

# Monitoring and Modeling Terrestrial Arthropod Diversity on the Kenai National Wildlife Refuge

Matthew L. Bowser<sup>1,2,3</sup> and John M. Morton<sup>1,2,4</sup>

**ABSTRACT:** *The primary purpose of the Kenai National Wildlife Refuge (KENWR) is to “conserve fish and wildlife populations in their natural diversity,” where “fish and wildlife” explicitly includes arthropods. To this end, we developed a Long Term Ecological Monitoring Program (LTEMP), a collaborative effort with the USDA Forest Inventory and Analysis (FIA) Program. In 2004 and 2006, we sweep-netted terrestrial arthropods on 255 100m<sup>2</sup> circular plots systematically distributed at 5-km intervals over the 805,000 ha KENWR. These samples yielded 15,136 specimens, which were sorted to families and to species when possible. The comprehensive spatial coverage of the LTEMP sampling design provided spatial data suitable for species distribution modeling, but we sought to improve upon this design by explicitly accounting for imperfect detection. We proposed a rotating panel design where each site would be visited once every ten years. Imperfect detection would be accounted for by spatial sub-sampling within plots. Using Monte-Carlo simulation, we assessed the proposed design of LTEMP for accurately monitoring changes in arthropod species distributions over time. Our simulations demonstrated that, for species that are likely to be collected in a single 50m<sup>2</sup> sweep net sample where they are present, the proposed LTEMP sampling design should provide accurate estimates of species distributions and local rates of colonization and extinction over the long-term. In order to document the landscape-scale patterns of arthropod diversity over the KENWR, we modeled arthropod family richness using random forest regression. Arthropod family richness data were obtained from LTEMP and topographic, temporal, and productivity variables were obtained from GIS datasets. The resulting map explained 22% of variation of diversity of sweep net samples. Highest diversity was predicted on the margins of coastal wetlands and in productive hardwood and mixed forests; lowest diversity was predicted at barren alpine sites.*

**KEYWORDS:** occupancy modeling, species distribution modeling, random forest regression, interpenetrating panel design, Monte-Carlo simulation.

---

<sup>1</sup>United States Fish & Wildlife Service; Kenai National Wildlife Refuge; PO Box 2139; Soldotna, Alaska 99669.

<sup>2</sup>University of Alaska Fairbanks; Biology & Wildlife Department; Fairbanks, Alaska, 99775.

<sup>3</sup>Matt\_Bowser@fws.gov

<sup>4</sup>John\_M\_Morton@fws.gov

## Introduction

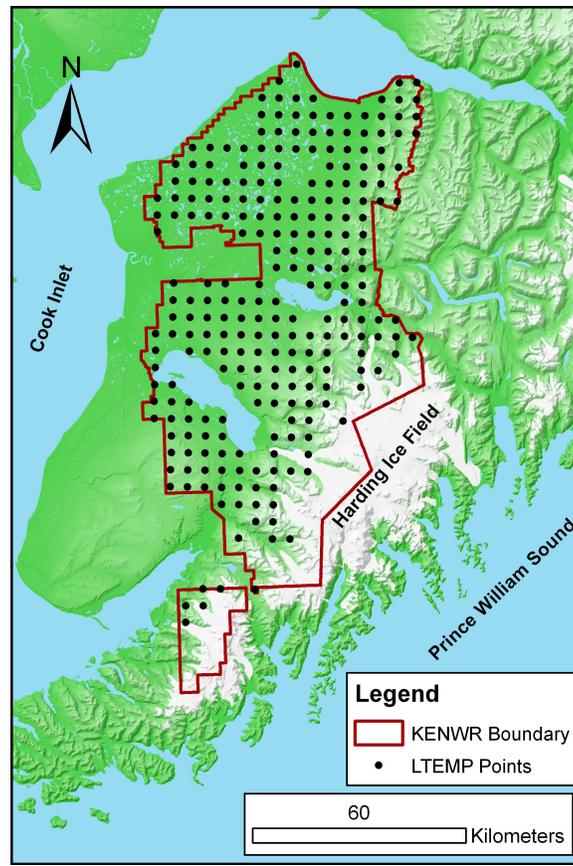
### Motivation: A Broad Conservation Mandate

The primary purpose of the Kenai National Wildlife Refuge (KENWR) is “to conserve fish and wildlife populations and habitats in their natural diversity,” where the term “fish and wildlife” was defined as “any member of the animal kingdom, including without limitation any mammal, fish, bird, amphibian, reptile, mollusk, crustacean, arthropod or other invertebrate,” (Alaska National Interest Lands Conservation Act (ANILCA) of 1980).

Most of the “fish and wildlife” that the KENWR is charged to conserve are arthropods. They represent 80% of described diversity worldwide (Kremen et al. 1993). In Alaska, an incomplete, unpublished checklist of terrestrial arthropods included 6,500 species (unpublished checklist by Derek Sikes, University of Alaska Museum, Fairbanks, Alaska), roughly an order of magnitude greater than the number of vertebrate species of the state. Consistent with the world and the region, arthropods represent most of metazoan diversity on the KENWR.

Arthropods are generally more important than vertebrates in terms of ecosystem functioning in most terrestrial systems (Wilson 1987). In many ecosystems, herbivory by arthropods has a greater influence on primary production than vertebrates (Schowalter 2006). Most plants require pollination by insects (Tepedino 1979). Arthropods also serve as important predators (Snyder and Evans 2006), parasites, fungivores (Hopkin 1997), and decomposers (Wallwork 1983). Even a single species of arthropod can alter the composition and functioning of a landscape. For example, an outbreak of the spruce bark beetle, *Dendroctonus rufipennis* Kirby, has recently removed the overstory of vast forests of mature white spruce (*Picea glauca* (Moench) Voss) on the KENWR, resulting in proliferation of understory grasses (Boucher and Mead 2006). Many arthropod species, including the spruce bark beetle, are essential to the natural functioning of the KENWR.

Despite the mandate of ANILCA to conserve the natural diversity of arthropods, the fact that they represent most of biodiversity, and their immense ecological importance, little progress had been made toward finding out what arthropod species occur on KENWR. Insects, spiders, and other arthropods have been overlooked because of somewhat circular reasons: (1) none of the species on KENWR were known to be of conservation concern, (2) arthropod species are assumed to be free from risk of extinction unless contrary evidence exists and (3) they are thought to be prohibitively difficult to learn about or conserve.



