



OPEN

## Modeling Eastern Russian High Arctic Geese (*Anser fabalis*, *A. albifrons*) during moult and brood rearing in the 'New Digital Arctic'

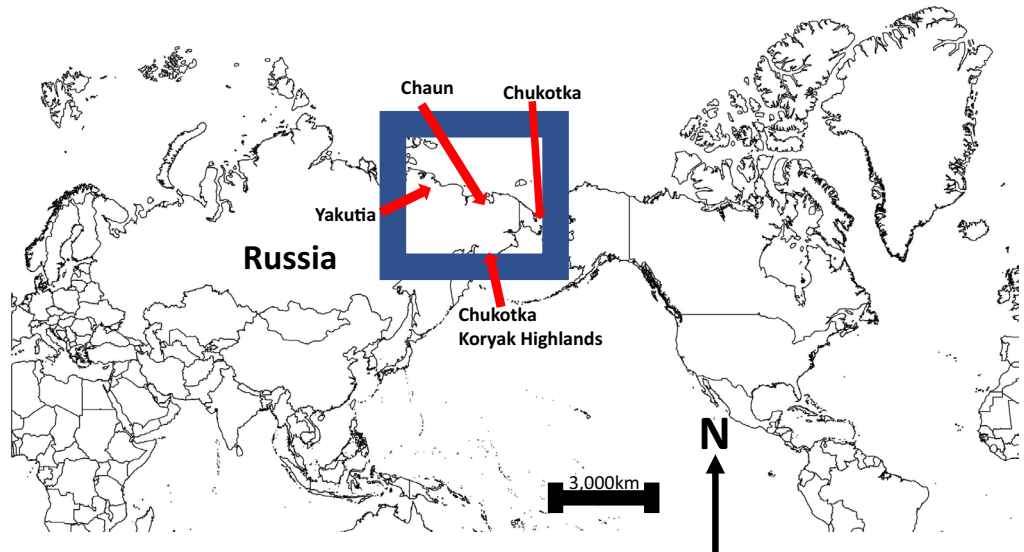
Diana Solovyeva<sup>1</sup>, Inga Bysykatova-Harmey<sup>2</sup>, Sergey L. Vartanyan<sup>3</sup>, Alexander Kondratyev<sup>1</sup> & Falk Huettmann<sup>4</sup>✉

Many polar species and habitats are now affected by man-made global climate change and underlying infrastructure. These anthropogenic forces have resulted in clear implications and many significant changes in the arctic, leading to the emergence of new climate, habitats and other issues including digital online infrastructure representing a 'New Arctic'. Arctic grazers, like Eastern Russian migratory populations of Tundra Bean Goose *Anser fabalis* and Greater White-fronted Goose *A. albifrons*, are representative examples and they are affected along the entire flyway in East Asia, namely China, Japan and Korea. Here we present the best publicly-available long-term (24 years) digitized geographic information system (GIS) data for the breeding study area (East Yakutia and Chukotka) and its habitats with ISO-compliant metadata. Further, we used seven publicly available compiled Open Access GIS predictor layers to predict the distribution for these two species within the tundra habitats. Using BIG DATA we are able to improve on the ecological niche prediction inference for both species by focusing for the first time specifically on biological relevant population cohorts: post-breeding moulting non-breeders, as well as post-breeding parent birds with broods. To assure inference with certainty, we assessed it with 4 lines of evidence including alternative best-available open access field data from GBIF.org as well as occurrence data compiled from the literature. Despite incomplete data, we found a good model accuracy in support of our evidence for a robust inference of the species distributions. Our predictions indicate a strong publicly best-available relative index of occurrence (RIO). These results are based on the quantified ecological niche showing more realistic gradual occurrence patterns but which are not fully in agreement with the current strictly applied parsimonious flyway and species delineations. While our predictions are to be improved further, e.g. when synergetic data are made freely available, here we offer within data caveats the first open access model platform for fine-tuning and future predictions for this otherwise poorly represented region in times of a rapid changing industrialized 'New Arctic' with global repercussions.

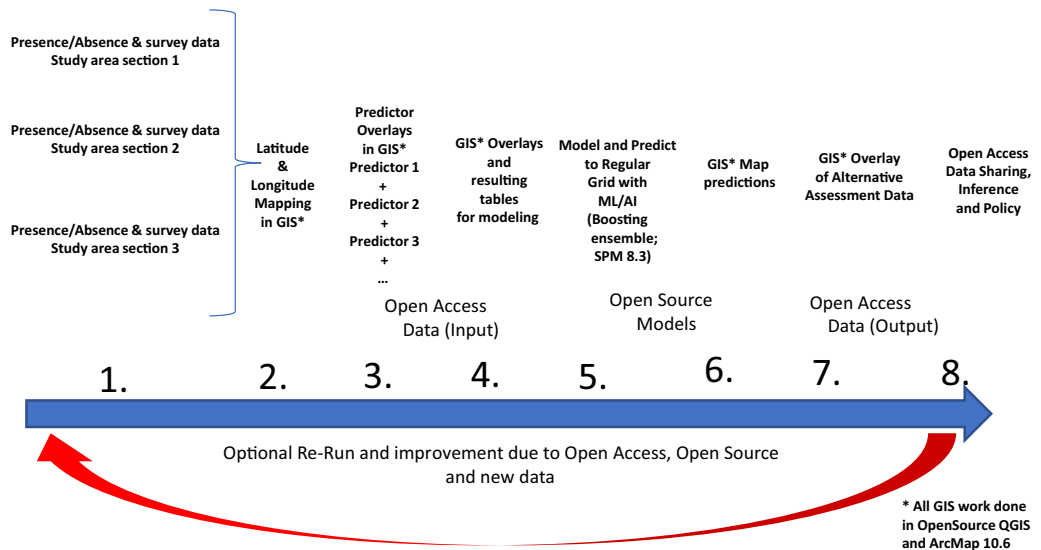
Adding to the natural global climate processes the human-caused global warming as well as new technology and industrialization also result into major shifts and dire ecological consequences giving rise to a 'New Arctic'<sup>1,2</sup>. The Russian Eastern Arctic—Yakutia and Chukotka—are part of this process and are also core zones supporting dense breeding populations of waterbirds in the circumpolar Arctic (Fig. 1). Waterbird populations in general, and geese and swans populations in particular, of Arctic Yakutia and Chukotka have been studied for a rather long time<sup>3–10</sup>. However, they are still comparatively poorly represented in the western literature thus far<sup>11–17</sup> show overall recent geese population estimates and outlines of their distribution in North East Asia; however this is based on expert opinion rather than on distribution models based on transparent data<sup>18</sup> for a review of expertism).

The 'New Arctic' is positioned in the Anthropocene, which also comes with digital data and the internet (for International Polar Years IPYs see<sup>19</sup> and<sup>1</sup>; examples shown in<sup>20–22</sup>. For this part of the world obtaining publicly available primary data of ground and aerial surveys with supporting information—in a modern format—is

<sup>1</sup>Institute of Biological Problems of the North, Far East Branch, Russian Academy of Sciences, Magadan, Russia. <sup>2</sup>Institute of Biological Problems of the Cryolithozone, Siberian Branch, Russian Academy of Sciences, Yakutsk, Russia. <sup>3</sup>North-East Interdisciplinary Scientific Research Institute N. A. Shilo, Far East Branch, Russian Academy of Sciences, Magadan, Russia. <sup>4</sup>EWHALE lab – Institute of Arctic Biology, Biology & Wildlife Department, University of Alaska Fairbanks (UAF), Fairbanks, Alaska, USA. ✉email: fhuettmann@alaska.edu



**Figure 1.** Circumpolar arctic and study area with general survey areas.



**Figure 2.** Workflow of this study.

very difficult despite some large species numbers and bird density and abundance<sup>10</sup>, though see some exceptions such as<sup>13</sup>. While environmental impact assessments are increasingly needed in the stressed area (e.g.<sup>20</sup>) and are to be based on best-available information<sup>23</sup>, internet resources such as GBIF.org or eBIRD are showing little data finds for this study area, nor are relevant bird banding data publicly shared and accessible for this region, yet. All while, the majority of birds on those flyways are clearly declining<sup>24</sup> (see<sup>25</sup> for songbirds,<sup>26</sup> for a highly endangered shorebird<sup>27</sup>, for Siberian crane<sup>17</sup>; for status of other geese; see<sup>28</sup> for historic perspectives on geese declines in Siberia, Asian wintering grounds etc.). Wilderness habitats along the flyway are also in an alarming down-trend, while urbanization is on the rise (see<sup>29</sup> for interior Asia flyway habitats). Excessive use of insecticides and pesticides, in addition to rampant poaching, has been described for many years<sup>28,30–32</sup>. Still, some *Anser* species are on the rebound and increase.

