Response to review. McCain, C. M. et al. “Unusually large upward shifts in cold-adapted, montane mammals as temperature warms" ECY19-1070 to Ecology

CMM et al. (Authors): *Responses to each point are in* *blue and italics*.

*CMM et al.: We appreciated the thoughtful and helpful reviewer comments. In light of these comments, we have …*

*Overall, we thank the reviewers for their time and comments, as they led to improved clarity and rigor of this manuscript.*

Dr. Madan Oli, Subject Matter Editor, Ecology

Although both reviewers were enthusiastic about your work, they also raised substantial issues. Consequently, we cannot accept the manuscript for publication in Ecology. However, we are willing to consider a substantially revised version for publication in the journal, assuming that you are able to modify the manuscript according to the recommendations.

Your revisions should address the specific points made by each reviewer. Reviewer 2 raises serious concerns regarding the Bayesian undersampling model, and recommends model testing using simulated (and/or real) data and sensitivity analyses. I also concur with Reviewers’ concerns arising from differences in sampling efforts, assumptions regarding historical vs. contemporary detection probabilities, and the need to incorporate measure(s) of uncertainties in parameter estimates. I believe that adequately addressing these and issues raised by the Reviewers would substantially improve the quality and impact of your manuscript.

*CMM et al.: We…*

**Reviewer: 1**

Comments to the Author

**Review of ECY19-1070. McCain, King, & Szewczyk. Unusually large upward shifts in cold-adapted, montane mammals as temperature warms.**

In this well-written manuscript, McCain and her colleagues provide a compelling report on responses by 42 rodents-range combinations and five shrew-range combinations to climate change in the Rocky Mountains of Colorado. This effort provides a strong and insightful comparison to abundant efforts in California (the “Grinnell Resurveys”). Unlike the California efforts, historic sampling in the Rockies did not provide a series of comprehensively sampled sites (e.g., the Yosemite Transect or the Lassen Transect pursued by MVZ researchers in the early 20th Century). The more dispersed and perhaps opportunistic historic sampling in the Rockies (“patchily distributed” on line 152) required McCain and Co. to apply a different quantitative framework, which they appear to have done quite effectively. The mean elevational change – 122 m upwards – is a frightening statistic; more frightening is the mean 337 m shift of higher elevation species (which presumably are more cold-adapted, and less well adapted to warming temperatures). The manuscript is suitable succinct and provides ample supporting information in supplementary appendices.

No manuscript passes review without SOME comments, of course. I have inserted a number of minor comments or suggestions in the pdf of the ms, which I will return to the editors. Some key comments or questions follow.

General.

I note “rodent-range” and “shrew-range” above because the authors actually sample only 4 shrews, one of which is in two ranges, hence 5 shrew-range combinations. Similarly, 31 rodents. I strongly suggest changing the text to avoid misdirection.

I do wonder why the authors did not include any larger species – lagomorphs, small carnivores. I presume this is a function of their trapping methodology (ye olde Sherman trap is not good for these larger species, and would need to be complemented with Tomahawks or another cage-trap); and, these often require efforts other than trapping. Still, this seems an unfortunate omission.

One substantial omission (to me, at least) is the lack of information on sampling efforts. What traps were used, how many, for how long, and how were they laid out (grids, lines, webs)? What bait was used? This is easy to include, but critical if we are to evaluate the efficacy of their sampling (I trust their sampling; my comment is aimed at ensuring that readers less familiar with this team don’t question these efforts).

I am somewhat surprised that the authors generally pigeon-hole all elevational ranges as shifts, effectively downplaying expansions and contractions. Fortunately, both expansions and contractions are much less common in these data than are bona fide shifts. The authors comment on <i>Tamias quadrivittatus</i> in the San Juans, which contracts downward, yet they refer to this as a downward shift in their analyses. On the other hand, <i>Tamias minimus </i>in the San Juans is treated as a downward shift, when in fact this species expanded both uphill and downhill (albeit clearly more in the latter direction). In the bigger picture, I suppose that these are minor issues (the trends are quite evident, and I believe that these two chipmunks are the only “offending” species in this dataset), and the authors do note contractions in the Discussion (but not expansions). I don’t have a clear recommendation for this issue, but I find myself mildly bothered by it.

