



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

Sumithra Sriram¹, Falk Huettmann²

¹College of Computing, Georgia Institute of Technology, Atlanta Georgia 30332, USA ²EWHALE lab, Inst. of Arctic Biology, Biology & Wildlife Dept., University of Alaska Fairbanks (UAF), Fairbanks Alaska 99775, USA

Correspondence to: Sumithra Sriram (sumithrasriram@gatech.edu)

Abstract. Peregrine falcons (*Falco peregrinus*) are among the fastest members of the animal kingdom, and they are probably the most widely distributed raptors in the world; their migrations and habitats range from the tundra, mountains and some deserts to the tropics, coastal zones and urban habitats. Habitat loss, conversion, contamination, pesticides and other anthropogenic pressures are all known factors that have an adverse effect on these species. However, while peregrine falcons were removed from the list of endangered species due to rebounding populations linked with the DDT ban in many nations of the world, no accurate global distribution models have ever been developed for good conservation practice and in an open

15 access data framework.

Here we used the best-available open access peregrine falcon data from the Global Biodiversity Information Facility (GBIF.org) to obtain the first publicly available global distribution model for peregrine falcons. For that purpose, we compiled over a hundred high resolution global GIS layers (1km pixel size) that incorporated various variables such as biological, climatic, and socio-economic predictors allowing to analysis habitat relationships in a holistic fashion and to

- 20 build a generalizable model. These value-added layers have also been made available by us for the global public, free of charge, for further use and consumption in any modeling effort wanted (https://scholarworks.alaska.edu/handle/11122/7151). We created data extraction explicit in space and time also with an open source python script tool as well as with ArcGIS (via the GUI) on a PC. The obtained data cube (global, 1km pixel, 104 GIS layers) was 'mined' with the Salford Predictive Modeler (SPM) software suite, which offers one of the best platforms for data mining, to build the prediction model for
- 25 robust inference. We found that peregrine falcons are widely urbanized occurring in coastal areas and also associated with riparian zones. This is the first model ever obtained using 104 predictors on a 1km scale predicting the potential ecological niche of falcons around the world. While our model might show uncertainty for parts of Siberia, Russia, it has an assessed global accuracy of over 95% and hence provides the currently best possible public available global prediction model for peregrine falcons, based on all available empirical data. Overlaid with the national parks of the world we found that most
- 30 peregrine hotspots are actually located outside of protected areas warranting more protection efforts while global change unfolds. Finally, a nationwide assessment of the presence points taken from GBIF allows for insight as to the many signatory





nations that are still in violation of the open access data sharing requirement set by the Convention of Biological Diversity (CBD) and the Budapest and Berlin Declaration.

1 Introduction

- 5 Predictive modeling explicit in space and in time has been used in ecology to generate distribution patterns for thousands of species to help model their ecological niches (Guisan and Zimmermann, 2000; Barry and Elith, 2006, Drew et al. 2011). These niches rely on data used and available. They are defined by a set of biotic and abiotic factors, and various methods have been explored for inference and the identification of relevant predictors (Breiman, 2001; Hastie et al., 2001; Peterson, 2001). Understanding these defining elements and concepts is important for successful conservation and species management, especially now, with the rapidly changing climate (Walther et al., 2002), shifting biomes and a large multitude
- of human impacts (Halpern et al., 2008). Species distribution modeling, both temporal and spatial, plays an important role in monitoring, managing and conserving

species distribution modeling, both temporal and spatial, plays an important role in monitoring, managing and conserving species effectively (Cushman and Huettmann, 2010; Drew and Perera, 2010). It is a convenient and effective method of dataanalysis, where identification of known ecological niches of species provides an insight into the presence or suitability of

- 15 presence in remote (un-sampled) areas. For example, the fundamental niche was modelled for three quintessential Arctic bird species, and distribution patterns spanning over 200 years (from 1900-2100) were designed to predict the suitability of survival and changing spatial concentration corresponding to the changes in climate (Booms et al., 2011). By now, there have been many such distribution models that have put forth the ecological niches for various species over the years (Drew et al. 2011). Generally, the models built deal with the estimation of the fundamental ecological niche of the species in
- 20 question. This can be achieved using either a mechanistic approach or a correlative approach (Soberon and Peterson, 2005). The mechanistic approach involves the physical modeling of the direct response of individuals and their metabolism, etc. to parameters such as temperature and humidity, and then using GIS to identify regions of positive fitness. The correlative approach, followed in this study, deals with using various predictor variables and building a predictive model using various supervised machine learning algorithms. On a global level, this is rather powerful because 'global correlates' can be
- 25 identified, pursued and tested further. These correlations are not biased but are very powerful for inference. Using many data are therefore essential.

The ecological niche has been used and classified in many ways (Soberon and Nakamura, 2009; Cushman and Huettmann, 2010) - the Grinnellian niche of a species is determined by its habitat and its behavioral adaptations; the Eltonian niche is classified according to the foraging activity of the species; the Hutchinsonian niche, which is the most generic form of niche,

30 takes into account the various diverse environmental conditions and resources the individual requires to survive (Bruno et al., 2003). The range of such biotic and abiotic conditions that define the requirements of survival of a particular species is the fundamental niche. The complete set of locations that fit the requirements of the fundamental niche is the potential niche,

Searth System Discussion Science Signate Data



and the set of locations where the species is actually found is the realized niche. Most models deal with the prediction of potential niches. Such modeling, however, has not been so prevalent with species that are as wide spread as Peregrine falcons. Global analysis platforms and computational solutions do not exist yet to deal with such large data, worldwide and on a 1 km pixel size, all as open access and open source for fast analysis. While the constraints were always put on species data, here we try to promote the opportunities on the 104 GIS predictors as open access and open source code to actually

5

1.1 Biology of peregrine falcons

operate such a data cube effectively, e.g. on a local PC.

Peregrine falcons, though known for their widespread habitat range and adaptability, also migrate long distances connecting the winter with their nesting areas. Some of these nesting areas have been in use for over hundreds of years, and probably

- 10 longer (Newton, 1979). Peregrine falcons can be classified into 19 subspecies depending on their geographic locations (Cade and Digby, 1982). Table 1 illustrates these known subspecies with their corresponding regions of occurrence. Poaching and hunting of this species has been ongoing for millennia, e.g. in falconry; perhaps it was somewhat sustainable even. But in the mid-20th century, the peregrine falcons were critically endangered globally, and were even close to extinction in North America, due to the excessive use of DDT and other chemical pesticides that led to their death or
- 15 reproductive failure due to the thinning of their egg shells (Newton et al., 2008). From the eventual ban on the use of organochlorine pesticides onwards, and with widespread reintroduction of these species and protection under various national and international legislations, they have since made a strong recovery in many parts of the world (Tordoff and Redig, 2001; Jacobsen et al., 2007). Known for their speed and broad geographical availability due to their adaptability, they are probably the most frequently used raptors for falconry. The detrimental effect of the pesticides aside, this aspect of
- 20 human pursuit has made these falcons more vulnerable, being pursued and poached by egg collectors and falconry thieves alike with the ever-present demand in the Middle-east. Some of the international protection policies that have been established to ensure the protection of these species have been listed in Table 2. Apart from international regulations, individual countries have declared their own additional laws that protect these species from harm, a few of which are listed in Table 3.

