

# Modeling Eastern Russian High Arctic Geese (*Anser fabalis*, *A. albifrons*) during moult and brood rearing with Machine Learning, Open Access GIS and 24 years-long Big Data: Good Evidence for the New Digital Arctic of 2021

**Diana Solovyeva**

Russian Academy of Sciences

**Inga Bysykatova-Harmey**

Siberian Branch of the Russian Academy of Sciences

**Sergey L. Vartanyan**

North-East Interdisciplinary Scientific Research Institute N. A. Shilo, Russian Academy of Sciences

**Alexander Kondratyev**

Institute of Biological Problems of the North, Far East Branch, Russian Academy of Sciences, Magadan, Russia

**Falk Huettmann** (✉ [fhuetmann@alaska.edu](mailto:fhuetmann@alaska.edu))

University of Alaska Fairbanks (UAF)

---

## Research Article

**Keywords:** Russian High Arctic, breeding, moulting, Tundra Bean Goose *Anser fabalis* and Greater White-fronted Goose *A. albifrons*, Geographic Information Systems (GIS). Open Access Model Predictions

**Posted Date:** July 7th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-637037/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Many polar species and habitats are now affected by man-made global climate change and underlying infrastructure; it makes for a New Arctic. Arctic grazers, like Eastern Russian migratory populations of Tundra Bean Goose *Anser fabalis* and Greater White-fronted Goose *A. albifrons*, are affected along the entire flyway in East Asia, namely China, Japan and Korea. Here we present the best-available long-term 24 years digitized GIS data for the breeding study area (East Yakutia and Chukotka) and its habitats with ISO-compliant metadata. Further, we used seven publically available GIS predictor layers to predict the distribution for these two species within the tundra habitats. We are able to improve on the ecological niche prediction inference for both species by focusing for the first time specifically on biological relevant aspects: post-breeding moulting non-breeders, as well as post-breeding parent birds with broods. We then assessed it with 4 lines of evidence including alternative best-available open access field data from GBIF.org as well as compiled literature and found a good model accuracy in support of our evidence for a robust inference of these new findings. Our predictions indicate a relative index of occurrence (RIO) based on the quantified ecological niche showing more realistic gradual occurrence patterns and that 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 the first open access model platform for fine-tuning and future predictions for this otherwise poorly represented region in times of a highly changing industrialized 'new' arctic with global repercussions.

## Introduction

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 a 'New Arctic' (<sup>1,2</sup>). The Russian Eastern Arctic – Yakutia and Chukotka – are part of this process and 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>3,4,5,6,7,8,9,10</sup>). However, they are still comparatively poorly represented in the English literature thus far (<sup>11,12,13,14,15</sup>). For overall recent geese population estimates and outlines of their distribution in North East Asia see <sup>16</sup>. and <sup>17</sup>.; but this is based on expert opinion rather than on distribution modelling with underlying and 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</sup>., <sup>21</sup>., <sup>22</sup>.). Obtaining publically available primary data of ground and aerial surveys with supporting information - in a modern format - for this part of the world is very difficult despite the huge 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 publically shared and accessible for this region, yet. All awhile, 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 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.,31.,32.</sup>). Still, some *Anser* species are on the rebound and increase.

Put Fig. 1 app here (Circumpolar arctic and study area)

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 elgasí*' population. We consider populations attributed to the West Pacific branch of the East-Asian-Australasian Flyway (see definition of 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.

The arctic breeding grounds for these species lack 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 as pristine, they are not due to the overgrazing by abundant domestic reindeer (<sup>38.,39.,15.</sup>). Climate change and associated permafrost thawing and 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>); it's the New Arctic (<sup>2.</sup>) which happens to be digital <sup>1.,20.</sup> and with such 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 (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).

Progress on this topic 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 and Open Access baseline can help here to document and to 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 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</sup>, <sup>27</sup> for applications) we developed a workflow on how data can be compiled, cleaned, employed in GIS and modeled, subsequently assessed for performance, and be used and interpreted for inference.

### 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</sup>, <sup>37</sup>, <sup>46</sup>). Despite many decades of publications and efforts, the study area is still poorly and not systematically surveyed for its birds with no centralized database readily available for geese species or their habitats. A coordinated research design or species atlas with models does not yet exist for the study area (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 here we were attempt to compile for the first time representative area species data sets for the two study species in their summer grounds. We focus 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 field surveys performed by authors along the rivers and on lake in the tundra during the period of 1997–2020 by the methods of motor-boat surveys and aerial surveys. Visual surveys from moving motor-boat, from foot ground surveys around lakes and from 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</sup>, <sup>51</sup>). Their habitat requirements include food availability and safety from predators and people. That's because flightless geese, traditionally co-evolved with and hunted by people for centuries during their wing-moult, are 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 change at least for one month except for minor local small-distance movements.