The authors use 4-letter abbreviations in the key Fig. 3, and refer readers to Appendix S2 for full names. I would add these abbreviations to Appendix S2. Also, 4-letter abbreviations comprised of the first 2 letters of the genus and species names are common, with perhaps the sole real benefit that a second letter helps to cement the genus when reviewing (e.g., “Pint” could be a species of <i>Peromyscus</i>, <i>Perognathus</i>, or <i>Phenacomys</i>). Yeah, this is very minor. In your case, the only problem with 4-letter codes would be the two voles, <i>Microtus mogollonensis</i> and <i>M. montanus</i>; I would use MIMG and MIMO to avoid confusion. Of course, this is pretty trivial, and I defer to the authors per their preference.

Note that “<i>Reithrodontomys</i>” is missing the terminal “s” in Appendix S2, and <i>Peromyscus nasutus</i> is misspelled as “nastutus”.

Introduction.

I was surprised that the authors do not reference Kelly & Goulden 2008 (Rapid shifts in plant distribution with recent climate change. PNAS 105:11823-11826). This seems quite suitable in the second paragraph of the Introduction, but of course this is the author’s decision to make.

Methods.

In the 2nd paragraph, the authors claim to have “amassed specimen data from all museums with Colorado mammal specimens”. I suspect there may be some lesser known or very small collections with RM specimens that were missed. This is truly trivial, but perhaps hedge bets with a caveat here?

Authors state that “Nomenclature was updated following Wilson and Reeder . . .” I believe they mean updated FROM W&R. W&R (and now the Handbook of Mammals of the World) provide a baseline for contemporary mammal taxonomy, but <i>Ictidomys</i>, as one example, was not used in W&R. Hence, the authors appear to have updated from W&R. An alternative would be to just use HMW, which includes three volumes of relevance (V. 6 & 7 on rodents, V. 8 on insectivores) and I believe uses the taxonomy presented herein.

At the end of the 2nd paragraph, the authors cite Armstrong’s 1972 treatise on Colorado mammals; I haven’t carefully scrutinized Armstrong 1972 against the later, second edition (Armstrong et al. 2011), but the authors cite both in this ms, and I don’t understand why they don’t include the latter at this point, as part of their baseline for known distributions of mammals.

Line 140. “At each of these 32 anthropogenically undisturbed sites . . .” While this may be true in relative terms, I find myself questioning this. No grazing, mining, timber extraction, fire suppression, and so on? Truly undisturbed? Perhaps clarify to avoid readers worrying that you might be overstating the purity of these sites.

Lines 149-151. The opening statement is readily confused (I was initially confused, at least). As stated, it appears to state the impossible – the documented range IS the range from the highest to the lowest points where an animal was recorded. What the authors INTEND to state here is that they assume that species extend beyond these documented records (hence the application of the Bayesian approach). I would suggest re-stating this along the lines of “We assumed that species were undetected at locations above and below the highest and lowest documented records, as species are not always detected even when they are present (Mackenzie et al. 2002).” Perhaps that is more verbose than these authors favor, but the approach makes the point less prone to misinterpretation.

For species that exhibited range contractions, “the shift direction [inferred by the authors] is the range limit with the largest shift.” I suppose that this generally works OK, although I would simply call these upward contractions or downward contractions, rather than upward or downward shifts. The authors note that only in “rare cases” are contractions similar in both directions; they exemplify with the most extreme case of <i>T. quadrivittatus</i> in the San Juan Mountains (which is SO slightly biased. I agree that this is rare; however, it seems that an upward contraction is distinct from an upward shift, and should be considered so here.

Lines 208-211. You state that range losses should be distributed across the gradient more or less evenly. However, isn’t it true that high elevation species would be less likely to show losses (versus extirpation, I suppose) simply because there is less available to lose?

Line 236. Does this sentence have extra, unnecessary words?

Results.

Lines 261-262. This sentence ("The changes in the range limits . . .") is awkward to me. Do you mean to refer to “differences” (e.g., between empirical and Bayesian-modeled values) rather than “changes”?