25 **1.2 Conservation efforts**

The Convention of Biological Diversity (CBD) plays a central role in modern times also using digital opportunities. It is therefore considered here in more detail. Specifically it deals with online data aspects of biodiversity conservation affecting world-wide conservation management. CBD is an important multilateral treaty signed for now by 196 parties from around the world for sustainable development. One of the key points still discussed at the 10th Conference of Parties (COP), held in

30 Japan, is the issue regarding sharing of data on biodiversity (Balmford, 2005). Often, the areas that are richest in biodiversity are also the ones that lack the resources for conservation, and enough data for analysis often is unavailable to make good decisions. Hence it is important for scientists handling databases, in public office and such funding, to make this data

Searth System Discussion Science Signate Data



available for all users and researchers around the world to help build robust and collaborative methods of conservation that will ensure holistic benefits (Huettmann, 2011; Resendiz-Infante and Huettmann, 2015). In this study we also examine the empirical data that is available for peregrine falcons, when placed against the predictive models obtained, in order to find the countries that are in good compliance with this agreement of data sharing.

- 5 By now, the peregrine falcon is one of the best known examples for 'synurbization', the adaptation of wild animals to the rapidly invasive urban conditions, since its reintroduction (Luniak, 2004). The increased number of urban pigeons presents a central role in this discussion as prey species and in populated settings. Modeling, predicting and studying the distribution pattern of such a global and adaptive species can give useful insights into ecological aspects such as the effects of globalization on biodiversity and wilderness habitats. So far, one will find several distribution maps put forth by various
- 10 organizations, showing generic but often conflicting habitat regions for these birds, none of which really carry relevant and compliant metadata, scientific accuracy metrics, are not repeatable and are not available for a repeatable scientific assessment in a useable GIS format (Huettmann, 2004; Huettmann 2015b; Zuckerberg et al., 2011). The global distribution patterns of peregrine falcons have not been studied in detail for conservation purposes since its

removal from the U.S. list of endangered species in 1999 and the rebound due to DDT ban and breeding programs. Modern

- 15 study methods have not been employed, yet. In this study, we investigate steps to achieve the first global distribution model for peregrine falcons, using over a hundred compiled open access predictors that include climatic, biological and socioeconomic factors to represent a more holistic set of factors that can have an effect on the survival and suitability of the species in the region. We consider the species – *Falco peregrinus* (Taxonomic Serial No.: 175604), which encompasses all the subspecies, to build a general global niche (whole year round). Further, here we try to present a software open source
- 20 analysis platform done in python code for such analysis cases for generic uses of this data cube readily to be used by the global public for their own purposes. We believe this is a rather large progress because such 104 data layers do not exist yet in a readily available GIS format as provided here. It allows to demand for best-possible holistic views in any habitat study.

2 Methods

2.1 Training data (presence and pseudo-absence)

- 25 We used the 'presence only' data for peregrine falcons from the Global Biodiversity Information Facility (GBIF.org). As per Convention of Biodiversity (CBD), it is the one-stop open access international data warehouse for species occurrence. Many nations confirmed their participation by signature and ministerial support. GBIF represents currently the largest known empirical data about peregrine falcons in the public realm, which includes information such as the geo-coordinates of the location, the date, and the organization that reported the record (e.g. a sighting or a specimen). This raw data had to be
- 30 filtered for accurate and duplicate records, for records with incorrect geo-referencing and for records with ambiguous data to finally obtain 60,261 unique presence points, less than half the size of the raw database. Once the presence points were established, we then plotted 35,800 evenly distributed random points of pseudo-absence over land masses (excluding





Antarctica), using the "Create Random Points" tool in ArcGIS. This was done to obtain a representative pseudo-absence data layer for the world.



Figure 1. Global distribution of Peregrine falcons (presence points)

- 5 Another major aspect of this study was the first-time compilation and public delivery of over hundred global GIS layers at a 1 km x 1 km resolution from various open source projects for use as predictor variables for such models. We followed the initial work by Ohse et al. (2010) and Herrick et al. (2013). The range of predictor layers used is presented in detail in Appendix Table 1 and they include the following:
 - Climatic data such as mean monthly temperatures, mean monthly precipitation, mean monthly solar radiation, and global aridity index
- 10
- Bioclimatic variables (bio 1 bio 19) as defined by WorldClim
- Digital Elevation models (DEMs) and other variables derived from them such as slope and aspect
- Variables pertaining to biodiversity such as species richness of birds, mammals, amphibians and plants, annual average potential evapo-transpiration
- Quantitative indices indicating the effect of humans on biodiversity such as the Human Influence Index (HII), Human Footprint level and Last of the Wild
 - Proximity measures to coast, rivers and roads, which were calculated using the Euclidean Distance tool in ArcMap
 - Socio-economic factors such as Gross Domestic Product (GDP), human population density and count, Infant mortality, literacy rate, life expectancy, and trade and night light pollution.
- Variables such as density of livestock population such as pigs and poultry were also included, which prove to be highly influential factors when dealing with species that live in close contact with humans.





5

These variables were gathered from various open source projects, were re-projected and re-sampled for better alignment (e.g. for coastline and for each other) allowing us to deliver this value-added data product. The global layers are available for open access in GeoTiff format (LZW compressed), in WGS-84 projection, with a resolution of 1 km x 1 km in our public repository dSPACE UAF library and Google Drive (available upon request from the authors). This dataset has a size of 37.5 GB. They can also be easily converted to/from ESRI grid and ASC formats. Using such a wide range of predictors helps us

to start explore and recognize the hidden but so far unknown but driving factors that influence the species. In predictive models, having a complete description of the ecological niches is essential and reduces uncertainty whereas parsimony fails (Elith et al., 2006; Guthery et al., 2005).

Usually, the compiled layers were loaded into ArcMap and then overlaid with the presence and pseudo-absence points compiled. Using the Extract Multi-values to Points tool on ArcMap, the appropriate values for all these layers at the aforementioned points were extracted, and this compilation of data was used for creating the distribution model. But here we developed a second and open source approach and making it available to the global public for their empowerment and to use these data more effectively for their own purposes: The extraction of values from this data cube can also be done using python and its supporting libraries. Python is rapidly becoming one of the most popular languages used for machine learning

- 15 and any advanced analysis (Harrington, 2012). The extensive libraries and packages available are programmed to do most of the 'heavy lifting' and provide efficient models and solutions, enabling users to concentrate on the problem at hand rather than the modeling specifics. It also gives users the powers to determine the predictors that are used to build predictive models (as per Leo Breiman 2001). The script, that is available for access for the global audience, can be used as a generic template to handle big data on small machines as well (IBM PCs here). This, when combined with other useful multi-
- 20 processing libraries can be used to scale up performance when run on the cloud or clusters, as needed for in-time applications for instance.

2.2 Modeling approach

We used primarily the TreeNet algorithm in SPM7 provided by Salford Systems Ltd (https://www.salford-systems.com/) to build the distribution model. We also tested RandomForest in comparison. These algorithms have been widely used for modeling by data mining of ecological data for conservation management (Craig and Huettmann, 2009). They are all known to generate highly accurate models for both regression and classification and are also pretty robust when dealing even with faulty data and outliers (Fernández-Delgado et al., 2014). SPM also gives the user the flexibility of controlling the parameters of the models. The classification models were trained to predict the relative index of occurrence (RIO) of peregrine falcons in any given region of the world using the presence and the pseudo-absence points with all attributes from

30 the data cube. We used the 'balance' class weight option to balance the unequal presence and pseudo-absence sample sizes and kept all others at 'default' (a setting known to perform very strong).





2.3 Display of prediction surfaces

Next, using the Create Fishnet tool in ArcMap, we generated a global layer with an equally spaced point lattice grid with a 1 km x 1 km resolution, bounded by the continental landmass. This was then overlaid with the hundred layers of predictors, and the technique that was used to extract the values of these variables to the presence points was also used to achieve the same with these points in the fishnet layer. This set of points was then 'scored' using the classification predictor model built

5 same with these points in the fishnet layer. This set of points was then 'scored' using the classification predictor model built in order to obtain the global distribution of the relative index of occurrence of peregrine falcons. We then used the Inverse Distance Weighting (IDW) tool in ArcMap to create the raster surface for the predicted RIOs of presence.

