ELSEVIER

GfÖ
GfÖ Ecological Society of Germany,
Austria and Switzerland

Basic and Applied Ecology 43 (2020) 77–85

Basic and
Applied Ecology

www.elsevier.com/locate/baae

# Testing prediction accuracy in short-term ecological studies

Connor M. Wood[a,d,*], Zachary G. Loman[a], Shawn T. McKinney[b,c],
Cynthia S. Loftin[b]

[a] *Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, Maine, USA*
[b] *U.S. Geological Survey, Maine Cooperative Fish and Wildlife Research Unit, Orono, ME, USA*
[c] *Rocky Mountain Research Station, Fire Sciences Laboratory, Missoula, MT, USA*
[d] *Department of Forest and Wildlife Ecology, University of Wisconsin - Madison, Madison, Wisconsin, USA*

Received 31 January 2019; accepted 19 January 2020
Available online 28 January 2020

---

## Abstract

Applied ecology is based on an assumption that a management action will result in a predicted outcome. Testing the prediction accuracy of ecological models is the most powerful way of evaluating the knowledge implicit in this cause-effect relationship, however, the prevalence of predictive modeling and prediction testing are spreading slowly in ecology. The challenge of prediction testing is particularly acute for small-scale studies, because withholding data for prediction testing (e.g., via *k*-fold cross validation) can reduce model precision. However, by necessity small-scale studies are common. We use one such study that explored small mammal abundance along an elevational gradient to test prediction accuracy of models with varying degrees of information content. For each of three small mammal species, we conducted 5000 iterations of the following process: (1) randomly selected 75 % of the data to develop generalized linear models of species abundance that used detailed site measurements as covariates, (2) used an information theoretic approach to compare the top model with detailed covariates to habitat type-only and null models constructed with the same data, (3) tested those models' ability to predict the 25 % of the randomly withheld data, and (4) evaluated prediction accuracy with a quadratic loss function. Detailed models fit the model-evaluation data best but had greater expected prediction error when predicting out-of-sample data relative to the habitat type models. Relationships between species and detailed site variables may be evident only within the framework of explicitly hierarchical analyses. We show that even with a small but relatively typical dataset ($n = 28$ sampling locations across 125 km over two years), researchers can effectively compare models with different information content and measure models' predictive power, thus evaluating their own ecological understanding and defining the limits of their inferences. Identifying the appropriate scope of inference through prediction testing is ecologically valuable and is attainable even with small datasets.

***Keywords:*** Elevational gradient; Expected prediction error; Model validation; Scale dependency; Small mammals

---

## Introduction

Applied ecology is based on an assumption that a management action affecting one variable, such as a habitat manipulation or species translocation, will result in a pre-

---

[*] Corresponding author. Present address: 1630 Linden Drive, Russell Labs, Madison, WI, 53706, USA.
*E-mail address:* cwood9@wisc.edu (C.M. Wood).

https://doi.org/10.1016/j.baae.2020.01.003

dicted outcome in another variable. When the ecological understanding conceptualizing that cause-effect relationship is incomplete, the outcome may differ from the prediction, and the consequences can be costly and even irreversible (Hiers, Jackson, Hobbs, Bernhardt, & Valentine, 2016), for example, as is illustrated by the history of deliberate introductions for pest control (Pearson & Callaway, 2003). In the face of such consequences, quantifying the predictive capacity and transferability of ecological research is prudent (Evans, Norris, & Benton, 2012). Correct predictions based on theories demonstrate scientific understanding of a cause-effect relationship, however, systematically testing anticipatory predictions is not yet widespread in ecology (Houlahan, McKinney, Anderson, & McGill, 2017).

The rarity of evaluating model predictions in ecology reflects the prevalence of explanatory rather than predictive modeling (e.g., Brudvig, 2017). Under the typical quantitative framework, data are collected, models are fit to those data, and a top-ranked model is identified (Burnham & Anderson, 2010). The result is an explanatory model that provides the most accurate description of the data possible. Competing explanatory models represent ecological hypotheses, so the accuracy of the predictions they make simply reflects support for those hypotheses. In contrast, predictive models generate anticipatory predictions about future states on the assumption that underlying hypotheses are correct (Shmueli, 2010; Mouquet et al., 2015). Developing predictive models requires partitioning data into model training and testing sets (i.e., *k*-fold cross-validation), bootstrapping, or other methods of resampling to evaluate models' anticipatory predictive capacity (Shmueli, 2010). This paradigm is particularly valuable in the face of increasing ecological and climatic novelty, because future states may lack contemporary analogs and thus lie beyond the scope of explanatory modeling (Allen, Fontaine, Pope, & Garmestani, 2011; Evans et al., 2012; Bradford, Betancourt, Butterfield, Munson, & Wood, 2018).

Yet withholding data during model training increases the leverage of outliers and potentially reduces precision and statistical power, while the logistical and financial constraints on data acquisition further discourage the exclusion of hard-earned data. This problem is compounded by the fact that most ecological studies span short durations and limited spatial extents relative to salient ecological processes (Hastings, 2004), and withholding data from these "small" datasets for model validation is generally an anathema. However, evaluating model predictions with data withheld during the model training phase ultimately decreases uncertainty that creates risk and potentially leads to irreversible consequences (Shmueli, 2010; Evans et al., 2012). For this reason, evaluating models' predictive abilities is considered a critical aspect of structured decision making (Martin, Runge, Nichols, Lubow, & Kendall, 2009) and adaptive management (Allen et al., 2011; McFadden, Hiller, & Tyre, 2011).

