While the idea to develop a global map of peregrine falcon distribution is a worthy goal, in my opinion, this paper has several major problems that make the results unreliable without major revisions.

**The source data are non-representative**

The data used in the analysis were collected from a variety of sources in a non-random way that leads to a dataset that is not representative of the entire distribution of peregrine falcons worldwide. Peregrines were more likely to be observed and reported in certain countries and certain locations. Because no effort appears to be made to account for the potential biases these data can introduce, this paper does not actually model the distribution of peregrine falcons. Instead, this paper models the distribution of peregrine *sightings that have been reported to GBIF*, which is a potentially very different distribution from the actual global distribution of peregrines.

By comparing data collected non-randomly from certain countries with random locations spread across the globe, the authors created a situation in which country-specific socioeconomic factors can erroneously be given explanatory power, whereas environmental variables that biologists know to be important in affecting peregrine distribution are not included. We know peregrine falcons do not select areas based on human infant mortality rates. Although it is possible that there is a strong unexplained spurious correlation with high predictive power between the two variables, a more plausible explanation is that peregrine falcons are more likely to be observed and reported in countries with low infant mortality rates. Therefore, it seems likely that the country-specific socioeconomic factors dominating the model actually are related to levels of reported sighting rates rather than to global occurrence. Many of the areas they determine have low probability of occurrence are remote areas with low human populations and, hence, little opportunity for online reporting. The authors discuss the fact that reporting varied dramatically among countries but do not explain how they dealt with this issue. The countries that they identify as lacking in data ("Russia, China, Brazil, and some African nations") correspond well with the few areas they identify as lacking peregrines (Figure 7), even though there is ample evidence that peregrines nest widely across Russia and into China. This disparity provides more evidence that these country-specific differences in reporting were erroneously carried over into their distribution map.

**The resulting model does not appear to fit the testing data well**

Because the authors use a data-mining technique with 104 variables, they are able to adequately fit a model to the training data. However, because they do not deal with underlying limitations of the data or use appropriate biological variables known to be important to peregrines, the resulting model performs poorly with the testing data (Figure 9). Although the authors claim a 98% accuracy rate, Figure 9 really shows a poor fit to the predicted model (a good fit would look like Figure 6). Although availability is not shown on this histogram to assess the fit properly, the data appear to perform little better than random points (i.e., if 2% of the land mass is predicted to be unoccupied, then 2% of random points would be expected to fall in that area by chance). Even with the very low threshold for a location's being considered occupied (>0.01 predicted index), the testing data still cross several areas not expected to have peregrines based on their map (Figure 8). In addition, the measure of model fit needs to be

modified to take availability into account (e.g., if 98% of the area is identified as having peregrines, even random points would be expected to have a 98% accuracy rate using their metric).

**The predicted map in Figure 3 does not appear to correspond well with some other smaller-scale peregrine distribution maps**

Figure 3 should be compared with existing maps of peregrine falcon distribution, where available, to gain some sense of how well the models actually work. Distribution maps synthesize expert opinion and multiple data sources to summarize local distributions. If this new map differs greatly from other maps based on a variety of sources, that difference should be explained by the authors. Table 5 attempts to compare the map to a previous description of peregrine distribution in Russia, but it does not give enough information to compare the modeled distribution adequately with the previously described distribution. The first comparison is too general and deals with population density, not occupancy, the second comparison was done at a spatial resolution different from that of the model, and the third one does not support the modeling results. Peregrine falcons do migrate across much of the area in Russia identified as not used (Dixon, A., A. Sokolov, and V. Sokolov. 2012. The subspecies and migration of breeding Peregrines in northern Eurasia. FALCO 39), again suggesting that the model does not fit independent data.

**Combining different types of data**

Although the composition of the testing data is not adequately described, these data appear to include breeding, non-breeding, and migratory locations. How can those types of data be fit with the same model when birds are selecting for very different conditions at different times of their annual life-cycles? A bird flying over the landscape during migration is selecting for very different variables than a bird selecting a nesting location; consequently, lumping these life-history stages in analyses is unlikely to uncover useful patterns and will likely result in incorrect results. To be accurate and useful, three different maps (breeding, non-breeding, and migration) should be developed, analyzed, and combined.

**Other comments/questions**

- The justification for using the 0.01 cutpoint for determining occupied pixels is not adequately explained. This cutpoint seems to be excessively low and, therefore, would be expected to erroneously include many areas that actually are not used.

- Because the authors claim that peregrines select areas with large populations and high light pollution, it is not surprising that national parks do not contain the modeled habitat. However, we do know that many national parks contain peregrine falcon populations and that peregrines use areas without light pollution over much of their range. This latter point is especially true at high northern latitudes.

- Check Figure 10, a visual comparison Figure 7 to a map WDPA protected areas suggests that many protected areas do appear to be in areas with predicted occupancy by peregrines.

- Give units for pixel size column in Appendix A.

- Although the authors say they use a 1 km x 1 km grid for all of their variables, the text should make clear that many variables are country-specific variables.

- How can the same model be fit to mean October and November temperatures in both the Northern and Southern hemispheres?

- Antarctica was removed from the analysis; similarly, the Greenland ice sheet is not potential habitat and also should be removed from the analysis.

- The methods say that 60,261 presence points were used, but Table 6 says there were 578,256 presence points in GBIF just for 11 selected countries. What explains this discrepancy? Please explain specifically which data were selected and how they were screened.

**Conclusions**

Data-mining has many useful applications, but it still requires unbiased data (or existing biases have to be adequately accounted for), the explanatory variables chosen have to explain the underlying processes and be relevant to them, and the resulting model needs to show good agreement with independent data (testing data and previous distribution maps). As currently written, this paper does not meet those three criteria.