These values of RIOs were then extracted for the known presence points that were used to build the model. A frequency distribution histogram was obtained for the RIOs to show the range of indices that predict the presence of the species in the region, according to the model built. The error percentage of the model is then used to determine the cut-off threshold of the

indices to obtain a binary presence/absence prediction for peregrine falcons.

2.4 Accuracy assessment

The accuracy of the models obtained was assessed using the Relative Operating Characteristic (ROC) - Area under the curve (AUC) metric, as is commonly used (Pearce and Ferrier, 2009). The ROC consists of a graph of a binary classifier that plots

- 15 the true positive rate against the false positive rate. We assumed AUC scores less than to 0.7 indicate low accuracy, between 0.7 and 0.9 to indicate moderate accuracy, and scores higher than 0.9 for high accuracy (Swets, 1988). We also obtained another set of the few publicly available Open Access presence points from MoveBank (www.movebank.org) that we used to validate our models. Though there were over fifteen datasets that were listed for Peregrine Falcons, only two of them were available for public access, and none of them were shared with GBIF. Extracting the predicted RIOs at these points and
- 20 plotting their frequency distribution allowed us to examine the accuracy of our model for many regions. The workflow in its entirety is illustrated in a flowchart shown in Figure 2.











Figure 2. Flowchart illustrating the process of building and validating the distribution model

3 Results

5

This is the first investigation making 104 value-added predictor layers publically available. As an example and case study, we employed it to peregrine falcons. This is also the first global distribution model for peregrine falcons with over hundred predictors, based on empirical data and using data mining. We found that the species, being globally widespread, is adjustable to a varied range of environments. Using such a vast number of predictors allowed us to obtain as the best-possible general model and inference.



Figure 3. Global heat-map of the predicted presence of Peregrine falcons, showing the relative index of occurrence (RIO) from 0 to 1.

Based on a RIO, the global heat-map of the predicted distribution of peregrine falcons obtained from the model is shown in Figure 3. This agrees with the general consensus that the peregrine falcons are known to be found on all continents, except Antarctica (Jefferies and Hickey, 1969). Due to their high adaptability, they are found in a wide range of biomes, except the extremes of the Amazon, interior Greenland and deserts like the interior Sahara. This model was produced with a global





perspective, considering all the sub-species of the peregrine falcons together and all-year round, in order to establish a quantified common ecological niche description for their presence around the world.

3.1 Deliveries of data and GIS layers

One of the important contributions of this paper is the compilation and global sharing of over a hundred global GIS layers.
5 We therefore present it as a result and details. These predictors have a resolution of 1km x 1km, that was compiled from various open source projects (details shown in appendix table 1). These data are available for free access in a good format at (https://scholarworks.alaska.edu/). They are all compiled (re-projected and re-sampled) so that the rasters are aligned correctly, and can be directly used for any other model without changes. We provided metadata with those layers.

3.2 Model performance

10 Both the models built, using the Random Forests and the TreeNet algorithms, resulted in a ROC of greater than 97%. However, on comparison by us, we assessed that the TreeNet algorithm resulted in a model that was more realistic and based on the ground truth data we had at hand (see below, limited Movebank web portal data). The results shown in this paper are derived from this model.

3.3 Variable importance

- 15 The top predictors for the classification model, as shown in Table 1, made it apparent that the most relevant predictors that drive the suitability of the region to peregrine falcons world-wide are actually socio-economic factors. These have never been described before (Kaufman, 2000). The ecological niche of the peregrine falcon is determined by a multivariate set of predictors; it is clearly not parsimonious. The population count and night light pollution index indicate the intensity of urbanization in the area, while the infant mortality rate and life expectancy indicate the status of development in the region.
- 20 High October temperatures are also important. These are all new predictors that should be studied and interpreted further on a local scale. Taken together, peregrines seem to favor warm coastal urban areas with low infant mortality and high light pollution. The species richness of birds and the density of poultry in region are also subsequent predictors as falcons are birds of prey, and as this species sit on the apex of the food chain. The graphs for the dependence of presence on the top four predictor variables are shown in Figure 4. Arguably, on a global scale, the peregrine falcon links with urban areas much
- 25 more than with remote wilderness.







5 Figure 4. Partial dependence graphs for top 4 predictor variables. The RIO dependencies plotted against (a) Infant mortality rate, measured as the number of deaths of infants (under one year) per 1000 live births (b) Population count (in thousands) (c) Night light pollution, indexed from 0-255 as increasing intensity of lights (d) Mean temperature of the month of October measured here as (10 x OC) as given in WorldClim (Hijmans et al., 2005)

3.4 Validation of the peregrine model

10 The classification model obtained from the TreeNet algorithm had an AUC of 98%, showing this model with a highly accurate performance metric. Figure 5 shows the associated ROC curve obtained for the model. The predicted indices obtained from the model were extracted for the presence points, and the frequency distribution of these values, shown in Figure 6, was plotted as a histogram in R (https://www.r-project.org/). The normal curve obtained for this distribution shows that most of the points are correctly classified with a high value of predicted occurrence.







Figure 5. ROC curve for the model obtained using TreeNet algorithm.



5 Figure 6. Frequency distribution of predicted vs reality Peregrine Falcon occurrence for TreeNet model for training data

Given that there is a 2% error in the global model that was obtained using the TreeNet algorithm, we can argue that everything above the bottom 2% of the predicted probabilities of the presence points indicate the presence of peregrine falcons in the area. Using the frequency distribution histogram shown in Figure 6, and like done in Kandel et al. (2015) the threshold thus calculated was 0.01, which was finally used to generate the binary presence/absence distribution map as





shown in Figure 7. Models in machine learning do not have to be symmetrical because no logistic function is used. As a matter of fact, nature is never symmetrical nor linear or logistic and our models are based on 'recursive partitioning' trees (Breiman, 1984).



5 Figure 7. Interpreted best-available predicted binary Presence/Absence prediction map of peregrine falcons with a cut-off of 0.01

Though the use of pseudo-absence points can affect the strength of evidence obtained from the ROC curve of the model, we used a balanced weight in model building, and the prediction of presence is validated, as the 'data contamination' occurs only in the absence points considered. The 'balance weight' setting is a specific feature in SPM and makes it very powerful for 'real' data mining! The AUC hence gives a close estimate to the model that can be achieved when using true absence points (Barbet-Massin, et al., 2012). To assess this potential shortcoming, further validations are performed.

Unfortunately we were not able to locate or obtain other public data for peregrine falcons; they are usally not made freely available by peregrine falcon investigators. But we located two datasets from Movebank (otherwise a closed website), tracking peregrine falcons with geolocators in the western hemisphere and in Europe.

Confronting our model with real world data, for the validation of the model using ground truthing, we used this other set of 7,800 presence points that were independent of the points used to train the model for validation. These test points overlaid with the presence/absence threshold map proposed is shown in Figure 8. Extracting the predicted RIOs of these points results in the frequency distribution histogram shown in Figure 9. Using the assumed threshold of 0.01, we see that the only 141 points out of 7,800, are falsely classified, resulting once more into a 98% accuracy, which is in good consistence with the 2% error in the predicted model. Considering that these are large-scale model predictions such high global accuracies must

20 seem as highly remarkable, we find.