We assessed an approach to prediction testing with a small-scale, short-term ecological field study that evaluated hypotheses about the variation in small mammal abundance along elevational gradients in the northeastern United States. Our objective was to conduct iterative *k*-fold validation on a small dataset ($n = 28$ points) to determine whether such an approach might be a viable method of rigorous prediction testing for studies with small datasets (Fig. 1). Importantly, we compared models with detailed ecological variables, a simple categorical variable representing whole plant communities, and null models to test the value of detailed site covariates in terms of increased prediction accuracy. We predicted that the detailed models would have the greatest prediction accuracy. This approach provided a framework with which to identify the limitations of our understanding within our system while we explored systematic prediction testing within the constraints of small-scale studies. This approach may help ecologists maximize the efficacy of their resources and better gauge the extent of their ecological understanding.

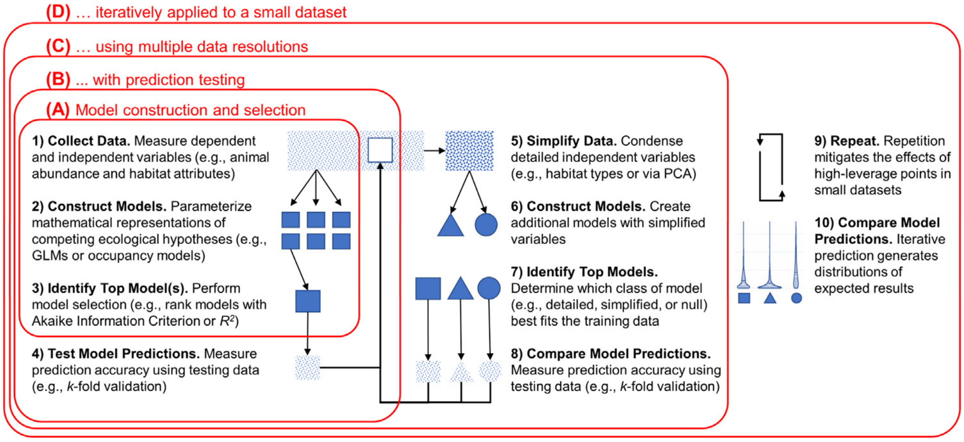
## Materials and methods

### Study site and sampling

We conducted small mammal and vegetation surveys in the Appalachian Mountains of the northeastern United States (approximately 44.65° N, 70.75° W) during June – August of 2014 and 2015 (Fig. 2). Dominant plant communities in the region were differentiated primarily by climatic factors associated with changes in elevation (Kupfer & Cairns, 1996). Deciduous forest, predominantly sugar maple (*Acer saccharum*) and beech (*Fagus grandifolia*) was found at low elevations (0−600 m ASL). Coniferous forest, dominated by red spruce (*Picea rubens*) and balsam fir (*Abies balsamea*), was found at intermediate elevations (200–1100 m ASL). Alpine tundra, characterized by grasses, black spruce (*P. mariana*), and exposed rock, was found at high elevations (1100–1800 m ASL).

We focused on three species, the woodland jumping mouse (*Napaeozapus insignis)*, the southern red-backed vole (*Myodes gapperi)*, and the deer mouse (*Peromyscus maniculatus*), because together they represented 89 % of the total individuals captured in the study (Wood, McKinney, & Loftin, 2017).

We established ten transects that spanned the three dominant plant communities and overall elevational range of 450−1460 m ASL (Fig. 2), and we randomly located a trapping grid at 10 sites within both deciduous and coniferous forest and at eight sites in alpine tundra ($n = 28$). Trapping grids were 90 × 90 m with 10 rows of 10 Sherman Live Traps (Sherman Trap Co., Tallahassee, FL) spaced 10 m apart and baited with oats and peanuts and supplied with cotton balls for nesting material. We checked traps on four consecutive mornings (7:00–10:00) and three consecutive evenings (16:30–19:00). We identified all individuals to species and applied a uniquely numbered ear tag (Kentucky Band and Tag Co., Newport, KY).



**1) Collect Data.** Measure dependent and independent variables (e.g., animal abundance and habitat attributes)

**2) Construct Models.** Parameterize mathematical representations of competing ecological hypotheses (e.g., GLMs or occupancy models)

**3) Identify Top Model(s).** Perform model selection (e.g., rank models with Akaike Information Criterion or $R^2$)

**4) Test Model Predictions.** Measure prediction accuracy using testing data (e.g., *k*-fold validation)

**5) Simplify Data.** Condense detailed independent variables (e.g., habitat types or via PCA)

**6) Construct Models.** Create additional models with simplified variables

**7) Identify Top Models.** Determine which class of model (e.g., detailed, simplified, or null) best fits the training data

**8) Compare Model Predictions.** Measure prediction accuracy using testing data (e.g., *k*-fold validation)

**9) Repeat.** Repetition mitigates the effects of high-leverage points in small datasets

**10) Compare Model Predictions.** Iterative prediction generates distributions of expected results

**Fig. 1.** The primacy of explanatory modeling in ecology (A) has recently waned as prediction testing has recently experienced a Renaissance in ecology (B, C), but predictive modeling has largely been confined to projects with large datasets. We propose that when time or resources constrain data collection, iteratively testing the predictive power of models with different information contents still may allow researchers to robustly evaluate the predictive power of their models (D).