Table 1

Description and details of data used for the relative index of occurrence (RIO) in this study for two goose species.

Species and sample size details	Location and subregion in Study Area	Survey focus	Source	Time period	Observer	Comment
<i>Anser albifrons</i> ( <i>n</i> = 219;	Yakutia	Brood & Non-breeders	Boat	2017, 2018	IB	
<i>Broods present</i> = 142, <i>Broods absent</i> = 77;	Chaun	Brood & Non-breeders	Boat	2002 til 2019	DS, AK	
<i>Non-breeders present</i> = 95	Chukotka	Brood & Non-breeders	Aerial	2002,	AK	
<i>Non-breeders absent</i> = 124)	Chukotka: Koryak Highland	Brood & Non-breeders	Aerial	1997	AK	
<i>Anser fabalis</i> ( <i>n</i> = 593;	Yakutia	Brood & Non-breeders	Boat	2017, 2018	IB	
<i>Broods present</i> = 213, <i>Broods absent</i> = 380;	Chaun	Brood & Non-breeders	Boat	2002 til 2019	DS, AK	
<i>Non-breeders present</i> = 303,	Chukotka	Brood & Non-breeders	Aerial	2002,	AK	
<i>Non-breeders absent</i> = 290)	Chukotka: Koryak Highland	Brood & Non-breeders	Aerial	1997	AK	
b) Testing Data						

Table 2

Predictor Importance Rank in model for two *Anser* species: brood-rearing parents with broods and moulting non-breeders.

Species	Source	Match with predictive model	Comment
<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 3

Data known to be available for the study area and of great use for this work but not publically made available (Note: Best professional practices, the International Polar Year, Migrator Bird Treaties, and most national funding schemes make Open Access data sharing mandatory).

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	100	100	90
Annual NDVI	82	65	89	100
Human Footprint	32	24	34	34
Elevation (ETOPO1)	100	73	92	99
* Data are readily available but show little coverage and information for the study area, yet.				

The data obtained during this geese census are a relative index of occurrence (RIO) and consist of decimal latitude, decimal longitude, observation time (24h) and date (day, month and year) as well as species presence and abundance and whether the birds were 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). The surveyed areas and their rivers and lake areas are listed in Appendix I a.

## Model assessment data

In addition to internal model metrics, it is important to assess models with alternative information, to confront models with reality<sup>52</sup>. Here we tried to inquire about the two geese species data that were not collected as part of the project. We located a dataset from GBIF.org and the literature. We also found other data sets like MOVEBANK and many research project data but those were not available in an Open Access format and thus could not be used by us.

## GIS data

Despite many decades of geological and geophysical survey work, modern GIS data layers for the study area is 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 Sriram and Huettmann unpublished (<https://essd.copernicus.org/preprints/essd-2016-65/>) and added open access layers as habitat predictors.

We selected GIS layers that are meaningful or that are 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 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 Appendix II 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 10.3 and OpenSource QGIS, and then overlaid them with GIS layers for the study area. The study area has a date line (180 degrees longitude located app. between Russia and Alaska) and 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</sup>.; see <sup>18</sup>. for applications). This was achieved by using Salford Predictive Modeler (SPM 8.3). We used exploratory models and their metrics, e.g. ROC. Model diagnostics are presented in the appendix. As this is the first model of its kind we did not emphasize specific questions of autocorrelation (for justification and conceptual details see <sup>54</sup>.). After creating a grove file in SPM with mostly default settings (known to perform best <sup>18</sup>.) to capture the actual model in a software format, we scored a app. 5km point lattice and obtained pixel-based predictions. Those 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) We used an internal ROC of the exploratory models as readily provided by SPM. 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 (Russian) 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 the models how they perform, using best-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. The data processing and GIS work was done in a computer lab in Russia and Alaska/US and all 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 publically-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 what was previously known; arguably, the distribution of this species is not as crisp as presented and assumed elsewhere.

broods, and b) post-breeding nonbreeders for the study area). Best-available GBIF

presence location for this species are overlaid for assessment.

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.

location for this species are overlaid for assessment.

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.

### **Model performance details and assessment**

#### Model performance details

Our models achieved good to very good accuracy. Predictors most strongly center around an interaction between climatic metrics like summer precipitation, temperature, as well as elevation and landcover categories, added by NDVI (detail shown in Appendix 4). 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 we identified NDVI as a powerful predictor with a positive relationship (Appendix 4). 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 (Appendix 4). Together with specific with 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 (Appendix 4). 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 predictor with a negative relationship (Appendix 4). Together with specific landcover classes it indicates where moulting flocks can be found in the study area.

#### Model assessment

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, 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 assessment, more independent but less specific for parents with broods and non-breeders, is 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).