Figure 8. Presence/Absence map (with a cut-off of 0.01) overlaid with test presence points used for validation



Figure 9. Frequency distribution of predicted vs reality Peregrine Falcon occurrence for TreeNet model for test data

5

3.5 Efficiency of Peregrine Falcon Conservation

The predicted RIOs of this model were overlaid with a layer of global national parks (UNEP-WCMC, 2015), to assess the occurrence of these birds within the protected lands. We observed that less than 3% of these points were actually situated within national parks. These protected areas include regions on the west coast of America, as well as Alaska, England, South

10 Africa and Madagascar, a large concentration in south-east Australia, and a few in south-east Asia. The conservation and land protection efforts for peregrine falcons have not been much active since its removal from the red list in 1999. Though many hand-raised and released habituated birds have been rapidly adapting themselves to the increasingly urbanizing space,





more conservation is needed along their nesting sites to maintain their healthy population in the wild. By now, wilderness seems to play a small role for this species.

Next, the polygons that represent the national parks around the world were interpreted as sampling points at a 1 km x 1 km resolution. The predicted ROI (Relative Occurrence indices) obtained from the model was then extracted for these points to

5

access if the areas of high ROIs are included under the plots demarked for conservation purposes. The frequency distribution histogram for the above, shown in Figure 10, clearly shows that the global national parks currently in place, just cover areas where the ROI of peregrine falcons are extremely low. The normal curve obtained is starkly different from the one desired for a conservation that maximizes peregrine falcon coverage and protection. This analysis clearly shows that the current system of national parks provides little or no protection to the peregrine falcons, globally.



Figure 10. Frequency distribution of the national park points plotted against the predicted indices of the model

All our model layers and data are made publicly available with metadata. We invite everybody to use, assess and improve them further, e.g. with the open source code we provide also. Considering this is the first global model for this species we believe this presents a solid achievement, a baseline, conservation progress and a template and role model how public science can be delivered worldwide. We hope this catches on further for progress of this species, habitats, human well-being

15

10

and the world (Huettmann, 2011).

4 Discussion

This study is able to present the first global distribution model prediction of Peregrine falcons, based on open access data, open source code, and data mining with machine learning using more than a hundred diverse predictors for global consumption. Choosing a few predictors *a priori* and restricting the range to a small range of commonly used climatic and

20 consumption. Choosing a few predictors *a priori* and restricting the range to a small range of commonly used climatic and biological variables is a common practice in species distribution modeling, albeit not making good use of all its potential and





5

opportunities yet. Starting the model unbiased and involving as many predictors as possible is however important, as the model might identify various unexpected influencing factors that might not be apparent to the modeler and expert. The model thus builds quantitative maps and predicts the global distribution pattern of Peregrine falcons, with highest possible accuracy (97% in most cases and assessments). Though the model has a high accuracy rate, we acknowledge that the predicted regions of presence are 'just' the year-round potential niche of the species. This might vary slightly from the realized niche, as predation, prey, disturbance, as well as intense urbanization, climate change and other anthropological factors have already disturbed the delicate balance. Further fine-scale assessments and re-runs should be performed at a smaller scale to obtain an accurate prediction of the realized niche as needed and for more local studies. Here we provide a true macro-ecology perspective but which allows for subsequent study at smaller scales as needed.

10 4.1 Model Uncertainty

The binary distribution pattern obtained (Figure 7) agrees with the general assumption that the falcons are globally widespread and can be found in all kinds of ecosystems. Exceptions are the particularly extreme ones, like the interior tropical rainforests in South America and the Sahara in Africa. Looking closely, it can also be seen that areas close to the coast and along rivers are consistently predicted as hot spots, specifically in the Siberian and the Amazonian regions. This

- 15 can be attributed to the specific layers that define the proximity to the coast and the rivers, and it provides there good knowledge (peregrine falcons are known to breed in and to use those areas). The pattern predicted around Russia however, probably remains less certain. Though the presence of peregrine falcons is a known fact around the area, there is no publically accessible data for Russia available for verification or to better the model built. When the presence/absence map is overlaid with the test presence points used for validation (shown in Figure 8), it can
- 20 be seen that the incorrectly classified points occur in the Amazonian region as well as in the far west corner of Russia. Since there were no publicly available presence points in Russia to validate our model with, we used the observations recorded by Rogacheva (1992) in her book "The Birds of Central Siberia" to compare and contrast our predicted potential niche in that region (shown in Figure 11). Some of the observations recorded in the book match well with the predictions made by the model. These are described in Table 5.







Figure 11. Presence/Absence binary map with a cut-off of 0.01 for the Russian region

We conclude that Russia, a GBIF member nation, is still not sharing many data on peregrine falcons (specifically for Siberia and Russian Far East) with the global community, but that our model so far predicts in good agreement with Siberian bird references.

4.2 Open access sharing of data

5

On analyzing the presence data that is available for open access on GBIF and MoveBank, it is apparent that there are a few countries such as Russia, China, Brazil and some African nations that have not really volunteered their data for public access. We find this is in violation of the data sharing requirements as set by the Convention of Biological Diversity (CBD)

- 10 and its signatory nations (such as Russia, Brazil and many African nations). This unfortunately also includes bird banding records, e.g. as maintained by EURING and BirdLife International. The availability of such data is very important though from the conservation perspective, as ecologists and organizations working towards conservation management need all possible data at hand to make suitable decisions. In North America for instance, such data are usually made readily available (Huettmann pers com.; Beiring, 2013).
- 15 Conservation agencies are truly mandated to contribute to those efforts for sound management of a public resource but fail. Other than GBIF, we have almost no data to use for those species. Table 6 shows the list of the top ten contributors to the open-access presence data of peregrine falcons that was obtained from GBIF. Though all the countries are CBD signatories, the list makes it apparent that such biodiversity data is prevalently available for developed countries that have the resources to make conservation a possibility. The countries with the most diversity are under-represented in these databases, and there
- 20 is a dire need for data sharing and open access of this data to build more robust models to help make decisions for conservation (Huettmann, 2015a; Huettmann, 2015b). It is a clear reflection of the North-South gradient described by (Rosales, 2008).





The study allows for a first comprehensive overview about the current distribution pattern of the peregrine falcons around the world. The model developed is robust and accurate, as shown by the validations performed. It is meant to be tested, used and improved further by any investigator of interest. The conservation status of these falcons, when assessed with the national parks, shows clearly the lapse in the conservation measures taken to protect this particular species. The ever increasing process of urbanization has made these raptors to adapt to living side-by-side with humans. A wilderness species became urbanized, worldwide. While one might easily argue that peregrine falcon populations are on the way up, they are arguably not really wilderness icons (anymore) nor does the future outlook look impressive, e.g. more people and climate change on the rise while the wilderness resources are widely used up by now. Wild peregrine falcons, as we initially know

them for millennia, might actually be gone by then.

10 References:

5

15

Anon: Check-list of North American birds, American Ornithologists' Union, New York., 1910.

Balmford, A.: ECOLOGY: The Convention on Biological Diversity's 2010 Target, Science, 307(5707), 212–213, doi:10.1126/science.1106281, 2005.

Barbet-Massin, M., Jiguet, F., Albert, C. H. and Thuiller, W.: Selecting pseudo-absences for species distribution models: how, where and how many?, Methods in Ecology and Evolution, 3(2), 327–338, doi:10.1111/j.2041-210x.2011.00172.x, 2012.

Barry, S. and Elith, J.: Error and uncertainty in habitat models, Journal of Applied Ecology, 43(3), 413–423, doi:10.1111/j.1365-2664.2006.01136.x, 2006.

Beiring, M.: Determination of valuable areas for migratory songbirds along the East-Asian Australasian Flyway (EEAF), and an approach for strategic conservation planning. Unpublished Master Thesis, University of Vienna, Austria, 2013.

Booms, T., Lindgren, M. and Huettmann, F.: Linking Alaska's Predicted Climate, Gyrfalcon, and Ptarmigan Distributions in Space and Time: A Unique 200-Year Perspective., Gyrfalcons and Ptarmigan in a Changing World, doi:10.4080/gpcw.2011.0116, 2011.