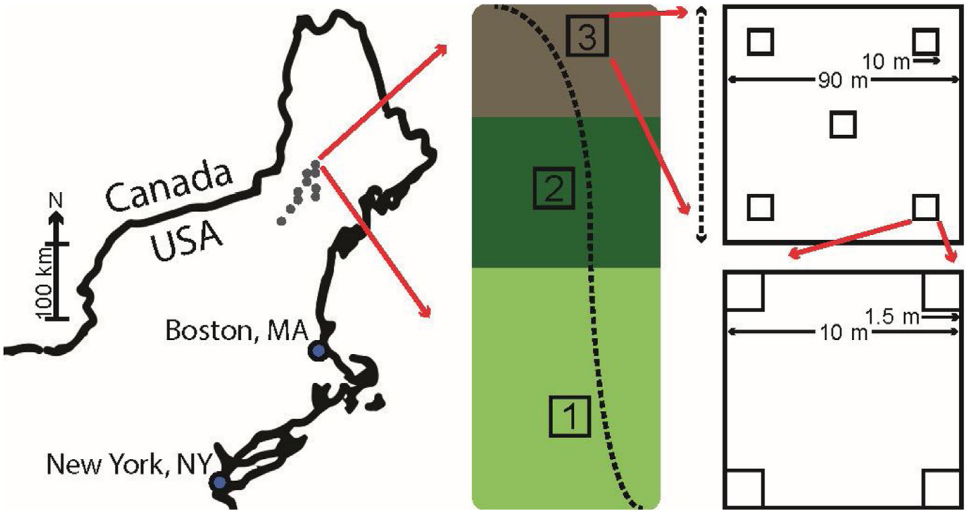


**Fig. 2.** Transect locations in the northeastern United States used to sample small mammals and habitat characteristics in 2014 and 2015. Insets show the overall study design with one transect spanning three habitat sites in ascending order of elevation: deciduous forest (1), coniferous forest (2), and alpine tundra (3). Each site comprised a 90 × 90 m trap grid, five 10 × 10 m habitat survey plots, and twenty 1.5 × 1.5 m subplots.

We also established five 10 × 10 m habitat survey plots systematically located within trap grids, with four subplots at the corners of each plot (Fig. 2). Subplot values were averaged to create a plot-level score for each variable; these scores were averaged to a site-level variable score; and, the site scores were independent variables in our models. The mean values of each variable in each plant community type are listed in Table 1.

## Data analysis

We conducted 5000 iterations of the following *k*-fold validation process of developing and testing competing generalized linear models of focal species' abundance. The full dataset ($n = 28$ sites) was randomly split into four equal groups ($k = 4$) consisting of three model training groups that were used to build the models ($n_{total} = 21$) and a fourth model testing group used to evaluate the models' predictions ($n_k = 7$). A preliminary sensitivity analysis indicated that doubling the number of iterations incurred computational delays without affecting results. In general, the number of iterations that is "sufficient" will depend on the size of the dataset and the amount of environmental variation it encompasses.

We first developed and compared detailed models of focal species abundance (Minimum Number Alive (MNA); *sensu* (Krebs, 1966)) with the model training data. For each species we identified four habitat variables that were highly correlated with its abundance, had a plausible biological explanation, and minimized multicollinearity among the other variables (see Table 1 for descriptions). Variables in the