Taken the evidences together, overall, we therefore think that the methodology shown (Fig. 2 for workflow) and results presented are a good start 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 while urgency in the study area (<sup>1., 2., 39., 40.</sup>) warrants good and immediate action though.

## 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>.). We further 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. They are of great use and relevance as quantified baselines for bird monitoring, range estimates and subsequent population estimations and conservation management of the assumed flyways for those strata.

The habitat GIS layers are also the first of their kind compiled for this species, 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>.).

Our models show 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's 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).

The patterns we found are robust, and as judged by the agreement with our four lines of alternative evidence. However, these patterns are new and 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.

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 a weak predictor in our model probably because our study area is among the least populated in the world (some industrial footprint does exist though). Interestingly, in another Arctic nesting *Anser* species, the Lesser White-fronted Goose *Anser erythopus*, 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).

available)

We would like to emphasize that our studied populations of both species are of the West Pacific Flyway, what means 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.</sup> and <sup>62.</sup> and <sup>62.</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>655</sup>. and <sup>63</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.

broods, and b) moulting non-breeders for the study area).

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</sup>,<sup>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 in time, in space and in the biology, e.g. for subspecies, timings and physiology, as it was started here. 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 publically 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.

## **Declarations**

## **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 Cryolitozone, 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. This is EWHALE lab publication #251.

## **References**

# Tables

Table 1: Description and details of data used for the relative index of occurrence (RIO) in this study for two goose species.

## a) Training Data

Species and sample size details	Location and subregion in Study Area	Survey focus	Source	Time period	Observer	Comment
<i>Anser albifrons</i> (n=219;	Yakutia	Brood & Non-breeders	Boat	2017, 2018	IB	
<i>Broods present</i> =142,	Chaun	Brood & Non-breeders	Boat	2002 til 2019	DS, AK	
<i>Broods absent</i> =77;	Chukotka	Brood & Non-breeders	Aerial	2002,	AK	
<i>Non-breeders present</i> =95	Chukotka:	Brood & Non-breeders	Aerial	1997	AK	
<i>Non-breeders absent</i> =124)	Koryak Highland					
<i>Anser fabalis</i> (n=593;	Yakutia	Brood & Non-breeders	Boat	2017, 2018	IB	
<i>Broods present</i> =213,	Chaun	Brood & Non-breeders	Boat	2002 til 2019	DS, AK	
<i>Broods absent</i> =380;	Chukotka	Brood & Non-breeders	Aerial	2002,	AK	
<i>Non-breeders present</i> =303,	Chukotka:	Brood & Non-breeders	Aerial	1997	AK	
<i>Non-breeders absent</i> =290)	Koryak Highland					

Table 1: Description and details of data used for the relative index of occurrence (RIO) in this study for two goose species.

## b) Testing Data

Species	Source	Match with predictive model	Comment
<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 2. Predictor Importance Rank in model for two *Anser* species: brood-rearing parents with broods and moulting non-breeders.

