modeling low- or high-intensity sampling data (Appendix S1: Figs. S8-S9).

We found that modeling low-intensity tracking data using a large weighted available sample for the logistic likelihood produced equivalent predictions as when using the Poisson likelihood (Fig. 2). The exception was a small, but consistent difference in predictions between likelihoods when using LASSO with additive and pairwise interactions. Predictive equivalence between likelihoods breaks down substantially for all model fitting strategies when the available sample is too small (Appendix S1: Figs. S10-S11). Notably, fitting the correct structural model with too small available sample reduced τ up to 0.23, R^2 up to 0.22, and increased MAE up to 0.17.

By binning the low-intensity tracking results, we found that low (0-25%) and high relative intensity of selection (75-100%) were universally easier to predict (Figs. A12-A14). Low-intensity of selection predictions using LASSO produced a τ ranging from 0.62 to 0.70 for additive only and 0.70 to 0.82 for additive and pairwise interactions. The corresponding R^2 ranged from 0.68 to 0.78 for additive models and 0.85 to 0.90 for additive and interaction models, while the MAE ranged from 0.035 to 0.045 for additive models and 0.01 to 0.02 for additive and pairwise interactions. The medium relative intensity of selection categories (25-50% and 50-75%) were generally comparable to one another and much worse in terms of τ, R^2 and MAE relative to the high and low bins (Figs. A12-A14).

Empirical deer RSF modeling indicated that predictions were very different when optimizing using AICc with additive variables and LASSO with additive and pairwise interactions (Fig. 1, Appendix S1: Figs. S15-S20). Across individuals, the mean τ, R^2, and standard deviation of prediction difference was 0.56 (range, 0.36-0.70), 0.37 (range, 0.07-0.71), and 2.70 (range, 0.29-21.82), respectively. Comparing within-sample predictive performance, the LASSO always outperformed the AICc strategy by improving the deviance explained by a mean of 2.60 times (range, 1.59-5.40) across individuals. Comparing out-of-sample predictive performance, the LASSO generally outperformed the AICc strategy by improving the deviance

by a mean of 1.75 times (range, 0.78-6.15) across individuals.

Discussion

We found that common model building strategies for RSF analyses (models of additive variables compared using AIC) led to highly inconsistent and sub-optimal predictions. A substantial gain in predictive accuracy and reliability can be made by adopting a continuous statistical regularization framework that leverages the power of cross-validation and efficient and stable computational algorithms. LASSO improved predictions in terms of τ, R^2, and mean absolute error across all sample sizes for modeling low- and high-intensity sample data. Further, we found that considering all pairwise-interactions with LASSO led to improved predictions, despite the increased estimation complexity. Perhaps an important but unsurprising finding was that predictions were best for the most strongly selected and avoided areas when maps were binned. This is critical, because many studies seek to identify habitat vs. non-habitat for species, which requires high resolution at the mid-ranges of the RSF, which might be difficult to achieve.

Our results highlight the equivalence between the logistic and Poisson likelihoods in approximating the weighted distribution, which has been discussed elsewhere, but is perhaps not appreciated by practitioners. When using the logistic likelihood, care needs to be taken to use a large available sample (Northrup et al. 2013) that is weighted with a large number (Fithian and Hastie 2013) to ensure a proper approximation of the weighted distribution (Eqn. 1). Otherwise, coefficients and predictions could be poor (example code is provided in Data S1). Our work also highlights an important finding of Warton and Shepard (2010), that because RSFs are point process models and can be fit in a generalized linear modeling framework, all the tools available to fitting such models can be used. This includes the efficient and robust algorithms that have been developed for continuous statistical regularization.

There are many regularization techniques that could be highly useful in optimizing RSFs for predictive performance. One alternative to LASSO is ridge regression, which also shrinks coefficients continuously, can accommodate extreme multicollinearity, but does not

have variable selection properties (Hastie et al. 2009). Ridge shrinks larger coefficients more than smaller ones, while LASSO shrinks them uniformly. LASSO also tends to remove one of two highly correlated variables, while ridge will shrink their coefficients towards one another. Which performs better depends on the number of variables considered, the true distribution of small and large effects, and whether there are many hypothesized variables that have no effect (Hastie et al. 2009). The generalization of ridge and LASSO is the elastic net (Zou and Hastie. 2005). Elastic net can accommodate extreme multicollinearity and lead to sparse interpretable models; however, for our context we found LASSO and elastic net to perform equivalently (Appendix S2). Flexible cross-validation along with ridge, LASSO, elastic net, and more are available in the R language and can be implemented using the 'glmnet' package (Friedman et al. 2010); example code is provided for each in Data S1. For researchers that want to relax assumptions of linearity, generalized additive models (Hastie et al. 2009) or boosted regression trees (Elith et al. 2008) are an option.

