CC: [erik.blomberg@maine.edu](mailto:erik.blomberg@maine.edu)

Ref.: Ms. No. CONDOR-20-074  
Improved status and trend estimates from the North American Breeding Bird Survey using a Bayesian hierarchical generalized additive model  
The Condor: Ornithological Applications  
  
Dear Authors,  
  
Two reviewers and Associate Editor Blomberg have provided thoughtful reviews of the manuscript, including its potential value to the literature about how to model population status and trends from BBS data. However, there were a number of concerns that you should address in a revision.

Many of the concerns are relatively straightforward. Other comments address the balance of the manuscript with regard to presenting and interpreting model results vs. advocacy of the approach and use and implementation of the R package. My feeling is that you have hit the balance relatively well in this version but that there are a number of places where addressing the reviewer comments will improve the tone and presentation.

For your guidance, reviewers' comments are appended below. If we have included any attached files, there will be a link at the end of this email to download the attachment, or you can go to your manuscript page and under Action Links you can choose View Attachments.  
  
If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript. Please use "track changes" in Word in the revised manuscript or otherwise make your changes from the original manuscript visible through the use of a colored font (i.e. red) so the editors and reviewers can readily see your revisions. Please include a clean copy of your manuscript in addition with your submission..

Your revision is due by Jun 28, 2020.  
  
To submit a revision, go to <https://www.editorialmanager.com/condor/> and log in as an Author.  
  
Username: ASmith-493  
Password: [click here to reset your password](https://www.editorialmanager.com/condor/l.asp?i=100162&l=C4A5DH1R)  
  
You will see a menu item call Submission Needing Revision. You will find your submission record there.  
  
Please keep in mind that if your paper is accepted and you are an AOS member, there will be no page charges for publishing your article. Non-AOS corresponding authors will be charged $100 per page (not counting appendices).   
   
So today might be a good day to join the AOS, <http://www.americanornithology.org/content/aos-membership-join-renew-update>.  
  
Yours sincerely,  
  
Catherine A Lindell, Ph.D.  
Editor-in-Chief  
The Condor: Ornithological Applications  
  
Associate Editor Comments: Thanks to the authors for submitting this work. I’ve now received two thorough reviews of the manuscript, and I’ve reviewed it myself. Both reviewers are well-versed in the specific analytical techniques the authors employed, and they offer a number of constructive comments. I didn’t see any fatal flaws with respect to the analyses, but rather suggestions around presentation of methods and interpretation of results that I think the authors should consider very carefully in moving forward with the manuscript. I won’t reiterate the more specific issues raised, except to say that in reading through the reviewer comments nothing jumped out at me as an unreasonable or unfounded suggestion, so I recommend the authors consider them all very carefully.  
  
More generally though there was some concern among both reviewers that the GAM-based modelling approach you present does not represent a dramatic leap forward in comparison to at least one of the alternative modelling approaches (DIFFERENCE). At the very least you didn’t provide compelling enough evidence to demonstrate a clear difference. To my mind that does not mean your approach can’t still be published as an alternative method, but I do think it may justify toning down some of the strong advocacy for the approach, as suggested by Reviewer 2. I don’t see any reason why you can’t present an alternative and equivocal approach, without necessarily claiming it’s the one clearly superior option.  
  
My only other big-picture comment on the manuscript has to do with what feels to me like a bit of tension between whether you envision this paper as a statistical methods paper evaluating the relative strengths of a variety of methods for analysis of BBS data (i.e. the focus is on the models), or if it is a paper describing use and implementation of your R package. Clearly you emphasize the former to a much greater degree, but in many places you focus on the utility of the package. You often present results in a ‘for example’ framework, e.g. relying on the barn swallows as a case study. I think this works for demonstrating the utility of the package, but it doesn’t necessarily illustrate the more generic performance qualities of each model type. Some of your other assessments that synthesize across species, such as in figure 4, provide a better general assessment of the modeling approaches. To be clear I think both presentation of the relative merits of the modelling approaches and presentation of the package are worthwhile objectives, and fit within the scope of a paper Condor could publish. So this comment is not meant to stifle either per se, but rather I would suggest making both objectives clear in the manuscript, and tailoring your presentation of results towards each. To that end, if the package provides tools for working with BBS data in a way that makes the analyses more accessible to a larger audience, you might actually be able to play up that element more. But again, it depends on your specific objectives for this particular manuscript.  
  