**Figure 1:** Map of the Kenai Peninsula showing the Kenai National Wildlife Refuge and LTEMP sampling locations.

The obvious first step is an inventory, a survey of the arthropod species that are present on the KENWR. However, a static inventory is inadequate because the composition of the fauna of the KENWR is changing and is expected to continue to change over the coming decades due to the responses of arthropods to a warming climate and due to additional introductions of exotic species. Mobile species are already moving up-slope and pole-ward as the climate warms (Gottfried et al., 1999; Parmesan 1996, 2006; Parmesan et al., 1999; Parmesan and Yohe 2003; Walther et al. 2002; Wilson et al. 2005), tracking their climatic preferences. As this process of re-distribution continues, some areas will become refugia (Gottfried et al. 1999) while others will serve as corridors (Hannah et al. 2002). While the most vagile species may respond rapidly, less motile species may not re-distribute themselves quickly enough to track climate. Habitat loss and fragmentation of suitable habitat will further exacerbate this problem, sometimes preventing species from moving to more suitable areas. Many species are expected to be lost as this process continues unless actions are taken to facilitate the re-distribution of species (Williams et al. 2005). Positioned as we are at the beginning of the re-distribution of species due to accelerating climate change, the

need for accurate documentation of current distributions of species and subsequent monitoring of species distributions is increasingly being recognized (Guisan and Thuiller 2005; Magness et al. 2008).

Long Term Ecological Monitoring Program (LTEMP) of the KENWR, a grid-based, multi-species, collaborative, inventory and monitoring framework undertaken through a memorandum of understanding between the USDA Forest Service Forest Inventory and Analysis (FIA) program (Figure 1), provided ample spatial data for documenting distributions of species (e.g. Magness et al. 2008). However, due to the sampling methods used, imperfect detection could not be accounted for.

When a species is not detected perfectly by field methods there is a possibility of reporting false absences (i.e., recording a species as absent where it was in fact present, but not observed). This leads to bias in estimates of the probability of occurrence of a species (MacKenzie et al. 2003, 2006). This issue of imperfect detection is especially relevant in a monitoring context. Unless it is accounted for, changes in the observed distribution of a species may be attributable to either (1) changes in the area occupied by a species or (2) changes in the likelihood that a species is detected due to changes in abundance, changes in seasonal phenology, etc.

### **Accounting for Imperfect Detection Using Occupancy Models**

Substantial literature exists on the subject of accounting for imperfect detection (see MacKenzie et al. 2006 for a review). Occupancy models explicitly account for imperfect detection using repeated survey data to obtain unbiased estimates of presence/absence metrics (MacKenzie et al. 2003, 2006). Multi-season occupancy models can also be used to estimate local rates of colonization and extinction, potentially some of the most relevant monitoring metrics given expected distribution shifts and potential expansions of exotic species.

In order to accurately monitor species distributions of arthropods over time, we propose to modify the field methods of LTEMP so that imperfect detection can be explicitly accounted for.

### **Diversity as a Measure of Ecosystem Health**

Species richness can be defined as the number of species distributions that overlap with the area of interest. Since ecosystem stability (McCann 2000), ecosystem function (Hooper et al. 2005) and resilience (Chapin et al. 2000) are generally correlated with diversity, species richness is itself considered to be a good indicator of the health of a system (Magurran 1988). Under this premise, taxon richness is often used as an indicator in environmental assessment studies

(e.g. Bechtel and Copeland 1970; Egloff and Brakel 1973; Wu 1982; Roth et al. 1994; Karr and Kimberling 2003).

## Objectives

Our first objective was an initial inventory of the arthropod fauna of the KENWR. Our second objective was to develop methods for monitoring arthropod distributions over the long-term. Our third objective was to document landscape-scale patterns of arthropod biodiversity on the KENWR.

## Methods

### Sampling Methods

**Study Area and Sampling Frame:** Located in south-central Alaska, the 805,000 ha KENWR covers much of the western Kenai Peninsula (Figure 1) and consists mainly of boreal forest, lowland wetlands, and alpine habitats. The FIA program imposed a rectangular grid of sampling sites across the Kenai Peninsula with 4.8 km spacing between the sites. Of the sites within KENWR, 255 sites that did not fall on water or ice were included in the sampling frame of LTEMP. Sampling design and methods are described in more detail in Morton et al. (2009).

**Plot Design and Field Methods:** FIA crews established four circular, 5.64 m radius (100 m<sup>2</sup>) sub-plots at each sampling site, with one central sub-plot centered on the site coordinates and three additional circular plots arranged in a triangle around the central sub-plot (Burkman 2005). LTEMP methods were focused on the central plots and excluded the three auxiliary plots.

Over the field seasons of 1999–2002, FIA field crews surveyed all of the 176 sites that had been determined to be forested. In 2004 and 2006, KENWR field crews sampled vegetation on the remaining 80 sites. Common to both FIA and KENWR field methods was collection of presence/absence data for all vascular plant species on the central 5.64m radius, circular plots.

Sampling of terrestrial arthropods was conducted by KENWR field crews over the field seasons of 2004 (152 sites) and 2006 (103 sites). Concurrent sampling of birds determined the seasonal (June 7–30) and daily (04:40 to 10:54 hours) sampling windows. A single sweep net sample was taken at each plot. For each sample, the collector swept a 30 cm diameter aerial insect net quickly back and forth over all vegetation and other substrates within reach over the entire circular plot. The contents of the net were then emptied into a vial of 80–90% ethanol.

**Sorting and Identification:** Arthropod specimens were processed and archived using appropriate curation methods. All arthropods present in the samples were sorted into orders and most were sorted to families. Species identifications were made whenever possible and many specimens were shipped to over 30 specialists for expert determinations. With the exception of small amounts of material lent to various systematists, all specimens remain in the arthropod collection of the KENWR (international collection coden: KNWR).

## **Occupancy Modeling of Species Distributions**

**Proposed Sampling Design:** The proposed long-term sampling design for LTEMP is a rotating panel (interpenetrating panel) design where 51 sites (20%) would be sampled every other year so that each site would be visited once every ten years. Each panel would be representative of the KENWR as a whole.

Detection histories suitable for correcting for imperfect detection are most often obtained by visiting each site multiple times within a season. Because surveying all sites multiple times in one season would be cost-prohibitive, increase damage to the sites through trampling, and increase the likelihood of inadvertently introducing exotic species to the sites, we propose to modify arthropod collecting methods so that detection probabilities can be estimated from spatial subsamples taken on a single visit to a site. In the place of a single sweep net sample over the 100 m<sup>2</sup> circular plot, the plot would be split into two semicircles along a north-south axis and a separate sweep net sample would be taken from each 50m<sup>2</sup> semicircle. These spatial subsamples would provide a detection history suitable for occupancy modeling.

**Monte-Carlo Simulations:** Our general approach for evaluating LTEMP sampling regimes was to assess the performance of the proposed designs through Monte-Carlo simulation. These simulations were designed to answer the question of whether or not occupancy metrics could be estimated well; they were not designed to actually model species distributions. In the same way that a mean can be considered a special case of linear regression where there is only an intercept and an error term, the simple occupancy models we used estimated only occupancy metrics (occupancy, detection probability, and, in multi-season cases, rates of colonization and extinction) without considering additional variables (e.g. possible covariates) that would normally be included to produce species distribution models. These simulations were designed to answer the questions of interest without unnecessary complexity.