To set a baseline research conservation example for the ‘New Arctic’, here we focus on two key hunted goose species important for the local and indigenous people and their habitats of concern: Tundra Bean Goose *Anser fabalis serrirostris* (hereafter TuBEGO; *Taxonomic Serial Number TSN* from itis.gov 175024), and Greater White-fronted Goose *A. albifrons* TSN 175020 (hereafter GWFG). These birds do nest in the high Arctic, stop-over in agriculture landscapes in the boreal zone and winter in natural and agriculture habitats in the East Asia<sup>17,33,34</sup>. According to<sup>17</sup> TuBEGO is part of the ‘A4 Eastern Tundra Bean Goose *Anser fabalis serrirostris*’ population, whereas GWFG is part of the ‘C3 Tule Greater White-fronted Goose *Anser albifrons elgasi*’ population. We consider populations attributed to the West Pacific branch of the East-Asian-Australasian Flyway (see definition of

Species and sample size details	Location and subregion in Study Area	Survey focus	Source	Time period	Observer	Comment
<b>(a) Training data</b>						
<i>Anser albifrons</i> (n = 219; Broods present = 142, Broods absent = 77; Non-breeders present = 95 Non-breeders absent = 124)	Yakutia	Brood & Non-breeders	Boat	2017, 2018	IB	
	Chaun	Brood & Non-breeders	Boat	2002 till 2019	DS, AK	
	Chukotka	Brood & Non-breeders	Aerial	2002,	AK	
	Chukotka: Koryak Highland	Brood & Non-breeders	Aerial	1997	AK	
<i>Anser fabalis</i> (n = 593; Broods present = 213, Broods absent = 380; Non-breeders present = 303, Non-breeders absent = 290)	Yakutia	Brood & Non-breeders	Boat	2017, 2018	IB	
	Chaun	Brood & Non-breeders	Boat	2002 till 2019	DS, AK	
	Chukotka	Brood & Non-breeders	Aerial	2002,	AK	
	Chukotka: Koryak Highland	Brood & Non-breeders	Aerial	1997	AK	
<b>Species</b>	<b>Source</b>	<b>Match with predictive model</b>	<b>Comment</b>			
<b>(b) Testing data</b>						
<i>Anser albifrons</i>	GBIF presence (n = 63)	Good	These data are 'just' occurrences but indicate presence absence in the landscape			
	Compiled literature presence absence (n = 14)	Good	These data are also 'just' presence/absence but carry some attributes on broods and non-breeders (not shown here). The source is from the literature and first-time presented as a GIS layer			
<i>Anser fabalis</i>	GBIF presence (n = 17)	Good	These data are 'just' occurrences but indicate presence absence in the landscape			
	Compiled literature presence absence (n = 18)	Good	These data are also 'just' presence/absence but carry some attributes on broods and non-breeders (not shown here). The source is from the literature and first-time presented as a GIS layer			

**Table 1.** Description and details of data used for the relative index of occurrence (RIO) in this study for two goose species. Subregions are shown in Fig. 1.

this branch in<sup>33,34</sup> because these populations are known to have increased for both species through global change and due to a switch to agricultural habitats on wintering grounds in Korea and Japan<sup>35,36</sup> for rough estimates of wintering grounds)). Many of those areas are coastal with agriculture, and the specific human populations are on the increase.

These species are of relevance because their arctic breeding grounds are still virtually free of a dense road network<sup>37</sup> or settlements; however recent proposals for the developing of mining and infrastructure (including supporting roads) may be critical for impacting summer habitats of both geese species (see <http://ecoline-eac.com/proekty/peschanka/deposit.html>). While the arctic grazing systems are often considered pristine, they are not due to the overgrazing by abundant domestic reindeer<sup>15,38,39</sup>. And man-made climate change and associated permafrost thawing, and even fires add other man-made features now. The Anthropocene -its characteristics and problems—is clearly found in the Arctic and Arctic plain<sup>40</sup> and along the species flyway<sup>25</sup>; it's the 'New Arctic'<sup>2</sup> which also happens to be digital<sup>1,20</sup>, can use Machine Learning methods<sup>18,20,27</sup> and has such processes, interactions and opportunities<sup>23</sup>.

Recent summer distribution data for these species—explicit in time and space<sup>41</sup>—are lacking, are not compiled and are not available in a good useable or digital format with metadata to understand them, thus far (compare with<sup>20</sup>). A subsequent open access model prediction for these species and their specific metrics does not yet exist but can be powerful for progress (e.g.<sup>42,43</sup>; but see<sup>27</sup> for Siberian Crane<sup>44</sup>, for Lesser White-fronted Goose; and<sup>45</sup> for concepts and workflows).

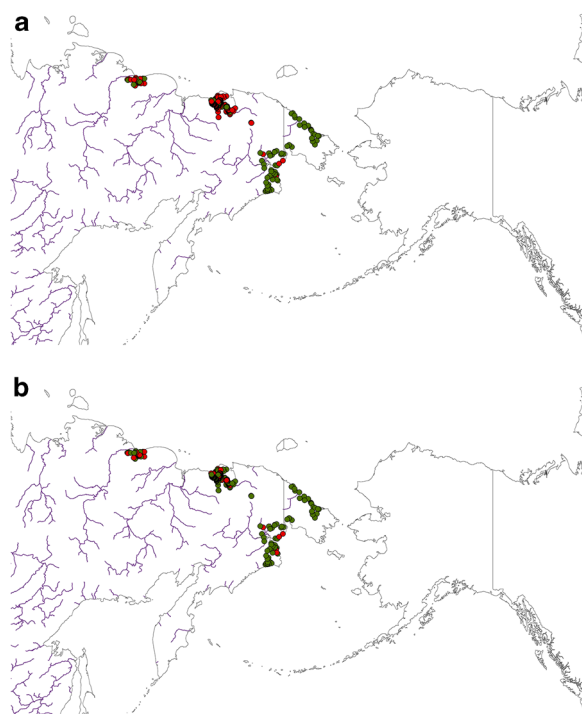
Using these Arctic geese allows for progress on this topic, which is important while development and massive changes push northwards and into the interior wilderness areas of this world and into the New Arctic<sup>38</sup> (for status see<sup>1,37</sup>). A solid study for those two species and Open Access baseline can help here to set the stage, document and address conservation problems in a best-available scientific manner for betterment along the flyway, in the stop-over and wintering areas, as well as for protected area questions (e.g.<sup>46</sup>). A distribution model quantifies bird-habitat relationships for the area for those geese for the first time, and might be helpful in forecasts of changes related to climate and other drivers of populations<sup>18</sup>.

## Methods

**Workflow.** Following best-practice<sup>41</sup> (see<sup>45</sup> as well as<sup>20,27</sup> for applications) we developed a workflow. It shows how the data can be compiled, cleaned, employed in GIS and model predicted, subsequently assessed for performance, and be used and interpreted for inference; general model details and concepts are shown in<sup>18</sup> and Fig. 2.