Lines 288, 292, 298. The values presented here cannot be AICc weights, since Akaike weights are (typically) normalized to sum to 1.0. Are these the actual AICc values?

Discussion.

Lines 319, 322, and 325. Here you repeat the basic regression & Chi-square stats that were given in the final paragraph of Results. I leave it to you to decide, but this doesn’t seem necessary.

Line 403. I don’t know that stating the percentage of the gradient is informative or ecologically insightful. It is the absolute change that really matters, isn’t it?

Line 404. You state that if this trend towards upward shifts continues, available habitat will decline to the point of risking local extinction. This would be much more impactful if you could estimate how long these trends need to continue to face such risks. This will differ across species, and of course you cannot know for certain at what point the numbers will decline to the point of oblivion or of stochastic “blinking” out, but if they have moved uphill 337 m in four decades, how soon will they hit boulder fields, or the top of the mountains? Perhaps this puts you on uncomfortable territory, but it does seem to me that the fact that “if this continues” these species are in trouble is brutally self-evident, hence perhaps trivial without some qualification or elaboration.

The observations presented here pretty much match ideally with expectations based on the most fundamental of physiological models. That is, cold-adapted species move more than less cold-adapted species. I feel that the closing lines in the Discussion could be more impactful by noting that these data exemplify, and highlight, concerns that CC scientists have been expressing for years. The facts match our fears – this is not good news.

I like, but do not favor, the closing line (“Take heed—it is only going to continue.”). In the broader context, this “feels” trivial to me, perhaps just too mild. In fact, it is likely to get much worse. However, to avoid any misperceptions by readers, perhaps delete this and close merely with the observation outlined just above – we are documenting the dramatic changes that were predicted with Bush2 opted to ignore climate change. These data make clear that things are moving, literally, and biotic communities as we know them may not be available to our children when they grow up. Pretty sobering to me.

Lit. Cit.

I am not reviewing the full Lit Cit but I note a few inconsistencies.

Armstrong 1972. Is this proper Ecology style? This is indeed this monograph, but the book is 416 pages long. This citation seems to imply solely page 416.

Cary 1911. Similar – NAF reports generally are treated as monographs. Hence, North American Fauna 33:1-256?

Erb et al. 2011. Check for italics in scientific names (<i>Ochotona princeps </i>in this case).

Hall 1981. This has two volumes. I defer to the EIC for how best to treat this in Ecology.

Figure Legends.

Lines 569-570. Since the legend for panel C is clearly intended to explain panel C, you probably could delete "is shown in panel C".

At first review, I inferred panel C to indicate that lighter shades (e.g., in panels a and b) referred to shifts, while darker shades referred to contractions. I don't know that you can avoid this misinterpretation (it helps, of course, when the legend is adjacent to the figure), but I found myself trying to discern which of the blue or red bars in panels a and b were lighter vs. darker.

I hope that these comments are useful as the authors revise this very interesting manuscript.

Douglas A. Kelt

**Reviewer: 2**

Comments to the Author

Review for Ecology

Title: Unusually large upward shifts in cold-adapted, montane mammals as temperature warms

Authors: McCain, C. M., S. R. B. King, and T. M. Szewczyk

General Comments

This manuscript examines changes to the elevational ranges of 47 small mammal species along two ranges of the Rocky Mountains, attributing any shifts to changes in climate. The authors use the locations of 4580 georeferenced specimens collected between 1886 and 1979 to establish historical range limits for species. These are compared to 7444 records from both museum collections and systematic sampling efforts post 2005. The data are used in a custom-built Bayesian model that attempts to account for “undersampling” to quantify changes in species occurrence in 50 m elevational bands. Model output is then used to test for significant expansions or contractions at elevational limits. Small mammals were quite variable in their responses – about half shifted upwards, while the remainder either shifted downwards, exhibited stasis, or showed range collapse. Shifts also differed between lower and upper range limits. Changes are interpreted as responses to climate warming.