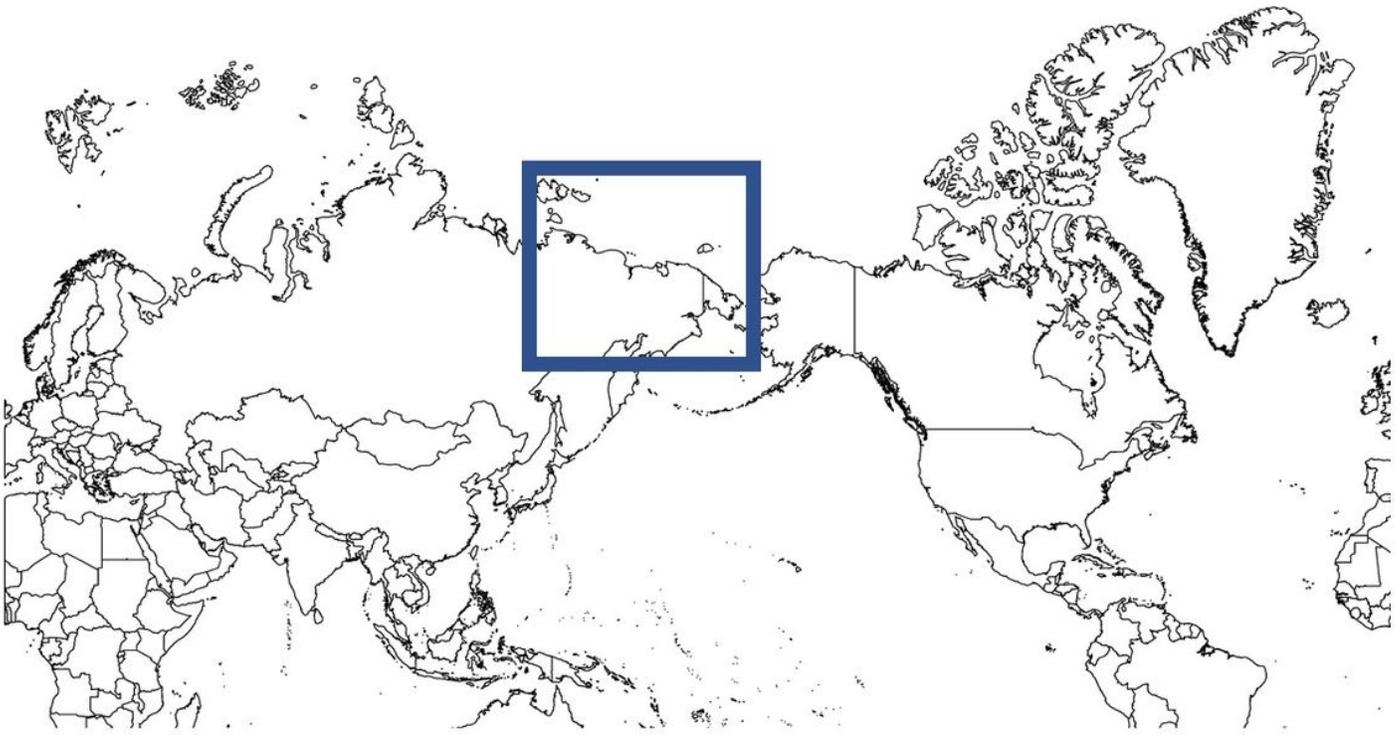
Predictor Name	Importance Rank in Model			
	Tundra Bean Goose: Non-breeders	Tundra Bean Goose: Brood	Greater White-fronted Goose: Non-breeders	Greater White-fronted Goose: Brood
Global Landcover	85	67	93	93
Mean Temperature in July	83	61	76	84
Mean Precipitation in July	60	100	100	90
Annual NDVI	82	65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

Table 3: Data known to be available for the study area and of great use for this work but not publically made available (Note: Best professional practices, the International Polar Year, Migrator Bird Treaties, and most national funding schemes make Open Access data sharing mandatory).