Predictor name	Importance rank in model			
	Tundra Bean Goose: Non-breeders	Tundra Bean Goose: Brood	Greater White-fronted Goose: Non-breeders	Greater White-fronted Goose: Brood
Global Landcover	85	67	93	93
Mean Temperature in July	83	61	76	84
Mean Precipitation in July	60	<b>100</b>	<b>100</b>	90
Annual NDVI	82	65	89	<b>100</b>
Human Footprint	32	24	34	34
Elevation (ETOPO1)	<b>100</b>	73	92	<b>99</b>

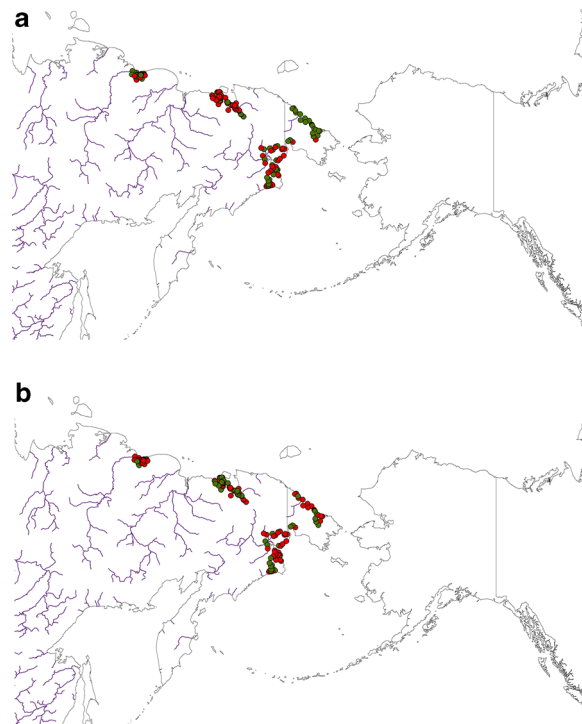
**Table 2.** Predictor Importance Rank in model for two *Anser* species: brood-rearing parents with broods and moulting non-breeders (The top-2 predictors for each species strata are presented in bold. Human population density shows little variation and relevance for the study area and is not shown).



**Figure 3.** Best-available compiled raw data of Tundra Bean Goose (*Anser fabalis*) presence/absence for (a) Brood, and (b) Non-breeders in the study area. For both figures presence is shown in red and absence in green. Map created by FH with OpenSource QGIS and ArcGIS Desktop 10.6 academic license.

**Bird data.** *Model training data.* The Russian Arctic is vast, and it presents the biggest landmass for this circumpolar ecosystem—arctic, boreal, terrestrial and marine alike<sup>27,37,46</sup>. Despite many decades of publications and efforts, our study area located in the Eastern Arctic with an extent of c. 3000 km East–West and c. 2500 km North–South is still poorly and not systematically surveyed for its birds with no centralized database readily available for geese species or their habitats and for specifics of breeding and post-breeding times. A coordinated research design or species atlas with models does not yet exist for this region (compare with<sup>47</sup> for Yukon<sup>48</sup>, and<sup>49</sup> for Alaska, or for Sweden see: <https://www.ebba2.info/contribute-with-your-data/national-coordinators/sweden/>; <https://pecbms.info/country/sweden/>).

In the absence of such data, for the first time we were attempt to compile representative area species sampling data sets for the two study species in their summer grounds. We focused on post-breeding, namely brood-rearing, and moulting time—mostly July–August—which has never really been spatially described, yet in a coherent fashion. Field data were collected during surveys performed by authors along the rivers and on lakes in the tundra during the period of 1997–2020 using motor-boat and aerial surveys. Visual surveys from moving motor-boat, foot ground surveys around lakes and aerial surveys were combined (Table 1). Only data on the presence and number of flightless geese, (moulting adults and brood-rearing groups) were used for this study. The flightless period is a critical time in the annual cycle of geese<sup>50,51</sup>. Their habitat requirements include food availability and safety from predators and people. That habitat requirement exists because flightless geese traditionally co-evolved with, and were hunted by, people for centuries. And during their wing-moult they are



**Figure 4.** Best-available compiled raw data of Greater White-fronted Goose (*Anser albifrons*) presence/absence for (a) Brood, and (b) Non-breeders in the study area. For both figures presence is shown in red and absence in green. Map created by FH with OpenSource QGIS and ArcGIS Desktop 10.6 academic license.

extremely sensitive to all kinds of human-induced disturbance. Being flightless, the geese stay stationary during that moulting period, and thus, their spatial distribution does not really change at least for one month except for minor local small-distance movements.

The sampling data obtained during this geese census are a relative index of occurrence (RIO) and contain decimal latitude, decimal longitude, observation time (24 h) and date (day, month and year). We also included species presence and abundance, and we categorized birds as either: (1) moulting non-breeders or (2) breeding pairs with goslings. These data, two data sets for each of the two species, were put into an OpenSource datasheet using CSV format and were GIS-mapped (Fig. 1; Table 1). The surveyed areas and their rivers and lake areas are listed in Supplement 3.

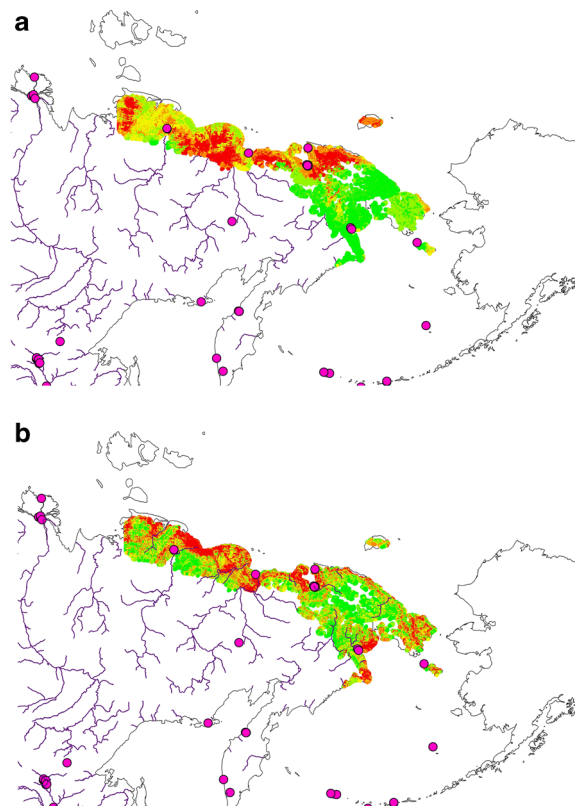
**Model assessment data.** In addition to internal model metrics, it is important to assess models with alternative information, to compare models with reality<sup>18,52,53</sup>. Therefore, we used alternative cleaned data from GBIF.org (DOIs <https://doi.org/10.15468/dl.up4kmu>; <https://doi.org/10.15468/dl.xwnkqe>) and the literature to test our models (see Supplement 1). We also found other data sets like MOVEBANK, Bird Banding Center data and many research project data mentioned in publications but those were not available in an Open Access format and thus could not be used (*sensu*<sup>19,41</sup>).

**GIS data.** Despite many decades of geological and geophysical survey work, modern GIS data layers for the study area are not really available, e.g. as needed as predictors in a raster format with known errors, a valid geographic projection and ISO-compliant metadata to understand them for a scientific purpose. We therefore followed data from Sriram and Huettmann (unpublished; <https://essd.copernicus.org/preprints/essd-2016-65/>) and added those open access layers as habitat predictors.

We selected GIS layers that are biologically meaningful or that are habitat use proxies and available for the prediction of the ecological niche for the two species during summer. Our models focus only on the Arctic tundra, and we used the CAVM map (<https://www.caff.is/flora-cfg/circumpolar-arctic-vegetation-map>) to exclude other habitat types where geese are not occurring, e.g. forests.