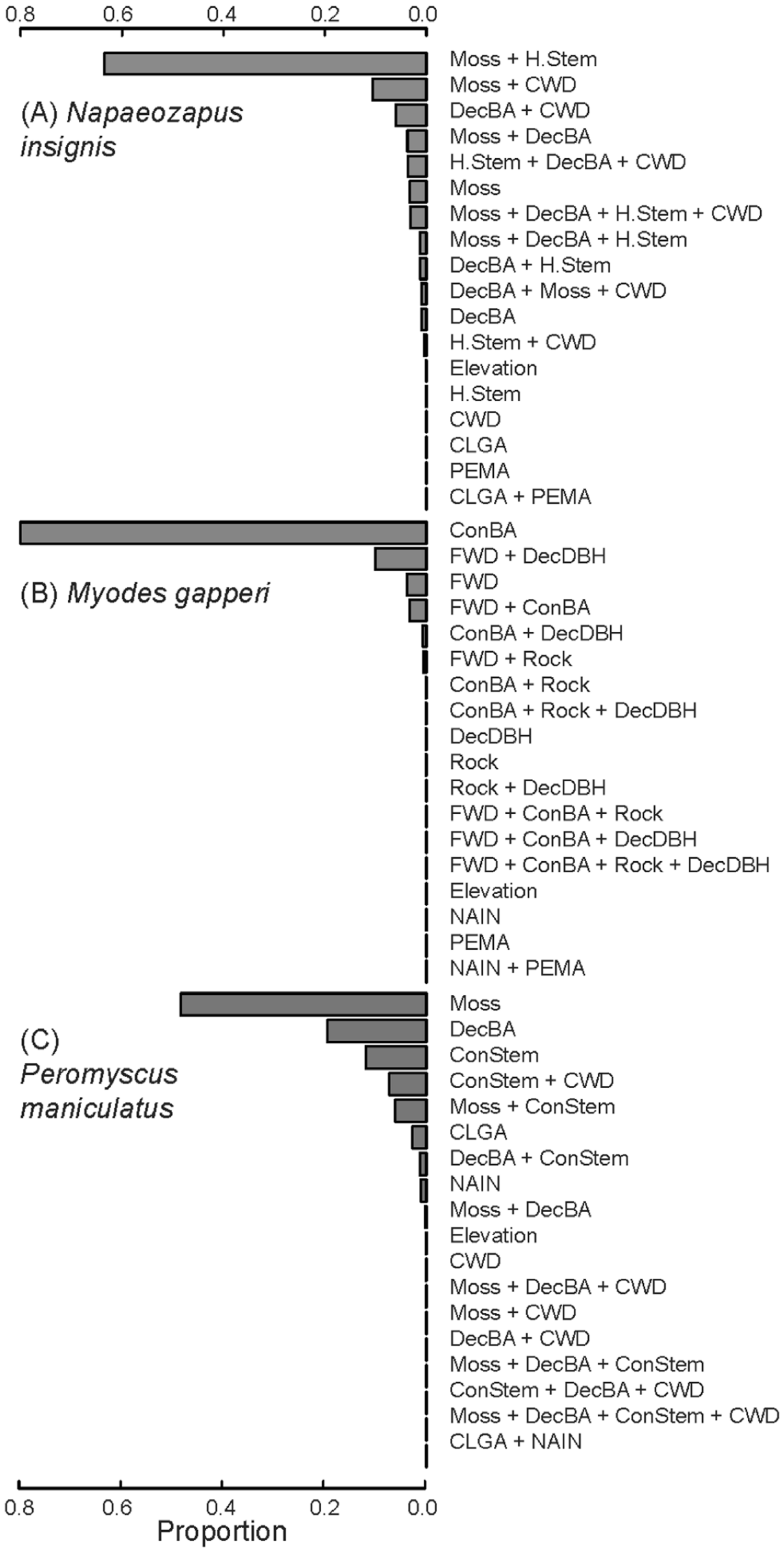


**Fig. 3.** Structure of candidate detailed models of (A) *Napaeozapus insignis*, (B) *Myodes gapperi*, and (C) *Peromyscus maniculatus* abundance sorted by the proportion of iterations of the random subsampling process ($n = 5000$) in which they were the top model ($\Delta AICc = 0$). Variables are described in Table 1. All *N. insignis* and *P. maniculatus* models included a year term.

**Table 1.** Habitat variables measured in each of five plots (10 m × 10 m) and 20 subplots (1.5 m × 1.5 m) at each site, variable abbreviations, measurement methods, and precision. Diameter at breast height (DBH, 1.37 m height; minimum DBH =2.5 cm) of deciduous (Dec) and coniferous (Con) trees was used to calculate basal area ($\text{m}^2\,\text{ha}^{-1}$); Dec and Con were limited to trees with a DBH ≥ 6 cm to represent potentially mature (seed-producing) individuals during the modeling process. Subplot ground cover focused strictly on the ground; only the estimated area of the stem at ground level of a widely branching plant was recorded. Plot-level means and standard errors are provided for all applicable variables in each plant community types (deciduous forest, coniferous forest, and alpine tundra).

| Plot metrics | Abbreviation | Description | $\bar{x}$ (SE) (Decid.) | $\bar{x}$ (SE) (Conif.) | $\bar{x}$ (SE) (Alpine) |
|---|---|---|---|---|---|
| Tree species | | Diameter ≥ 0.025 m | NA | NA | NA |
| Tree diameter at breast height | Dec/Con/Tot | ± 0.005 m | Dec: 1.141 (0.13) | Con: 1.52 (1.36) | Tot: 0.17 (0.07) |
| Canopy cover | CanCov | Densiometer (±3 %) | 93.02 (2.18) | 87.40 (1.36) | 10.07 (5.64) |
| Canopy height | CanHt | Hypsometer (±1.5 m) | 19.14 (0.66) | 11.55 (0.79) | 0.82 (0.50) |
| Slope | Slope | Hypsometer (±2°) | 11.27 (1.80) | 19.08 (1.86) | 13.75 (2.54) |
| Aspect | Asp | GPS (±1°) | NA | NA | NA |
| Elevation | El | GPS (±3 m) | 544.73 (21.92) | 927.95 (31.39) | 1220.64 (39.84) |
| Date | Yr/Jul | Year/Julian date | NA | NA | NA |
| Plot ground structure | | | | | |
| Large woody debris | LWD | Count; ø > 0.25 m | 1.40 (0.34) | 3.00 (0.60) | 0.00 (0.00) |
| Coarse woody debris | CWD | Count; ø 0.10–0.25 m | 7.24 (1.70) | 11.50(1.59) | 1.25 (0.53) |
| Fine woody debris | FWD | Count; ø 0.025–0.10 m | 14.46 (1.48) | 23.60 (2.47) | 9.25 (2.67) |
| Subplot metrics | | | | | |
| Woody stems at 1 m | W.Stem1 | Count | 2.23 (0.20) | 2.77 (0.54) | 1.55 (0.56) |
| Woody stems at 0.25 m | W.Stem25 | Count | 8.05 (0.72) | 7.22 (1.42) | 8.00 (1.63) |
| Herbaceous stems at 0.25 m | H.Stem | Count | 3.80 (0.83) | 1.63 (0.47) | 0.03 (0.03) |
| Subplot ground cover | | | | | |
| Leaf litter and FWD | LL | Ocular estimate ±3 % | 85.29 (2.17) | 65.93 (4.07) | 15.19 (4.11) |
| Herbaceous plants | Grass | Ocular estimate ±3 % | 2.00 (0.35) | 1.16 (0.15) | 1.35 (0.46) |
| Woody stems | W.Stem0 | Ocular estimate ±3 % | 3.98 (0.72) | 4.31 (0.75) | 14.17 (4.95) |
| Moss and lichen | Moss | Ocular estimate ±3 % | 6.20 (1.16) | 25.01 (4.21) | 41.69 (5.41) |
| Logs > 0.25 m | Log | Ocular estimate ±3 % | 1.10 (0.31) | 1.69 (0.19) | 0.03 (0.03) |
| Exposed rock | Rock | Ocular estimate ±3 % | 1.63 (0.76) | 0.59 (0.20) | 27.02 (8.53) |
| Water | Water | Ocular estimate ±3 % | 0.37 (0.20) | 0.12 (0.06) | 0.02 (0.02) |