\* Data are readily available but show little coverage and information for the study area, yet.

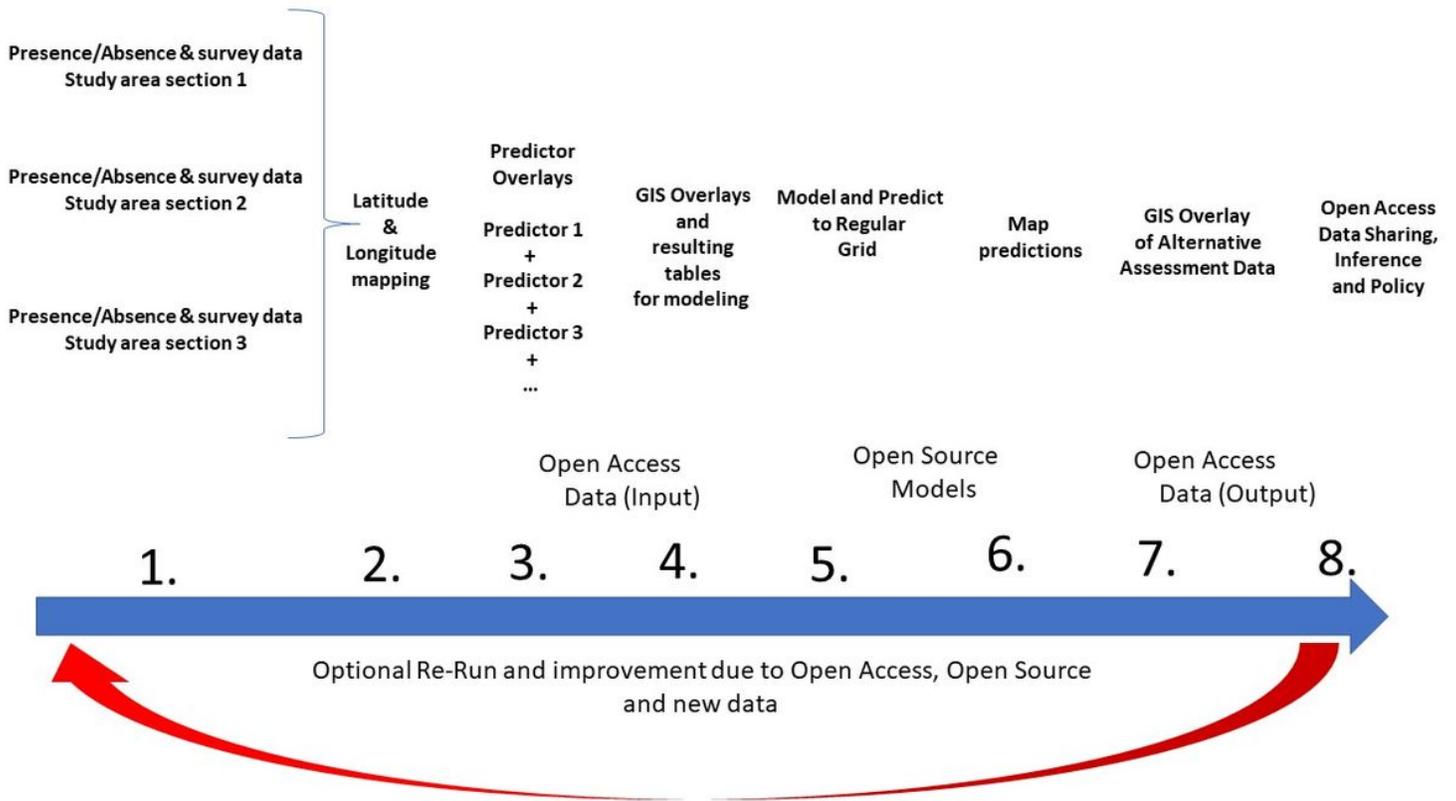
<b>Data set name</b>	<b>Content</b>	<b>Source</b>	<b>Comment</b>
Movebank	Geolocations	Various funders	Data blocked behind login
Goose tagging	Locations	China	Data not made publically available regardless of publication
Bird Banding	Banding location, resighting and recovery	National Bird Banding Center	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 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