The following seven predictors were used for the study area: Global Landcover, Mean Temperature in July, Mean Precipitation in July, Annual NDVI, Human Footprint, Elevation (ETOPO1) and Human Density. A list of those GIS maps and their details can be seen in Supplement 2 and GIS files are available for free download and further use from sources mentioned.

**Data processing.** We followed the workflow outlined in the beginning of this section (Fig. 2). We used ASCII CSV data and imported them into ArcGIS desktop 10.6 and OpenSource QGIS 3.16, and then overlaid them with GIS layers for the study area with external layout edits. The study area has a date line (180 degrees

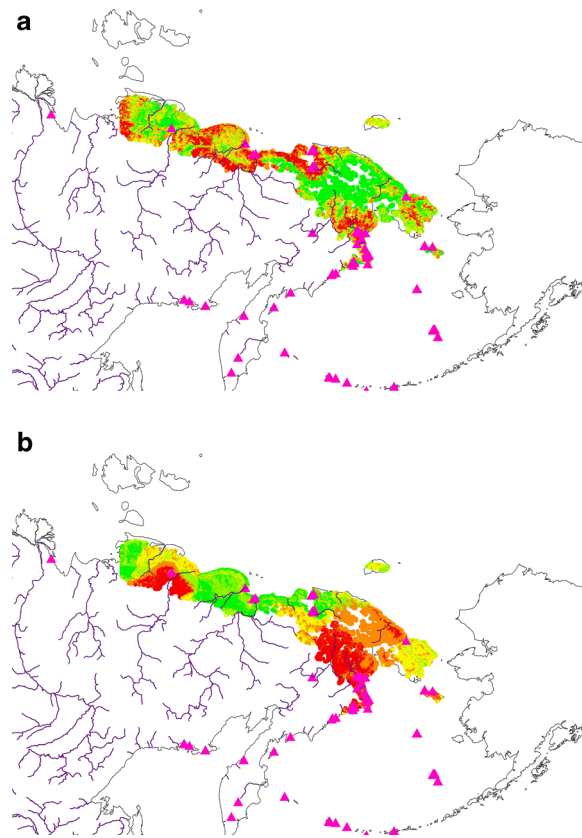


**Figure 5.** Predictions of Tundra Bean Goose (*Anser fabalis*) (a) brood-rearing parents with broods, and (b) post-breeding nonbreeders for the study area). The map predictions are presented as a ‘heatmap’ of the relative index of occurrence (RIO): red is a high RIO and green is a low RIO. Best-available GBIF presence location for this species are superimposed for assessment; they are shown as pink points. Map created by FH with OpenSource QGIS and ArcGIS Desktop 10.6 academic license.

longitude located app. between Russia and Alaska). In addition, we used the Mercator geographic projection with a Pacific meridian using decimal latitude and longitude (WGS84). We then exported data from the GIS as a table for subsequent model-predictions presented in the next section. These steps are generally used in<sup>20</sup> and in a more detailed way applied in<sup>45</sup> and<sup>20,27</sup> as a proof of concept for the area.

**Predictive modeling.** Here we are following a widely-used concept of inference from predictions<sup>52,53</sup> (see<sup>18,42</sup> for applications), employed for n-dimensional ecological niche models<sup>20,42</sup>. This was achieved by using Minitab-Salford Predictive Modeler (SPM 8.3; <https://www.minitab.com/en-us/products/spm/>). We employed TreeNet (Stochastic Gradient Boosting<sup>18</sup>; see<sup>27</sup> for an application and example of the algorithm; see<sup>42,43</sup> for general performance assessments of the algorithm as being among the most suited and powerful). To find the best solution we started with exploratory models and their metrics, e.g. confusion matrix and ROC, to be improved sequentially. We used default model settings (known to perform best<sup>18</sup>) with ‘balanced sampling weights’, tenfold Cross-Validation, a node depth of 10, and 2 as the minimum sample size for terminal tree branches; we used 400 trees to assure an optimal solution was found. Model diagnostics are presented in the appendices; see<sup>18</sup> for general applications. As we employ non-parametric machine learning techniques we are less concerned with autocorrelation. Also, this is the first model of its kind and we did not emphasize specific questions of autocorrelation. Stochastic Gradient boosting is robust to data with autocorrelation; for justification, conceptual details and lack of a problem see for instance<sup>34</sup>). After creating a grove file in SPM to capture the actual model in a software format, we scored an approx. 5 km point lattice and obtained pixel-based predictions. We used that conservative scale to overcome GIS data inaccuracies inherent in many of the currently available Arctic data, e.g. coastline location and digital elevation models (DEM). Those lattice points then were mapped for the study area and a GIS legend was fit to visualize the RIO.

**Model assessment.** Our model was assessed in four ways for evidence: (i) Based on the 24 year presence and absence data we used an internal ROC of the exploratory models as readily provided by SPM and its confusion matrix<sup>42,43</sup>. (ii) For a deeper assessment we also overlaid the model surface with the training and absence data for each species and the two data sets (moulting non-breeders, broods) allowing for a visual assessment of the generalization achieved. (iii) We further used the alternative assessment data -GBIF.org and compiled (Rus-



**Figure 6.** Predictions of Greater White-fronted Goose (*Anser albifrons*) (a) parents with broods, and (b) moulting nonbreeders for the study area). The map predictions are presented as a ‘heatmap’ of the relative index of occurrence (RIO): red is a high RIO and green is a low RIO. Best-available GBIF presence location for this species are overlaid for assessment; they are shown as pink points. Map created by FH with OpenSource QGIS and ArcGIS Desktop 10.6 academic license.

sian) literature—which we also overlaid for a visual assessment of the predicted ecological niche for this species. Lastly, (iv) we then compared our findings with the literature and other sources for this species.

Overall, all of these four assessments allow us to get a generic confirmation of how well models perform, using all of the best publicly available data as lines of evidence.

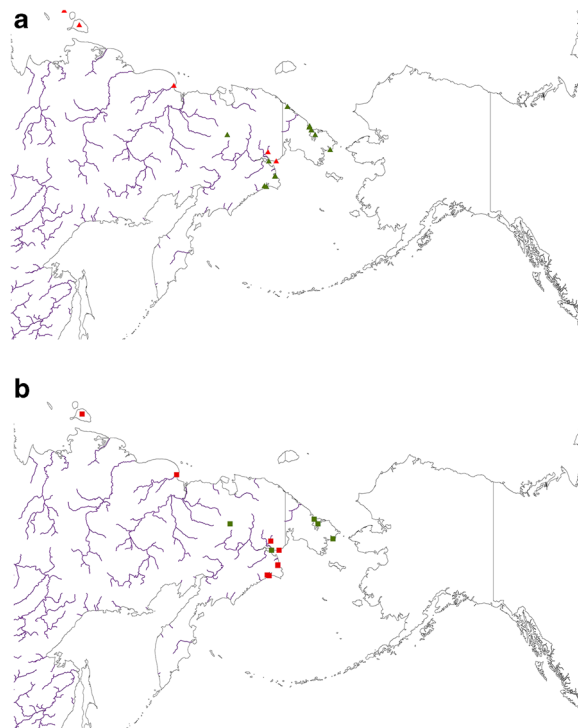
**Ethical statement.** All methods in this workflow and as presented were performed in accordance with the relevant guidelines, regulations and ethics committees by the authors and their institutions involved (see author list and affiliations for details). This research is entirely based on non-intrusive surveys, data compilation and digital data analysis; no specimen were collected. Field work was done in Russia remotely with binoculars and ‘naked eye’ in the field, following their national regulations accordingly. All presented map products and associated shapefiles were done with OpenSource QGIS 3.16 (<https://www.qgis.org/en/site/forusers/download.html>) and ArcGIS Desktop 10.6 under the University of Alaska Fairbanks (UAF) academic campus license by FH/EWHALE lab. The data processing and GIS work and edits were done in computer labs in Russia and Alaska/US and all applicable data are presented and available Open Access for a transparent and repeatable approach including ISO-compliant metadata.

