Ecological Informatics 78 (2023) 102364

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

# Model-based prediction of a vacant summer niche in a subarctic urbanscape: A multi-year open access data analysis of a 'niche swap' by short-billed Gulls

Falk Huettmann [a,*,1], László Kövér [b], Richard Robold [a,c], Mark Spangler [a], Moriz Steiner [a,d]

[a] EWHALE Lab, Institute of Arctic Biology, Biology & Wildlife Department, University of Alaska Fairbanks, Alaska, United States of America
[b] Department of Nature Conservation, Zoology and Game Management, University of Debrecen, Böszörményi Str. 138, 4032 Debrecen, Hungary
[c] University of Potsdam, Am neuen Palais 10, 14469 Potsdam, Germany
[d] IUCN Small Mammal Specialist Group (SMSG) and IUCN Species Survival Commission (SSC), IUCN, Rue Mauverney 28, 1196 Gland, Switzerland

A R T I C L E I N F O

*Keywords:*
Short-billed (common/mew) Gull *Larus canus*
Geographic information system (GIS)
Open access data
Big data
Socio-economics
Multi-year field work
Machine learning ensemble predictions
(RandomForest Treenet CART MARS)

A B S T R A C T

Gulls belong to the seabird group, and are widely considered to be useful ecological indicators. Increasing urbanization throughout the Anthropocene has led to the rise of the urban landscape ('urbanscape') as a globally dominant habitat. Though historically less developed, similar urbanization patterns are emerging in the boreal forest, the world's largest forest ecoregion. Here we evaluate the ecological position of migratory Short-billed Gulls (*Larus canus*) in a subarctic urbanscape. We attempt to investigate a summer niche swap theory, in which gulls annually fill a vacant niche otherwise occupied by common ravens in winter. For that investigation we conducted seasonal plot surveys for Short-billed Gull presence from 2013 to 2016 in the Fairbanks, AK municipality. All data were made publicly available. We compiled them in a open source and ESRI geographic information system (GIS) platform and then added 68 open-access predictor layers, including socio-economic U. S. Census data. We trained, tested, and evaluated the performance of an ensemble of machine learning models, resulting in predictions of gull-abundance hotspots and coldspots, at 100-m resolution for inference. We find that Short-billed Gulls prefer the synergy of industrial areas near man-made water bodies, impervious surfaces, gravel pits, strip malls, transfer sites (garbage dumps) and some young forest vegetation. This study is a first-known attempt to utilize a blended 'Big Data' approach, in combination with traditional multi-year field-based data collection and alternative model assessments, in order to characterize an urban seabird niche. Our findings, and the digital infrastructure herein, provide an interdisciplinary baseline for potential applications in urban planning and monitoring the spread of disease reservoirs.

## 1. Introduction

Seabirds are important ecological indicators (Piatt et al., 2007). While not always perfect, they can often get easily monitored, and largely reflect the intricate, less observable processes of the larger ecosystems in which they are fully embedded (Durant et al., 2009). Gulls are an inherent part of this polyphyletic species group. While they are usually associated with oceans and coasts, some gull species are more closely associated with freshwater and occur farther inland (e.g. Kirk et al., 2008). The traditional view of gulls as being tied to oceans ("*sea gulls*") is changing in the public eye as an increasingly high number of gulls are documented in urbanized and in-land environments (e.g. Rock, 2005), foraging, roosting and nesting there (Auman et al., 2008, 2011; Belant, 1997; Huig et al., 2016; Pais de Faria et al., 2021; Zelenskaya, 2019, 2021). They use urbanized coastal regions and industrial ports (Huettmann et al., 2000), and inland ecosystems, including agricultural lands, roof tops (Jiménez et al., 2023; Rock, 2005), developed mountain resorts and road-side communities (Huettmann pers. com. for Denali National Park region). This trend is part of an ongoing coevolution of synanthropic species with humans in the Anthropocene (Langley, 2021; Louise, 2020; Ouled-Cheikh et al., 2021; Schilthuizen, 2018). This coevolution has resulted in several species becoming globally favored and, thus, they are globally on the rise (e.g., many species of cockroaches and mosquitoes, Norwegian rat, deer, red fox; Bonnefoy et al., 2008).

* Corresponding author.
*E-mail address:* fhuettmann@alaska.edu (F. Huettmann).
[1] Co-authorship in alphabetic order.

https://doi.org/10.1016/j.ecoinf.2023.102364
Received 2 August 2023; Received in revised form 28 October 2023; Accepted 1 November 2023
Available online 10 November 2023
1574-9541/

Birds are part of that trend and avian examples include mynas, sparrows, swallows, pigeons, (Hansen and Huettmann, 2020, Barua, 2022; Carlen and Munshi-South, 2021), owls, corvids (Kövér et al., 2015) and gulls, among others.

A main driver behind the rise of synanthropic species is human food subsidies, a byproduct of increasing industrialization and globalization. Gulls are increasingly found in urbanscapes, but are particularly associated with impervious surfaces where food waste is found in abundance, such as near garbage dumps and associated strip malls and supermarkets (Baltensperger et al., 2013; Burger, 1981; Burger and Gochfeld, 1983; Gabrey, 1997; see Weiser and Powell, 2011 for Alaska). The known implications of large congregations of gulls resulting from human food subsidies include for instance altered behavior, microbiome, parasites and disease load (Fuirst et al., 2018; Goumas et al., 2020). Other factors include stress, death and morphological changes, e.g. Benmazouz et al., 2023 for body size changes.

With urbanization increasing globally, a pattern of synanthropic species is also beginning to emerge in the boreal forest - the largest consistent forest ecosystem in the world. Historically, the boreal forest has had few trails and roads and has been sparsely settled (e.g. Betts et al., 2022; Huettmann and Young, 2022), but habitat fragmentation, industrialization and the development of impervious surfaces are increasingly on the rise (Bocharnikov and Huettmann, 2019; Johnson et al., 2015). Subsequently, gulls have already entered the urbanized boreal forest regions, though little has been documented yet with regards to this habitat shift and over time (Sinclair et al., 2011).

Alaska is located in the far northwestern corner of the boreal forest in North America, and the Fairbanks municipality represents a major urban hub (population approximately 90,000) in that region (Fig. 1). While studies of avian populations in the urbanscape are few for those regions (Hedblom and Murgui, 2017), common ravens (*Corvus corax*) were previously investigated in Fairbanks over a five-year period during the winter and summer months (Baltensperger et al., 2013; Beaulieu, 2022). Annual occurrence patterns, as well as winter roosts, were affiliated with fast food and supermarket parking lots. These studies documented a generalabsence of ravens in downtown Fairbanks from May thru August, a time when migratory gulls – specifically Short-billed Gulls (*Larus canus*)[2] – were observed to occur in their place. This observational pattern by coauthor FH started an investigation of a niche swap (shift) theory for the study area. Generally, we investigate whether for same locations in downtown Fairbanks ravens appear to be replaced in summer with gulls.