## Figures



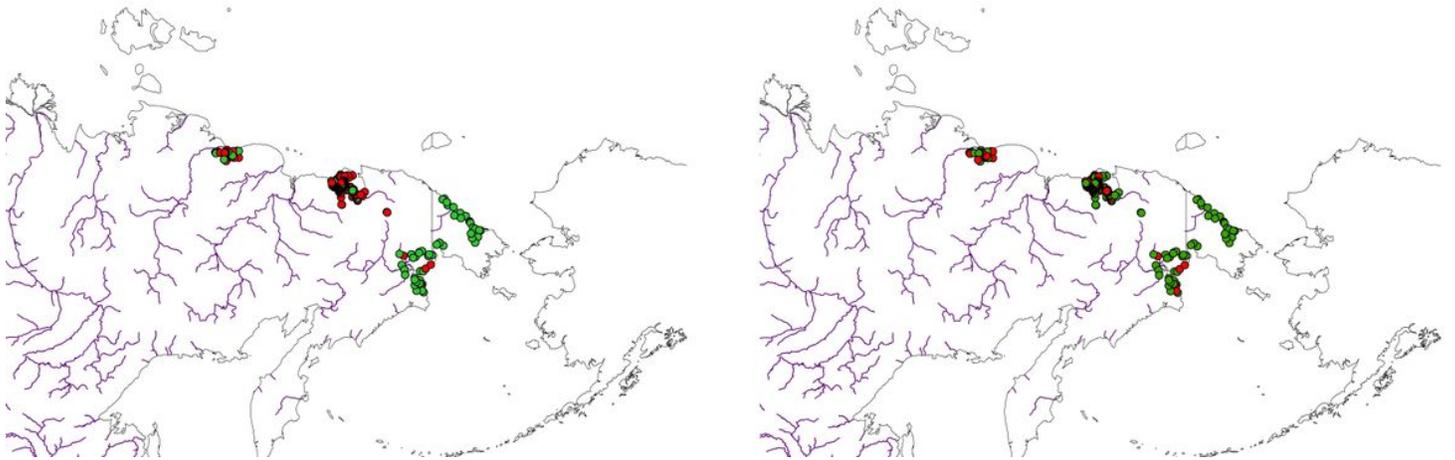
**Figure 1**

Circumpolar arctic and study area



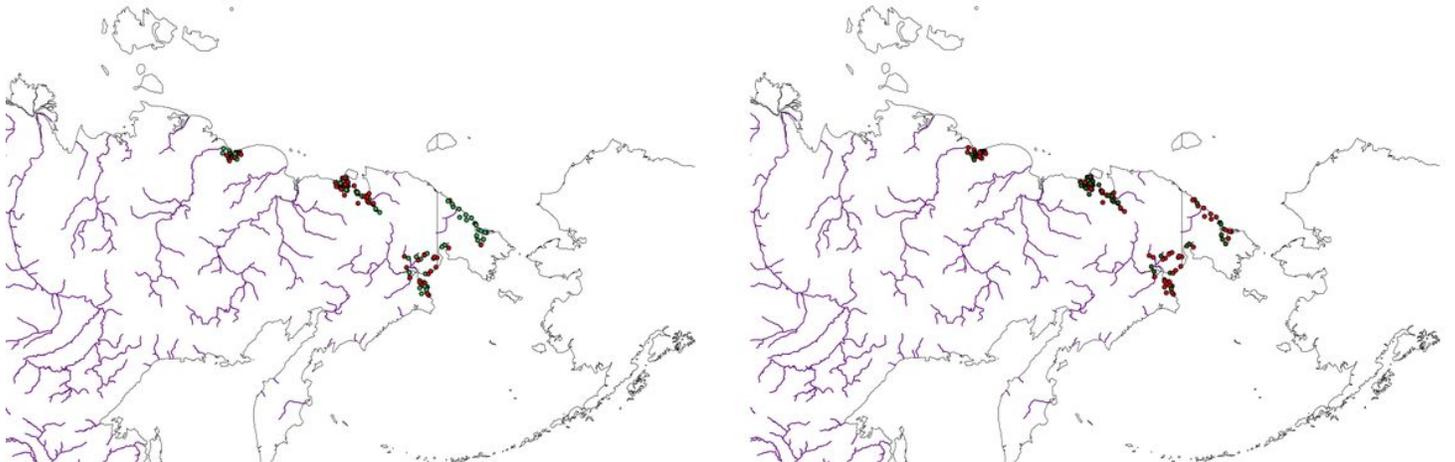
**Figure 2**

workflow



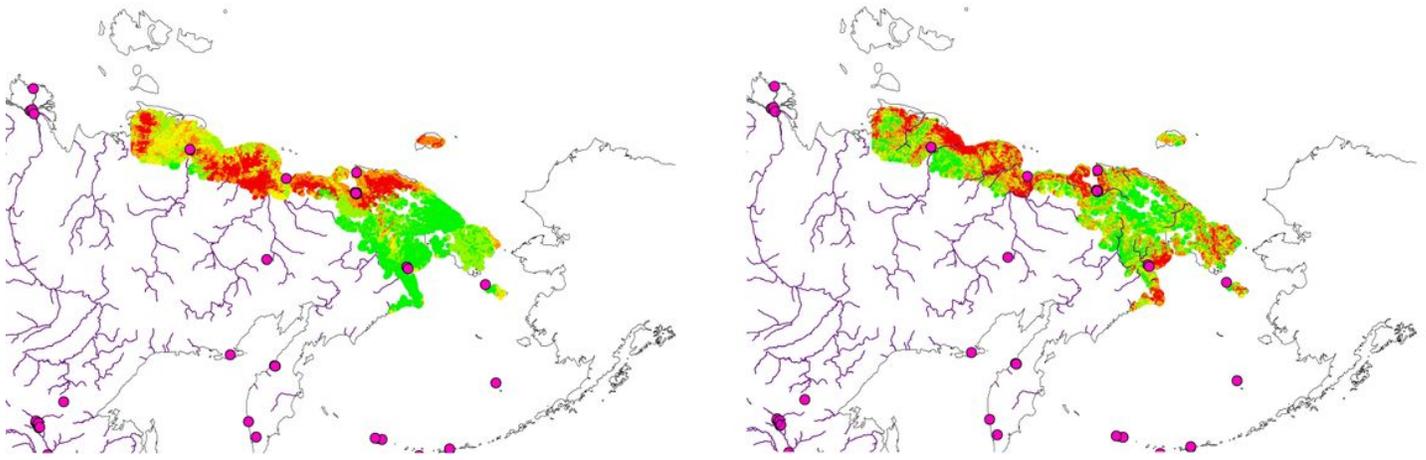
**Figure 3**

Best-available compiled raw data of *Anser fabalis* presence/absence for a) Brood, and b) Non-breeders in the study area