detailed models for *N. insignis* were the proportion of ground covered by moss, the basal area ($\text{m}^2\,\text{ha}^{-1}$) of deciduous trees with a diameter at breast height (DBH) > 6 cm, the number of coarse woody debris stems, and the number of herbaceous stems ≥ 25 cm tall. The 6-cm threshold for DBH was used as the approximate minimum diameter for a mature tree in the study area. For *M. gapperi* we chose the variables basal area of coniferous trees with a DBH > 6 cm, the mean DBH of deciduous trees, the number of fine woody debris stems, and the proportion of ground covered by rock. For *P. maniculatus* we chose the proportion of ground covered by moss, the basal area of deciduous trees with a DBH > 6 cm, the number of coarse woody debris stems, and the number of coniferous trees with a DBH ≥ 2.5 cm. We constructed generalized linear models of each species' abundance (minimum number alive; MNA) as a function of: (1) all first-order combinations of the variables we selected (Fig. 3); (2) site elevation, a common proxy for climate; and, (3) the abundance of the other two focal species. All models, including plant community type (PCT) and null (described below), included an offset term to account for variation in trap effort among sites, and the *N. insignis* and *P. maniculatus* models included a year term to account for substantial interannual population fluctuation (Wood et al., 2017). We compared the detailed models ($n = 18$ per species) with the Akaike Information Criterion with a correction for small sample size ($\text{AIC}_c$) and ranked them by subtracting the smallest score from all others ($\Delta\text{AIC}_c$), allowing us to identify the top-ranked detailed model for each species (i.e., the model with the most support from the data used to construct it) (Burnham & Anderson, 2010).

Next, we compared the top detailed model to two simplified models of each species' abundance representing successive reductions in data complexity and effort required to gather model covariates relative to the detailed models. Plant community type models treated plant community type as a single categorical variable with three classes: deciduous forest, coniferous forest, and alpine tundra. Null models contained only an intercept term equal to the average abundance of that species. We ranked each species' top detailed model, PCT model, and null model with $\Delta\text{AIC}_c$.

We evaluated model performance by first recording which of the 16 detailed model structures was most frequently best fit. Next, we calculated the proportion of the 5000 iterations in which each class of model best fit the model-evaluation data $\Delta\text{AIC}_c = 0$). Finally, we quantified prediction accuracy with a quadratic loss function (i.e., expected prediction error) by comparing predicted and observed abundance for all seven withheld model testing data points. This approach incorporated a penalty for increased variance in predictions, rather than being based solely on absolute accuracy (Shmueli, 2010).

We used R version 3.0.3 (R Core Development Team, 2014) and the packages MASS (Ripley et al., 2013), MuMIn (Bartoń, 2015), and ggplot (Wickham & Chang, 2016) for these analyses. See the Supplementary material for R code.

## Results

### Trapping

*Napaeozapus insignis* abundance was greatest in deciduous forest in both years, with a substantial peak in 2014 (Table 2). *Myodes gapperi* abundance was greatest in coniferous forest in both years (Table 2). In 2014, *P. maniculatus* abundance was substantially greater in deciduous forest than in coniferous forest and alpine tundra, however, in 2015, abundance was similar among the three habitats (Table 2). Overall abundance of each species was greater in 2014 than 2015. For all species, abundance patterns were driven more by habitat differences than by spatial autocorrelation. See Supplementary material for complete trapping records of all 10 captured species.

### Top detailed models

The top detailed model of *N. insignis* abundance was most frequently [moss + number of herbaceous stems] ($\Delta AICc = 0$ in 64 % of 5000 iterations); the next most frequent top model was [moss + coarse woody debris] (11 % of iterations; Fig. 3A). The top detailed model of *M. gapperi* abundance was consistently [basal area of coniferous trees] (82 % of iterations); the next closest model was [DBH of deciduous trees + fine woody debris] (8 % of iterations; Fig. 3B). The top detailed model of *P. maniculatus* abundance was most frequently [moss] (49 % of iterations), although [basal area of deciduous trees] and [number of conifer stems] also performed well (21 % and 12 % of iterations, respectively; Fig. 3C).