It is important to recognize the potential inferential costs of an optimal predictive modeling approach with correlated variables. We found among all simulation scenarios that the sign of estimated coefficients are not reliable in terms of evaluating whether a resource is being selected or avoided. There is a necessary trade-off between prediction and understanding when modeling, such that a single modeling approach will unlikely be optimal for both purposes. When use of spatial predictions from RSFs for conservation and management decision making is a priority, regularization techniques that optimize predictions using cross-validation should be employed.

Acknowledgments

The mule deer data was collected by Colorado Parks and Wildlife (CPW) and we thank C.R. Anderson Jr. for graciously allowing us to use these data as well as the CPW staff that helped collect the data and the numerous funding sources of the research conducted in the Piceance Basin of Colorado. We would also like to thank two anonymous reviewers for improving the manuscript. This work was supported by the USDA National Institute of Food and

Agriculture, Hatch Formula project 1017848.

References

Aarts, G., Fieberg, J., & Matthiopoulos, J. 2012. Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution* **3**:177-187.

Bickel, P. J., Li, B., Tsybakov, A. B., van de Geer, S. A., Yu, B., Valds, T. et al. 2006. Regularization in statistics. *Test*, **15**, 271-344.

Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. 2002. Evaluating resource selection functions. *Ecological Modelling* **57**:281-300.

Burnham, K. P., & Anderson, D. R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York..

Calcagno, V., & de Calcagno, C. 2010. glmulti: an R package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software* **34**:1-29.

Chetkiewicz, C. L. B., & Boyce, M. S. 2009. Use of resource selection functions to identify conservation corridors. *Journal of Applied Ecology* **46**:1036-1047.

Coates, P. S., Casazza, M. L., Ricca, M. A., Brussee, B. E., Blomberg, E. J., et al. 2016. Integrating spatially explicit indices of abundance and habitat quality: an applied example for greater sagegrouse management. *Journal of Applied Ecology* **53**:83-95.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carr, G., ... & M*ü*nkem*ü*ller, T. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**:27-46.

Elith, J., Leathwick, J. R., & Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology, **77**:802-813.

Fithian, W., & Hastie, T. 2013. Finite-sample equivalence in statistical models for presence-only data. *The Annals of Applied Statistics* **7**:1917-1939.

Friedman, J., Hastie, T., & Tibshirani, R. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**:1-22.

Gerber, B. D., Kendall, W. L., Hooten, M. B., Dubovsky, J. A., & Drewien, R. C. 2015. Optimal population prediction of sandhill crane recruitment based on climatemediated habitat limitations. *Journal of Animal Ecology* **84**:1299-1310.

Gerber, B. D., Hooten, M. B., Peck, C. P., Rice, M. B., Gammonley, J. H., Apa, A. D., & Davis, A. J. 2018. Accounting for location uncertainty in azimuthal telemetry data improves ecological inference. *Movement Ecology* **6**:14.

Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* **84**:2809-2815.

Hastie, T., Tibshirani, R. & Friedman, J. 2009. The elements of statistical learning. Springer, New York.

Hastie, T., & Fithian, W. 2013. Inference from presence-only data; the ongoing controversy. *Ecography* **36**:864-867.

Hebblewhite, M., Miquelle, D. G., Murzin, A. A., Aramilev, V. V., & Pikunov, D. G. 2011. Predicting potential habitat and population size for reintroduction of the Far Eastern leopards in the Russian Far East. *Biological Conservation* **144**:2403-2413.

Hooten, M. B., Johnson, D. S., McClintock, B. T., & Morales, J. M. 2017. Animal movement: statistical models for telemetry data. CRC Press, Boca Rotan.

Johnson, D. S., Thomas, D. L., Ver Hoef, J. M., & Christ, A. 2008. A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, **64**, 968-976.

Lele, S. R., & Keim, J. L. 2006. Weighted distributions and estimation of resource selection probability functions. *Ecology* **87**:3021-3028.