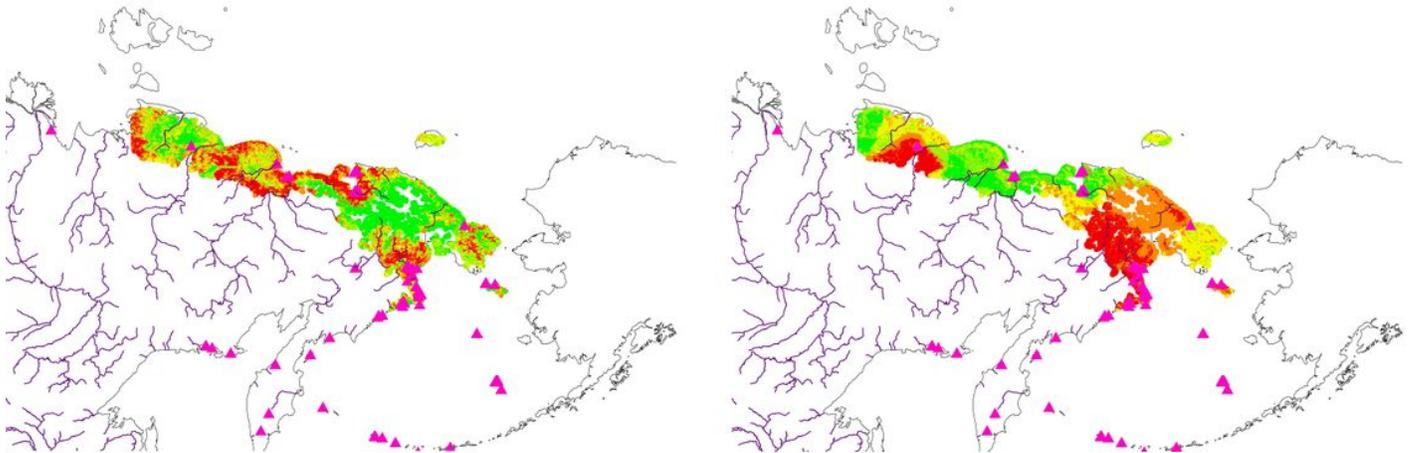
**Figure 4**

Best-available compiled raw data of *Anser albifrons* presence/absence for a) Brood, and b) Non-breeders in the study area



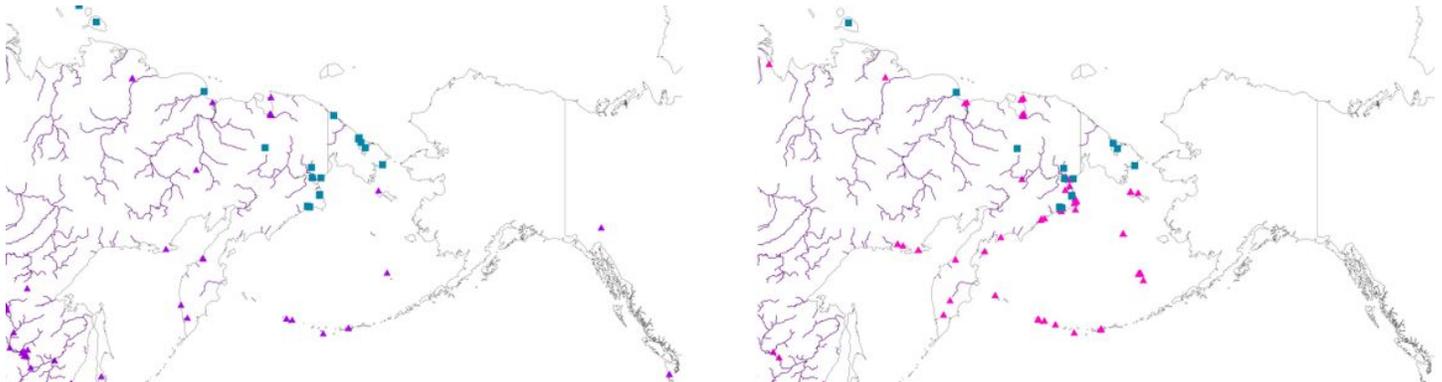
**Figure 5**

Predictions of Tundra Bean Goose a) brood-rearing parents with broods, and b) post-breeding nonbreeders for the study area). Best-available GBIF presence location for this species are overimposed for assessment.



