I have had a real go at the MS. I was unable to save it with track changes (too many things moved around, I think), so this version just shows comments and a few recent edits. There weren’t many changes made to what you sent me, but I have numerous comments and questions. For parts missing from the draft you sent (abstract and discussion) I added in text from the latest previous version I had. I edited abstract, but discussion almost untouched. I’ll get to this separately in next week or so.

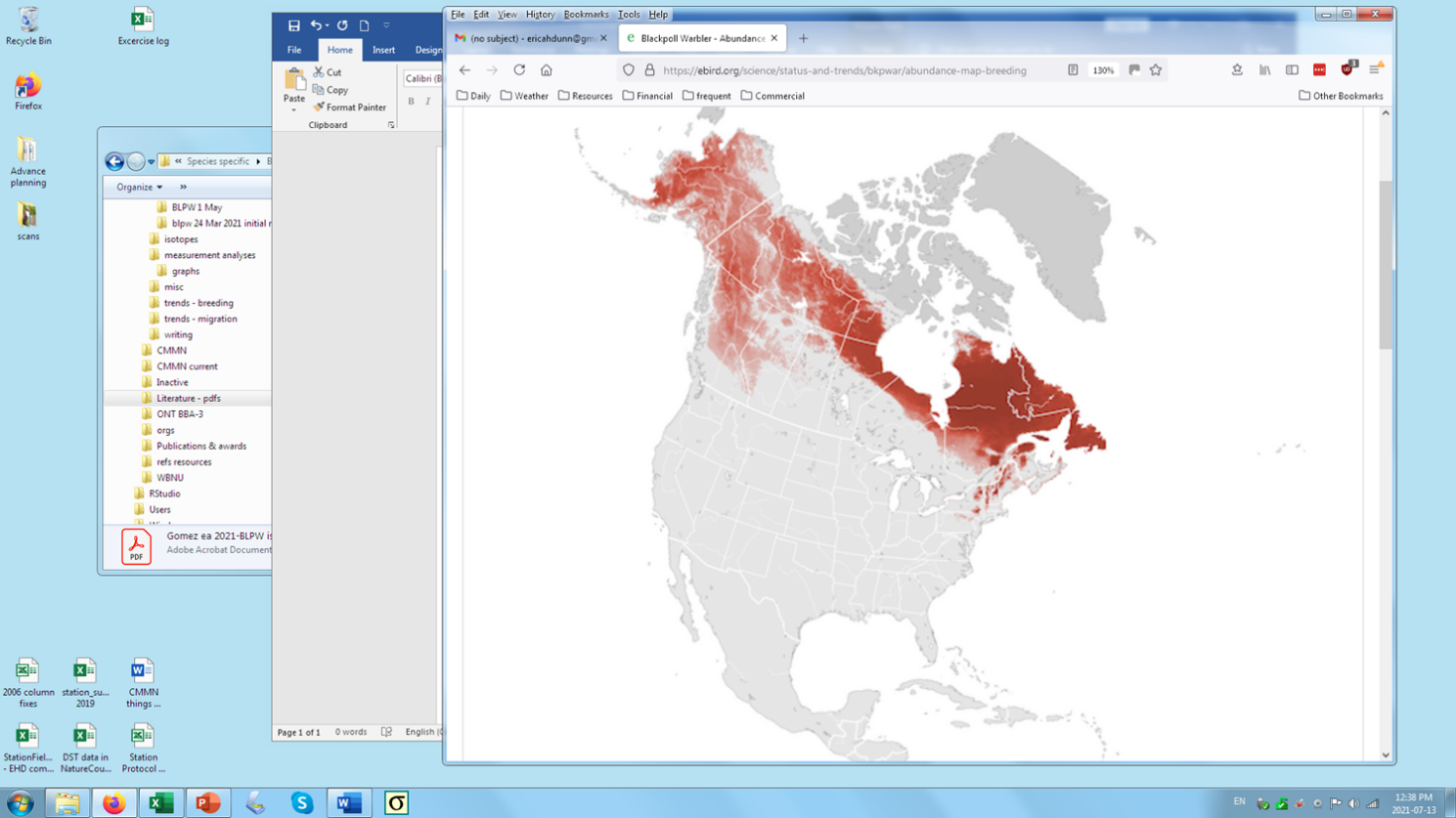
The things that took me most time were the figures, tables and appendices. The versions you sent have a lot of material that needn’t be included, some that is out of date and some that is duplicated in different files. I have combined seasons into single figures for clarity and simplicity. Consider also whether it would be useful to separate Supplementary Info into separate Appendices: a) Diagnostics (goodness of fit, estimates of bias), tests of model sensitivity to changes in parameters; and b) Additional output.(“expanded results”) -- though I’m not sure I’d know what should go where. I wonder if all figures in appendices are useful. Unless each one illustrates a point made in text, we should probably consider deleting.