I have just a few other more specific comments.  
  
Line 13. Is ‘BBS models’ an appropriate term? To my mind BBS describe the data or the survey design, but the models are more general in that they could be applied to any number of time series data? ‘model types fit to BBS data’ is probably more accurate, although I realize this is more verbose.  
  
  
Line 135. “…analyze the BBS data.” Rather than the BBS per se.  
  
Line 136. So the data point is the number of birds counted summed across a BBS route? Rather than the mean number per point? Maybe this is obvious for those who commonly work with BBS data, but its probably worth being specific.  
  
Line 302. Do you mean the slopes were less linear, rather than and less linear?  
  
Please feel free to contact me with questions about this review. Erik Blomberg, Associate Editor. [erik.blomberg@maine.edu](mailto:erik.blomberg@maine.edu)  
  
  
Reviewer #1: The paper presents a model for smooth trend analyses of the Breeding Bird Survey. The model combines the main components of previous BBS models with additive model components. GAMs are frequently used to estimate population trends, and conceptually the model presented here is a fairly straightforward extension of previous models used in the analysis of bird trends. The model, however, seems well designed and applying these techniques on a large data set like the BBS is not necessarily trivial. The paper also presents an extensive comparison of different models for the BBS using cross-validation, and provides an R-package that I believe many will find useful.  
  
Line 4-6. Not clear what you are saying here.  
  
Line 72. GAMs are not well suited to abrupt change points, although obviously better than a linear model.  
  
Line 134. How were the species selected?  
  
Line 144. Here you say four BBS in addition to the GAMs, but on line 135 you say two alternative models.  
  
Line 147-152. More details on how the parameters are modeled are needed here. E.g., are the omegas fixed or random, are the epsilon independent across strata, time and route, etc.  
  
Line 153-154. How did you treat the df parameter of the t-distribution?  
  
Line 159. Consider matrix notation here.  
  
Line 167. If sigma controls shrinkage towards a linear term something seems to be missing here. Setting sigma\_B to 0 would give Delta = 0, i.e. no trend. The smooth model here looks like the thin plate spline as given in Crainiceanu et al, but with the linear component missing.  
  
Line 182. Why only 13 knots?  
  
Line 190. Are these independent across strata and years?  
  
Line 220. Better motivation for this approach is needed. If you really want to test the ability of the different models to capture population change it would also make sense to leave out entire years from the folds and let the models interpolate the missing year(s).  
  
Line 233-234. 'k' should be 'v'.  
  
Line 236. Define X\_i.  
  
Line 241. Here you are indexing by -i, not -v?  
  
Line 249. It seems reasonable to also include the fold in this model, perhaps nested in year and stratum.  
  
Line 274. Relative to what? I.e. what is the baseline?  
  
Line 279. It's not obvious why you would want to include observer-route effects here. E.g. if observers get better over years including observer effects would bias the trend estimates.  
  
Line 281. The correction factor needs further explanation.  
  
Line 288. What is alpha?  
  
Line 357-360. In what way are the GAM trends more useful and have better accuracy than the DIFFERENCE trends? Your results suggest that the linear trend model (SLOPE) performs worse, which is expected, but that exactly how the non-linear trends are modeled plays less of a role (similar performance of DIFFERENCE and GAMYE). I wouldn't be surprised if a piecewise linear model, used e.g. in the TRIM software that provides trend estimates for European countries, would also perform similarly. The main point seems to be that the non-linearity matters, not the details of the specification.  
  
Fig 4. Can you provide any guidance on how to interpret the scale? E.g., how meaningful is a difference in elpd of 0.01?  
  