While data were lacking on this topic, here we investigate this indicated niche pattern and try to provide the first quantified description using modern methods explicit in space and time. In addition to traditional year-long field-based sampling methods, including opportunistic repeat surveys in downtown Fairbanks, we also employ a complex set of high-resolution open access GIS data layers as model predictors for describing the ecological niche of the gulls, including 'proxies' (e.g. Russo, 2011) and socioeconomic U.S. Census data. Our analyses follows Huettmann, 2011 and Huettmann and Arhonditsis (2023), and it relies on GIS and non-parametric methods of machine learning ensemble predictions as a robust platform for inference (Breiman, 2001). These methods are based on a powerful, but rarely-used, field research design using model-predicted inference with alternative assessments explicit in space and time over the years based on different lines of evidence for a consolidated inference (Humphries et al., 2018). This is progress and departure from traditional studies which tend to use parametric and parsimonious approaches to inference, but which are known to be biased, limited and often faulty when used indiscriminately (Humphries et al., 2018 and citations within). The datasets and analyses described here characterize a 'Big Data' approach. This term can have many meanings. For avian research, especially in Interior Alaska, the use of >2 years of data and >60 GIS predictors, including complex and socio-economic predictors can justifiably be considered a 'Big Data' approach in avian research and conservation (Huettmann et al., 2018; Karmacharya et al., 2020). It moves towards a more holistic and inclusive approach (*sensu* Naess and Jickling, 2000, Huettmann, 2015).

Put Fig. 1: Study area with roads, transfer stations, restaurants, railroads and water bodies.

## 2. Methods

### 2.1. Field data collection

#### 2.1.1. Presence/absence surveys

In order to investigate the ecological niche of the Short-billed Gull in interior Alaska in summer we carried out an exhaustive opportunistic survey in the Fairbanks municipality between May to June 2013. In this survey, we established eighty ($n = 80$) georeferenced plots for the study area (Fig. 2). Within each plot, we recorded all Short-billed Gull observations over a 5-min period. Additionally, we performed a 360-degree scan with unconstrained detection widths (not used in this study). We conducted repeat surveys in May of 2014 and 2015, revisiting a subset ($n = 50$) of the previously established plots (Fig. 3).

The data for the three survey years in summer (~early nesting period for Short-billed gulls) were compiled by plot. The combined dataset lists each individual sighting, as well as detection width, when feasible. Though the summer months serve as the nesting season for Short-billed gulls, and may influence detection probability, it should be kept in mind that the population also consists of quite a high number of nonbreeding and failed individuals often found in clusters and roosts. Using 'presence' starts such investigations and allows to capture this species across all behaviors, e.g. nesting, roosting and foraging strata.

#### 2.1.2. Abundance surveys

During summer 2017, from April through July, we conducted repeat surveys at a single location to document change in abundance over time throughout the nesting season. The selected site – a grocery store parking lot in a commercial center of Fairbanks – represents a known presence location. Additionally, it is a documented key area for common ravens in winter (Baltensperger et al., 2013), allowing for an investigation of an urban niche swap theory between these two avian predators. We surveyed the site for Short-billed Gull occurrence using a 360-degree sweep count. The site was revisited 16 times, loosely at 2–10 day intervals throughout the nesting season allowing for an abundance index throughout summer for a presence plot. This location presented a core area for wintering ravens in the study area (details in Baltensperger et al., 2013).

Put Fig. 2 app here: a) 50 survey location data (black dots) for Short-billed Gulls (*Larus canus*) during May and June 2014 and 2015. b) 50 survey location data (presence = purple dots, absence = green dots) for Short-billed Gulls (*Larus canus*) during May and June 2014 and 2015. These surveys are the model training data.

Put Fig. 3: app here:a) U.S. Census blocks (in red) as well as 80 survey location data (black dots; presence/absence surveys) for Short-billed Gulls (*Larus canus*) during May 2013. b) Presence (pink dots)/absence (green dots) for 80 survey location for Short-billed Gulls (*Larus canus*) during May 2013. These point data serve as the external model assessment data.

### 2.2. GIS data compilation and processing

#### 2.2.1. Predictor layers

Following Baltensperger et al. (2013), we utilized the Fairbanks North Star Borough (FNSB) GIS Web Services database (https://www.

[2] The Mew Gull *Larus canus* was split 2021 into two species, a Kamchatka Common Gull and the Short-billed Gull in North America (https://www.sibleyguides.com/2021/07/mew-gull-is-now-two-species-how-to-identify-common-gull-and-short-billed-gull/). This study deals with the latter species.