Manly, B. F. J., McDonald, L. L., Thomas, D. L., Mcdonald, T. L., and Erickson, W. P. 2002. Resource selection by animals: statistical design and analysis for field studies. Second edition. Kluwer Academic Publishers, Dordrecht, The Netherlands.

McLoughlin, P. D., Morris, D. W., Fortin, D., Vander Wal, E., & Contasti, A. L. 2010. Considering ecological dynamics in resource selection functions. *Journal of Animal Ecology* **79**, 4-12.

Montgomery, R. A., Redilla, K. M., OrtizCalo, W., Smith, T., Keller, B., & Millspaugh, J. J. 2018. Evaluating the individuality of animalhabitat relationships. *Ecology and evolution*, **8**:10893-10901.

Morris, D. W. 2003. Toward an ecological synthesis: a case for habitat selection. *Oecologia* **136**:1-13.

Morris, L. R., Proffitt, K. M., & Blackburn, J. K. 2016. Mapping resource selection functions in wildlife studies: Concerns and recommendations. *Applied Geography* **76**:173-183.

Muff, S, Signer, J, and Fieberg, J. Accounting for individualspecific variation in habitatselection studies: Efficient estimation of mixedeffects models using Bayesian or frequentist computation. Journal of Animal Ecology. https://doi.org/10.1111/1365-2656.13087

Northrup, J. M., Hooten, M. B., Anderson, C. R., & Wittemyer, G. 2013. Practical guidance on characterizing availability in resource selection functions under a useavailability design. *Ecology* **94**:1456-1463.

Northrup, J. M., Anderson Jr, C. R., & Wittemyer, G. 2015. Quantifying spatial habitat loss from hydrocarbon development through assessing habitat selection patterns of mule deer. *Global Change Biology*, **21**:3961-3970.

Northrup, J. M., Anderson, C. R., Hooten, M. B., & Wittemyer, G. 2016. Movement reveals scale dependence in habitat selection of a large ungulate. *Ecological Applications* **26**:2746-2757.

Reid, S., & Tibshirani, R. 2014. Regularization paths for conditional logistic regression: the clogitL1 package. *Journal of statistical software*, **58**:12.

Shmueli, G. 2010. To explain or to predict? *Statistical Science*, 25:289-310.

Stone M. 1977. An asymptotic equivalence of choice of model by cross-validation and Akaikes criterion. *Journal of the Royal Statistical Society Series B* **39**:44-7.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* **1**:267-288.

Warton, D. I., & Shepherd, L. C. 2010. Poisson point process models solve the pseudo-absence problem for presence-only data in ecology. *The Annals of Applied Statistics* **4**:1383-1402.

Zou, H., & Hastie, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B* **67**:301-320.

Figure 1. The true RSF and exemplars of RSF predictions by varying the model building and fitting strategies for simulated low-intensity (a) and high-intensity (b) sampling data, and comparative predictions from three mule deer (row) from the piceenace basin in Colorado, USA with AIC predictions in column one and LASSO predictions in column two (c).

Figure 2. Measures of agreement between the true RSF and predictions from modeling low-intensity sampling data using the logistic (□) and Poisson likelihood (○) by sample size across model building and fitting strategies. Agreement is measured by Kendall's τ, R^2, and mean absolute error. For logistic models, the available sample were all locations of the landscape weighted by 1000. Note that the y-axis is different for the bottom figure and Prop (%) is the proportion of the landscape used at least once.

Figure 3. Measures of agreement between the true RSF and predictions from modeling high-intensity sampling data using conditional logistic regression by sample size across model building and fitting strategies. Agreement is measured by Kendall's τ, R^2, and mean absolute error. The y-axis labels: CM is 'Correct Model', 'L-Add' is LASSO with additive variables, 'L-Int' is LASSO with additive and pairwise interactions, and 'AIC-Add' is Akaike's Information Criterion (with a correction for small sample size) with additive variables.