### Comparing model types

Several patterns were consistent across species. For example, the top detailed models almost always fit the training data better than the plant community type or null models ($\Delta AIC_c = 0$ in 79–98%, 5000 iterations for all species). However, the top detailed models consistently had the greatest expected prediction error i.e., quadratic loss function, meaning that they made the lowest prediction accuracy. The PCT models consistently had the lowest expected prediction error i.e., made the best anticipatory predictions and the null models preformed slightly worse but still made substantially better predictions than the top detailed models (Fig. 4).

## Discussion

Making accurate predictions is perhaps the best method of demonstrating scientific understanding (Houlahan et al., 2017), and predictive modeling is increasingly valuable for conservation in the face of ecological and climatic novelty (Allen et al., 2011; Evans et al., 2012; Bradford et al., 2018). However, prediction testing can be particularly challenging with small datasets. We have demonstrated an approach to testing models' anticipatory predictions and subsequently evaluating knowledge gained using data collected in a study whose small scale is well-represented in the contemporary ecological literature. We employed a novel approach to prediction testing with limited data in which we iteratively assessed the relative model fit (with $AIC_c$) and predictive power (with a quadratic loss function) of models based on detailed variables, categorical plant community type, and average abundance. Our results have implications for small mammal community change in the northern Appalachian Mountains and demonstrate the value of rigorous prediction testing in ecological studies conducted at similarly small scales.

We used a modest ecological study, which entailed a single two-person crew surveying for 2.5 months across 125 km during two successive years, to explicitly measure the level of understanding possible by testing our top models' ability to predict randomly selected out-of-sample data. We found that detailed models of small mammal abundance constructed with site-specific environmental variables requiring intensive survey effort (i.e., one quarter to one third of the working hours) and that included bottom-up biotic interactions, interspecific competition, and various aspects of habitat structure, consistently fit the observed data better. Counter to our prediction, however, the plant community models made the best anticipatory predictions (i.e., had the lowest expected prediction error; Fig. 4). Reducing the information content of our models did not significantly reduce their predictive power, suggesting that the resources expended to acquire the more detailed information about habitat structure and biotic interactions did not necessarily improve our understanding of this system.

In the context of structured decision making, this would suggest that decision thresholds (Martin et al., 2009) in management actions related to small mammals would not require detailed site covariates. Instead, the projected climate change-driven 71–100 % decrease in coniferous forest and corresponding increase in deciduous forest in our study area (Tang & Beckage, 2010) would be a more salient ecological resolution. Our models indicate that *M. gapperi* is likely to become less abundant while *N. insignis* and *P. maniculatus* become more abundant, as the extents of their respective core habitats change. This could lead to increased instability

**Table 2.** Small mammal abundance (minimum number alive) in each plant community type across years.

| | 2014 | | | 2015 | | |
|---|---|---|---|---|---|---|
| | Deciduous forest | Coniferous forest | Alpine tundra | Deciduous forest | Coniferous forest | Alpine tundra |
| *Napaeozapus insignis* | 112 | 25 | 2 | 31 | 6 | 0 |
| *Myodes gapperi* | 24 | 87 | 24 | 27 | 99 | 3 |
| *Peromyscus maniculatus* | 97 | 20 | 26 | 22 | 17 | 12 |

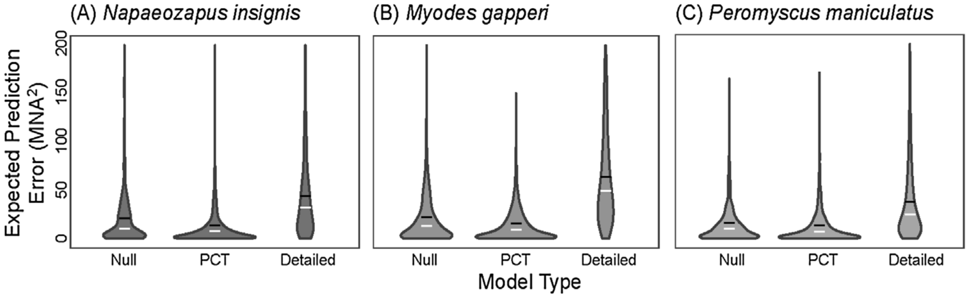


**Fig. 4.** Quadratic loss between predicted and observed abundance (Minimum Number Alive) of (A) *Napaeozapus insignis*, (B) *Myodes gapperi*, and (C) *Peromyscus maniculatus* based on 35,000 predictions of out-of-sample data; greater expected prediction error values indicate worse model performance. Bar width corresponds with the probability density; median and mean values are indicated with white and black dashes, respectively.

(i.e., population fluctuations) in the small mammal community (Wood et al., 2017). If small mammal abundance in the region were monitored through time, these predictions could be iteratively refined and, potentially, continually improved (Dietze et al., 2018), and management decision thresholds revised accordingly (Allen et al., 2011).

Model evaluation can reveal the transferability of predictions in space and time (Bahn & McGill, 2013; Sequeira et al., 2016; Yates et al., 2018). However, it is important to consider each step of the model validation process. First, define the purpose of the model, the performance criteria, and the context. Second, examine the mathematical operation of the models, their relationship to the relevant ecological theory, and the assumptions and characteristics of the input data (Rykiel, 1996). If understanding the relationships between focal species and fine-scale, site-specific attributes (e.g., interspecific competition or habitat structure) is an important objective, sampling within a single habitat type may minimize large-scale variation. Alternatively, if a study spans multiple habitat types, as ours did, the ecological importance of highly detailed covariates for each site may be difficult to discern without a carefully designed hierarchical approach to both sampling and analysis. Implementation of model testing largely has been confined to datasets with hundreds or thousands of points, yet, field studies that are limited in time or space, such as ours, can nonetheless result in datasets that are appropriate for rigorous prediction testing (Fig. 1D) to address the essential question: what level of understanding can we expect from such a study?