I admire and support the authors’ use historical data combined with contemporary resurveys to understand long-term changes in small mammals. The manuscript develops a new modeling approach for inferences on range shifts that attempts to correct for the effects of sampling differences, which have long been known to plague comparisons between historical and contemporary data (Tingley and Beissinger 2009, TREE 24:625-633). While the method is creative and could have wider application, it is also untested and has several important shortcomings discussed below. I wish I could be more supportive, but I feel there are too many weaknesses with the approach as presented in this manuscript (perhaps due to space limitations?) to be confident yet in the validity of its inferences. If this method is going to work and be widely applied, which would be an important advance, the authors need to rigorously demonstrate the model’s behavior and veracity (e.g., through a paper in Methods in Ecology and Evolution).

The Bayesian undersampling model needs to be better grounded to statistical theory, should be tested with both simulated and real data sets, and probed with sensitivity analyses. Starting with the latter, how sensitive is the model to the choice of elevational band width? The authors chose a fine elevation band-width of 50 m. Would the results hold if they used 100 or 200 m band widths? Extensively testing the model’s ability to detect changes with simulated data sets seems like an obvious thing to do. The model is being asked to detect changes at the edges of distributions, which is challenging since few individuals are likely to occur there. Finally, the model should be evaluated with real data. The authors have apparently collected the type of data that should enable them to use occupancy models to conduct robust tests of range contractions -- that is sites where species were found historically but are no longer occupied. Results from the Bayesian undersampling model to quantify the absence of a species at contemporary sites could be compared to results from an occupancy model using the probability of false absence test (Tingley and Beissinger 2009), which is a robust method to test for elevational change of small mammals (Moritz et al. 2008, Rowe et al. 2014) and birds (Tingley et al. 2012).

The Bayesian undersampling model does not explicitly account for differences in sampling effort (i.e., trap nights per site or elevational band), or the differential sampling of historical sampling localities in elevational bands. Instead, the basis for inference is built around the estimation of (pj), the probabilities that a randomly observed individual from an elevational bin belongs to a species j. Small mammals are notorious for year-to-year changes in abundances. These fluctuations have strong influence on community composition (species’ proportions), much more so than on site-level occupancy (presence or absence). I’d guess that annual fluctuations in abundances would be likely to influence the probability that a randomly observed individual from an elevational bin belongs to a species (pj). This could be tested with real or simulated data.

A key way the model tries to account for differences in species-level detection in the calculation of pj is by using measures of individual detection probability (δj) from capture-recapture methods. A mean value without a confidence interval is given for each species in the Appendix S2, but there is no documentation for how it was estimated (Lines 173-175). Many papers have been written about estimating detection probability from capture-mark-recapture data, and the type of documentation and analysis required to produce robust estimates. A reader really needs this information to evaluate the veracity of the detection estimate. Moreover, the uncertainty around the mean detection probability could be incorporated into the Bayesian model.

Possibly the biggest shortcoming of the approach used in the Bayesian undersampling model is the assumption that individual detection probability (δj) will be the same for historic and contemporary sampling. This seems unlikely, given differences in trapping methods, sampling effort (trap nights per site or band), weather, abundances, etc. between contemporary and historical surveys. I don’t know how to deal with this problem. One might be able to reconstruct the historic capture histories at the species level, at which point standard occupancy model would be more useful.

I feel the manuscript also needs to do a better job of dissecting the changes at the range limits. Shifts are likely occurring independently at upper and lower elevation limits for many species. If a lower limit contracts but an upper limit does not expand, the manuscript interpreted this as upward movement (e.g., Fig. 3 Front Range: Clat, Smon, Mflav). But this is not the same as an upward movement resulting from a shift both the lower limit and upper limit that pushes the species further upslope. Looking at Figure 3, it is difficult to see clear evidence for downward movement of many species (e.g., Front Range: Ovar, Itri, Pnas). In these low to mid-elevation species, it seems the upper limit contracted, an unexpected response from climate warming. As a result of the mixing of elevational limit change patterns, Figure 3C feels meaningless without a little clearer connection to which range limit is changing (or not). It is hard to draw conclusions about the causes of range shifts presented in the manuscript. The authors assume that ranges would shift in response to warming (Lines 97-98 and 103-105), but no down-scaled, local measures of climate change are presented to relate to the elevational changes sampled. Did the climate change the same way throughout the elevational bands at each range? This was not the case in the Sierras, where some elevations and locations warmed more than others and some even cooled (Tingley et al. 2012). Precipitation change was also quite variable at sites along elevational transects in the Sierras. The authors indicate that the San Juans got wetter while the Front Range dried. It would be useful to see whether the variation in species’ responses could be attributed to different levels and different dimensions of climate change that are associated with elevational limits.