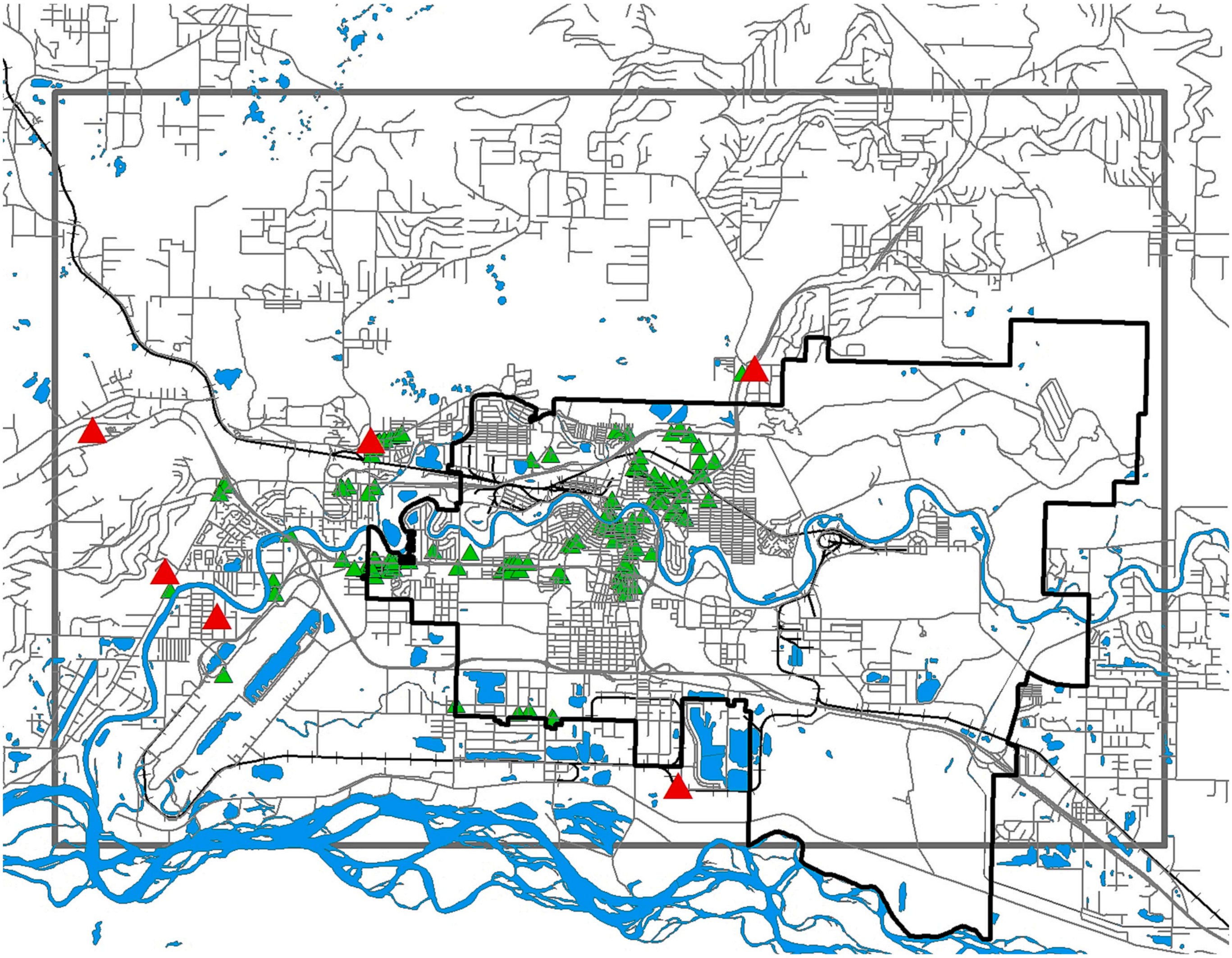


**Fig. 1.** Study area (black squared frame) with roads (thin grey lines), transfer stations (red triangle), restaurants (green triangles), rail roads (thick black lines) and water bodies (blue). The Fairbanks city region is shown with the rugged black polygon. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fnsb.gov/438/Get-FNSB-GIS-Data) to acquire spatial layers characterizing various measures of urban habitat (Table 1). This GIS database is publicly accessible and carries a University of Alaska Fairbanks (UAF) campus research license. The data are projected in Alaska Stereographic (ESRI:102633 NAD 1983 SPCS Alaska 3 (Feet)). We used the Distance (Euclidean) tool (Spatial Analyst) in ArcMap v8.1 to compute proximity grids for the restaurant ($n$ = 138), transfer station ($n$ = 6), water (polygon), railroad (polyline), and road (polyline) layers (Baltensperger et al., 2013). These grids serve as proxies for feature-related impacts and have been documented to act as powerful predictor variables (e.g. Humphries et al., 2018; Russo, 2011). Each predictor essentially presents a hypothesis tested in concert to explain gull occurrence. Using many predictors allows for a 'correlational science' where strong correlations present habitat associations. Classic predictors like elevation and climate are not used here because the study area is relatively flat, with no significant elevational ruggedness or gradient, and climate variations are quite small.

Put Table 1 app. here: GIS layers used.

In addition to the FNSB data layers described above, we acquired spatial layers for soil type (Ohse et al., 2009), vegetation class (Tanana Valley State Forest, in Steiner and Huettmann, 2023), and from the U.S. Census 2000 dataset (https://live.laborstats.alaska.gov/cen/index.html). The latter socio-economic dataset includes many attributes, with a resolution determined by census block (Fig. 2). The selected attributes are presented in Table 2 (https://live.laborstats.alaska.gov/erg/occmanual2010.pdf; https://www2.census.gov/geo/pdfs/reference/glossry2.pdf; https://www.rand.org/content/dam/rand/www/external/labor/projects/dnors/pubs/pdfs/DNORS_contextual_block_listing.pdf).

Put Table 2 app. here: US Census 2000 layer details, those are predictors included and tested individually in the ensemble model as part of the 68 predictors.

### 2.2.2. *Data cubes*

We imported all of the compiled predictor layers into GIS with the combined 2014–15 Short-billed Gull presence/absence survey dataset. We used the 'Extract Multi Values to Points' (Spatial Analyst) and 'Spatial Join' (Analysis) tools to add each of the predictor variables to the survey dataset. The resulting data cube contained 2 × 50 rows (50 per year) and 68 associated predictor columns (Booms et al., 2010, 2011, Ohse et al., 2009). The same was done for the 2013 survey dataset, resulting in a data cube of 80 rows and 68 predictor columns respectively.

To extrapolate model predictions across the entire study area, we created a lattice grid vector layer with a point spacing = 100 m ($n$ = 31,244 points). Using the tools and processes described above, we added