## Results

We were able to compile for the first time find the best-available long-term data for the two species and their two metrics—brood locations and non-breeders—for the study area of Eastern Yakutia and Chukotka. Also, we were able to obtain the best publicly-available assessment data in a digital format explicit in space and time.

Further, our findings show the first achieved predictions and their assessments for post-breeding moult of Tundra Bean Goose and Greater White-fronted Goose for non-breeders and parents with brood (Table 2; Figs. 5 and 6).

**Species: Tundra Bean Goose.** The moulting non-breeders are primarily distributed in coastal areas of Yakutia and Western Chukotka, thus inhabiting coastal plains. A low occurrence is predicted in the eastern study area, and the birds are more or less absent in the mountains of the interior Chukotka, wider inland and along the coast of northern Bering Sea (Fig. 5a). This shows a more nuanced and complex distributional picture than



**Figure 7.** Best available occurrence data from the literature for (a) Tundra Bean Goose (*Anser fabalis*; shown as triangles) and (b) Greater White-fronted Goose (*Anser albifrons*; presented as squares). For both figures presence is shown in red and absence in green. Map created by FH with OpenSource QGIS and ArcGIS Desktop 10.6 academic license.

Data set name	Content	Source	Comment
Movebank	Geolocations	Various funders	Most data generally blocked behind login
Goose tagging	Locations	China	Most data generally not made publically available regardless of publication
Bird Banding	Banding location, resighting and recovery	National Bird Banding Center	EURING, nor the EU, is explitley not sharing geo-referenced data in GBIF
(International) expedition sighting records	Documented locations of presence and absence	Many researchers, institutions and NGOs worldwide	Those various data were collected and exist for over 50 years in the study area
*Citizen Science data	Documented locations of presence and absence	Many tourists, naturalists, governmental employees and researchers worldwide	iNaturalist, eBIRD etc. are growing rapidly

**Table 3.** Data known to be available for the study area and of great use for this work but not publically made available or useable (Note: Best professional practices, the International Polar Year, Migrator Bird Treaties, and most national funding schemes make Open Access data sharing mandatory; see <sup>19</sup> for a reality assessment and as found here). \*Data are readily available but show little coverage and information for the study area, yet.

what was previously known; arguably, the distribution of this species is not as crisp as presented and assumed elsewhere.

For the parents with broods the above pattern shows even stronger, with the parents and broods primarily occurring in the western section of the study area. It is noteworthy that the parents with broods are absent along the coastline and are found more inland, primarily Yakutia Arctic and around the wider Chaun Bay region, while Chukotka Peninsula is widely free of this cohort (Fig. 6b).

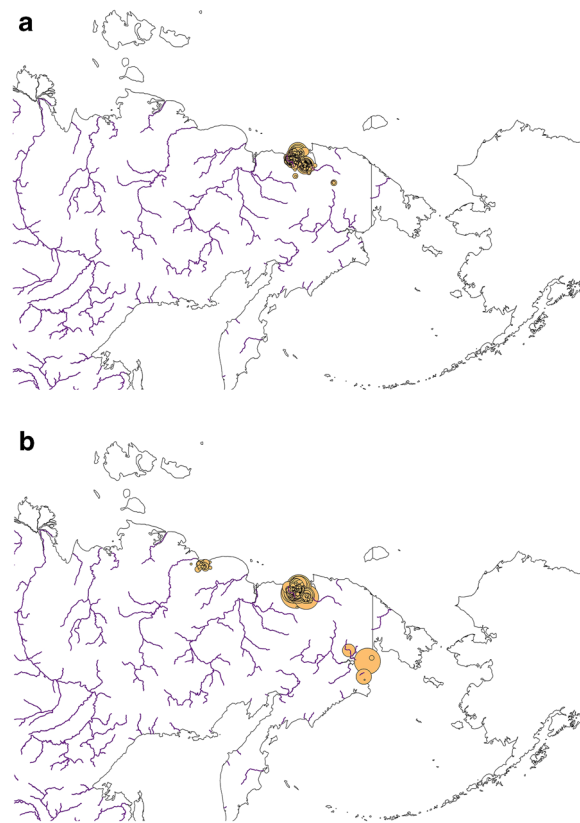
It is noteworthy that the non-breeders are not really overlapping with the parents with broods; the latter concentrate in the western section of the study area and more inland.

**Species: Greater White-fronted Goose.** The moulting non-breeders are widely dispersed in the study area but seem to avoid the mountain habitats, e.g. inner parts of the Chukotka Peninsula and parts of Yakutia.

For the actual parents with broods it shows an almost opposite pattern, where the species is found in the interior, specifically in Chukotka and in Yakutia.

The patterns are hardly overlapping and are somewhat complementary to each other. There are two distinct patches, leaving a coastal area free of this species.





**Figure 8.** Field survey numbers for Tundra Bean Goose (*Anser fabalis*) (a) broods, and (b) moulting non-breeders for the study area). Yellow circles are scaled ranging from 0 to 100 individuals. Map created by FH with OpenSource QGIS and ArcGIS Desktop 10.6 academic license.

**Model performance details and assessment.** *Model performance details.* Our models achieved good to very good accuracy (details shown next section). Predictors most strongly center around an interaction between climatic metrics like summer precipitation, temperature, as well as elevation and landcover categories, added by NDVI (Table 2; detail shown in Supplement 5). While the human footprint showed a smaller role, those trends were upwards indicating that those geese are somewhat affiliated with the human footprint.

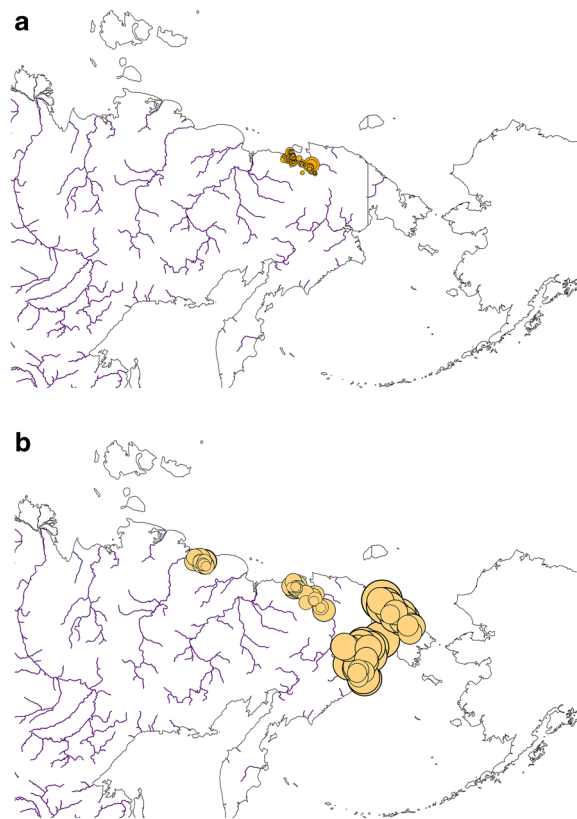
For Tundra Bean Goose broods in the multivariate context we identified NDVI as a powerful predictor with a positive relationship (Table 2; details shown in Supplement 5). Together with lower elevations below 150 m it indicates where brood-rearing habitats can be found in the study area. For non-breeders we found precipitation in July as a powerful predictor with a positive relationship (Table 2; details shown in Supplement 5). Together with specific arctic coastal landcover classes it indicates where moulting areas can be found in the study area.