Using a quadratic loss function to thoroughly test expected prediction error of anticipatory ecological models can reveal spatial and temporal constraints to the inferences supported by a given dataset (Shmueli, 2010; Houlahan et al., 2017). This is a critical aspect of model validation, because applying models beyond the bounds of their reliable transferability may produce spurious results (Sequeira et al., 2016). Identifying the resolution of peak model performance may suggest the scale at which animals perceive and interact with their environment (Kie, Bowyer, Nicholson, Boroski, & Loft, 2002; Fuller & Harrison, 2010). For example, our finding that microhabitat variables such as plot-scale vegetation and substrate characteristics were no more effective than larger-scale variables such as plant community type in predicting small mammal abundance is consistent with existing literature and illustrates this effect of data resolution (Orrock, Pagels, McShea, & Harper, 2000; Coppeto, Kelt, Vuren, Wilson, & Bigelow, 2006; Fauteux, Imbeau, Drapeau, & Mazerolle, 2012). This phenomenon reinforces the importance of sampling at the appropriate spatiotemporal scales and the value of conducting both sampling and analyses in an explicitly hierarchical framework when studying small mammal habitat use and community assembly (Kelt, Meserve, Patterson, & Lang, 1999; Williams, Marsh, & Winter, 2002). The resampling process may also be sensitive to the amount of variation in the dataset, and datasets containing greater environmental variation are likely to require more extensive resampling (i.e., more iterations of random resampling) to ensure that the full range of environmental conditions has been tested.

We have shown that even with limited datasets it is possible to quantify model uncertainty and thus measure the gain in ecological understanding. Limitations of small datasets can be quantified by iteratively comparing relative model fit and testing prediction accuracy of random subsets of the data (Fig. 1). Applying that process to models with different information content can identify the relevant data resolution for the questions of interest, which in turn can help optimize resource allocation in subsequent studies. Small-scale studies will continue to be important contributions to ecology; systematically testing the predictions they generate will explicitly bound the scope and depth of understanding they may yield.

## Authors' contributions

CMW and STM conceived and designed the study; CMW collected the data; CMW, ZGL, STM, and CSL collaborated in the analysis and wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## Data accessibility

Small mammal capture records (10 species), detailed habitat measurements, and the R script used to iteratively identify the top detailed model, rank the null, PCT, and top detailed model, make predictions based on those models, and analyze those outputs are available on Zenodo.org (DOI: https://doi.org/10.5281/zenodo.887359).

## Acknowledgments

We thank Bill Halteman, Mac Hunter Jr., and Jacquelyn Gill for advice and guidance, Aaron Black and Kyle Shute for assistance in the field, the Maine Department of Agriculture, Conservation, and Forestry for logistical support, and D. Levesque and several anonymous reviewers whose input improved the manuscript. This research was funded by the Maine Outdoor Heritage Fund (project 141-01-04), funds provided by the University of Maine and Maine Department of Inland Fisheries and Wildlife though the Cooperative Agreement with the U.S. Geological Survey Maine Cooperative Fish and Wildlife Research Unit, and the University of Maine Graduate Student Government. This research was conducted with approval of the University of Maine Institutional Animal Care and Use Committee for protocol number A2014-04-03. Any use of trade, firm, or product names is for descriptive purposes only and does not constitute endorsement by the U.S. Government.

## References

Allen, C. R., Fontaine, J. J., Pope, K. L., & Garmestani, A. S. (2011). Adaptive management for a turbulent future. *Journal of Environmental Management*, *92*(5), 1339–1345. http://dx.doi.org/10.1016/j.jenvman.2010.11.019

Bahn, V., & McGill, B. J. (2013). Testing the predictive performance of distribution models. *Oikos*, *122*(3), 321–331.

Bartoń, K. (2015). *Multi-model inference (version 1.15.6)*. CRAN.

Bradford, J. B., Betancourt, J. L., Butterfield, B. J., Munson, S. M., & Wood, T. E. (2018). Anticipatory natural resource science and management for a changing future. *Frontiers in Ecology and the Environment*, *16*(5), 295–303. http://dx.doi.org/10.1002/fee.1806

Brudvig, L. A. (2017). Toward prediction in the restoration of biodiversity. *Journal of Applied Ecology*, *54*(4), 1013–1017. http://dx.doi.org/10.1111/1365-2664.12940

Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer.