For each scenario considered, we (1) generated large numbers of simulated datasets that conformed to the proposed LTEMP monitoring design and had known parameter values, (2) fitted occupancy models using program PRESENCE (Hines 2007), and (3) compared the estimates obtained to the known parameter

values. We wrote scripts in the R programming language (R core development team 2008) scripts to generate datasets, write input files, fit occupancy models using PRESENCE, and extract results.

We generated datasets suitable for occupancy modeling in the following way. For a given set of scalar values of occupancy ( $\Psi$ , the proportion of the area occupied by a species) and detection probability ( $p$ ), we first generated a list of occupancy states at  $n$  sites by specifying that the occupancy states at all sites were independently and identically distributed realizations of a Bernoulli process with a rate of  $\Psi$  using the `rbern` function in the Rlab add-on library (Boos et al. 2006). Similarly, we specified that detection events in each of the two sweep net samples from each site were independently and identically distributed realizations of a Bernoulli process with a rate of  $p$ . Multiplying the detection states by the occupancy states yielded detection histories with the specified parameter values of  $\Psi$  and  $p$ .

For single-season scenarios, we considered nine values of occupancy (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9) and the same nine values of detection probability. For each combination of parameters, we generated 100 datasets (a total of  $9 \times 9 \times 100 = 8,100$  datasets). Using the same simulated datasets, we obtained estimates from program PRESENCE and naïve estimates of occupancy, where naïve estimates are simply the proportion of sites where a species is observed in any survey.

In order to assess how well rates of extinction and colonization could be monitored over the long-term by the proposed sampling design, we intentionally selected parameter values that were realistic. We set initial occupancy to 0.7, detection probability to 0.8, the local rate of extinction ( $\epsilon$ ) to 0.1, and the local rate of colonization ( $\gamma$ ) to 0.05. Multi-season data were generated by simulating the Markovian processes of colonization and extinction over a specified number of seasons. For each site that was occupied at season  $t$ , there was a probability  $\epsilon$  (the local rate of extinction) that it would become unoccupied by season  $t + 1$ ; sites unoccupied in season  $t$  had a probability  $\gamma$  (the local rate of colonization) of becoming colonized by season  $t + 1$ . One hundred datasets were generated conforming to a population with the specified parameter values over 20 years. For each simulated population, estimates of the local rates of colonization were obtained using program PRESENCE every time the population would be sampled by the proposed LTEMP rotating panel design. Bowser (2009) provided a more detailed explanation of the methods used for generating datasets and fitting occupancy models using program PRESENCE, including the R scripts used.

**Table 1:** Families included in analyses.

Acanthosomatidae	Culicidae	Nabidae
Achilidae	Curculionidae	Otitidae
Acrididae	Delphacidae	Phlaeothripidae
Agromyzidae	Diapriidae	Phoridae
Anisopodidae	Dolichopodidae	Pipunculidae
Anobiidae	Drosophilidae	Platygasteridae
Anthocoridae	Dryinidae	Psilidae
Anthomyiidae	Dryomyzidae	Psyllidae
Anthomyzidae	Elateridae	Pteromalidae
Aphelinidae	Empididae	Pythidae
Aphididae	Encyrtidae	Rhagionidae
Apidae	Entomobryidae	Scarabaeidae
Argidae	Ephydriidae	Scathophagidae
Asteiidae	Eulophidae	Scelionidae
Bethylidae	Eurytomidae	Sciomyzidae
Bibionidae	Formicidae	Scirtidae
Braconidae	Heleomyzidae	Sclerosomatidae
Cantharidae	Hemerobiidae	Sepsidae
Carabidae	Hypogastruridae	Simuliidae
Ceraphronidae	Ichneumonidae	Sminthuridae
Chamaemyiidae	Isotomidae	Sphaeroceridae
Chloroperlidae	Lathridiidae	Staphylinidae
Chloropidae	Lauxaniidae	Stratiomyidae
Chrysomelidae	Leiodidae	Syrphidae
Chrysopidae	Lithobiidae	Tabanidae
Cicadellidae	Lycidae	Tachinidae
Clusiidae	Lygaeidae	Tenthredinidae
Coccinellidae	Micropezidae	Tephritidae
Coenagrionidae	Miridae	Thripidae
Coniopterygidae	Muscidae	Torymidae

Throughout the simulation analyses, program PRESENCE often yielded fitted values of zero or one for at least one of the parameters estimated. When this occurred, it appeared as if one parameter was fixed at either zero or one while fitting was performed on the remaining variable(s), leading to erroneous results. In these situations, program PRESENCE issuing the warning, “numerical convergence was not reached,” in its output. The failure rate of a set of simulations was calculated by dividing the number of failed simulations by the total number of simulations and was expressed as a percentage. All other

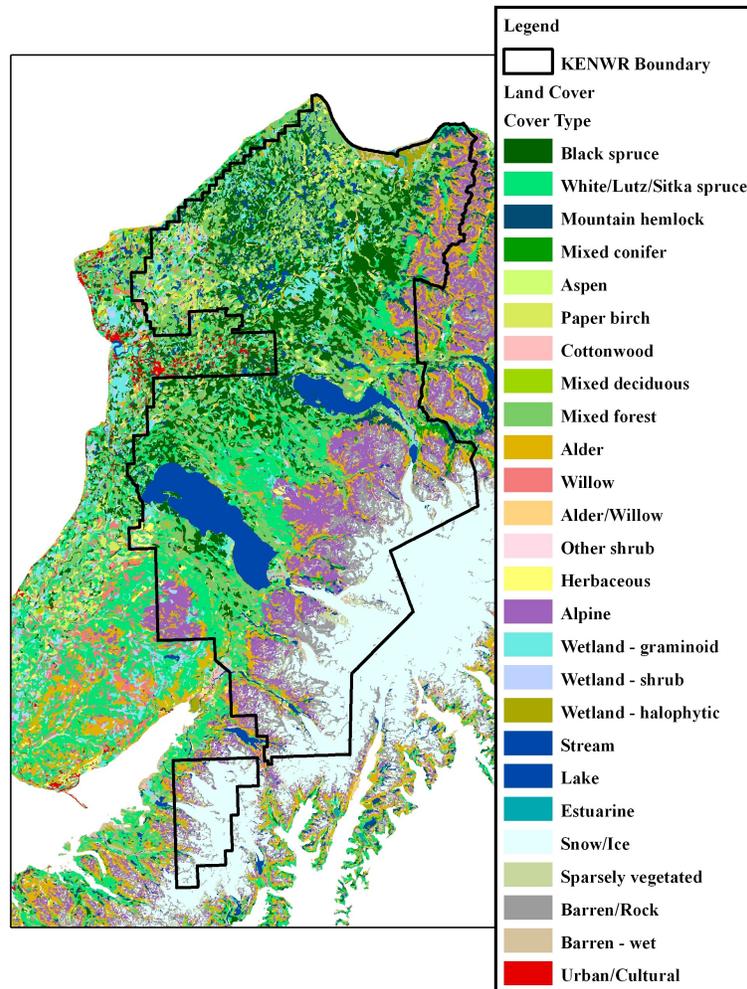
summaries presented below exclude results from simulations where PRESENCE failed to converge, summarizing only simulations where PRESENCE converged successfully.