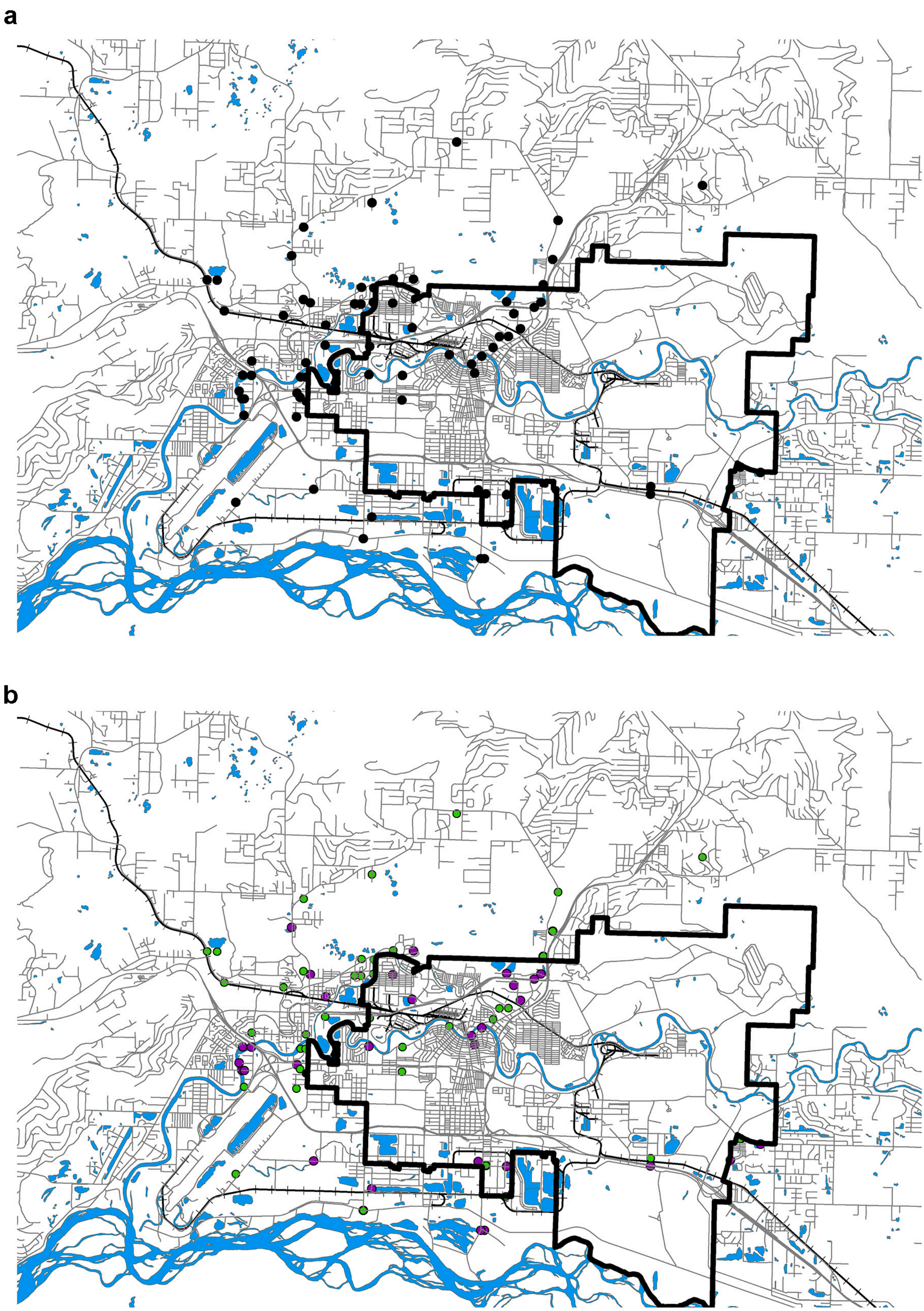
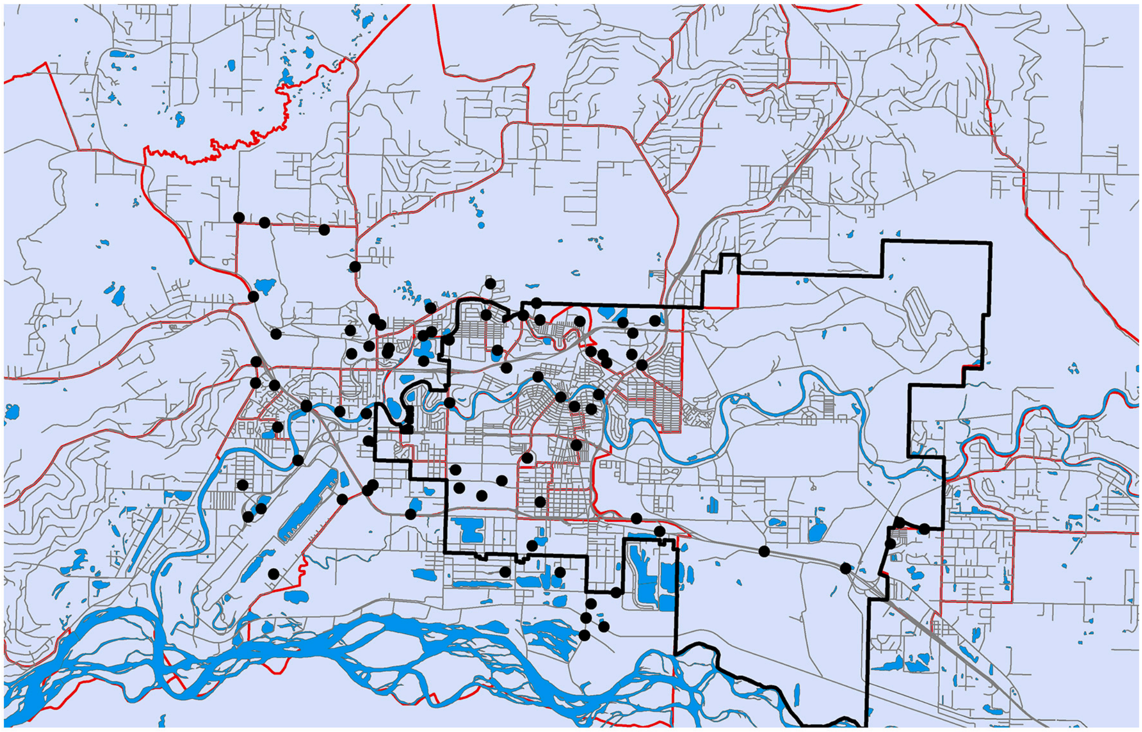



**Fig. 2.** a) U.S. Census blocks (in red) as well as 80 survey location data (black dots; presence/absence surveys) for Short-billed Gulls (*Larus canus*) during May 2013. b) Presence (pink dots)/absence (green dots) for 80 survey location for Short-billed Gulls (*Larus canus*) during May 2013. These point data serve as the external model assessment data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a