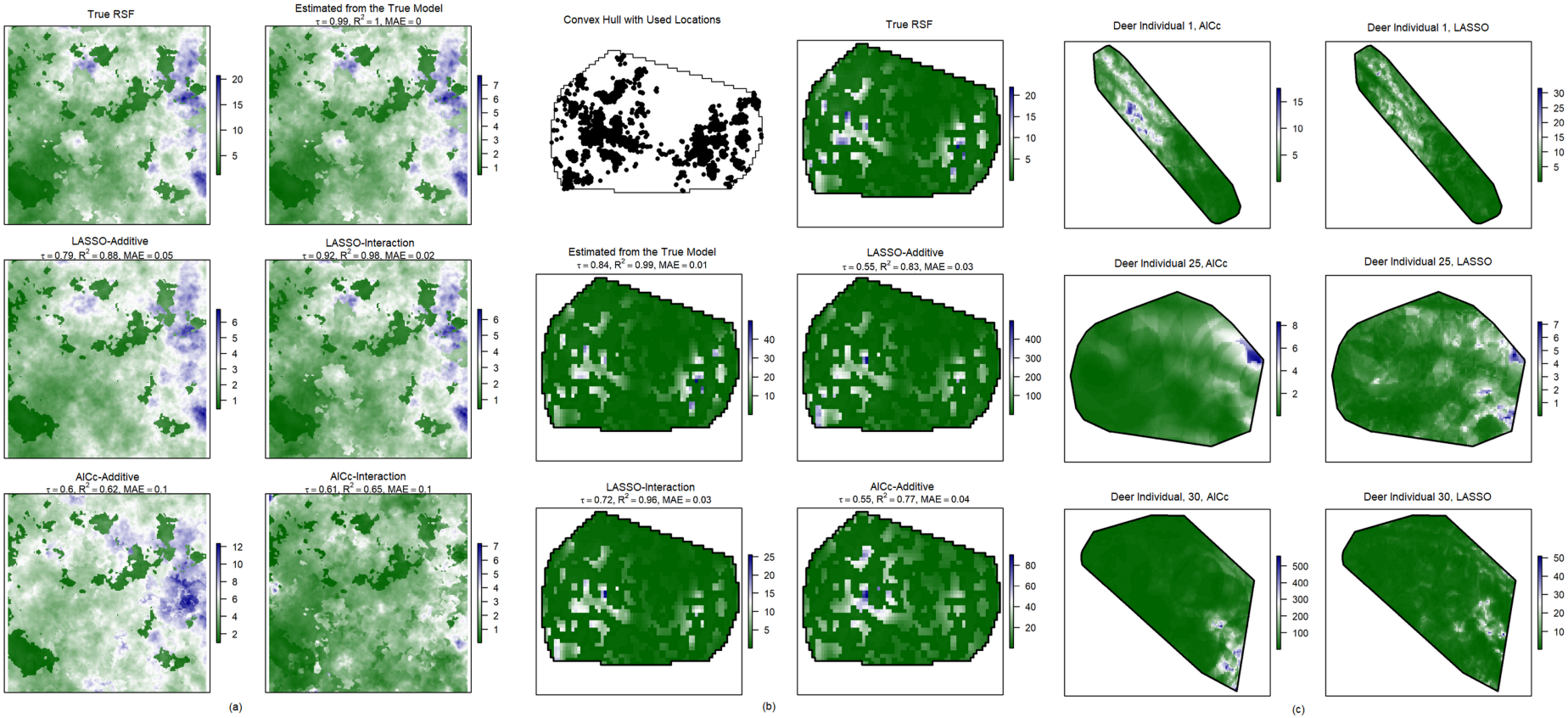
True RSF
Estimated from the True Model
τ = 0.99, R2 = 1, MAE = 0
LASSO-Additive
τ = 0.79, R2 = 0.88, MAE = 0.05
LASSO-Interaction
τ = 0.92, R2 = 0.98, MAE = 0.02
AICc-Additive
τ = 0.6, R2 = 0.62, MAE = 0.1
AICc-Interaction
τ = 0.61, R2 = 0.65, MAE = 0.1
(a)
Convex Hull with Used Locations
True RSF
Deer Individual 1, AICc
Deer Individual 1, LASSO
Estimated from the True Model
τ = 0.84, R2 = 0.99, MAE = 0.01
LASSO-Additive
τ = 0.55, R2 = 0.83, MAE = 0.03
Deer Individual 25, AICc
Deer Individual 25, LASSO
LASSO-Interaction
τ = 0.72, R2 = 0.96, MAE = 0.03
AICc-Additive
τ = 0.55, R2 = 0.77, MAE = 0.04
Deer Individual, 30, AICc
Deer Individual 30, LASSO
(b)
(c)