### Random Forest Regression of Arthropod Family Richness

In order to create a continuous raster of arthropod family richness over the KENWR, predictions were made by random forest regression, a machine learning algorithm with high predictive accuracy (Breiman, 2001). For this exercise, taxonomic breadth was kept as broad as possible, although some groups were excluded due to practical constraints. Ninety families (Table 1) were included in analyses. Of the 255 sweep net samples available, two were rejected for the purposes of this regression analysis. In one case, a logistical error had caused spatial misalignment of the arthropod sample with other data and in another case, an avalanche dramatically altered a plot immediately before arthropods were sampled.

**Table 2:** Variables included in Random Forest regression.

Variable Name	Number of Variables	Description
<b><u>Spatial</u></b>		
latitude	1	
longitude	1	
<b><u>Topographic</u></b>		
elevation	1	Elevation from the Digital Elevation Model (DEM).
slope	1	Derived from DEM.
aspect	1	Derived from DEM.
curvature	1	Derived from DEM.
distance_ocean	1	Distance to the ocean.
<b><u>Climate</u></b>		
precipitation	13	Annual (precipitation_annual) and monthly (precipitation01-12) precipitation from the PRISM model.
temperature	13	Annual (temperature_annual) and monthly (temperature01-12) temperature from the PRISM model.
accumulation	1	Accumulated surface runoff based on PRISM annual precipitation data and DEM.
<b><u>Vegetation</u></b>		
NDVI	1	Normalized difference Vegetation Index calculated from 2002 LandSat 7 imagery.
land cover	1	Vegetation cover classes (Figure 2).
<b><u>Temporal</u></b>		
day	1	Julian day.
hour	1	Hours since midnight.
<b><u>Historic</u></b>		
years_post_fire	1	Years since last fire.
<b>Total</b>	<b>39</b>	



**Figure 2:** Kenai Peninsula land cover classification.

**Extraction of Covariates:** A 100 m × 100 m resolution prediction grid was imposed over the KENWR. Values of covariates that were available as raster datasets were extracted by resampling (bilinear interpolation) using ArcMap (Table 2). Covariates available as vector datasets were converted to raster datasets conforming to the prediction grid. For all covariates, values from pixels in which LTEMP sampling locations fell were extracted for use in fitting of random forest regression models.

Elevation data were extracted from a USGS digital elevation model (DEM) resampled from 30 m × 60 m to 30 m × 30 m resolution. Topographic variables were calculated from the DEM. The normalized difference vegetation index (NDVI) was calculated from a 30m × 30m resolution mosaic of LandSat 7 imagery taken in 2002 and made available through the Multi-Resolution Land Characteristics Consortium. Climate parameters were extracted from the 2 km resolution PRISM (Parameter-elevation Regressions on Independent Slopes

Model) raster datasets produced by the Spatial Climate Analysis Service at Oregon State University (SCAS/OSU). This dataset is currently the best climate coverage available for Alaska (Simpson et al. 2005). Vegetation types were extracted from a recent vegetation classification of the Kenai Peninsula (KENWR, unpublished data).

Random forest regressions were fitted using the randomForest package (Liaw and Wiener 2002) for R. For all random forest regressions, 5,000 trees were built. To select an optimal value of the ‘mtry’ parameter (the number of predictors randomly selected for consideration at each node), we ran random forest regressions for all values of mtry between 1 and 50 and selected the value of mtry that yielded the highest value of pseudo-R-squared and the lowest mean squared error. A random forest regression model was then fitted using this optimal value of the mtry parameter. We used this model to make predictions at all 100 m × 100 m pixels over the KENWR. For prediction, the time was set at median observed values from observed LTEMP data (June 18 at 8:00 am).

## Results

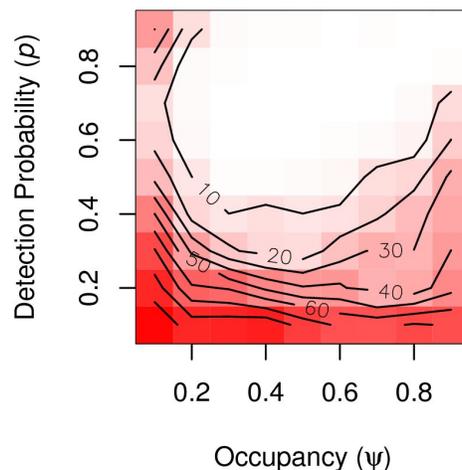
### Material Collected

The 255 sweep net samples yielded a total of 15,136 specimens, of which 9,961 were of the 90 families included in analyses. Over half of the arthropods collected (56%) were Diptera. Hemiptera (26%), Collembola (11%), and Hymenoptera (7%) also comprised substantial fractions of the specimens considered.

The abundance and frequency of specimens from each family varied greatly. Culicidae, with a total of 3,697 individuals collected and a frequency of 0.76 (i.e., collected at 76% of sites), was the most abundant and frequently collected family. Aphididae, Sminthuridae, Cicadellidae, Muscidae, Delphacidae, Empididae, Ichneumonidae, Simuliidae, Braconidae, Lauxaniidae, Phoridae, Biobionidae, Anthomyiidae, and Cantharidae were also relatively abundant and frequently collected. Ephydriidae were abundant locally, one site on the margin of Chickaloon Flats yielding 133 of the 135 Ephydrid specimens collected, but they were generally infrequent, collected at only three sites. In contrast, Diapriidae were relatively common but were usually represented by few individuals at each site. Many taxa were rarely encountered and were represented by few specimens. Seventeen families (Isotomidae, Chloroperlidae, Achilidae, Nabidae, Phlaeothripidae, Anobiidae, Carabidae, Lathridiidae, Lycidae, Pythidae, Scarabaeidae, Scirtidae, Clusiidae, Dryomyzidae, Sepsidae, Aphelinidae, and Bethyidae) were represented by singletons.