b

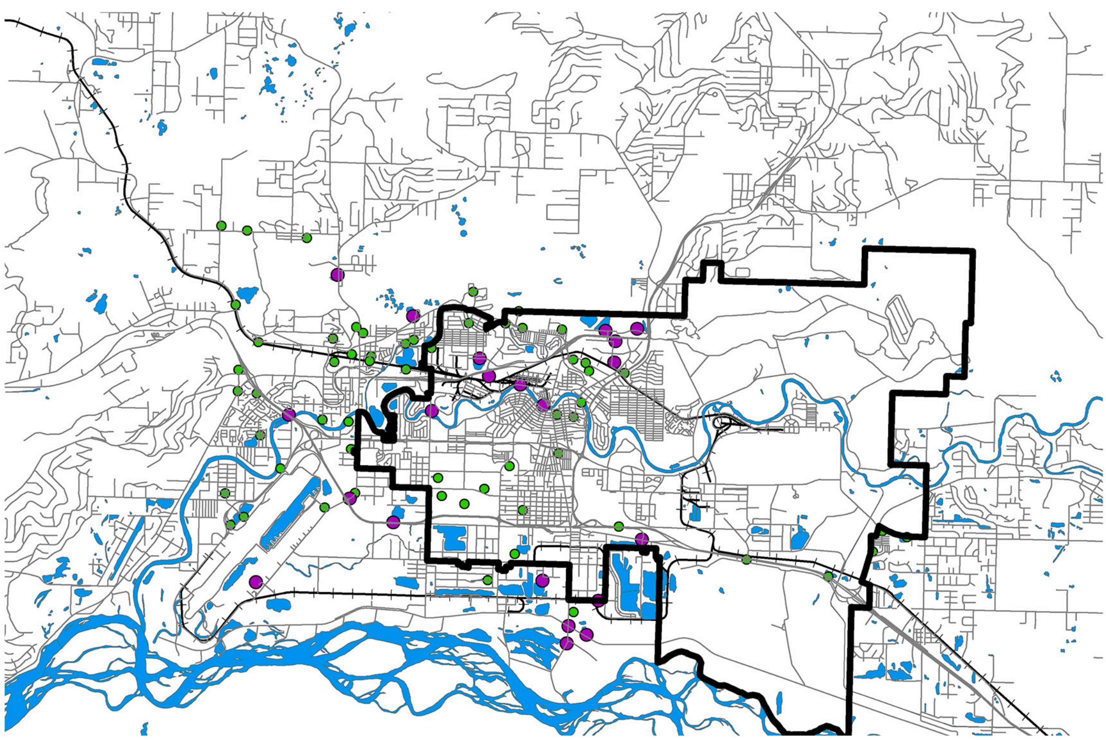


**Fig. 3.** a) 50 repeat survey location data (black dots) for Short-billed Gulls (*Larus canus*) in the study area during May and June 2014 and 2015. b) 50 survey location data (presence = purple dots, absence = green dots) for Short-billed Gulls (*Larus canus*) during May and June 2014 and 2015. These surveys are the model training data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

each of the predictor variables to the lattice point dataset. We used the 'Add XY Coordinates' tool (Data Management in ArcGIS) to provide spatial reference and subsequent GIS mapping. The resulting data cube contained 31,244 georeferenced rows and 68 associated predictor columns matching the earlier data cube.

Finally, we created a 2013–2015 pooled dataset to visually explore hotspots and coldspots of Short-billed Gull abundance (=numbers of birds seen).

**Table 1**
GIS layers used for predictors.

| Predictor layer | Source | Comment |
|---|---|---|
| Distance to Transfer Station | Fairbanks Burrough GIS, ARCGIS | Value-added map product by FH |
| Distance to Road | Fairbanks Burrough GIS, ARCGIS | Value-added map product by FH |
| Distance to Water | Fairbanks Burrough GIS, ARCGIS | Value-added map product by FH |
| Distance to Restaurant | Fairbanks Burrough GIS, ARCGIS | Value-added map product by FH |
| Distance to Railway | Fairbanks Burrough GIS, ARCGIS | Value-added map product by FH |
| Vegetation name | Forest Layer Tanana Valley State Forest | |
| Soil type | Alaska Clearing house | |
| US Census 2000 (60 attributes) | Alaska US Census portal | See Table 2 for details |

### 2.3. Machine learning ensembles

We used the combined 2014–2015 Short-billed Gull presence/absence data cube to train an ensemble of machine learning (ML) models in Salford Predictive Modeler (SPM) v8.3 (Formula 1). We used several established and leading ML algorithms (Hegel et al., 2010; Fernández-Delgado et al., 2014 for overview) and combined them in an averaged ensemble to obtain the best-possible prediction and overall synergy from each algorithm best as possible (*sensu* J. Friedman '*many weak learners make for a strong learner*'; Friedman, 2002, Hastie et al., 2009. See also Hardy et al., 2011, Fox et al., 2017 and Boulanger-Lapointe et al., 2022 for applications). That approach will overcome stand-alone approaches such as linear regression, which do not offer unique solutions, e.g. Ascombe's Quartet, and have difficulty to meet all required parametric assumptions (McArdle, 1988). The specific algorithms and their settings are described below.

**Formula 1**

$$\begin{aligned}\text{Presence/Absence} \sim \text{Distance to Road} &+ \text{Distance to Transfer Station}\\ &+ \text{Distance to Railways}\\ &+ \text{Distance to Water}\\ &+ \text{Distance to Restaurants}\\ &+ \text{Soil} + \text{Vegetation}\\ &+ \text{U.S.2000 census}\end{aligned}$$

#### 2.3.1. CART decision trees

We performed a tree-based CART Decision Tree analysis using all 68 predictors and PA01 as the target (response; also used for all other algorithms). The settings in SPM were maintained as default (fraction of cases sampled at random set at 0.1, the best tree is selected to be "Minimum cost tree regardless of size", tree type set to "Gini", and with all predictors having the same weight). Limits were set to five minimum samples for a node split and two minimum samples for terminal nodes. This adjustment was made to allow for a more sophisticated decision tree, accommodating the high number of predictors.

#### 2.3.2. CART ensembles and bagger

We performed a CART Ensembles and Bagger analysis using all 68 predictors. These analyses belong to the tree algorithm family, but combine trees using a simple "bagging" concept (Breiman, 2001). The same settings were used as for the CART Decision Trees analysis.

#### 2.3.3. TreeNet (stochastic gradient boosting)

We performed a TreeNet analysis using default settings, 600 trees, and balanced weights. TreeNet algorithms employ stochastic gradient boosting, a proven approach in data mining with high predictive accuracy. TreeNet algorithms are also tree-based, with enhanced robustness via a 'boosting' concept (Friedman, 2002). As such, TreeNett is an algorithm of choice for similar data mining analyses (Booms et al., 2010, 2011, Ohse et al., 2009, Baltensperger et al., 2013).

#### 2.3.4. RandomForest (bagging)

We performed a RandomForest analysis using 200 trees and balanced weights. Similar to CART Ensembles and Bagger, RandomForest belongs to the tree algorithm family, but then is enhanced via 'bagging' (Breiman, 2001). Prior studies have shown RandomForest to be a superior choice for predictive accuracy in classification problems (Mi et al., 2017) and inference (Breiman, 2001).

#### 2.3.5. MARS regression splines

In addition to the tree-based machine learning algorithms described above, we also conducted a MARS analysis, based on regression splines. We used the same combined 2014–2015 presence/absence data cube and applied the MARS Regression Splines analysis to the 68 predictors to analyze which predictors best inform the summer niche (and a possible shift) of Short-billed Gulls in a subarctic urbanscape. Settings were retained as default except for the testing method, which was set to "Fraction of cases selected at random as 0.1 - exact". Additionally, limits were set to 30 maximum basis functions, and speed factor to 3/5 (medium accuracy vs time). All predictors were included in the analysis. The MUNAME and VEG_NAME predictors were assigned as categorical variables, all other predictors as continuous variables. "PA01" was set as the target variable. Here the target type was set to "Regression", though the outcome does not noticeably change if set to "Classification/Logistic Binary".