  
Reviewer #2: General comments:  
  
1. A good deal of the introduction and results of this manuscript focus on concerns about the SLOPE model. Its lack of fit relative to other models have been described by Link et al. (2020), who documented that the DIFFERENCE model is superior for the vast majority of species covered by the BBS, and most recent analyses releases from USGS incorporate their model selection results. Given that, I am not sure how important it is to discuss the deficiencies of the SLOPE model in such detail. Also, the presentation of the SLOPE model in the introduction is misleading. The authors criticize that slope parameter as overly restrictive. Although there is an obvious analogy in structure between the slope model and the smooth models described here, the slope parameter in the SLOPE model is just a structural element of the year effects model. The goal of the SLOPE model is estimation of year effects, and inference of change over time from the model is always based on the year effects. I know of no one who interprets that slope parameter as a descriptor of population trajectory. That is an important conceptual nuance: Although the GAMYE model could be viewed as a variation in SLOPE model in which the slope parameter is replaced by a smooth, the smooth part of the GAMYE is the focus of interest in that model is a way that it is not in the SLOPE model.  
  
2. The authors display clear preference for GAM models, and all of their discussion reflects this bias, including their final recommendation for use of GAMYE. Even the title reflects this tendency for strong advocacy. I would say such a recommendation is premature, based on an analysis of a handful of species. There are many details there, such as the spatial associations and the predefined complexity (number of knots), that need additional discussion.  
  
3. There is another viewpoint on the relative merits of GAM models relative to DIFFERENCE models. Many users of these data prefer year effects, and consider associations over space to not be a desirable component of an analysis (i.e., they want regional results to only reflect their regions). While the GAM-year effects model provide year effects, they are clearly influenced by the shape of the GAM trajectories. The DIFFERENCE model seems to work quite well for analysis of BBS data, and that came out clearly in the analyses presented in this manuscript.  
  
4. There are many places in this manuscript where the authors cite informal sources, such as non peer reviewed manuscripts in online archivies or onlines sets of results, as a source for analysis methods. This has the effect of making these methods appear to be supported by peer review literature while in practice this manuscript is the first place in which they have encountered peer review. I think these sources need to be clearly distinguished from peer reviewed method publications and the methods need additional rationalization.  
  
Specific comments:  
  
  
l. 55-56. I don't agree with this. Most of these uses can be well-addressed using derived statistics from year effects models.  
  
l. 57. For the USGS 2018 analysis, the "best" model as suggested by Link et al (2020) was used. The slope-random year effect model is only used for a few species.  
  
l. 60-61. These models are conservative when estimating trends? I have never seen an indication of that, and that is clearly not a result from Fewster et al 2000.  
  
l. 62-67. This is an overstatement. The slope parameter in that model does not necessarily constrain trend estimates, as they are based on year effects. This can occur, but generally does not.  
  
l. 70-73. This is true, and they can take any form ranging from linear to year effects, depending on how you define the structure. You do not do any evaluation of that in the approach described in this manuscript.  
  
l. 73-74. Even in the context of complex, multiscale models with random effects? That seems unlikely.  
  
l. 75-76. It is certainly legitimate to criticize the slope-year effects model for not fitting the data well, but it is very misleading to imply that the slope parameters are the goal of inference in the slope-year effects model. That is not true, and I have never heard of them used in that way. Instead, the slope parameter in that model is a tool employed for fitting the year effect model; the year effects are the goal of inference and these slope parameters are a component of the year effect parameterization. Of course they change over time, they represent the framework on which the year effects are estimated. To suggest that these slope parameters are comparable to the smooths is incorrect. For example, the statement in l. 76, is true (that is a slope parameter), but that is irrelevant unless someone mistakes it as an exact representation of population change over time. But, even so, why is it a concern of it changes as you add more data?  
  
l. 81-84. Do the authors have evidence that priors in this model are unduly influencing our view of population trajectories? This is a straw man criticism, in my view, and the comment "the user of published trend estimates has no clear way to discern its influence" really misstates the significance of this criticism. Incidentally, the phrase "population trends are shrunk to zero" is a misleading use of the word "trend." What is a population trend here? It is a yearly change?  
  
l, 91-98. It is certainly true that there may be virtues in sharing information over space in smooths, but it is certainly untrue that this can only be accomplished in the context of smoothing analyses. Also, there is a great of complexity inherent in "allowing the model to share information on the shape of a species population trajectory across a species range."  
  