## Monte-Carlo Simulations

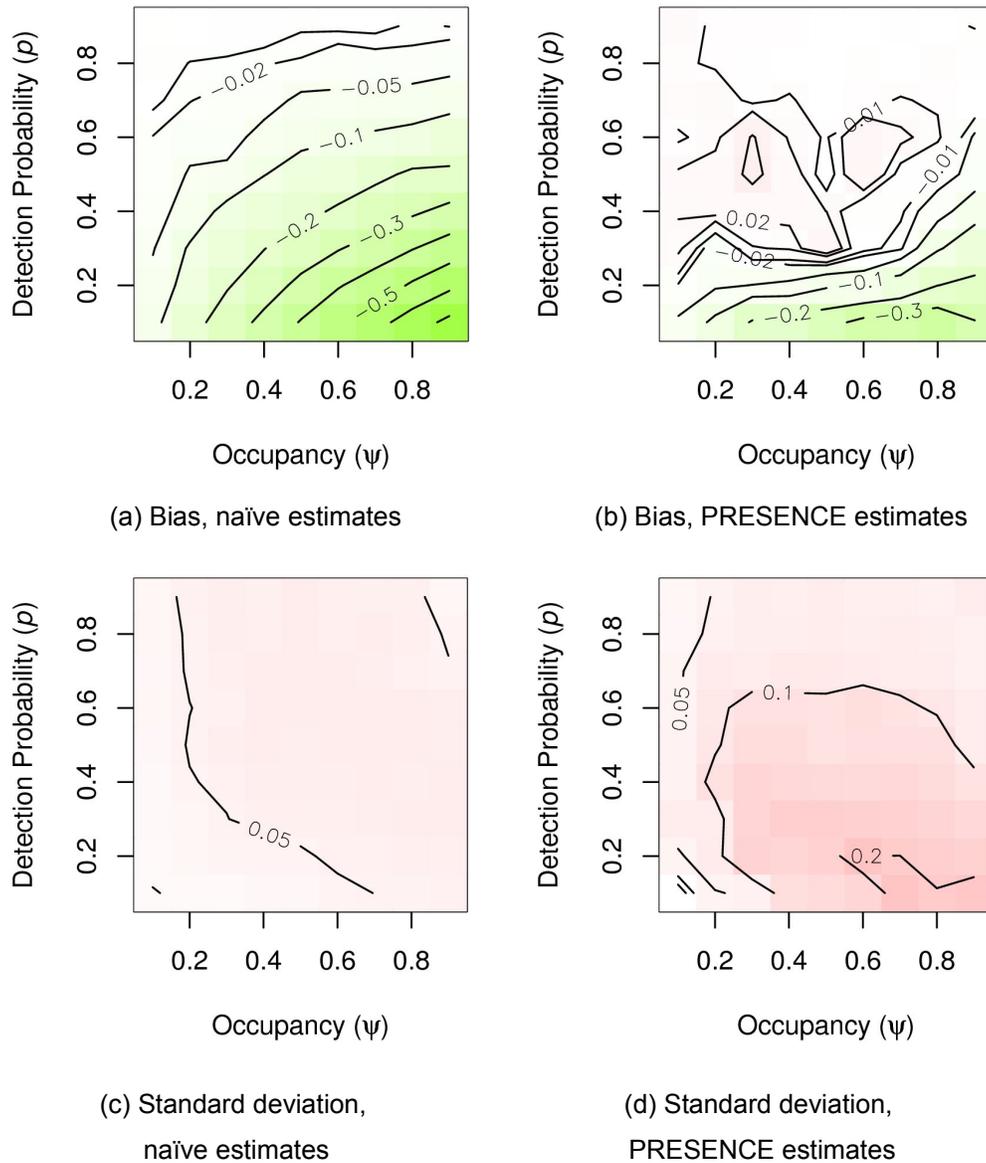
**Short-Term Results:** Using data from sampling of a single panel (a single sampling season), program PRESENCE had a high failure rate, especially when detection probability was low (Figure 3). The average failure rate was 24 failures out of every 100 simulations, with a high of 98 failures out of 100 simulations when both occupancy ( $\Psi$ ) and detection probability ( $p$ ) were 0.1. Program PRESENCE was most reliable when  $\Psi$  was between 0.3 and 0.7 and  $p$  was between 0.6 and 0.8.



**Figure 3:** Failure rate of Program PRESENCE over a range of values of  $\Psi$  and  $p$  for the rotating panel design. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of either  $\Psi$  or  $p$  with a value of 1 or 0. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 100%.

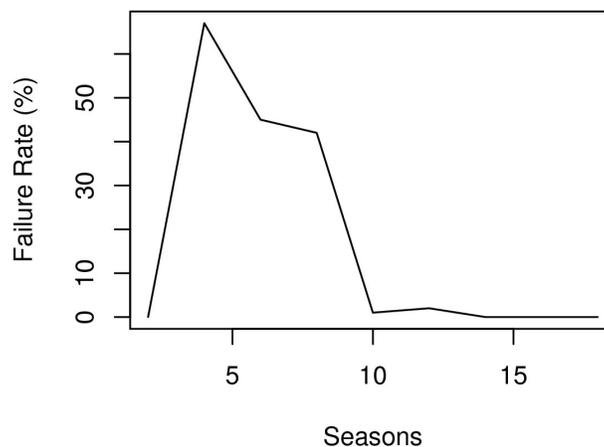
Naïve estimates of occupancy were nearly always biased low, with the magnitude of the bias increasing as  $p$  decreased and as  $\Psi$  increased (Figure 4(a)), with a maximum magnitude of average bias of -0.725. Only when detection probability was highest (0.9) were naïve estimates essentially unbiased. Program PRESENCE nearly always yielded less biased estimates of occupancy than the naïve estimates when it did not have convergence failures (Figure 4(b)). It produced generally unbiased estimates of occupancy as long as detection probability was greater than 0.3; when  $p$  was less than 0.3, estimates of  $\Psi$  tended to be biased low. The magnitude of bias was greatest when  $p$  was at its minimum value (0.1). Near this extreme of the parameter space, estimates of  $\Psi$  were highly variable.

Naïve estimates of occupancy were more precise, even if less accurate, than estimates obtained by program PRESENCE (Figure 4(c) and (d)). Standard deviation of naïve estimates was generally less than 0.05; standard deviation of estimates obtained from program PRESENCE were generally greater than 0.1 when detection probability was less than about 0.6.

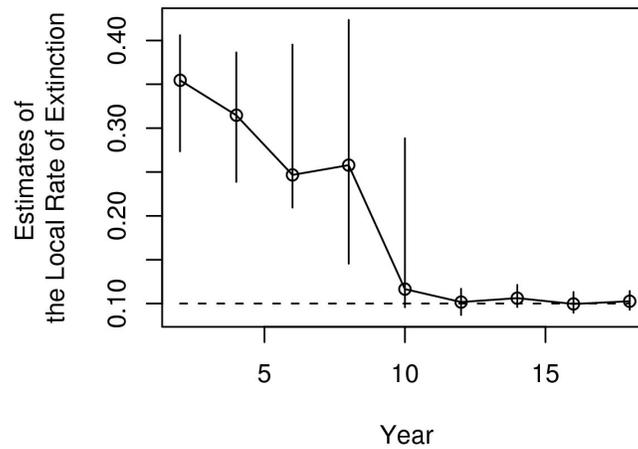


**Figure 4:** Bias and standard deviation of naïve estimates and PRESENCE estimates of occupancy over a range of values of occupancy ( $\Psi$ ) and detection probability ( $p$ ) for a single-season estimate (i.e., a survey of one panel). The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