Breiman, L.: Classification and regression trees, Wadsworth International Group, Belmont, CA., 1984.
Breiman, L.: Machine Learning, Machine Learning, 45(3), 261–277, doi:10.1023/a:1017934522171, 2001.
Bruno, J. F., Stachowicz, J. J. and Bertness, M. D.: Inclusion of facilitation into ecological theory, Trends in Ecology & Evolution, 18(3), 119–125, doi:10.1016/s0169-5347(02)00045-9, 2003.

Cade, T. J. and Digby, R. D.: The falcons of the world, Comstock/Cornell University Press, Ithaca, NY., 1982.

30 Craig, E. and Huettmann, F.: Using "Blackbox" Algorithms Such AS TreeNET and Random Forests for Data-Ming and for Finding Meaningful Patterns, Relationships and Outliers in Complex Ecological Data, Developing New



Methodologies Through Pattern Discovery and Recovery Intelligent Data Analysis, 65-84, doi:10.4018/978-1-59904-982-3.ch004, 2009.

Cushman, S. A. and Huettmann, F.: Introduction: Ecological Knowledge, Theory and Information in Space and Time, Spatial Complexity, Informatics, and Wildlife Conservation, 3–18, doi:10.1007/978-4-431-87771-4_1, 2010.

5 Dementiev, G. P.: On The Shaheen Falco Peregrinus Babylonicus., Ibis, 99(3), 477–482, doi:10.1111/j.1474-919x.1957.tb01961.x, 2008.

Döttlinger Hermann: The black shaheen falcon (Falco peregrinus peregrinator Sundevall 1837), its morphology, geographic variation and the history and ecology of the Sri Lanka (Ceylon) population, The Author, Schweitenkirchen., 2002.

Drew, C. A. and Perera, A. H.: Expert Knowledge as a Basis for Landscape Ecological Predictive Models, C.A. Drew,
 Yo. Wiersma and F. Huettmann (eds) Predictive Species and Habitat Modeling in Landscape Ecology, 229–248,
 doi:10.1007/978-1-4419-7390-0_12, 2012.

Ellis, D.H., Sabo, B.A., Fackler, J.K. and Millsap, B.A.: Prey of the Peregrine Falcon (Falco peregrinus cassini) in southern Argentina and Chile, Journal of Raptor Research 36.4, 315-319, 2007.

15 Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D: Do we need hundreds of classifiers to solve real world classification problems, J. Mach. Learn. Res 15.1, 3133-3181,2014.

Golicher, D., Ford, A., Cayuela, L. and Newton, A.: Pseudo-absences, pseudo-models and pseudo-niches: pitfalls of model selection based on the area under the curve, International Journal of Geographical Information Science, 26(11), 2049–2063, doi:10.1080/13658816.2012.719626, 2012.

20 Guisan, A. and Zimmermann, N. E.: Predictive habitat distribution models in ecology, Ecological Modelling, 135(2-3), 147–186, doi:10.1016/s0304-3800(00)00354-9, 2000.

Guthery, F. S., Brennan, L. A., Peterson, M. J. and Lusk, J. J.: Invited Paper: Information Theory In Wildlife Science: Critique And Viewpoint, Journal of Wildlife Management, 69(2), 457–465, doi:10.2193/0022-541x(2005)069[0457:itiwsc]2.0.co;2, 2005.

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'agrosa, C., Bruno, J. F., Casey, K. S., Ebert, C., Fox, H. E., Fujita, R., Heinemann, D., Lenihan, H. S., Madin, E. M. P., Perry, M. T., Selig, E. R., Spalding, M., Steneck, R. and Watson, R.: A Global Map of Human Impact on Marine Ecosystems, Science, 319(5865), 948–952, doi:10.1126/science.1149345, 2008.

Harrington, P.: Machine learning in action, Manning Publications, Shelter Island, NY., 2012.

30 Hastie, T., Friedman, J. and Tibshirani, R.: Unsupervised Learning, The Elements of Statistical Learning Springer Series in Statistics, 437–508, doi:10.1007/978-0-387-21606-5_14, 2001.

Herrick, K. A., Huettmann, F. and Lindgren, M. A.: A global model of avian influenza prediction in wild birds: the importance of northern regions, Veterinary Research, 44(1), 42, doi:10.1186/1297-9716-44-42, 2013.

Science Scienc



5

10

20

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, International Journal of Climatology, 25(15), 1965–1978, doi:10.1002/joc.1276, 2005.

Huettmann, F.: Constraints, suggested solutions and an outlook towards a new digital culture for the oceans and beyond: experiences from five predictive GIS models that contribute to global management, conservation and study of marine wildlife and habitat, in: Vanden Berghe, E. et al. (Ed.) Proceedings of 'Ocean Biodiversity Informatics': an international conference on marine on biodiversity data management Hamburg, Germany. IOC Workshop Report, 202, VLIZ Special Publication 37: 49-61, 2004.

Huettmann, F.: Serving the Global Village through Public Data Sharing as a Mandatory Paradigm for Seabird Biologists and Managers: Why, What, How, and a Call for an Efficient Action Plan, TOOENIJ The Open Ornithology Journal, 4(1), 1–11, doi:10.2174/1874453201104010001, 2011.

Huettmann, F.: A Short Presentation and Discussion of Digital Research Data Sets Central America, and for the Field Schools of La Suerte (Costa Rica) and Ometepe (Nicaragua), in Central American Biodiversity: Conservation, Ecology, and a Sustainable Future, edited by F. Huettmann, pp. 77–104, Springer, New York., 2015a.

Huettmann, F.: On the Relevance and Moral Impediment of Digital Data Management, Data Sharing, and Public Open
 Access and Open Source Code in (Tropical) Research: The Rio Convention Revisited Towards Mega Science and Best
 Professional Research Practices, in Central American Biodiversity: Conservation, Ecology, and a Sustainable Future,
 edited by F. Huettmann, pp. 77–104, Springer, New York., 2015b.

Jacobsen, F., Nesje, M., Bachmann, L. and Lifjeld, J. T.: Significant genetic admixture after reintroduction of peregrine falcon (Falco peregrinus) in Southern Scandinavia, Conservation Genetics, 9(3), 581–591, doi:10.1007/s10592-007-9373-4, 2007.

Jefferies, D. J. and Hickey, J. J.: Peregrine Falcon Populations: Their Biology and Decline., The Journal of Applied Ecology, 6(3), 514, doi:10.2307/2401519, 1969.

Kandel, K., Huettmann, F., Suwal, M. K., Regmi, G. R., Nijman, V., Nekaris, K., Lama, S. T., Thapa, A., Sharma, H. P. and Subedi, T. R.: Rapid multi-nation distribution assessment of a charismatic conservation species using open access

25 ensemble model GIS predictions: Red panda (Ailurus fulgens) in the Hindu-Kush Himalaya region, Biological Conservation, 181, 150–161, doi:10.1016/j.biocon.2014.10.007, 2015.

Kaufman, K.: Birds of North America, Houghton Mifflin, New York., 2000.

Longmire, J. L., Ambrose, R. E., Brown, N. C., Cade, T. J., Maechtle, T. L., Seegar, W. S., Ward, F. P. and White, C. M.: Use of Sex-Linked Minisatellite Fragments to Investigate Genetic Differentiation and Migration of North American

30 Populations of the Peregrine Falcon (Falco peregrinus), Experientia Supplementum DNA Fingerprinting: Approaches and Applications, 217–229, doi:10.1007/978-3-0348-7312-3_15, 1991.

Luniak, M: Synurbization–adaptation of animal wildlife to urban development, In Proceedings of the 4th International Symposium on Urban Wildlife Conservation, Tucson, 50-55, 2004.