For Greater White-fronted Goose broods we found precipitation in July as a powerful predictor, but with a negative relationship (Table 2; details shown in Supplement 5). Together with somewhat higher elevations around 300 m it indicates where brood-rearing geese occur in the study area. For non-breeders we identified elevation as a powerful 4th predictor with a negative relationship (Table 2; details shown in Supplement 5). Together with specific landcover classes it indicates where moulting flocks can be found in the study area.

*Model assessment details.* For robust inference and evidence, we actually used four pathways to assess the performance of our data-based model predictions for Tundra Bean Goose and Greater White-fronted Goose and their post-breeding non-breeders and parents with brood. The first is the internal aspatial ROC metric that comes with the exploratory model data itself. It shows a ROC of 82% (Tundra Bean Goose non-breeders), 85% (Tundra Bean Goose broods), 91% (Greater White-fronted Goose non breeders) and 94% (Greater White-fronted Goose broods) for both species and their metrics. The ROC is based on the confusion matrix from the binary presence and pseudo-absence of the two survey data used for each of the two species (see Figs. 3, 4, 5 and 6). Those assessments indicate already a rather good model on the training data.

The second—more thorough—assessment is based on a visual match of the predictions with their training data on a map, allowing us to provide evidence of a good general match of the pattern predicted (see Figs. 5 and 6) for the two species and their metrics.

The third and fourth assessments, more independent but less specific for parents with broods and non-breeders, are based on the GBIF.org data and the compiled literature references for the species and its ecological niche overall in summer, less though for the brood and the non-breeders (see Table 1b; Fig. 7). But at least on a generic level it shows a very high match for the models (compare with Figs. 5, 6).



**Figure 9.** Field survey numbers for Greater White-fronted Goose (*Anser albifrons*) (a) broods, and (b) moulting non-breeders for the study area. Yellow circles are scaled ranging from 0 to 100 individuals (in b) a log scale was used ranging from 0 to 10,000). Map created by FH with OpenSource QGIS and ArcGIS Desktop 10.6 academic license.

Taken the evidences together, overall, we therefore think that the methodology shown (Fig. 2 for workflow) and results presented are a good start for inference and offer us presentable validity, allowing to move next into thorough abundances and population trend models. Arguably, better data, e.g. more explicit, more extensive, and ideally corrected for detectability coming from a proper research design (see<sup>55</sup> for an example) will allow for fine-tuning our findings further<sup>1,2,39,40</sup>.

## Discussion

For the study area of the Russian Eastern Arctic, this study is the first that compiled ISO-documented digital long-term data explicit in space and time for Tundra Bean Goose and Greater White-fronted Goose (Compare with<sup>56</sup>). While the true ranges remain unknown, here we provided important steps for two keystone species for the flyway. We tried to achieve the first digital model workflow and approaches for the species in Russia. We further tried to advance knowledge for this species by focusing on the post-breeding time and moult locations for (1) parents with broods, indicating also nesting habitats because non-flying goslings cannot move far from their nests, as well as (2) non-breeders away from the nest. Those data fill a gap in existing databases (e.g.<sup>57</sup>) and they are more specific than the generic ecological niche in summer and hand-drawn maps (e.g.<sup>58,59</sup>) for each species. In the wide absence of public information on these specific questions the data are part of the global arctic research legacy and the findings should be of good use and relevance for the study area and flyway as quantified baselines for bird monitoring, range estimates and subsequent population estimations and conservation management. Arguably, those deserve to be improved further and frequently with more data.

The habitat GIS layers are also the first of their kind compiled for this species for the public, the study area, provided in a modern digital grid format and made available free of charge in a documented form. Those data can be assessed and fine-tuned for more work as well (see Sriram and Huettmann unpublished for over 100 GIS layers to be used; see an application for the study area by<sup>27</sup>).

While this study has limitations, here we use an open access approach and we open all steps up for detailed review and scrutiny for model improvements; sensu<sup>53</sup>.

Our models are the first generation of such workflows and deserve careful use. However, they are assessed with 4 lines of evidence and allow for a subsequent inference. They show us a new, nuanced and complex species distribution pattern. They have little overlap of parents with brood vs nonbreeders indicating movements and specific staging sites; it is a new piece of information and needs more study. This biological mismatch is most pronounced for Greater White-fronted Geese. It shows that non-breeders and probably early failed-breeders,

stay apart from their breeding grounds commencing moult migration to the areas/habitats differing from the ones used by parents with broods. Generally, we found from our models that for both species' parents with brood retreat from the coast and then move more inland. Except for non-or-failed- breeding Greater White-fronted Goose we found that Eastern Chukotka is of less relevance for both species during the post-breeding times. The Greater White-fronted Goose is distributed on both sides of the Bering Sea, being a truly circumpolar species, while Tundra Bean Goose is an Eurasian species, not existing in North America (replaced by the Canada goose).

However, despite the four lines of evidence matching these patterns our findings are previously unknown as they are only partly in agreement with the coarser<sup>58,59</sup> maps and with<sup>17</sup>. In addition to showing more differentiated and realistic distribution patterns they also include highly preferable areas/habitats of populations migrating along the West Pacific Flyway.

In a multivariate context we found that climatic variables play a larger role for the presence of the two *Anser* species in post-nesting flightless times. We also found a positive relationship with NDVI (see also<sup>60</sup> for green wave and NDVI link) and with the Human Footprint. However, Human Footprint is currently a weak predictor in our model probably because our study area is among the least populated in the world (some physical industrial footprint does exist though). Interestingly, in another Arctic nesting *Anser* species, the Lesser White-fronted Goose *Anser erythropus*, the habitat suitability in the same study area decreases with human disturbance, reflecting the negative impacts of human presence there. Lesser White-fronted Goose (the species is Endangered with declining population<sup>17</sup>) are found to select mostly human-free sites among huge area of suitable summer habitats<sup>44</sup>, while abundant and increasing Tundra Bean Goose and Greater White-fronted Goose are found to utilize a wider set of habitats including areas close to human settlements. Co-occurrence with humans may be an occasional result of selecting areas close to large and medium-sized rivers.

Overall, our predictions and assessments could have been stronger if existing data we located to exist were actually made better available by the international community<sup>19,23</sup> see Table 3 for data that exist for the study area and study species).

We would like to emphasize that our studied populations of both species are of the West Pacific Flyway, what it means is that their wintering areas are in Korea and Japan. Trends of Greater White-fronted Goose populations are contrasting between the West Pacific and East-Asian Continental Flyways, with the birds of the latter all wintering in China. However, from our work we feel that such strict delineations might be somewhat inaccurate, as the more graduated prediction maps show (see for instance<sup>61–63</sup> for patterns). The inclusion of small-sized Lesser White-fronted Goose sharing summer and—in part—winter habitats with our study species poses another question of competition for the food resources to be studied in more detail<sup>44</sup>. More thought is to be given about their range, distribution and flyway memberships and 'straddling' while habitats and climates are changing so rapidly overthrowing evolved and assumed patterns.

In forthcoming work species abundances could be addressed to match for instance the overall flyway and winter estimates (for model concepts see<sup>63,65</sup>). But Figs. 8 and 9 make it clear from our additional survey data we compiled that numbers seem to be large when extrapolated to the ecological niche that we presented here.

As presented by<sup>27</sup> we find that approaches of data mining, predictive modeling done with an open access and open source concept are new, very promising, insightful and should be applied here more and with policy implementations. However, changes that are currently happening in the Arctic and its flyways are dramatic, shaping global processes and events, and it is unclear whether concerted policy actions even can mitigate them any time soon.

These findings matter because they help filling study gaps in time and space, as well showing the state of the art for these species, their habitats and scientific data. It is noteworthy that the species studied are also vectors for diseases, which in times of pandemics are of importance (e.g.<sup>64,65</sup>).