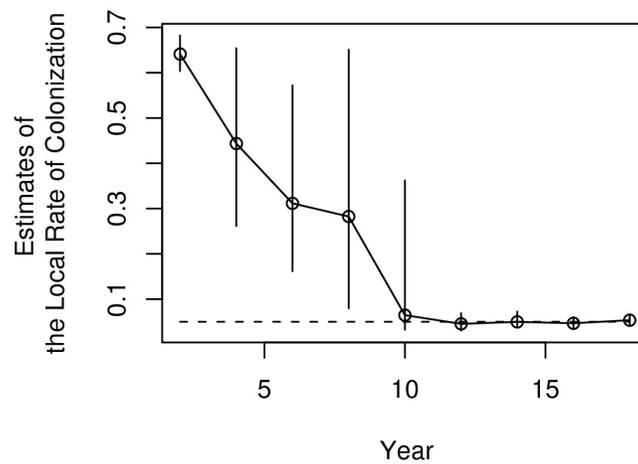
**Long-Term Results:** In these simulations, the rotating panel design did not yield reasonable estimates of local rates of colonization and extinction until some sites were visited twice, ten years into the sampling program. After this point, estimates of these change parameters were generally quite good. Estimates of occupancy were consistently unbiased (mean bias = 0.004), but not quite as precise as the desired maximum standard deviation of about 0.05 (SD = 0.062). Over time, the failure rate of program PRESENCE first rose dramatically from 0 at the second year to 67% at the fourth year, then remained high until the tenth year, when some sites were sampled a second time (Figure 5). After this point, the failure rate became negligible. Estimates of the change parameters  $\epsilon$  and  $\gamma$  were initially biased high (Figures 6 and 7). Mean bias at the second year was 0.25 for  $\epsilon$  and 0.61 for  $\gamma$ . This bias diminished slowly until the tenth year, when bias dropped quickly as some sites were sampled a second time. By the twelfth year and thereafter, estimates of these change parameters were essentially unbiased. After year twelve, precision was still not as good as is desirable, with standard deviations of about 0.1, but this large standard deviation was mostly due to a small number (2-9) of simulations each season where estimates of program PRESENCE were biased extremely high. When these outliers were removed, standard deviation of estimates of  $\epsilon$  and  $\gamma$  were less than 0.02.



**Figure 5:** Failure rate of program PRESENCE over 20 years of simulations. Values are the percentage of simulations in which PRESENCE failed.



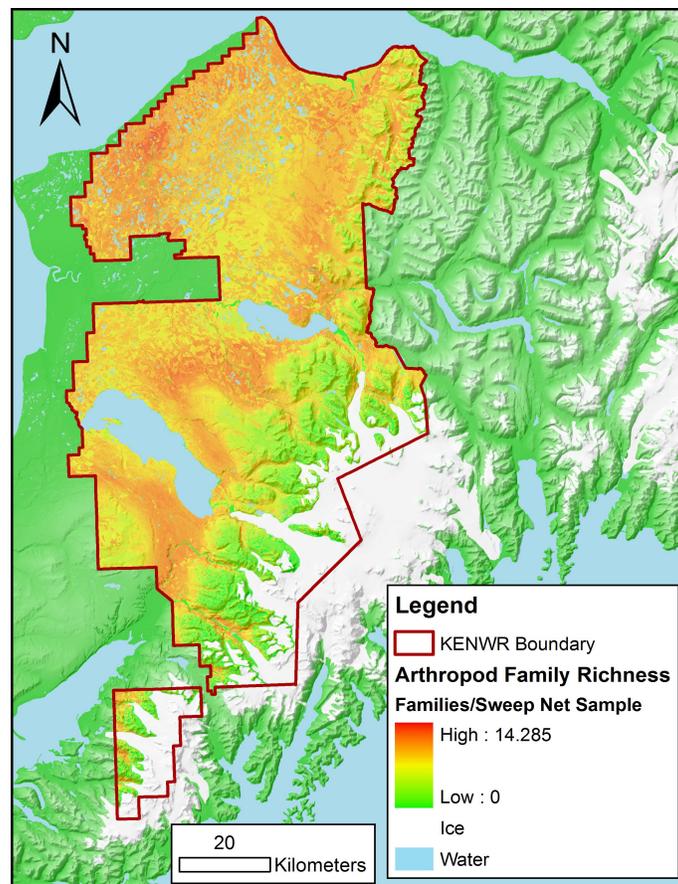
**Figure 6:** Estimates of the local rate of extinction over 20 years of simulations. Dashed line: true value (0.1). Circles and solid line: median of estimates with bars spanning from 25% to 75% quantiles.



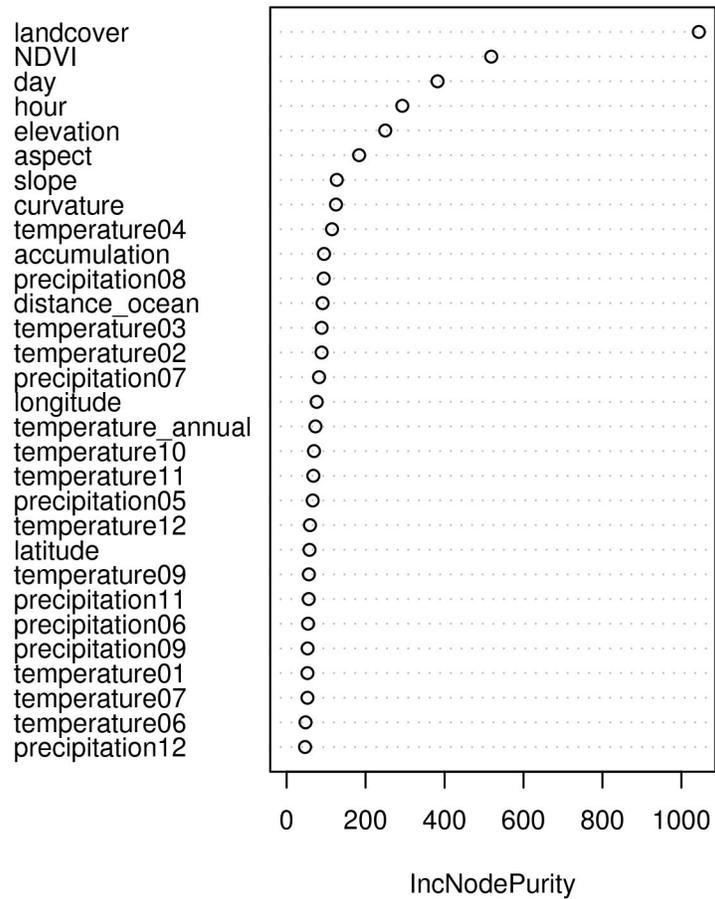
**Figure 7:** Estimates of the local rate of colonization over 20 years of simulations. Dashed line: true value (0.05). Circles and solid line: median of estimates with bars spanning from 25% to 75% quantiles.