Newton, I.: Population ecology of raptors, Buteo Books, Vermillion. S.D., 1979.

Earth System Discussion Science Signations



5

Newton, I., Bogan, J. A. and Haas, M. B.: Organochlorines and mercury in the eggs of British Peregrines Falco peregrinus, Ibis, 131(3), 355–376, doi:10.1111/j.1474-919x.1989.tb02785.x, 2008.

Ohse, B., Huettmann, F., Ickert-Bond, S. M. and Juday, G. P.: Modeling the distribution of white spruce (Picea glauca) for Alaska with high accuracy: an open access role-model for predicting tree species in last remaining wilderness areas, Polar Biology, 32(12), 1717–1729, doi:10.1007/s00300-009-0671-9, 2009.

Peakall, D. B. and Kiff, L. F.: Eggshell Thinning And Dde Residue Levels Among Peregrine Falcons Falco Peregrinus:A Global Perspective, Ibis, 121(2), 200–204, doi:10.1111/j.1474-919x.1979.tb04962.x, 2008.

Pearce, J. and Ferrier, S.: Evaluating the predictive performance of habitat models developed using logistic regression, Ecological Modelling, 133(3), 225–245, doi:10.1016/s0304-3800(00)00322-7, 2000.

Peterson, A. T.: Predicting Species' Geographic Distributions Based On Ecological Niche Modeling, The Condor, 103(3), 599, doi:10.1650/0010-5422(2001)103[0599:psgdbo]2.0.co;2, 2001.

Razafimanjato, G., Roland, L.-A. R. D., Rabearivony, J. and Thorstrom, R.: Nesting biology and food habits of the Peregrine Falcon Falco peregrinus radama in the south-west and central plateau of Madagascar, Ostrich, 78(1), 7–12, doi:10.2989/ostrich.2007.78.1.2.46, 2007.

15 Resendiz-Infante, C. and Huettmann, F.: Bird Conservation Status and Meaningful Socioeconomic Correlates in Central America: Results from an Open Access Data-Mining Approach for Parrots Using Machine Learning Indicate Serious Economic Problems, in Central American Biodiversity: Conservation, Ecology, and a Sustainable Future, edited by F. Huettmann, pp. 77–104, Springer, New York., 2015.

Rogacheva, H.: The birds of central Siberia, Husum Druck- und Verlagsgesellschaft, Husum., 1992.

20 Rosales, J.: Economic Growth, Climate Change, Biodiversity Loss: Distributive Justice for the Global North and South, Conservation Biology, 22(6), 1409–1417, doi:10.1111/j.1523-1739.2008.01091.x, 2008.

Soberon, J. and Nakamura, M.: Niches and distributional areas: Concepts, methods, and assumptions, Proceedings of the National Academy of Sciences, 106(Supplement_2), 19644–19650, doi:10.1073/pnas.0901637106, 2009.

Soberon, J. and Peterson, A. T.: Interpretation of Models of Fundamental Ecological Niches and Species' Distributional
 Areas, Biodiversity Informatics, 2, doi:10.17161/bi.v2i0.4, 2005.

Swets, J.: Measuring the accuracy of diagnostic systems, Science, 240(4857), 1285–1293, doi:10.1126/science.3287615, 1988.

Tordoff, H. B. and Redig, P. T.: Role of Genetic Background in the Success of Reintroduced Peregrine Falcons, Conservation Biology, 15(2), 528–532, doi:10.1046/j.1523-1739.2001.015002528.x, 2001.

U.S. Fish & Wildlife Service 1 800/344 WILD http://www.fws.gov May 2006
 UNEP-WCMC (2015). World Database on Protected Areas User Manual 1.0. UNEP-WCMC: Cambridge, UK.
 Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O. and Bairlein, F.: Ecological responses to recent climate change, Nature, 416(6879), 389–395, doi:10.1038/416389a, 2002.





Zuckerberg, B, Huettmann, F. and Friar, J.: Proper Data Management as a Scientific Foundation for Reliable Species Distribution Modeling. Chapter 3 In: C.A. Drew, Y. Wiersma and F. Huettmann (eds). Predictive Species and Habitat Modeling in Landscape Ecology. Springer, New York. Pp 45-70, 2011.





S No.	Subspecies	Taxonomic	Avibase ID	Region of occurrence
		Serial Number		
1	Falco peregrinus anatum	175605	6C7DCE6BBEF6D590	Throughout North America, from Northern Mexico to the tundra. Mainly found in the Rocky Mountains (Check- list of North American Birds, 1910)
2	Falco peregrinus babylonicus	824656	3AE2025EA8077AFB	East Iran and Mongolian Altai ranges (Dementiev, 2008)
3	Falco peregrinus brookei	824658	35E420F3FE11283C	Southern Europe
4	Falco peregrinus calidus	668605	03CD58057852EB3B	The arctic tundra region of Eurasia (Also known as the Siberian peregrine falcon)
5	Falco peregrinus cassini	668604	187F275EBF4473EC	Western South America (Ellis et al., 2007)
6	Falco peregrinus ernesti	824659	B18889654FF0B162	From Indonesia to the Philippines (Peakall and Kiff, 2008)
7	Falco peregrinus furuitii	824660	8FF6743395EEC5EC	Japan
8	Falco peregrinus japonensis	668606	A49EF0E2B7548301	North-east Siberia and Japan
9	Falco peregrinus macropus	824661	8786B805A9AE33B7	Tasmania and Australia (except the south-western part) (Peakall and Kiff, 2008)
10	Falco peregrinus madens	824662	DF92B0BBCD76E5AC	Cape Verde islands
11	Falco peregrinus minor	824r663	2723EFFCF7DB6CA2	Sparsely distributed in the sub-Saharan desert (Peakall and Kiff, 2008)

Table 1. The subspecies of peregrine falcons based on their geographic locations





_

12	Falco nesiotes	peregrinus	824664	3FE2CBE64F82F865	Fiji
13	Falco pealei	peregrinus	175606	15AB1C636B60F5D2	Pacific coast region of North America, including the Aleutian islands (Check- list of North American Birds, 1910)
14	Falco peregrin	peregrinus ator	824665	A7814DCB34C6D173	South Asia (Also known as the Indian peregrine falcon or Shaheen falcon) (Hermann, 2002)
15	Falco peregrin	peregrinus us	175609	F10DDB18594ACAA2	Europe and Northern Asia (Check-list of North American Birds, 1910)
16	Falco radama	peregrinus	824666	101DB29616B50A3D	MadagascarandComoros(Razafimanjato et al., 2007)
17	Falco submelat	peregrinus nogenys	<u>NA</u>	D2376241DC6822B8	South-west Australia
18	Falco tundrius	peregrinus	175608	F9F3B7A179D6F4DF	Northern tundra regions of North America, and the only known subspecies to nest in Greenland (Longmire et al., 1991)





Conservation policies	Comments
Convention on International Trade in Endangered Species (CITES)	This regulates the international trade of peregrine falcons and its subspecies
Memorandum of Understanding concerning the Conservation of Migratory Birds of Prey in Africa and Eurasia (Raptors MOU)	With the inclusion of UAE to Convention of Migratory Species (CMS) held at Bonn in 2016, the migratory falcons along the coastline in UAE now have an elevated conservation status
EC birds Directive (2009)	Apart from the protection of migratory species, this also specified the conditions under which hunting and falconry can be undertaken
Migratory Bird Treaty Act	It regulates the conservation of all migratory birds, including the peregrine falcons, in US, Canada and Mexico