Lastly, and as shown in<sup>20,25</sup> it should be feasible to create circumpolar and/or flyway predictions for the species of interest in order to tackle modern questions of Arctic and migratory species management. These predictions can be high resolution explicit in time, in space and in the biology, e.g. for subspecies, timings and physiology, as it was started here (see<sup>66</sup> for high resolution model options). While no meaningful large-scale tracking of high arctic species and ecological niche estimates exist yet, those data from Movebank, Bird banding and other efforts—if made publicly available—would contribute much to all efforts reported here, ideally, for future predictions during a still unabated man-made climate change with associated sustainable policy implications.

Received: 18 June 2021; Accepted: 27 October 2021

Published online: 11 November 2021

## References

- Huettmann, F. (ed.) *Protection of the Three Poles* 337 (Springer, 2012).
- Krupnik, I. & Crowell, A. L. *Arctic Crashes: People and Animals in the Changing North* (Smithsonian Institutional Press, 2020).
- Portenkorupnik, L. A. *Birds of Chukchi Peninsula and Wrangel Island, Part 1* 424 (Nauka, 1972) (in Russian).
- Vorobiev, K. A. *Birds of Yakutia* 336 (Nauka, 1963) (in Russian).
- Egorov, O. V. State of population of the waterbirds and other bird species in the Lena Delta and Yana-Indigirka Tundra by the materials of air-record. In *Nature of Yakutia and Its Protection* 124–127 (Yakutsk .Kn. Izd., 1965) (in Russian).
- Egorov, O. V. & Perfilov, V. I. Inventory of the population of the waterbirds from airplane in the North Yakutia. In *Methods of Inventory of Population* 71–78 (Yakutsk. Kn. Izd., 1970) (in Russian).
- Kishchinskii, A. A. & Flint, V. E. Waterbirds of the Near-Indigirka Tundra. In *Resources of the Waterbirds of the USSR, Their Reproduction and Use* 62–65 (Mosk. Gos. Univ., 1972) (in Russian).
- Krechmar, A. V., Andreev, A. V. & Kondratyev, A. Y. *Bird of Northern Plains* 228 (Nauka, 1991) (in Russian).
- Andreev, A. V. Monitoring of the Geese in the North Asia. In *Species Diversity and State of Populations of the Waterbirds in the North-East of Asia* 5–36 (SVNTs DVO RAN, 1997) (in Russian).

10. Krechmar, A. V. & Kondratyev, A. V. *Waterfowl in North-East Asia* 458 (NESC FEB RAN, 2006).
11. Kistchinski, A. A. Waterfowl in north-east Asia. *Wildfowl* **24**, 88–102 (1973).
12. Pearce, J. M., Eesler, D. & Degtyarev, A. G. Birds of the Indigirka River Delta, Russia: Historical and biogeographic comparisons. *Arctic* **51**(4), 361–370 (1998).
13. Hodges, J. I. & Eldridge, W. D. Aerial surveys of eiders and other waterbirds on the eastern Arctic coast of Russia. *Wildfowl* **52**, 127–142 (2001).
14. Hupp, J. W. *et al.* Moul migration of emperor geese *Chen Canagica* between Alaska and Russia. *Avian Biol.* <https://doi.org/10.1111/j.0908-8857.2007.03969.x> (2007).
15. Degtyarev, A. G. Monitoring of the Bewick's Swan in the tundra zone in Yakutia. *Contemp. Probl. Ecol.* **3**, 90–99. <https://doi.org/10.1134/S1995425510010157> (2010).
16. Stroud, D. A., Fox, A. D., Urquhart, C. & Francis, I. S. *International Single Species Action Plan for the Conservation of the Greenland White-fronted Goose (Anser albifrons flavirostris)* AEW Technical Series No. 45 (AEWA, 2012).
17. Fox, A. D. & Leafloor, J. O. *A Global Audit of the Status and Trends of Arctic and Northern Hemisphere Goose Populations*. (Conservation of Arctic Flora and Fauna International Secretariat). ISBN 978-9935-431-66-0. (2018).
18. Humphries, G., Magness, D. R. & Huettmann, F. *Machine Learning for Ecology and Sustainable Natural Resource Management* (Springer, 2018).
19. Carlson, D. A. Lesson in sharing. *Nature* **469**, 293. <https://doi.org/10.1038/469293a> (2011).
20. Huettmann, F., Artukhin, Y., Gilg, O. & Humphries, G. Predictions of 27 Arctic pelagic seabird distributions using public environmental variables, assessed with colony data: A first digital IPY and GBIF open access synthesis platform. *Mar. Biodivers.* **41**, 141–179. <https://doi.org/10.1007/s12526-011-0083-2> (2011).
21. Huettmann, F. & Ickert-Bond, S. *On Open Access, data mining and plant conservation in the Circumpolar North with an online data example of the Herbarium*. (University of Alaska Museum of the North Arctic Science, 2017). <http://www.nrcresearchpress.com/toc/as/0/ja> (2017).
22. Ickert-Bond, S. *et al.* New Insights on Beringian Plant Distribution Patterns. Alaska Park Science—Volume 12 Issue 1: Science, History, and Alaska's Changing Landscapes. (U.S. National Park Service, 2018). <https://www.nps.gov/articles/aps-v12-i1-c-10.htm>.
23. Huettmann F (2015) 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: Huettmann F (eds) *Central American Biodiversity: Conservation Ecology and a Sustainable Future*. Springer, New York, pp. 391–418
24. Zoeckler, C. Chapter 9: Status, Threat, and protection of Arctic Waterbirds. In *Protection of the Three Poles* (ed. Huettmann, F.) 203–216 (Springer, 2012).
25. Jiao, S., Huettmann, F., Guo, Y., Li, X. & Ouyang, Y. Advanced long-term bird banding and climate data mining in spring confirm passerine population declines for the Northeast Chinese-Russian flyway. *Glob. Planet. Change* **144C**, 17–33. <https://doi.org/10.1016/j.gloplacha.2016.06.015> (2016).
26. Zoeckler, C. *et al.* The winter distribution of the Spoon-billed Sandpiper *Calidris pygmaeus*. *Bird Conserv. Int.* **26**, 476–489 (2016).
27. Huettmann, F., Mi, C. & Guo, Yu. 'Batteries' in machine learning: A first experimental assessment of inference for Siberian Crane Breeding Grounds in the Russian High Arctic Based on 'Shaving' 74 predictors. In *Machine Learning for Ecology and Sustainable Natural Resource Management* (eds Humphries, G. *et al.*) 163–184 (Springer, 2018).
28. Rogacheva, E. V. *The Birds of Central Siberia* (Husum Druck-Verlag, 1992).
29. Chen, J. *et al.* Prospects for the sustainability of social-ecological systems (SES) on the Mongolian plateau: five critical issues. *Environ. Res. Lett.* **13**, 123004. <https://doi.org/10.1088/1748-9326/aaf27b> (2018).
30. Sethi, S., Goyal, S. P. & Choudhary, A. N. Chapter 12: Poaching, illegal wildlife trade, and bushmeat hunting in India and South Asia. In *International Wildlife Management: Conservation Challenges in a changing World* (John Hopkins University Press, 2019).
31. Huettmann, F. Chapter 6: Effective Poyang Lake Conservation? A local ecology view from downstream involving internationally migratory birds when trying to buffer and manage water from HKH with 'Modern' concepts. In *Hindu Kush-Himalaya Watersheds Downhill: Landscape Ecology and Conservation Perspectives* (eds Regmi, G.R. & Huettmann, F.) 99–112 (Springer, 2020).
32. Yong, D. L. *et al.* The State of Migratory Landbirds in the East Asian Flyway: Distributions, Threats, and Conservation Needs. *Front. Ecol. Evol.* **9**, 613172. <https://doi.org/10.3389/fevo.2021.613172> (2021).
33. Deng, X. Q. *et al.* Contrasting trends in two East Asian populations of the Greater White-fronted Goose *Anser albifrons*. *Wildfowl Special Issue* **6**, 181–205 (2020).
34. Li, C. *et al.* Population trends and migration routes of the East Asian Bean Goose *Anser fabalis middendorffii* and *A. f. serrirostris*. *Wildfowl* **70**(6), 124–156 (2020).
35. Shimada, T., Mori, A. & Tajiri, H. Regional variation in long-term population trends for the Greater White-fronted Goose *Anser albifrons* in Japan. *Wildfowl* **69**, 105–117 (2019).
36. Wilson, R. E., Ely, C. R. & Talbot, S. L. Flyway structure in the circumpolar greater white-fronted goose. *Ecol. Evol.* <https://doi.org/10.1002/ece3.4345> (2018).
37. Bocharnikov, V. & Huettmann, F. Wilderness Condition as a Status Indicator of Russian Flora and Fauna: Implications for Future Protection Initiatives. *Int. J. Wilderness* **25**, 26–39 (2019).
38. Klein, D. R. & Magomedova, M. Industrial development and wildlife in arctic ecosystems: Can learning from the past lead to a brighter future? In *Social and Environmental Impacts in the North* (eds Rasmussen, R. O. & Koroleva, N. E.) 5–56 (Kluwer Academic Publishers, 2003).
39. Klein, D. R. *et al.* *Management and conservation of wildlife in a changing Arctic environment. Arctic Climate Impact Assessment (ACIA) ACIA Overview Report* 597–648 (Cambridge University Press, 2005).
40. Banerjee, S. *Arctic Voices: Resistance at the Tipping Point* (Seven Stories Press, 2012).
41. Zuckerberg, B., Huettmann, F. & Friar, J. Proper data management as a for reliable species distribution modeling. In *Predictive Species and Habitat Modeling in Landscape Ecology* (Drew, C. A. *et al.*) 45–70 (Springer, 2011).
42. Elith, J., Graham, C. & NCEAS Working Group. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 (2006).
43. Elith, J. *et al.* Presence-only and presence-absence data for comparing species distribution modeling methods. *J. Biodivers. Inform.* **15**, 69–80 (2020).
44. Tian, H. *et al.* Combining modern tracking data and historical records improves understanding of the summer habitats of the Eastern Lesser White-fronted Goose *Anser erythropus*. *Ecol. Evol.* <https://doi.org/10.1002/ece3.7310> (2021).
45. Ohse, B., Huettmann, F., Ickert-Bond, S. & Juday, G. 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 Biol.* **32**, 1717–1724 (2009).
46. Spiridonov, V. *et al.* Chapter 8 Toward the new role of marine and coastal protected areas in the Arctic: The Russian case. In *Protection of the Three Poles* (ed. Huettmann, F.) 171–201 (Springer, 2012).
47. Nixon, W., Sinclair, P. H., Eckert, C. D. & Hughes, N. *Birds of the Yukon Territory* (UBC Press, 2003).
48. Booms, T., Huettmann, F. & Schempf, P. Gyrfalcon nest distribution in Alaska based on a predictive GIS model. *Polar Biol.* **33**, 1601–1612 (2009).