## Regression Results

Random forest regressions incorporating only information that was available as continuous raster data explained only 22% of the variation of family richness, but produced a reasonable map of predicted family richness (Figure 8). The land cover classification was the variable most often used by the decision trees for predicting arthropod diversity, followed by NDVI. The regression predicted lowest arthropod family richness at barren alpine sites. Highest richness was predicted in coastal wetlands and productive hardwood and mixed forests the KENWR (Table 3).



**Figure 8:** Expected family richness of terrestrial arthropods predicted by random forest regression.



**Figure 9:** Importance of variables included in final random forest regression model. Variable names were defined in Table 2. IncNodePurity is the mean decrease in node purity, also known as the Gini splitting criterion (Breiman 2002).

## Discussion

### Utility of LTEMP for Monitoring Distributions

The proposed field methods of taking two spatial subsamples at each plot will allow for efficient monitoring of arthropod distributions on the KENWR while explicitly accounting for imperfect detection. Because precision was poor when detection probability was less than 0.6, the proposed methods would be most useful for species with detection probabilities of about 0.6 or greater. Although sampling of a single panel (51 sites) did not provide a large enough sample size

for modeling species distributions with good precision, the frequent sampling of this design would provide up-to-date information on a frequent basis while allowing a larger overall sampling frame than would be possible by sampling the same sites every season. Over the long term, the proposed methods would allow accurate estimation of species distributions and local rates of extinction and colonization.

**Table 3:** Predicted arthropod family richness by land cover type.

<b>Land Cover Type</b>	<b>Mean</b>	<b>S.D.</b>
Wetland - halophytic	9.03	0.90
Black cottonwood	8.88	1.17
Mixed forest	8.31	1.19
Paper birch	8.24	1.13
Wetland - shrub	8.15	1.44
Alder/Willow	8.05	1.54
Mixed deciduous	8.01	1.23
White/Lutz/Sitka spruce	7.99	1.33
Black spruce	7.79	1.09
Wetland - graminoid	7.76	1.24
Herbaceous	7.68	1.30
Urban/Cultural	7.54	1.34
Aspen	7.51	1.00
Willow	7.43	1.86
Mixed conifer	6.57	1.38
Alder	6.26	1.41
Other shrub	6.18	1.84
Mountain hemlock	6.06	1.16
Barren - wet	5.70	2.40
Alpine	4.76	1.26
Sparsely vegetated	3.69	1.48
Barren/Rock	3.33	0.85

In order to monitor distributions of arthropod less likely to be collected in individual sweep net samples (i.e., with detection probabilities  $< \sim 0.6$ ), more than two sweep net samples should be taken at each site. Increasing the number of surveys at each site is generally the most efficient way to improve precision of occupancy estimates, especially for species with low detection probabilities (MacKenzie and Royle 2005). In the case of the proposed LTEMP design, where sample size is already large ( $\sim 254$  sites) and the number of samples taken at each site is at its minimum value (2), increasing the number of sweep net samples is certainly the best way to improve precision and enable monitoring of a larger

number of arthropod species. Since the sweep net sampling method yielded information on relative abundance of arthropod species, estimates of occupancy metrics may also be improved by incorporating this information as was done by Royle and Nichols (2003).

The reason that rates of local colonization and extinction could not be estimated well over the first eight years of simulations was that the rotating panel design provided little information pertaining to these parameters until some sites were sampled a second time, ten years into the proposed sampling program. Precision and accuracy of estimates of these change parameters would best be improved by addition of a number of sites that would be sampled in consecutive seasons to furnish information about changes in occupancy states over the short-term.

### **Patterns of Diversity**

The good spatial coverage of the LTEMP sampling framework provided ample data for modeling arthropod family richness. This exercise yielded a map of coarse, landscape-scale patterns of arthropod diversity that will serve as a baseline for comparison with future arthropod diversity on the KENWR. It is also useful for putting small-area samples of arthropod diversity in the context of the KENWR as a whole.

Bowser (2009), using the same arthropod diversity data from LTEMP, found that temporal variable, productivity, and climate were the most important determinants of arthropod diversity on the KENWR. His analysis and the current analysis by random forest were consistent with McCoy (1990) in that, at about 60° latitude, maximum arthropod diversity occurred near sea level. The margins of coastal salt marshes and productive deciduous and mixed forest habitat types where random forest regression predicted highest arthropod diversity were relatively warm, productive sites; the alpine and barren habitat types where lowest arthropod diversity was predicted are relatively cold, unproductive sites. This apparent positive relationship between arthropod diversity and vegetative productivity is consistent with the species-energy theory of Wright (1983). He posited that, as the amount of energy moving through a system increases, the population sizes of living things supported by that system increase, reducing extinction rates.

### **Conclusions**

The excellent spatial representation of the LTEMP sampling framework provided ample spatial data for modeling over the landscape of the KENWR. By correcting for imperfect detection in the future, we will be poised to monitor for changes in the distributions of species over time. Although monitoring was mostly

discussed in the context of the responses of species to climate, these same monitoring methods should serve well for monitoring the spread of exotic species and the subsequent responses of native biota in terms of distribution shifts.

## Acknowledgments

This work would not have been possible without collaboration with the FIA program. Pat Doak and Diane Wagner (University of Alaska Fairbanks) greatly improved the quality of this work through their advice. Marathon Oil Corporation provided additional funding for this project. Charlotte Hockin (US Fish & Wildlife Service) performed a large part of the rough sorting of arthropod specimens. Todd Eskelin (US Fish & Wildlife Service) helped in various ways. Many experts helped with identifications.

## Literature Cited

- Bechtel, T. J.; Copeland, B. J. 1970. Fish species diversity indices as indicators of pollution in Galveston Bay, Texas. *Contributions in Marine Science*. 15:103-32.
- Boos, Dennis D.; Brooks, Atina D.; Nychka, Douglas. 2006. Rlab: Functions and Datasets Required for ST370 class. R package version 1.3. [Online]. Available: <http://www.r-project.org> [August 25, 2008].
- Boucher, Tina V.; Mead, Bert R. 2006. Vegetation change and forest regeneration on the Kenai Peninsula, Alaska following a spruce beetle outbreak, 1987-2000. *Forest Ecology and Management*. 227:233-246.
- Bowser, Matthew L. 2009. Terrestrial arthropod biodiversity on the Kenai National Wildlife Refuge, Alaska. M.Sc. thesis. Fairbanks, AK: University of Alaska Fairbanks. 141p.
- Breiman, Leo. 2001. Random Forests. *Machine Learning*. 45:5-32.
- Burkman, Bill. 2005. Forest Inventory and Analysis: Sampling and Plot Design. FIA Fact Sheet Series. USDA Forest Service. [Online]. Available: <http://fia.fs.fed.us/library/fact-sheets/data-collections/Sampling%20and%20Plot%20Design.pdf> [March 26, 2008].
- Chapin, III, F. Stuart; Zavaleta, Erika S.; Eviner, Valerie T.; Naylor, Rosamond L.; Vitousek, Peter M.; Reynolds, Heather L.; Hooper, David U.; Lavorel, Sandra; Sala, Osvaldo E.; Hobbie, Sarah E.; Mack, Michelle C.; Díaz, Sandra. 2000. Consequences of changing biodiversity. *Nature*. 405:234-242.
- Egloff, D. A.; Brakel, W. H. 1973. Stream pollution and a simplified diversity index. *Journal Water Pollution Control Federation*. 45:2269-2275.