Table 2. International laws protecting peregrine falcons





Conservation policies	Country	Comments
Endangered Species Conservation Act of 1969, USA (U.S. Fish & Wildlife Service, 2006)	USA	The predecessor of the Endangered Species Act of 1973. The species was then removed from the list of endangered species in 1999 after a successful reintroduction program
Banning of DDT by the Environmental Protection Agency (1972), USA (U.S. Fish & Wildlife Service, 2006)	USA	Carried under the authority of Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA)
The Peregrine Fund (1970) (http://www.peregrinefund.org/)	USA	Introduced in 1970 in the US when the peregrine falcons were at the brink of extinction, they have played a major role in the reintroduction of these species in North America
Wildlife and Countryside Act 1981, UK	UK	Offence is punishable by fines up to 5000 (pounds) and/or prison sentence of up to 6 months
Nature Conservation (Scotland) Act 2004	Scotland	Additional protection laws in Scotland
Red Data book of the Russian Federation (RDBRF)	Russia	Peregrine falcons feature in the red book that lists the endangered species in Russia
Species at Risk Act, Canada	Canada	This act provides legal protection for species included in the list
Individual State (and Northern Territory) Acts of Parliament, Australia	Australia	This act gives legal protection to all 32 raptors occurring in Australia, including peregrine falcons

Table 3. A selection of some local protection policies of Peregrine Falcons around the world





Variable		
Importance	Predictor	Score
Rank		
1	Infant Mortality	100.00
2	Population Count	55.39
3	Night Light pollution	47.28
4	Mean temperature of October	35.47
5	Koeppen Geiger	29.76
6	Species richness of birds	25.27
7	Cars	21.03
8	Mean temperature of November	20.51
9	Proximity to the coast	18.71
10	Globcover	17.83
11	Mean Solar Radiation of September	17.16
12	Proximity to roads	16.79
13	Human Influence index	13.41
14	Income per capita	12.26
15	Life expectancy	12.05

Table 4. Variable importance by percentage





 Table 5. Validation of the predicted model in the region around Russia by comparing and contrasting with the observations made by Rogacheva (1992) in the key reference "The Birds of Central Siberia"

S No.	Observations from "The Birds of Central Siberia"	Observations from the predicted model (Figure
	(Rogacheva, 1992)	11)
1	"Higher population densities of the Peregrine were	The model correctly predicts the two belts, one in
	always distributed in two wide belts, one in the north and	the north and the other in the south where higher
	the other in the south."	population for these falcons are commonly
		observed.
2	"The Peregrine requires open habitats richly supplied	The model predicts this characteristic of the
	with prey for hunting; therefore, it does not breed in	falcons, which can be seen by the predicted
	denser taiga and occurs in such areas only around large	presence along the rivers. This uses the
	lakes, along river valleys, and near large, open marshes."	"Proximity to rivers" layer that was used as one of
		the predictors.
3	"There are no peregrine falcons present in the Severnaya	The model classifies this correctly, except along
	Zemlya Archipelago."	the coasts, as the coastal areas generally boast
		suitable conditions for their presence.

5



 Table 6. Top 10 contributors for open-access presence data of peregrine falcons in GBIF and Movebank, ranked by

 the number of presence points contributed to the GBIF database. The surface area of the countries and the average

 ROI predicted are also listed for a more detailed assessment. (Russia was added for information due to relevance in

	the model discussion)						
			Number of unique	Number of open			
			presence points	access unique	Average ROI of		
Rank	Country	Surface area	reported in GBIF	presence points in	presence of		
	Country	(km ²)	(Number of	Movebank (Number	peregrine falcons		
			presence points per	of presence points	in the country		
			unit surface area)	per unit surface area)			
1	United States	9 831 510	228504	1792	0.53		
1	Sinted States	2,031,310	(0.024)	(0.0019)	0.55		
2	Norway	385 178	28631	2	0.47		
2	Ttorway	565,176	(0.088)	(0)	07		
3	Canada	9 984 670	28323	282	0.14		
5	Canada	J,J0 1 ,070	(0.0028)	(0)	0.14		
4	Australia	7,741,220	22585	0	0.33		
4	Australia		(0.0029)	.0029) (0)			
5	<u>Surviu</u>	504,781	7963	4	0.89		
5	Span		(0.016)	(0)			
6	Germany	357 170	5416	203	0.86		
0	Germany	557,170	(0.015)	(0.00057)	0.00		
7	United	243 610	5327	0	0.85		
1	Kingdom	243,010	(0.022)	(0)	0.05		
8	Mavico	1 964 380	5270	351	0.45		
0	MCAICO	1,904,300	(0.0027)	(0.000178)	0.43		
0	South Africa	1 210 000	4402	0	0.45		
7	South Annea	1,219,090	(0.0036)	(0)	0.45		
10	Franco	540.087	3264	407	0.81		
10	France	549,087	(0.0060)	(0.000744)	0.01		
33	Russian	17 008 250	288	653	0.017		
33	Federation	17,098,200	(0)	(0)			





Appendix A

Table 1. Index of the layers used as predictor variables (Type- Co: Continuous, Ca: Categorical)

Layer	Ty pe	Unit	File name	Original Projecti on	Original pixel size	Reference URL	Citation
Mean Tempe rature of each month (12)	Co	°C * 10	tmean_1 - tmean_1 2	WGS84	0.008333	http://www.worldclim.org/curre nt	Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.
Mean Precipi tation of each month (12)	Co	mm	prec_1 – prec_12	WGS84	0.008333	http://www.worldclim.org/curre nt	Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.
Solar Radiati on of each month (12)	Co	mm/da y	solrad1 - solrad12	WGS-84	0.025	http://www.cgiar-csi.org	Zomer RJ, Trabucco A, Bossio DA, van Straaten O, Verchot LV, 2008. Climate Change Mitigation: A Spatial Analysis of Global Land Suitability for Clean Development Mechanism Afforestation and Reforestation. Agric. Ecosystems and Envir. 126: 67-80.
Mean Cloud Cover	Co	Numbe r of days	cloud1 - cloud12	WGS84	0.008333	http://www.earthenv.org/cloud	"Wilson AM, Jetz W (2016) RemotelySensed High-Resolution Global CloudDynamics for Predicting Ecosystem and











Koepp en Geiger	Ca	_	kopgei	WGS84	0.5	http://koeppen-geiger.vu- wien.ac.at/shifts.htm
Cyclon e risk map	Co	1-10 risk factor	cy-clone	WGS84	0.0417	http://sedac.ciesin.columbia.edu /data/set/ndh-cyclone-hazard- frequency-distribution
Human Influen ce Index	Co	-	hii	GCS_Cla rke_1866	0.008333	http://sedac.ciesin.columbia.edu /wildareas/

Rubel, F., and M. Kottek, 2010: Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. Meteorol. Z., 19, 135-141. DOI: 10.1127/0941-2948/2010/0430.

Center for Hazards and Risk Research -CHRR - Columbia University, Center for International Earth Science Information Network - CIESIN - Columbia University, International Bank for Reconstruction and Development - The World Bank, and United Nations Environment Programme Global Resource Information Database Geneva -UNEP/GRID-Geneva. 2005. Global Cyclone Hazard Distribution. Frequency and Palisades, NY: NASA Socioeconomic Data Applications and Center (SEDAC). http://dx.doi.org/10.7927/H4CZ35 3K. Accessed DAY MONTH YEAR

Wildlife Conservation Society - WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data

GCS_Cla

rke_1866

WGS84

0.008333



Human

footpri

Last of

the

wild

ETOP

01

nt

Co

Ca

Co m

hf

ltw



and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4BP00 QC. Accessed DAY MONTH YEAR.

Wildlife Conservation Society - WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4M61 H5F. Accessed DAY MONTH YEAR.

Wildlife Conservation Society - WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Last of the Wild Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4348H 83. Accessed DAY MONTH YEAR.