Figure 1

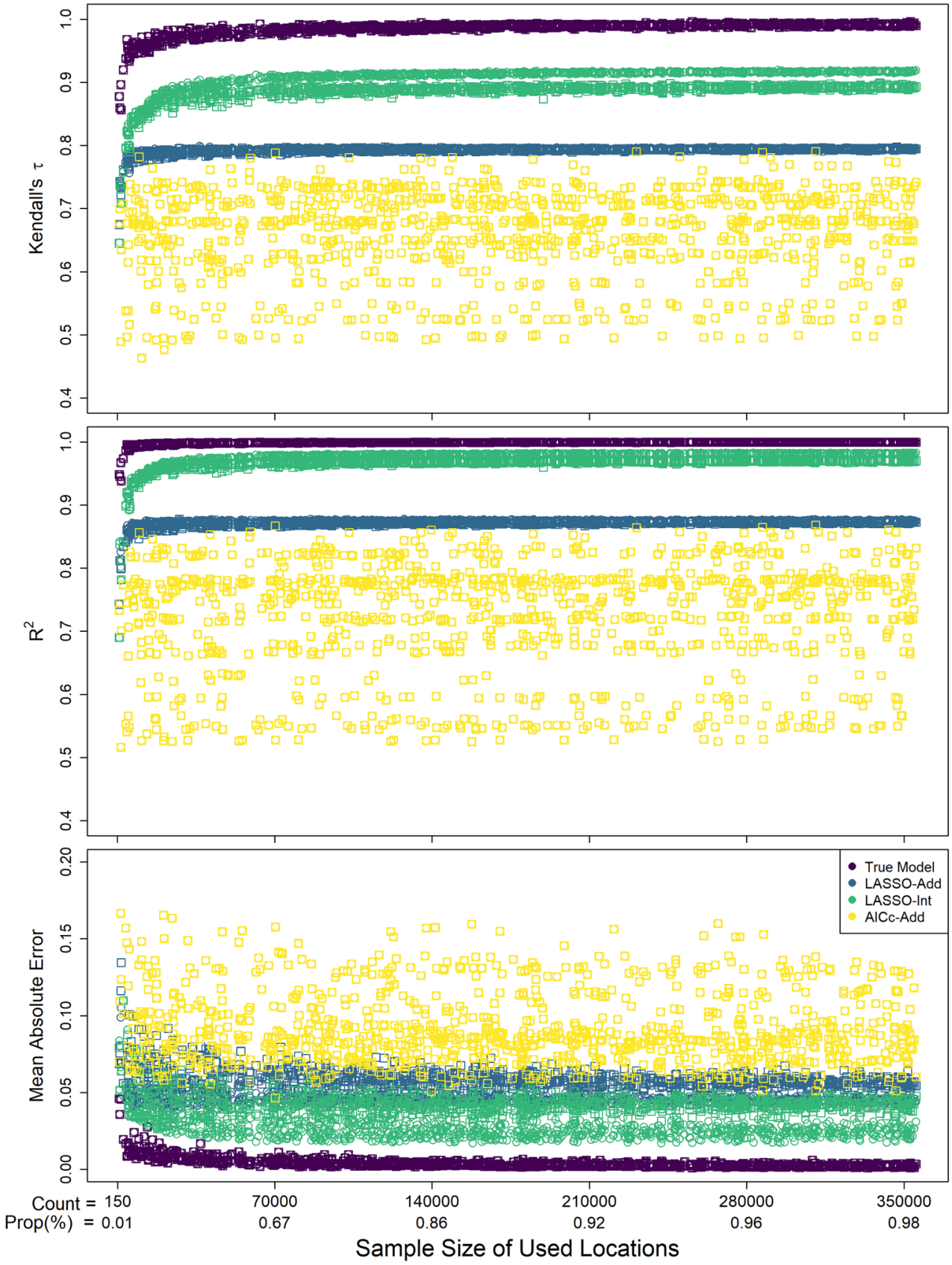
Kendall's τ
1.0
0.9
0.8
0.7
0.6
0.5
0.4
R²
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Mean Absolute Error
0.20
0.15
0.10
0.05
0.00
True Model
LASSO-Add
LASSO-Int
AICc-Add
Count = 150 70000 140000 210000 280000 350000
Prop(%) = 0.01 0.67 0.86 0.92 0.96 0.98
Sample Size of Used Locations