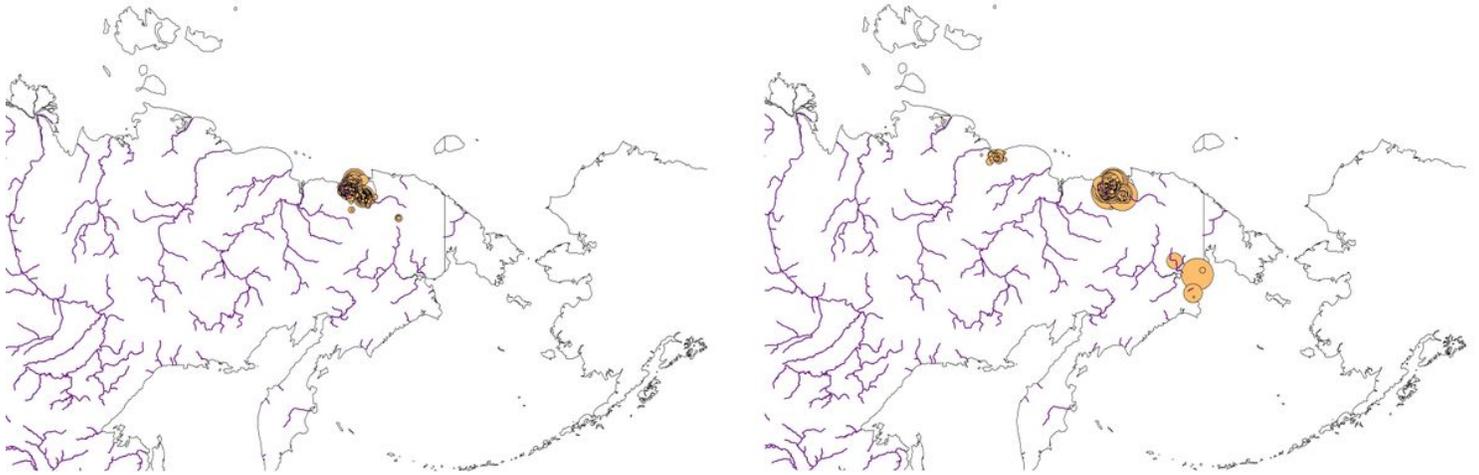
**Figure 6**

Predictions of Greater White-fronted Goose a) parents with broods, and b) moulting nonbreeders for the study area) Best-available GBIF presence location for this species are overimposed for assessment.



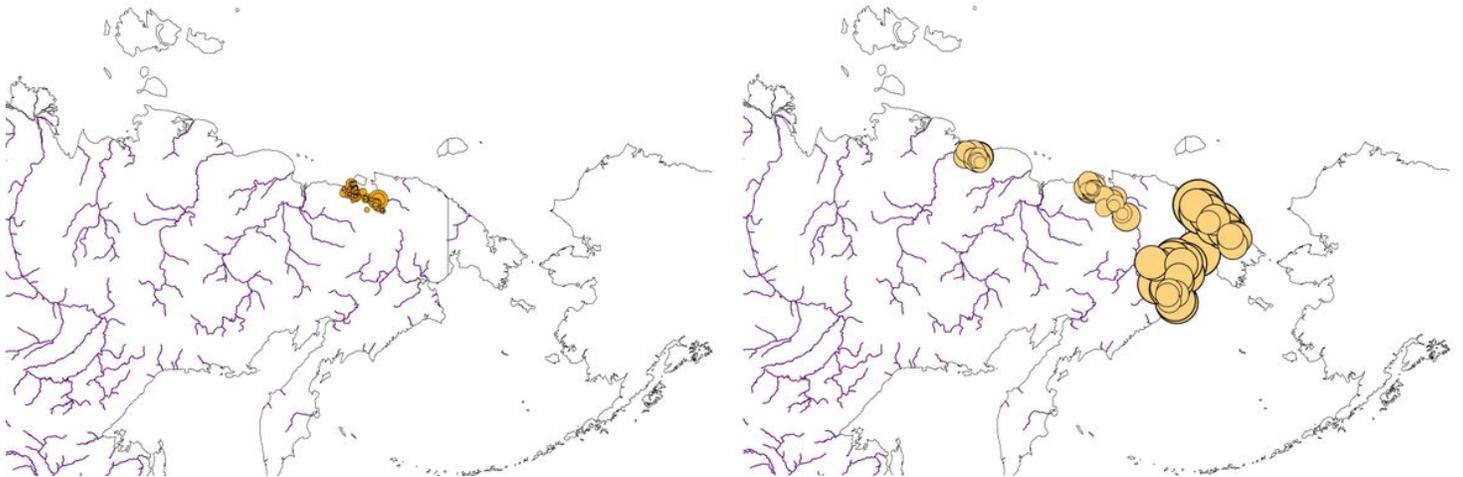
**Figure 7**

The third assessment, more independent but less specific for parents with broods and non-breeders, is 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; figure 7). But at least on a generic level it shows a very high match for the models (compare with figures 5, 6).



**Figure 8**

Survey numbers for Tundra Bean Goose a) broods, and b) moulted non-breeders for the study area



**Figure 9**

Survey numbers for Greater White-fronted Goose a) broods , and b) moulted non-breeders for the study area

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [HighArcticModelChaun2020FH7SciRepAppendixIIILocations.docx](#)
- [HighArcticModelChaun2020FH7SciRepAppendixV.docx](#)

- [Appendix4aModelOverlays.zip](#)
- [Appendix2BaseMapPacRim3merc.zip](#)
- [Appendix1b.zip](#)
- [APPENDICESMainOverviewFH1.docx](#)
- [Appendix1b.zip](#)
- [Appendix4bPredictions.zip](#)
- [HighArcticModelChaun2020FH7SciRepAppendixIIILocations.docx](#)
- [HighArcticModelChaun2020FH7SciRepAppendixIIList.docx](#)
- [HighArcticModelChaun2020FH7SciRepAppendixV.docx](#)