Amante, C. and B.W. Eakins, 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS

etopo1	WGS84	0.01667	https://www.ngdc.noaa.gov/mgg /global/

0.008333

ia.edu/wildareas

http://sedac.ciesin.columbia.edu

http://www.sedac.ciesin.columb

/wildareas/



T.T......



NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M

Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Count. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4X63J VC. Accessed DAY MONTH YEAR. Center for International Earth Science

Information Network - CIESIN - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Density. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4NP22 DQ. Accessed DAY MONTH YEAR.

Newsweek Demographic Data. World Thematic Map Data Description. ESRI Website. 2000. http://www.esri.com/data/online/wothdata.ht

ml (23 March 2000).

Populat ion Count	Co	Numbe r of persons	pcount	WGS84	0.04167	http://sedac.ciesin.columbia.edu /data/set/gpw-v4-population- count
Human Populat ion Densit y	Со	persons /km2	pden- sit1	WGS84	0.008333	http://sedac.ciesin.columbia.edu /data/set/gpw-v4-population- density
Human Literac y Rate	Co		literacy_ rate	WGS84	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy





Human Life expecta ncy	Co	years	Lifeexpc tl	WGS84	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	NewsweekDemographicData.WorldThematicMapDataDescription.ESRIWebsite.2000.http://www.esri.com/data/online/wothdata.html (23 March 2000).
GNP Growth	Co	%	gnp	WGS84	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	Newsweek Demographic Data. World Thematic Map Data Description. ESRI Website. 2000. http://www.esri.com/data/online/wothdata.ht ml (23 March 2000).
Income per capita	Co	\$	incperca p	WGS84	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	NewsweekDemographicData.WorldThematicMapDataDescription.ESRIWebsite.2000.http://www.esri.com/data/online/wothdata.html (23 March 2000).
Night Light Polluti on	Co	0-255 indicati ng lights	nlp	WGS84	0.043945	https://www.arcgis.com/home/it em.html?id=b3824432e4204db6 9ed48552c8c3b81f	ESRI, NASA - Visible Earth
Livesto ck - Pigs	Co	Animal density /km2	pigs	WGS84	0.05		Herrick, K. A., Huettmann, F., & Lindgren, M. A. (2013). A global model of avian influenza prediction in wild birds: the importance of northern regions. Veterinary Research, 44(1), 42. http://doi.org/10.1186/1297-9716-44-42
Livesto ck - 5	Co	Animal density	poultry	WGS84	0.05	35	Herrick, K. A., Huettmann, F., & Lindgren, M. A. (2013). A global model of avian





Poultry		/km2					influenza prediction in wild birds: the importance of northern regions. Veterinary Research, 44(1), 42. http://doi.org/10.1186/1297-9716-44-42
Species richnes s of Amphi bians	Co	Numbe r of species	amphibi ans	World_E ckert_IV	10km	http://www.biodiversitymapping .org/download.htm	Jenkins, CN, SL Pimm, LN Joppa (2013) Global Patterns of Terrestrial Vertebrate Diversity and Conservation. PNAS 110(28): E2602-E2610. doi: 10.1073/pnas.1302251110 (PDF)
Species richnes s of Birds	Co	Numbe r of species	birds	World_E ckert_IV	10km	http://www.biodiversitymapping .org/download.htm	(2013) Global Patterns of Terrestrial Vertebrate Diversity and Conservation. PNAS 110(28): E2602-E2610. doi: 10.1073/pnas.1302251110 (PDF)
Species richnes s of Mamm als	Co	Numbe r of species	mammal s	World_E ckert_IV	10km	http://www.biodiversitymapping .org/download.htm	Jenkins,CN,SLPimm,LNJoppa(2013) GlobalPatternsofTerrestrialVertebrateDiversityandConservation. PNAS110(28):E2602-E2610.doi: 10.1073/pnas.1302251110UDF)
Species richnes s of Plants	Co	Numbe r of species	plants	WGS84	Polygon	https://www.arcgis.com/home/it em.html?id=e69b61fd6acf434bb 4343d69d3eb9552	Plant Species Richness by ecoregion: Kier, G., J. Mutke, E. Dinerstein, T. H. Ricketts, W. Ku, H. Kreft, and W. Barthlott. 2005. Global patterns of plant diversity and floristic knowledge. Journal of Biogeography 32: 1107-1116.





Proxim ity to coast	Co	m	prox_co ast	WGS-84	0.008333	http://www.naturalearthdata.co m/downloads/10m-physical- vectors/10m-coastline/	
Proxim ity to river	Co	m	prox_riv er	WGS-84	0.008333	http://www.naturalearthdata.co m/downloads/10m-physical- vectors/10m-rivers-lake- centerlines/	
Proxim ity to road	Со	m	prox_roa ds	WGS-84	0.008333	http://sedac.ciesin.columbia.edu /data/set/groads-global-roads- open-access-v1/data-download	Cento Infor Univ Outro Geor Data NY: Appl (SEE WCT
Annual Averag e Potenti al Evapo- transpir ation	Со	mm	pet_he_y r	WGS-84	0.008333	http://www.cgiar-csi.org	Trab Poter and Geo- Spati

Center for International Earth Science Information Network - CIESIN - Columbia University, and Information Technology Outreach Services - ITOS - University of Georgia. 2013. Global Roads Open Access Data Set, Version 1 (gROADSv1). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4VD6 WCT.

Trabucco, A., and Zomer, R.J. 2009. Global Potential Evapo-Transpiration (Global-PET) and Global Aridity Index (Global-Aridity) Geo-Database. CGIAR Consortium for Spatial Information.





Global Aridity Index (multip lied by 10000)	Co	-	ai_yr	WGS-84	0.008333	http://www.cgiar-csi.org	Trabucco, A., and Zomer, R.J. 2009. Global Potential Evapo-Transpiration (Global-PET) and Global Aridity Index (Global-Aridity) Geo-Database. CGIAR Consortium for Spatial Information.
Globco ver	Ca		globcove r	WGS84 ellipsoid	0.002778	http://due.esrin.esa.int/page_glo bcover.php	Source Data: © ESA / ESA Globcover 2005 Project, led by MEDIAS-France/POSTE
Infant mortali ty	Co	Percent age	infmort	WGS-84	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	Newsweek Demographic Data. World Thematic Map Data Description. ESRI Website. 2000. http://www.esri.com/data/online/wothdata.ht ml (23 March 2000).
Sanitati on	Co	%	sanitatio n	WGS-85	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	Newsweek Demographic Data. World Thematic Map Data Description. ESRI Website. 2000. http://www.esri.com/data/online/woth/data.ht ml (23 March 2000).
Cars	Со	Thousa nds	cars	WGS-86	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	Newsweek Demographic Data. World Thematic Map Data Description. ESRI Website. 2000. http://www.esri.com/data/online/woth/data.ht ml (23 March 2000).
Export	Co	Million s	export	WGS-87	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	Newsweek Demographic Data. World Thematic Map Data Description. ESRI Website. 2000. http://www.esri.com/data/online/wothdata.ht ml (23 March 2000).





Import	Co	Million s	import	WGS-88	Polygon	https://nelson.wisc.edu/sage/dat a-and- models/atlas/data.php?incdatase t=Life%20Expectancy	NewsweekDemographicData.WorldThematicMapDataDescription.ESRIWebsite.2000.http://www.esri.com/data/online/woth/ata.html (23 March 2000).
Slope	Co	Degree	slope_de g	WGS-84	0.008333	Using ETOPO1	
Aspect	Co		aspect	WGS-84	0.008333	Using ETOPO1	
NDVI	Co	1-255 levels	ndvi	WGS-84	0.1	http://neo.sci.gsfc.nasa.gov/vie w.php?datasetId=MOD13A2_M _NDVI	Images by Reto Stockli, NASA's Earth Observatory Group, using data provided by the MODIS Land Science Team.