49. Aycrigg, J. *et al.* Novel approaches to modeling and mapping terrestrial vertebrate occurrence in the Northwest and Alaska: An evaluation. *Northwest Sci.* **89**, 355–381. <https://doi.org/10.3955/046.089.0405> (2015).
50. Carboneras, C. & Kirwan, G. M. Greater White-fronted Goose (*Anser albifrons*). In *Handbook of the Birds of the World Alive* (eds del Hoyo, J. *et al.*) (Lynx Edicions, 2013).
51. Carboneras, C. & Kirwan, G. M. Bean Goose (*Anser fabalis*). In *Handbook of the Birds of the World Alive* (eds del Hoyo, J. *et al.*) (Lynx Edicions, 2014).
52. Hilborn, R. & Mangel, M. *The Ecological Detective: Confronting Models with Data* (Princeton University Press, 1997).
53. Breiman, L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Stat. Sci.* **16**, 199–231 (2002).
54. Betts, M. G. *et al.* Comments on “Methods to account for spatial autocorrelation in the analysis of species distributional data: A review”. *Ecography* **32**, 374–378 (2009).
55. Fox, C. H. F. *et al.* Predictions from machine learning ensembles: Marine bird distribution and density on Canada’s Pacific coast. *Mar. Ecol. Prog. Ser.* **566**, 199–216 (2017).
56. Fox, A. D. *et al.* Twenty-five years of population monitoring: The rise and fall of the Greenland White-fronted Goose *Anser albifrons flavirostris*. In *Waterbirds Around the World* (eds Boere, G. *et al.*) 637–639 (The Stationary Office, 2006).
57. Dornelas, M. *et al.* BioTIME: A database of biodiversity time series for the Anthropocene. *Glob. Ecol. Biogeogr.* <https://doi.org/10.1111/geb.12729> (2018).
58. BirdLife International. *Anser albifrons*. The IUCN Red List of Threatened Species 2016: e.T22679881A85980652. <https://doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22679881A85980652.en>. (Accessed 28th March 2021) (2016).
59. BirdLife International. *Anser fabalis*. The IUCN Red List of Threatened Species 2018: e.T22679875A132302864. <https://doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22679875A132302864.en>. (Accessed 28 March 2021) (2018).
60. Wang, X. L. *et al.* Stochastic simulations reveal few green wave surfing populations among spring migrating herbivorous waterfowl. *Nat. Commun.* **10**, 2187 (2019).
61. Ruokonen, M., Litvin, K. & Aarvak, T. Taxonomy of the bean goose–pink-footed goose. *Mol. Phylogenet. Evol.* **48**, 554–562 (2008).
62. Kölsch, A. *et al.* Flyway connectivity and exchange primarily driven by moult migration in geese. *Mov. Biol.* **7**, 3 (2019).
63. Yen, P., Huettmann, F. & Cooke, F. Modelling abundance and distribution of Marbled Murrelets (*Brachyramphus marmoratus*) using GIS, marine data and advanced multivariate statistics. *Ecol. Model.* **171**, 395–413 (2004).
64. Herrick, K. A., Huettmann, F. & Lindgren, M. A. A global model of avian influenza prediction in wild birds: The importance of northern regions. *Vet. Res.* **44**(1), 42. <https://doi.org/10.1186/1297-9716-44-42> (2013).
65. Gulyaeva, M. *et al.* Data mining and model-predicting a global disease reservoir for low-pathogenic Avian Influenza (AI) in the wider Pacific rim using big data sets. *Sci. Rep.* **10**, 1681. <https://doi.org/10.1038/s41598-020-73664-2> (2020).
66. Robold, R. & Huettmann, F. High-Resolution Prediction of American Red Squirrel in Interior Alaska: A role model for conservation using open access data, machine learning, GIS and LIDAR. *PEERJ*. <https://peerj.com/articles/11830/> (2021).

## Acknowledgements

We appreciate the opportunity for this international collaboration. All data collectors, including the pilots, are specifically thanked. Fieldwork was funded by the Russian Institutes (Institute of Biological Problems of the North, Far East Branch, Russian Academy of Sciences, Magadan, Russia, Institute of Biological Problems of Cryolithozone, Siberian Branch of the Russian Academy of Sciences, Yakutsk, Russia). Some of the fieldwork was funded by the Institute of IB with RCCE CAS, prof Cao Lei. FH thanks H. Berrios, E. Huettmann, the team around Mr. Chrome as well as the late Mr. Thor and Minitab-Salford. This is EWHALE lab publication #251.

## Author contributions

D.S., I.B.H., S.V. and A.K. collected data in the field. GIS compilations were done by A.K. and F.H., data analysis and research concept was done by F.H. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-01595-7>.

**Correspondence** and requests for materials should be addressed to F.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021