l. 99-111. This is a bit misleading. Even "inherently smooth" trajectories require additional analyses and derived statistics to provide parameter estimates for these conservation uses. GAMS are simply an alternative set of trajectories, but a-priori are not better suited for additional summaries than appropriate summaries from year effects would be. The virtues described on l. 109-111 are not unique to smooths, and in fact are common to any reduced model.  
  
l. 123. I suggest you omit "the medium and long-term" from this sentence, or replace them with more specific terms.  
  
l. 124. This is vague. How can a GAM smooth be treated as a random effect?  
  
l. 127. Replace "sub-regions of" with "the strata." Why use 2 different terms for the same thing in 1 sentence?  
  
l. 127-130. What is Model "GS?" This sentence needs clarification, perhaps with a description of these HGAM and GS models from Pederson et al 2019. "Wiggliness" is a technical term?  
  
l. 136-139. Is this our only introduction to the BBS? Robbins et al. (1986) is a venerable, if outdated, choice of citation for the BBS.  
  
l. 140-142. Both your citations only reference leave-one-out cross validation, but does k-fold cross validation increase precision at the cost of bias?  
  
l. 144 is incorrected stated, as it appears to exclude the GAMS from the general formulation on l. 146. If this statement is correct, then the GAM analysis is not clearly defined.  
  
l. 163-166. I don't have a good intuition as to how an overall smooth that governs shape of strata smooths would work. Does this mean that they all have essentially the same shape, or can only fold in similar amounts (is that wiggliness?).  
  
l. 167. "first-degree polynomial" is a line.  
  
l. 175-179. I am always concerned when vague justifications such as "we have so far had good results...area of ongoing research" are stated in conjunction with a prior that constrains inference, and then users are asked to explore alternative priors. This suggests that the authors have concerns about the effect of this constraint on inference, and should be better explained here.  
  
l. 180-187. Typically, the number of knots would be expected to vary depending on the requirements of the data, and would be fit using a model selection procedure. You present a preset number of knots and rationalize that choice with a citation to an unreviewed document. Although this is not an unreasonable approach if one is merely interested in portraying pattern in the data, I think that if the smooth is to be used for inference one would want to evaluate this species-by-species.  
  
l. 186. This is misleading, as it implies that the smooth just misses extreme 2 year changes and captures all other relevant shape. The amount of knots affects the shape, and it doesn't just cause the smooth to not capture "shortest-term variation."  
  
l. 188-192. This approach makes perfect sense, as it is the familiar residual indices of Sauer and Geissler or the slope annual indices model, but with the slope replaced by the year effects. However, on another level, it doesn't. Presumably, the smooth and the annual indices are competing to explain the year effects, and including this parameter must influence the smooth. Its inclusion also makes and estimate of change based on the smooth problematic, as the smooth is now defined as trajectory with those year components extracted. There are also some complexities associated with the fact that the smooth is itself a spatially-extracted random effect while the year effects are done separate among stratum. That does not seem consistent.  
  
l. 195-195. This sentence is not consistent with l 144.  
  
l. 197. Is the slope parameter log-linear?  
  
l. 198. This is not a "linear-trend model." It is a year-effects model.  
  
l. 201, 204. Omit the word "main."  
  
l. 206. Is this correct: "follow a random walk from the first year of the time series?"  
  
l. 209-215. This is little more than an advertisement for a software package. Please provide more details of the fitting and diagnostics.  
  
l. 217-229. I believe that the properties of "v-fold" cross validation are poorly understood in the context of complex hierarchical models. Even though the authors state that the random subsetting "included some observations from every combination of strata and years," It is not clear to me how such a subsetting can be done for complex models applied over many strata without causing the model to be quite different in terms of random effects when applied to the subsets. That means that the cross validation becomes a de-facto evaluation of the stability of the model when applied to the subsets. Maybe this isn't a concern, but the tradeoff between bias and variance in cross validation is apparently an active topic of discussion in the cross-validation literature.  
  
l. 220-221. Please specify how this was done: "included some observations from every combination of strata and years"  
  
l. 227. Microsoft and Weston (2019) is an odd citation that appears incomplete.  
  