#### 2.3.6. Ensemble models and their assessments

We scored each of the five machine learning model outputs individually, also computing the individual algorithm receiver operating characteristic (ROC) curves, and calculated an average score for the combined ensemble (for details see also Boulanger-Lapointe et al., 2022). The scoring included x and y variables to map the findings in GIS. The ensemble model performance was tested against the alternative and independently-collected 2013 Presence/Absence data cube.

## 3. Results

### 3.1. Gull distribution and abundance

This study is the first that has compiled multi-year data (2013–2016) with high sample numbers from a research design on Short-billed Gulls and other birds in subarctic urban Alaska, all provided in an open access framework (see Figs. 2 and 3 and appendix; sensu Huettmann, 2011, 2015). The prime emphasize is on 'presences' and a model prediction inference for the Short-billed Gull during summer in the 'raven niche' (winter; details on the distribution pattern in the model section). However, abundance and multi-species data were also collected allowing for further evidence and analysis (see Discussion for further interpretation and research suggestions).

### 3.2. Machine learning ensemble predictions for short-billed Gull presence/absence

The output of the ensemble model expresses predicted Short-billed Gull presence/absence using an averaged relative index of occurrence (RIO). Rasterization of RIO values extrapolated to the lattice data cube allows visualization of a gull abundance "heatmap," with areas of high gull occurrence shown in warm colors (hotspots) and low gull occurrence shown in cool colors (coldspots), across the Fairbanks municipality (Fig. 4; see Fig. 5 for time series of such an urban presence hotspot within the model).

The model displays a close association of Short-billed Gull presence with soil type regions, water bodies, 'transfer stations' (garbage dumps), roads, and railroads. The association with restaurants was less strong

**Table 2**
Census layers and predictors and their units used for the census blocks (overlay polygons; for more details see URLs provided in the Methods).

| Name of predictor in US Census data set | Meaning and Units | Comment |
|---|---|---|
| POP2000 | Total number of people in the census block in year 2000 | This is essentially population density |
| WHITE | Total number of people who are white alone | Those ethnicities are based on The U.S. census categories |
| BLACK | Total number of people who are black alone | Those ethnicities are based on The U.S. census categories |
| AMERI-ES | Total number of people who are Native American alone | Those ethnicities are based on The U.S. census categories |
| ASIAN | Total number of people who are Asian alone | Those ethnicities are based on The U.S. census categories |
| HAWN-PI | Total number of people who are Hawaiian Pacific Island alone | Those ethnicities are based on The U.S. census categories |
| OTHER | Total number of people who are 'other' alone | Those ethnicities are based on The U.S. census categories |
| MULT_RACE | Number of Multirace Individuals in census unit | Those ethnicities are based on The U.S. census categories |
| HISPANIC | Total number of people who are Hispanic alone | Those ethnicities are based on The U.S. census categories |
| MALES | Total number of males | This uses the binary classification of the U.S. census but could be wider defined |
| FEMALES | Total number of females | This uses the binary classification of the U.S. census but could |
| AGE_UNDER5 | Number of people under age of 5 | |
| AGE_5_17 | Number of people aged 5–17 | |
| AGE_18_21 | Number of people aged 18–21 | |
| AGE_22_29 | Number of people aged 22–29 | |
| AGE_30_39 | Number of people aged 30–39 | |
| AGE_40_49 | Number of people aged 40–49 | |
| AGE_50_64 | Number of people aged 50–64 | |
| AGE_65_UP | Number of people aged 65 upwards | |
| MED_AGE | Median Age of people in the census unit | |
| MED_AGE_M | Median Age of males in the census unit | |
| MED_AGE_F | Median Age of females in the census unit | |
| HOUSEHOLDS | Number of households in the census unit | Household can be a dubious definition |
| AVE_HH_SZ | Average Household Size in census unit | |
| HSEHLD_1_M | Number of Males in Household | |
| HSEHLD_1_F | Number of Females in Household | |
| MARHH_CHD | Number of Married couples with Children per Household | |
| MARHH_NO_C | Number of Married couples without Children per Household | |
| MHH_CHILD | Number of male headed households with children | |
| FHH_CHILD | Number of female headed households with children | |
| FAMILIES | Number of Families in Census Unit | |
| AVE_FAM_SZ | Average Family Size | |
| HSE_UNITS | Household Units | |
| URBAN | Urban classification of census unit | |
| RURAL | Rural classification of census unit | |
| VACANT | Total number of vacant housing units | |
| OWNER_OCC | Occupied owner units | |
| RENTER_OCC | Renter occupied units | |
| POP_MILE | Population per mile2 | |
| HH_INCOME_ | Household Income | |
| HH_10K | Household Income up to 10 K | |
| HH10TO15K | Household Income 10-15 K | |
| HH15TO20K | Household Income 15-20 K | |
| HH20TO25K | Household Income 20-25 K | |
| HH25TO30K | Household Income 25-30 K | |
| HH30TO35K | Household Income 30-35 K | |
| HH35TO40K | Household Income 35-40 K | |
| HH40TO45K | Household Income 40-45 K | |
| HH45TO50K | Household Income 45-50 K | |
| HH50TO60K | Household Income 50-60 K | |
| HH60TO75K | Household Income 60-75 K | |
| HH75TO100K | Household Income 75-100 K | |
| HH100_125K | Household Income 100-125 K | |
| HH125_150K | Household Income 125-150 K | |
| HH150_200K | Household Income 150-200 K | |
| HH_200K | Household Income 200 K | |
| HHMEDIAN | Median Household Income | |
| PER_CAPITA | Per capita income | |
| Polygon area | | A geographic predictor to identify the polygon block, used for internal referencing but not as a model predictor |
| x-coordinate | | Like above |
| y-coordinate | | Like above |