Figure 2

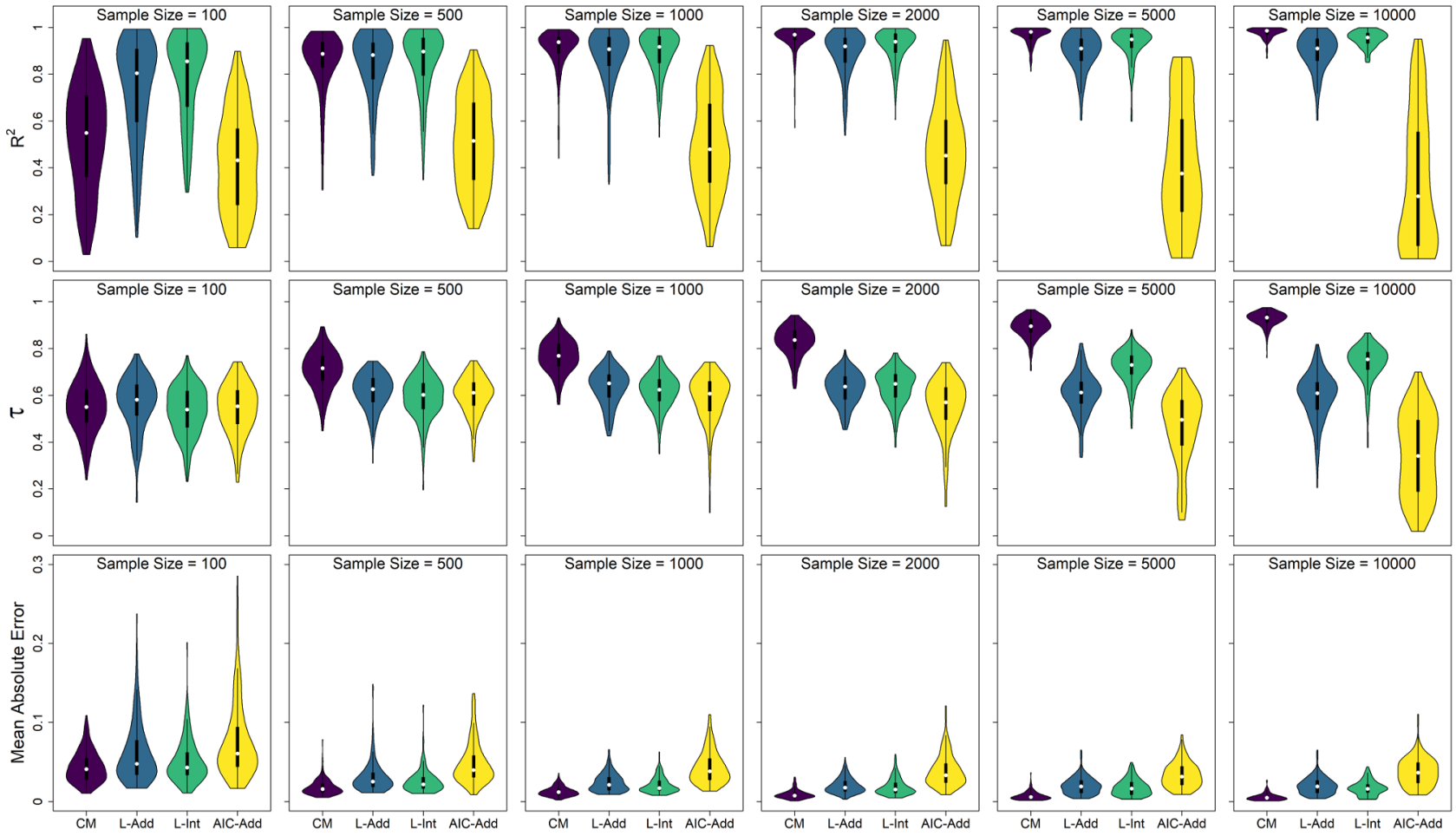
Sample Size = 100
Sample Size = 500
Sample Size = 1000
Sample Size = 2000
Sample Size = 5000
Sample Size = 10000
R2
τ
Mean Absolute Error
0
0.2
0.4
0.6
0.8
1
0.1
0.3
CM
L-Add
L-Int
AIC-Add

Figure 3