Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2016-65-EC1, 2017

Earth System Science Data
Open Access Discussions

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

D. J. Carlson (Editor)

ipy.djc@gmail.com

Received and published: 19 July 2017

The manuscript ESSD-2016-65, addressing global distribution of peregrine falcons, presents one of the most difficult decisions this journal has made.

The authors have assembled data from many sources. They have made all their sources explicit and all the data, databases, tools and algorithms reliably and openly accessible. In this openness they set a high standard that resonates with the ideals of this journal. They push the boundaries of ecology and indeed of earth system science.

Their work crosses a wide range of disciplines and thereby challenges the evaluation processes of this journal to provide competent reviewers and reviews across that range of topics.

Despite this breadth, this work has unfortunately failed to assure potential users of its fundamental quality. In particular it has not adequately addressed two concerns: that the distribution data themselves reflect a substantial observational bias and that the environmental data, particularly the meteorological data, do not support utilisation at 1 km resolution.

Reviewer 1 raised the resolution issue. The authors have not responded to this issue. As this editor knows, and as many papers in this journal demonstrate, the interpolation of any global environmental data to 1 km resolution remains far beyond the reach of the research community. In this journal one can find 1 km data for rainfall over the UK or very high resolution aerosol data over Germany, but no global variable of any type or source at 1 km. ECMWF, a leading global centre for atmospheric reanalysis, confronts severe validation challenges as it prepares global reanalyses at 11 km resolution. The authors fail to provide any rationale or validation for their interpolations. By this failure they invite, unfortunately, disdain for or dismissal of their work.

The issue of bias in detectability arose from two comments submitted during the open discussion. Although the authors submitted a reply, this editor feels that their reply took a somewhat orthogonal direction from the original concerns. One of the researchers who submitted an open comment has responded to my request for further comment as follows:

"Thank you for allowing me to respond to the author's comments. In my opinion, this is not a question of data-mining versus avian ecology, but an inappropriate application of data mining. My main criticism of the paper is that the peregrine locations are biased by country-specific differences in reporting rates. This bias in the location data results in country-specific socioeconomic variables being incorrectly influential in the model and

therefore, the resulting prediction map is also suspect. I do not believe the author's response adequately addressed this criticism.

The authors state "that cannot be so correct because we use best-available data worldwide" and "these are sites where the falcons where actually seen, presence; unbiased". Clearly, these assertions are not adequate proof that the locations are unbiased. Data can be the best-available and represent actual locations and still be biased by factors such as differential reporting in certain countries and areas. In fact, the manuscript states that certain countries had low reporting rates and a comparison of their map to other data sources indicate that these countries with low reporting rates also have erroneously low predicted levels of peregrine occurrence.

The authors claim their data mining methods are not sensitive to issues of biased data, but the only basis they give for this claim is citing Kadmon et al. (2004) but Kadmon et al. (2004) does not support their assertion. Kadmon et al. (2004) looks at vegetation in Israel to determine if the fact that much of the data was collected near roads changes the results of bioclimatic models. They conclude that the effect of the bias in data collection is present but small because, in Israel, habitat conditions near roads were similar to habitat conditions away from roads. So, they do not assert that the effect of biased data can be ignored; they examine the underlying data and conclude that the effect of this bias is low in their specific situation. Their bioclimatic models do not however, include distance to road as a variable because that would obviously be affected by roadside bias in the data collection and magnify the effect of that bias. This is however, essentially what this peregrine manuscript does; it uses data with a country-specific reporting bias and then models the results with country-specific predictors.

There is no basis to support the author's claim that these machine learning techniques are impervious to problems of bias in the reported locations. The bias can be minimised with careful model selection, but this paper makes no effort to do so. Without evidence that they overcame the bias in their data, the conclusions are very suspect and this paper should not be published. Perhaps an independent statistical expert can

be consulted on this question.

I also stand by my other criticisms of this manuscript: 1) their testing data does not actually fit their prediction map, the fact that 98% of testing data falls in areas of predicted peregrine use is not informative if 98% of the area is predicted peregrine range; 2) peregrines select for different factors for nesting, wintering, and migration and these cannot and should not be modelled with a single model; and 3) their map does not match independent maps of peregrine distribution (the authors misunderstood this criticism and only stated that Figure 3 and Figure 7 are similar)."

I thank that researcher for taking time to again address both the manuscript and the author responses.

I address these two issues in this comment to ensure an open process even in the most difficult cases. I have taken an editorial decision to reject the manuscript.

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