Coppeto, S. A., Kelt, D. A., Vuren, D. H. V., Wilson, J. A., & Bigelow, S. (2006). Habitat associations of small mammals at two spatial scales in the northern Sierra Nevada. *Journal of Mammalogy*, *87*(2), 402–413. http://dx.doi.org/10.1644/05-MAMM-A-086R1.1

Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., . . . & White, E. P. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*, *115*(7), 1424–1432. http://dx.doi.org/10.1073/pnas.1710231115

Evans, M. R., Norris, K. J., & Benton, T. G. (2012). Predictive ecology: Systems approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1586), 163–169. http://dx.doi.org/10.1098/rstb.2011.0191

Fauteux, D., Imbeau, L., Drapeau, P., & Mazerolle, M. J. (2012). Small mammal responses to coarse woody debris distribution at different spatial scales in managed and unmanaged boreal forests. *Forest Ecology and Management*, *266*(Supplement C), 194–205. http://dx.doi.org/10.1016/j.foreco.2011.11.020

Fuller, A. K., & Harrison, D. J. (2010). Movement paths reveal scale-dependent habitat decisions by Canada lynx. *Journal of Mammalogy*, *91*(5), 1269–1279. http://dx.doi.org/10.1644/10-MAMM-A-005.1

Hastings, A. (2004). Transients: the key to long-term ecological understanding? *Trends in Ecology & Evolution*, *19*(1), 39–45. http://dx.doi.org/10.1016/j.tree.2003.09.007

Hiers, J. K., Jackson, S. T., Hobbs, R. J., Bernhardt, E. S., & Valentine, L. E. (2016). The precision problem in conservation and restoration. *Trends in Ecology & Evolution*, *31*(11), 820–830. http://dx.doi.org/10.1016/j.tree.2016.08.001

Houlahan, J. E., McKinney, S. T., Anderson, T. M., & McGill, B. J. (2017). The priority of prediction in ecological understanding. *Oikos*, *126*(1), 1–7. http://dx.doi.org/10.1111/oik.03726

Kelt, D. A., Meserve, P. L., Patterson, B. D., & Lang, B. K. (1999). Scale dependence and scale independence in habitat associations of small mammals in southern temperate rainforest. *Oikos*, *85*(2), 320–334. http://dx.doi.org/10.2307/3546498

Kie, J. G., Bowyer, R. T., Nicholson, M. C., Boroski, B. B., & Loft, E. R. (2002). Landscape heterogeneity at differing scales: Effects on spatial distribution of mule deer. *Ecology*, *83*(2), 530–544. http://dx.doi.org/10.1890/0012-9658(2002)083[0530:LHADSE2.0.CO;2]

Krebs, C. J. (1966). Demographic changes in fluctuating populations of *Microtus californicus*. *Ecological Monographs*, *36*(3), 239–273. http://dx.doi.org/10.2307/1942418

Kupfer, J. A., & Cairns, D. M. (1996). The suitability of montane ecotones as indicators of global climatic change. *Progress in Physical Geography*, *20*(3), 253–272. http://dx.doi.org/10.1177/030913339602000301

Martin, J., Runge, M. C., Nichols, J. D., Lubow, B. C., & Kendall, W. L. (2009). Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications*, *19*(5), 1079–1090.

McFadden, J. E., Hiller, T. L., & Tyre, A. J. (2011). Evaluating the efficacy of adaptive management approaches: Is there a formula for success? *Journal of Environmental Management*, *92*(5), 1354–1359. http://dx.doi.org/10.1016/j.jenvman.2010.10.038

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., . . . & Loreau, M. (2015). Predictive ecology in a changing world. *Journal of Applied Ecology*, *52*(5), 1293–1310. http://dx.doi.org/10.1111/1365-2664.12482

Orrock, J. L., Pagels, J. F., McShea, W. J., & Harper, E. K. (2000). Predicting presence and abundance of a small mammal species: The effect of scale and resolution. *Ecological Applications*, *10*(5), 1356–1366.

Pearson, D. E., & Callaway, R. M. (2003). Indirect effects of host-specific biological control agents. *Trends in Ecology & Evolution*, *18*(9), 456–461. http://dx.doi.org/10.1016/S0169-5347(03)00188-5

R Core Development Team. (2014). *R: A language and environment for statistical computing (Version 3.0.3)* Retrieved from. http://www.R-project.org/

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2013). *Support functions and datasets for Venables and Ripley's MASS*. CRAN.

Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling*, *90*(3), 229–244.

Sequeira, A. M. M., Mellin, C., Lozano-Montes, H. M., Vanderklift, M. A., Babcock, R. C., Haywood, M. D. E., . . . & Caley, M. J. (2016). Transferability of predictive models of coral reef fish species richness. *Journal of Applied Ecology*, *53*(1), 64–72. http://dx.doi.org/10.1111/1365-2664.12578

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. http://dx.doi.org/10.1214/10-STS330

Tang, G., & Beckage, B. (2010). Projecting the distribution of forests in New England in response to climate change. *Diversity & Distributions*, *16*(1), 144–158. http://dx.doi.org/10.1111/j.1472-4642.2009.00628.x

Wickham, H., & Chang, W. (2016). *ggplot (Version 2.2.1)*. CRAN.

Williams, S. E., Marsh, H., & Winter, J. (2002). Spatial scale, species diversity, and habitat structure: Small mammals in Australian tropical rain forest. *Ecology*, *83*(5), 1317–1329. http://dx.doi.org/10.1890/0012-9658(2002)083[1317:SSSDAH]2.0.CO;2

Wood, C. M., McKinney, S. T., & Loftin, C. S. (2017). Intraspecific functional diversity of common species enhances community stability. *Ecology and Evolution*, *7*(5), 1553–1560. http://dx.doi.org/10.1002/ece3.2721

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., . . . & Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, *33*(10), 790–802. http://dx.doi.org/10.1016/j.tree.2018.08.001