The notes below address things I think should be investigated to test model sensitivity. We should do some work and add text presenting results of testing the model for sensitivity to changes in abundance weighting, or regional splits assigned to each station, or adding more stations to sample each region (Simulation paper by Crewe indicated that migration monitoring could meet monitoring targets in 40 years with 3 sites sampling each region, while doing so in 20 years would require 5-10 sites/region. Crewe, T.L., P.D. Taylor, D. LePage, A.C. Smith and C.M. Fran­cis. 2015. Quantifying regional variation in population trends using migration counts. Journal of Wildlife Management 80: 245-255.)

Lines 261-268; the paragraph indicating how relative abundance of BLPW in West and East was estimated.

In earlier version, you used BAM density estimates, modeled on point counts within the breeding range and modeling abundance in areas without data based on consideration of habitat suitability. While there are weaknesses with that approach, so too are there weaknesses with the eBird predictions. Both are affected by lack of bird counts in many parts of the breeding range. Also, there are notable differences between the two sets of predictions. BAM shows higher density in NF than in PQ than does eBird, and lower density in YK(?) and northern ON. AK is not shown on BAM map, so that can’t be compared. In an earlier version of your MS you used BAM, but didn’t indicate the relative size of population in the different regions (and in any case, we now have only 2 regions instead of 3). What is the relative abundance in West and East using BAM as compared to the values you found using eBird? If very different (and even if not), I’d like to see how much results are affected by using different relative weights. (NB. - Quality of relative abundance data for other boreal species will likely be better than for BLPW, but may as well test sensitivity of the method to weighting factor now.)

Abundance map from eBird





Density map from BAM project [https://drive.google.com/drive/folders/](about:blank) 1exWa6vfhGo1DNUL4ei2baDz77as7jYzY

(would be easier to see if reversed from negative to positive)

Lines 310-314 and your Appendix 5. Simulation of repeated feather collection

I recommend dropping this, as the tests you ran seem very unrealistic to me, and adding more monitoring sites in key locations seems a much more powerful approach to increasing trend precision (see Crewe paper notes above.) I do, however, like the idea of testing model sensitivity to different priors, and would suggest presenting results that test sensitivity to a) changing the split between regions (e.g. for spring, change all sites from TCBO through MBO to several higher levels of percentages of East, and see how much change is needed to alter results significantly); and b) Separately, test effect of different East-West breeding abundance estimate (eBird or BAM), as noted above. How much difference between density in the two strata would change outcome of model if nothing else is changed?

Although additional feather sampling is a good idea, the aim should be to have historic samples on hand for analysis once count data from several sites in a catchment zone start showing signs of population change. For example, the isotope shifts over time in BLPW at LPBO, combined with decline of BLPW in Massachusetts in 80’s but not until more recently at LPBO, indicate differential change in regional population trends; with decline in eastern population starting near coast and only later spreading to Great Lakes portion of the East stratum. Because the set of relevant circumstances would likely vary for every species, I don’t think it’s worthwhile to include your simulation. Nonetheless, continued feather collection for potential analysis in future when regional differences in population trend start to appear is a good idea, and is something we should recommend in discussion.