The analysis of what traits are associated with range shifts could be valuable, but the AIC analysis was not conducted or presented properly (see below).

Specific Comments:

Introduction - The Introduction rambles, is contradictory and oversimplified in places, and doesn’t do a strong job of setting up the study. The first paragraph is too general and not entirely representative of how knowledge about the effects of climate change has progressed over the past decade. Hyperbolic statements like “Essentially the planet and all its inhabitants are in a climate crisis of unknown magnitude” aren’t accurate, as we know: (1) there will be both winners and losers among species from climate change, (2) temperature and precipitation have changed differently in the same location, and can have contrasting effects on range shifts, and (3) that the climate is not changing the same way and at the same speed every place on the planet. Moreover, historical resurveys are now well established as a method to examine how biodiversity has responded to climate change and whether traits do a good job of explaining quantified (not modelled) change (see MacLean and Beissinger 2017, Global Change Biology 23:4094-4105, for a review of over 50 studies that paint a picture of mixed success of traits). Likewise, the second paragraph tries to summarize the broad literature on elevational shifts to determine both patterns and magnitude of upward or downward change, but a clear message doesn’t emerge. It’s not surprising because species’ responses have been quite variable, studies measure different kinds of change (mean elevation versus range limits), offer different metrics (e.g., absolute change versus change/decade), and have been conducted at locations with different rates and types (e.g., temp versus precip) of climate change. See Rapacuciuolo et al. (2014, Global Change Biology 20:2841-2855) for a focused comparison of heterogeneous elevational responses to climate change across plants, butterflies, birds and mammals in the Sierra Nevada based on historical resurveys, and Smith et al. (2019, Nature Climate Change 9, 787-794) for a range-wide comparison of pika responses.

Lines (L) 138-146 - Little information is provided about the trapping effort. How many nights were traps opened at a site and how many traps were set per site? What kinds of traps were used? How many live traps versus pitfall traps? Were they baited? Sorry if I missed these details.

L 159 – probability of sampling error is very vague and undefined. There are many processes that create “sampling error”, including differences in detectability, uncertainty in sampling location, differences in bias from sampling techniques, misidentification of species, etc. The manuscript does not indicate what kinds of error are accounted for with the Bayesian model and what are not modeled.

L. 200 – Why is elevational range midpoint important? What test is used to determine if a change in midpoint is significant?

L. 207-214 – The essence of the argument here is that upper range limit shifts reflect climate change, whereas lower range limit shifts reflect habitat change if a species is found lower on the elevational gradient. But if this is a “cold adapted fauna”, wouldn’t one expect lower range limits to be shifting upwards regardless of where on the mountain the species occurs?

L. 235-240 – The use of multiple stepwise regression to allow the computer to sort through combinations of potential models is not what Burnham and Anderson (spelled with an “on” and not and “en”) had in mind when they championed the information theoretic approach AIC. They advocated developing a set of a priori models based on hypotheses, implementing them, and then presenting the entire model set using AIC to compare the ability of models (hypotheses) to describe the data. Nowhere are these results presented to assess how well the final model described the data compared to other models. I know that’s not always feasible to conduct AIC analyses this way with multivariate data sets, because there can be many feasible models. I think a good approach is to present one model set composed of the ability of each single factor (trait in this case) to describe the data, and then present results from the all subsets of models (sometimes known as dredging). In both cases, the key metric is the model weight. For the degrading approach, use the cumulative model weight.

L. 249-259 – Missing from this presentation is a consideration of the samples in relation to the number of elevational bands. With the 50 m bands, approximately 45-50 bands represent the elevational gradient. How should we think about the intensity of the historical sampling of 97 specimens per species across these bands?

L. 288 – AICc weights represent a measure of how well a model describes the data compared to other models in the same data set. Weights are standardized to add up to 1. What is given here is not an AIC weight but apparently an AIC score, which is meaningless unless compared to other models in the same data set.

Steve Beissinger