l. 238-258. This hierarchical model for differences in fit is rationalized "to account for the imbalances in the BBS-data among years and regions." I am not sure what this means; the statement needs to specify what these imbalances are, and the method description also should explain how this model accomplishes that goal. In my view, any sort of shrinkage estimator for these observation-specific differences has the effect of discounting some observations and enhancing others, and the precision of the estimates is generally the criterion by which this is accomplished. Whether the analysis is a "robust estimation approach (l. 252-253, 257)" or if it disproportionately emphasizes regions with better data, is unclear from this description.  
  
l. 238-239. What is the practical difference between a sum and a mean?  
  
l,. 272-273. The R' does not remove the annual fluctuations from the trend estimate. I merely removes the portion of the annual estimates that are partitioned into the quasi-year effects.  
  
l. 279-281. This differs from earlier published approaches to this scaling, and it can be argued that it is not to be preferred to the existing scaling. It might be useful to mention that this is a topic of ongoing research.  
  
l 299-315 I think the results section requires revision to provide a more balanced view of the comparison of the smooth models to year effect models. As now written, the results focus on comparison of smooth models to the SLOPE model, and identify large differences in the early years of the time series. These are interesting results, but not surprising as the SLOPE model has been shown elsewhere to also differ from the DIFFERENCE model in those early years. The authors mention this, but not until l. 333-355 do they note that the DIFFERENCE model results are quite similar to the smooth model results.  
  
l. 303. The notion that the smooths tend to "better track the nonlinear patterns in the raw data" is not necessarily a virtue. The whole idea of analyzing BBS data is to distinguish real pattern in the populations from patterns in the raw counts. If we trusted those patterns, we would not be using these complicated models.  
  
l. 306-307. This is not surprising, in that both models share a similar form, and both models are structured to define an overall smooth from which regional smooths are derived. I think it is important to address the consequences of the consequences of the spatial dependences, as these models are defined by having an underlying structure from which the regional smooths are derived.  
  
l. 308-313. These comparisons focus on the slope model relative to the smooths. While I agree that there is still an ongoing interest in why the trajectories from the slope model differ from other models in the early years, these patterns are well known. The slope model also differs from the difference model.  
  
l. 316-332. Some of this material is better suited for the discussion section.  
  
l. 357-372. This is somewhat of an overstatement. I think that the authors need to focus on why smooths provide a better framework for estimation of population change. While the authors have shown that the SLOPE model does not fit as well as the smooths in the early years of the survey, they have not demonstrated that those models are superior to the DIFFERENCE model. The value of the smooth trajectory as a measure of trend is relevant, but of course the difference model provides the means for computing any alternative trend metric, including a smooth, without requiring the fitting of the smooth as a basic model.  
  
l. 373-399. The discussion of how trend is defined in conservation assessments has been going on for many years. This discussion advocates the use of smooths as a reasonable approach, and it is, but one gets the impression that the estimation procedure advocated in this manuscript are the only way to produce metrics based on smooths. That is not true. All possible estimates of trend can be computed as derived statistics from year effects models, and this manuscript does not establish that smooth-based estimation procedures are inherently superior to year effect models. It also does not acknowledge that every summary described here has advantages for a smooth-based estimation can (and has) been implemented by including smooths as derived statistics of year effects models. The discussion of whether a smooth-based analysis procedure is preferable to a year-effects based procedure dates back to James, Weidenfeld, and McCullouch's initial implementation of LOESS procedures for the BBS, and the discussions on these topics is still relevant.  
  
l. 422-428. It is unclear here whether you are citing Roberts et al. 2017 as a critique of v-fold or leave-one-out cross validation. It should be clarified that the issue is primarily a issue for v-fold. I am not convinced that the hierarchical model addresses this concern, as it does not completely model spatial and temporal dependencies in the data. "Reasonably confident" seems to be faint praise!  
  
l. 438-457. I find this discussion to be unconvincing. The appeal to "careful thinking" seems to be suggesting that "we are convinced, you should be, too," even though there is no evidence of clear superiority of DIFFERENCE or GAM models based on your analyses.  
  
John Sauer

*In compliance with data protection regulations, you may request that we remove your personal registration details at any time.* [*(Remove my information/details)*](https://www.editorialmanager.com/condor/login.asp?a=r)*. Please contact the publication office if you have any questions.*