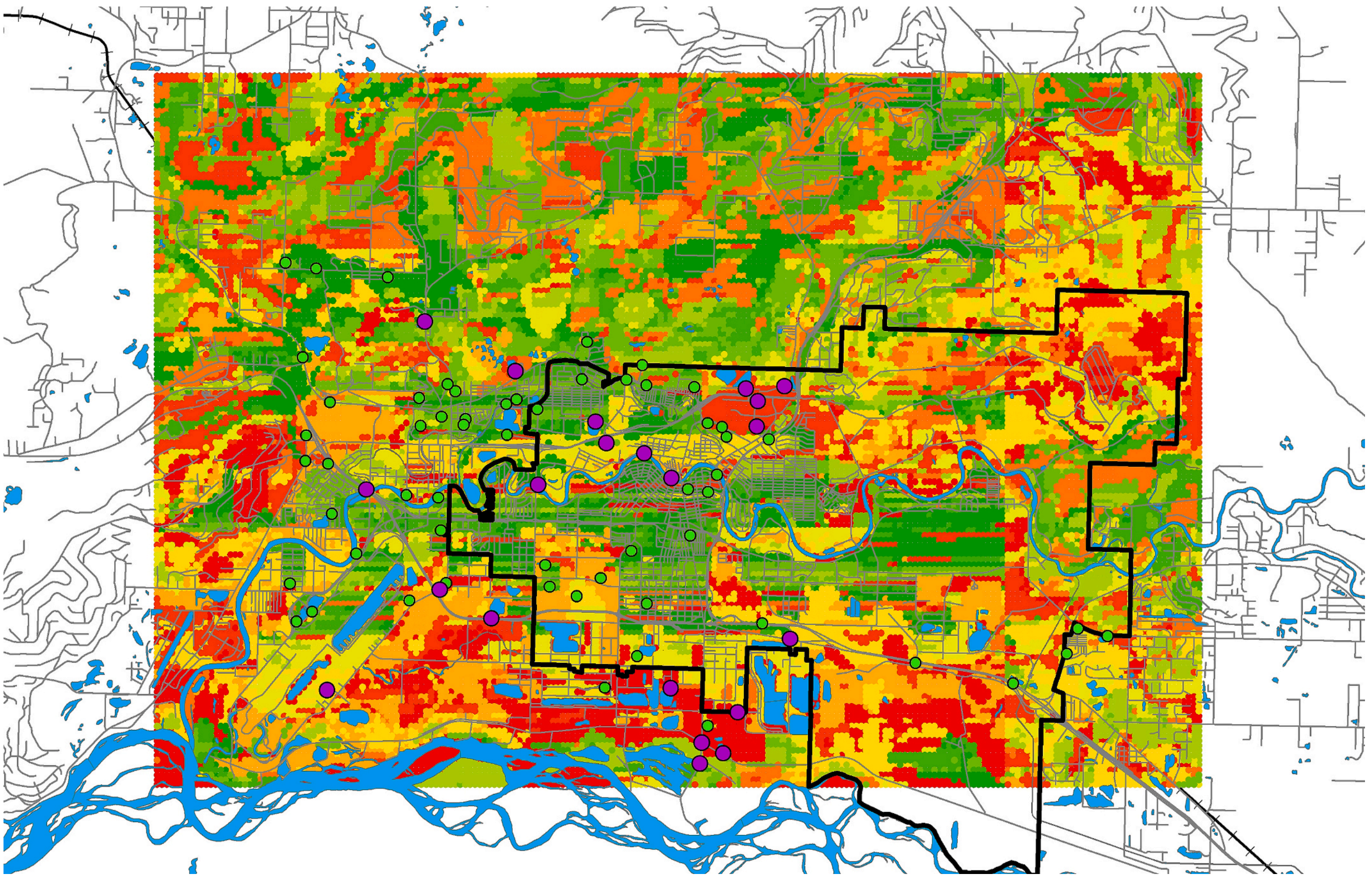


**Fig. 4.** Predicted RIO for Short-billed Gulls based on a machine learning ensemble. The training data are plotted on top (pink dots = presence; green dots = absence). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

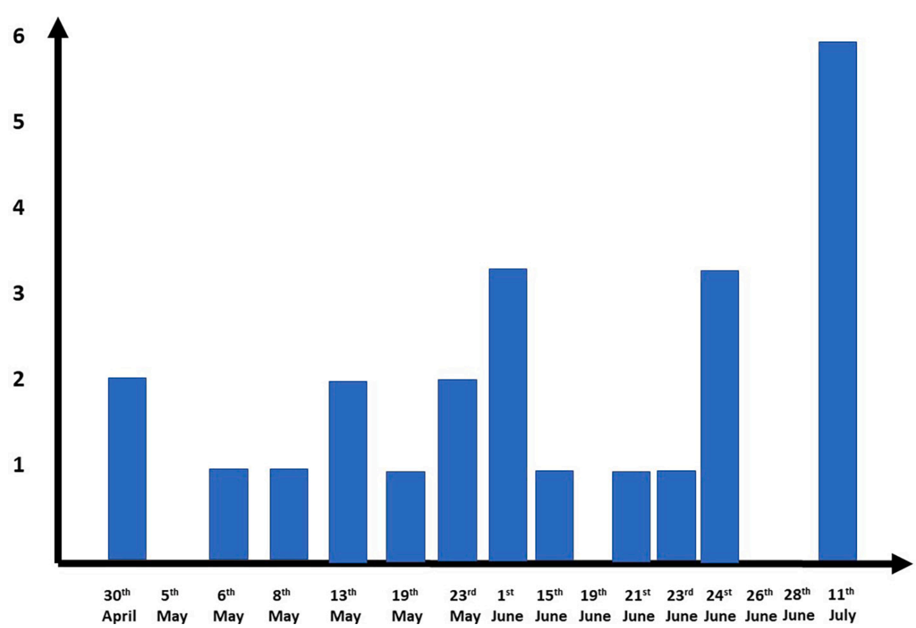



**Fig. 5.** Time series data for a 'presence site' in Fairbanks at a strip mall parking lot during summer 2017 that is a raven hotspot for the ecological niche in winter. The x- axis shows the survey date, the y-axis presents number of gulls observed displaying a consistent use.

than when compared with common ravens in winter (Baltensperger et al., 2013). Short-billed Gulls overall clearly associate with impervious surfaces, specifically in industrial zones and at strip malls (Fig. 5), whereas the typical boreal forest habitats are widely avoided.