- Gottfried, Michael; Pauli, Harald; Reiter, Karl; Grabherr, Georg. 1999. A fine-scaled predictive model for changes in species distribution patterns of high mountain plants induced by climate warming. *Diversity and Distributions*. 5:241-251.
- Guisan, Antoine; Thuiller, Wilfried. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*. 8:993-1009.
- Hannah, L.; Midgley, G.; Millar, D. 2002. Climate change-integrated conservation strategies. *Global Ecology & Biogeography*. 11:485-495.
- Hines, James E. Program PRESENCE. [Online]. Available: <http://www.mbr-pwrc.usgs.gov/software/presence.html> [August 25, 2008].
- Hooper, D. U.; Chapin, III, S. F.; Ewel, J. J.; Hector, A.; Inchausti, P.; Lavorel, S.; Lawton, J. H.; Lodge, D. M.; Loreau, M.; Naeem, S.; Schmid, B.; Stetälä, H.; Symstad, A. J.; Vandermeer, J.; Wardle, D. A. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*. 75:3-35.
- Hopkin, Stephen P. 1997. *Biology of the Springtails (Insecta: Collembola)*. New York: Oxford University Press. 330p.
- Karr, James R.; Kimberling, Diana N. 2003. A terrestrial arthropod index of biological integrity for shrub-steppe landscapes. *Northwest Science*. 77:202-213.
- Kremen, C.; Colwell, R. K.; Erwin, T. L.; Murphy, D. D.; Noss, R. F. & Sanjayan, M. A. 1993. Terrestrial arthropod assemblages: their use in conservation planning. *Conservation Biology*. 7:796-808.
- Liaw, Andy; Wiener, Matthew. 2002. Classification and regression by randomForest. *R News*. 2:18-22.
- MacKenzie, Darryl I.; Nichols, James D.; Hines, James E.; Knutson, Melinda G.; Franklin, Alan B. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*. 84:2200-2207.
- MacKenzie, Darryl. I.; Nichols, James D.; Royle, J. Andrew; Pollock, Kenneth H.; Bailey, Larissa L.; Hines, James E. 2006. *Occupancy estimation and modeling*. New York: Elsevier. 324p.
- MacKenzie, Darryl. I.; Royle, J. Andrew 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*. 42:1105-1114.
- Magness, Dawn R.; Huettmann, Falk; Morton, John M. 2008. Using random forests to provide predicted species distribution maps as a metric for ecological inventory & monitoring programs. pp. 209-229 in *Studies in Computational Intelligence 122: Applications of Computational Intelligence*. Springer Berlin. 424p.

- Magurran, Anne E. 1988. *Ecological diversity and its measurement*. Princeton, NJ: Princeton University Press. 179p.
- McCann, Kevin S. 2000. The diversity-stability debate. *Nature*. 405:228-233.
- McCoy, Earl D. 1990. The distribution of insects along elevational gradients. *Oikos*. 58:313-332.
- Morton, John M.; Bowser, Matthew; Berg, Edward.; Magness, Dawn; Eskelin, Todd. 2009. Long Term Ecological Monitoring Program on the Kenai National Wildlife Refuge, Alaska: An FIA Adjunct Inventory. in McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2008. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of agriculture, Forest Service, Rocky Mountain Research Station. [CD-ROM].
- Parmesan, Camille. 1996. Climate and species' range. *Nature*. 382:765-766.
- Parmesan, Camille. 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology and Systematics*. 37:637-669.
- Parmesan, Camille.; Ryrholm, Nils; Stefanescu, Constantí; Hill, Jane K.; Thomas, Chris D.; Descimon, Henri; Huntley, Brian; Kaila, Lauri; Kullberg, Jaakko; Tammaru, Toomas; Tennent, W. John; Thomas, Jeremy A.; Warren, Martin. 1999. Poleward shifts in geographic ranges of butterfly species associated with regional warming. *Nature*. 399:579-583.
- Parmesan, Camille; Yohe, Gary. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*. 421:37-42.
- R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. [Online]. Available: <http://www.R-project.org> [February 29, 2008].
- Royle, J. Andrew; Nichols, James D. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology*. 84:777-790.
- Roth, Dana S.; Perfecto, Ivette; Rathcke, Beverly. 1994. The effects of management systems on ground-foraging ant diversity in Costa Rica. *Ecological Applications*. 4:423-436.
- Schowalter, Timothy D. 2006. *Insect Ecology: An Ecosystem Approach*. 2nd ed. London: Academic Press. 576p.
- Simpson, James J.; Hufford, Gary L.; Daly, Christopher; Berg, Jared S.; Fleming, Michael D. 2005. Comparing maps of mean monthly surface temperature and precipitation for Alaska and adjacent areas of Canada produced by two different methods. *Arctic*. 58:137-161.

- Snyder, William E.; Evans, Edward W. 2006. Ecological effects of invasive arthropod generalist predators Annual Review of Ecology, Evolution, and Systematics. 37:95-122.
- Tepedino, J. V. 1979. The importance of bees and other insect pollinators in maintaining floral species composition. Great Basin Naturalist Memoirs.3:139-150.
- Wallwork, John A. 1983. Oribatids in forest ecosystems. Annual Review of Entomology. 28:109-130.
- Walther, Gian-Reto; Post, Eric; Convey, Peter; Menzel, Annette; Parmesan, Camille; Beebee, Trevor J. C.; Fromentin, Jean-Marc; Hoegh-Guldberg, Ove; Bairlein, Franz. 2002. Ecological responses to recent climate change. Nature. 416:389-395.
- Williams, P.; Hannah, L.; Andelman, S.; Midgley, G.; Araújo, M.; Hughes, G.; Manne, L.; Martinez-Meyer, E.; Pearson, R. 2005. Planning for climate change: identifying minimum-dispersal corridors for the Cape Proteaceae. Conservation Biology. 19:1063-1074.
- Wilson, Edward O. 1987. The little things that run the world (the importance and conservation of invertebrates). Conservation Biology. 1:344-346.
- Wilson, Robert J.; Gutiérrez, David; Gutiérrez, Javier; Martínez, David; Agudo, Rosa; Monserrat, Víctor J. 2005. Changes to the elevational limits and extent of species ranges associated with climate change. Ecology Letters. 8:1138-1146.
- Wright, David H. 1983. Species-energy: an extension of species-area theory. Oikos. 41:496-506.
- Wu, Rudolf S. S. 1982. Periodic defaunation and recovery in a subtropical epibenthic community, in relation to organic pollution. Journal of experimental marine biology and ecology. 64:253-269.