

Interactive comment on “A Global Model of Predicted Peregrine Falcon (*Falco peregrinus*) Distribution with Open Source GIS Code and 104 Open Access Layers for use by the global public” by Sumithra Sriram and Falk Huettmann

Anonymous Referee #1

Received and published: 2 April 2017

Difficult data set, provocative presentation, but potential highly useful. Attempts to set a very good open access example. Uncertain how it fits for this journal, but more uncertain how it would fit better anywhere else. With some changes and edits could be a very useful product. (Not easy to review without page numbers!)

From the text as provided - and following Figure 2 very closely - this reader understand four purposes:

1) Compile a clean presence / absence database for peregrine falcons; 2) Compile a diverse but consistent database of potential predictor factors; 3) Through iterative

C1

scoring (weighting) and modelling, develop a predictive indicator of falcon occurrence; and 4) Validate the predictive model by comparison with independent falcon distribution data.

Thinking of subsequent research users, this work should give them enhance capabilities to:

a) repeat the same analysis for a different species, avian or terrestrial? b) from these data, test alternative hypotheses or outcomes for falcons? c) from this basic outline and recipe, try to update the case for falcons using more recent predictor fields?

Thinking of resource managers, this work should provide them some tools to:

a) better monitor falcon populations and distributions? b) develop better or alternate conservation strategies?

(I suppose these researchers and resource managers constitute the “global public” in the title?)

Whatever their purposes, these subsequent readers and users should be able to take advantage of and rely on:

a) the clean presence / absence data (perhaps in two versions, one without and one with the pseudo-absences); b) the compiled and georeferenced predictor layers (perhaps a users could select some subset of those layers); and c) a complete recipe with links to all resampling, extraction, thresholding, etc. software tools (contained in this manuscript?).

Authors provide link to very good ScholarWorks site at University of Alaska. There one finds all the predictor layer files individually, plus a reference table (same as Appendix A) and a Python code. But the presence / absence files from GBIF seem missing? And, although the url references a ‘handle’ identifier system, this reader understands ‘handles’ as allowing changes to metadata or data covered by that identifier? A reliable location but flexible content? Not sure how the ScholarWorks system fits the DOI

C2

requirements of the ESSD journal? Authors or editors will know more about this than this reviewer? Should we have a separate more-permanent snapshot of the essential files, saved as a backup somewhere else?

How does a user get access to the presence / absence data? Back to GBIF for each use? But the authors describe (section 2.1) substantial reprocessing of GBIF data, to remove half of the raw data. A user needs access to that product.

The authors have provided a useful snapshot. But, as they describe in their text, the falcon populations, their relative health, and the global environment all evolve. The authors should perhaps make more emphasis of the transitional nature of these predictors. They make valid points about urbanisation, but based on quite old data and with little sense of what next? They do comment on conservation options, but mostly in retrospection? What new or different information would users get if the authors used more up-to-date source files. For example, economic and demographic data, mostly from a year 2000 publication (Newsweek / ESRI) which reports data from late 1990's. Climate data from well before 2005? Some data layers come from more recent publications, e.g. 2013. If, as I suspect, the authors have provided a remarkable product by assembling the best or most available - to them - data sources, they need to justify this in the text and to then comment about how they or others might update all or specific layers with fresher informations (e.g. new land use data sets from ESSD). This reader also wonders about the extensive incorporation of light, radiation and clouds in the predictor layers when only night-time light and September solar radiation turned out to have predictive value?

If interpolating any of these layers to 1 km by 1 km spatial resolution was easy, many of the sources would already have performed that downscaling. Here, authors describe a 'push-the-button' sort of downscaling technique, that takes all layers from original resolution to 1 km by 1 km. Apparently this happened through a 'create fishnet' tool in ArcMap? Users need more information about this downscaling step. For example, when this reader searched on '1 km' among the ESSD data sets, I found 1 km DEM for

C3

Antarctica, 1 km satellite source data for lake and river water temperatures in Europe and 1 km projections for rainfall in UK. One could imagine and hope that some future researchers would want to compare the UK 1 km rainfall patterns with the 1 km predictor fields produced here and to then compare the falcon distributions data likewise. But have these authors given those future users enough information to understand these downscaled data?

To my list of four purposes above, the authors seem to have added a fifth:

5) Evaluate and discuss national compliance with CBD conventions and data contributions to GBIF. That discussion, although valid, does not add to this manuscript and perhaps belongs elsewhere?

Several small comments but too hard to submit those without page numbers as reference. Reviewer or editor can address those in a subsequent version.

Interactive comment on Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-65, 2017.

C4