The three most important predictor variables and their correlations across the model ensemble providing the best-possible predictions in synergy consist of soil class (water, gravel, pile driver complex, and pit soil being the highest drivers of gull occurrence; Fig. 6), land/vegetation

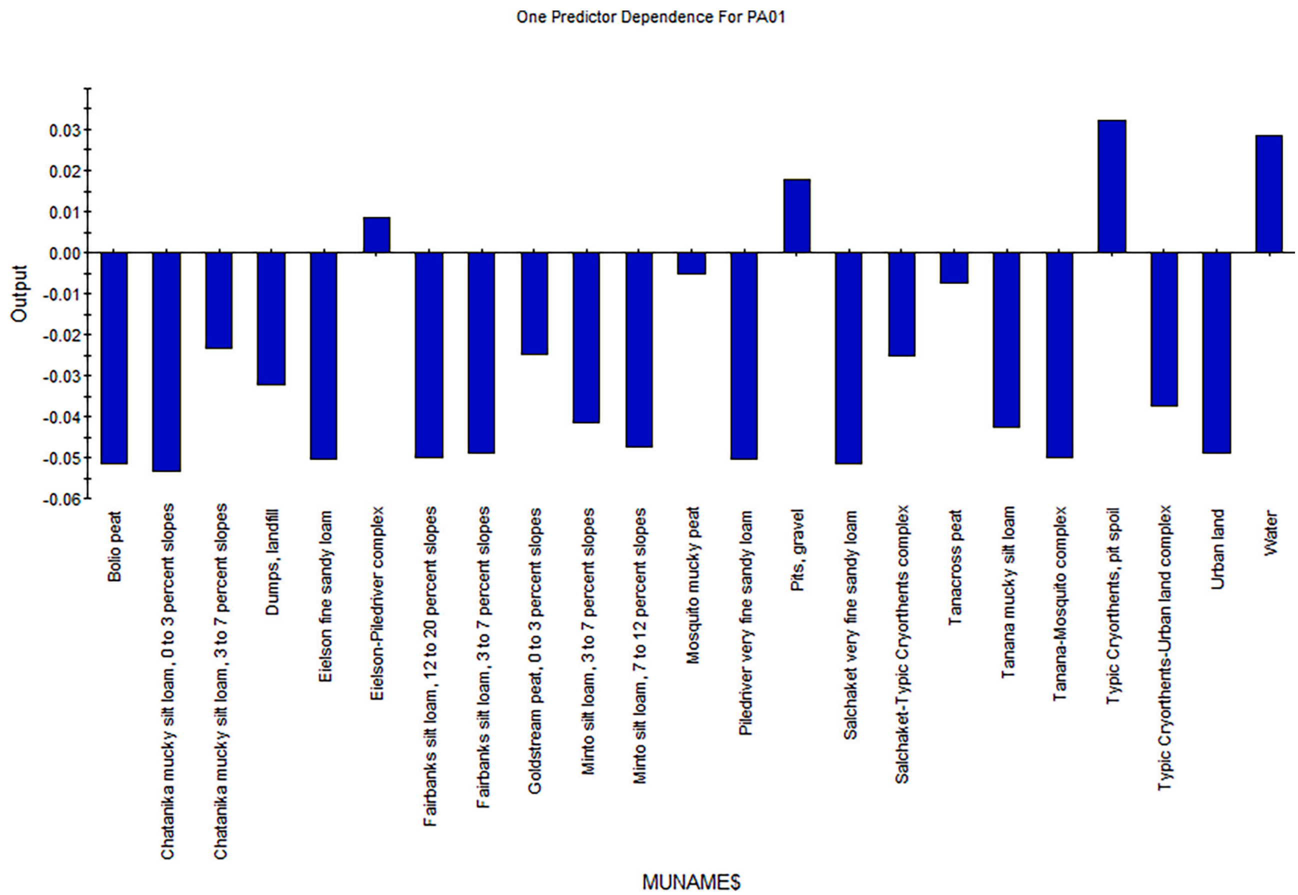



**Fig. 6.** Model Diagnostics Partial Dependence plot of first top-ranked predictor (soil).

cover (water, young burned/reproducing birch-aspen stands, and shrub being the highest drivers of gull occurrence; Fig. 7), as well as distance to water (<2000 m; Fig. 8). Additional important predictor variables include distance to transfer station and distance to road (Table 3; Fig. 11). At the scale of investigation, the socio-economic data were not primary predictor variables informing the model ensemble. It shows thus far that those predictors are overruled by the stronger above-mentioned predictor complex. It is further noteworthy that most typical boreal forest categories are avoided by gulls supporting a truly urbanized habitat selection instead. Presumably this reflects foraging and roosting (e.g. near water) as well as nesting (in shrub-related habitat and on the ground; e.g. https://www.allaboutbirds.org/guide/Short-billed_Gull/overview). While this is not a nesting study, arguably gulls move between nest locations and food provided, intermittent with social roosts. In urban Fairbanks, a gull hotspot within the boreal forest, most gulls are related either way to water, which is now a heavily managed man-made feature. The actual winter raven niche, e.g. at strip malls (Fig. 5), generically provide social attraction and food also in summer, but for gulls.

Put Fig. 4 app here: Predicted RIO for Short-billed Gulls based on a machine learning ensemble.

Put Fig. 5 app here: Time series data for a 'presence site' in Fairbanks during summer 2017 that is a raven hotspot for the ecological niche in winter.

Put Table 3 app here: Model Diagnostics importance rank of predictors from machine learning model.

Put Fig. 6 app here: Model Diagnostics Partial Dependence plot of first top-ranked Predictor (soil type).

Put Fig. 7 app here: Model Diagnostics Partial Dependence plot of second top-ranked Predictor (vegetation class).

Put Fig. 8 app here: Model Diagnostics Partial Dependence plot of third top-ranked Predictor (distance to water).

Put Fig. 9 app here: Model Diagnostics 3D Partial Dependence plot of top-ranked continuous predictor (distance to water vs distance to transfer site).

### 3.3. Model assessment

Model performance was assessed internally (ROC curves) and externally (field testing dataset). Internally, the model ROC curves for each individual algorithm range between 75 and 85% (Table 4). The average ROC value across the ensemble is 71%. Externally, we mapped the 2013 Presence/Absence data cube on the rasterized heatmap. A visual assessment can be made by comparing actual presence/absence data points to predicted RIO scores, with a general alignment of the two, indicating a robust, moderate-high model performance (Fig. 10).

Put Table 4 app here: ROC values for algorithms of the machine learning ensembles.

Put Fig. 10: a) Model assessment data of alternative 80 locations for spring 2013 (Fig. 3) overlaid over the predicted RIO for Short-billed Gulls based on a machine learning ensemble (Fig. 4). b) Map as below but binary classification (black is presence, grey is absence).

## 4. Discussion

Here we present on the fist ever compiled open access data and analysis for specific subarctic gulls (*sensu* Huettmann, 2011, 2015, Huettmann and Arhonditsis, 2023). Same as done in the previous raven study (Baltensperger et al., 2013, here we focus on 'presence' data as a synergetic view across behaviors allowing subsequently for more fine-tuned assessments. We focus on the vacant summer niche and Short-billed Gulls within using model learning data, predicted ecological

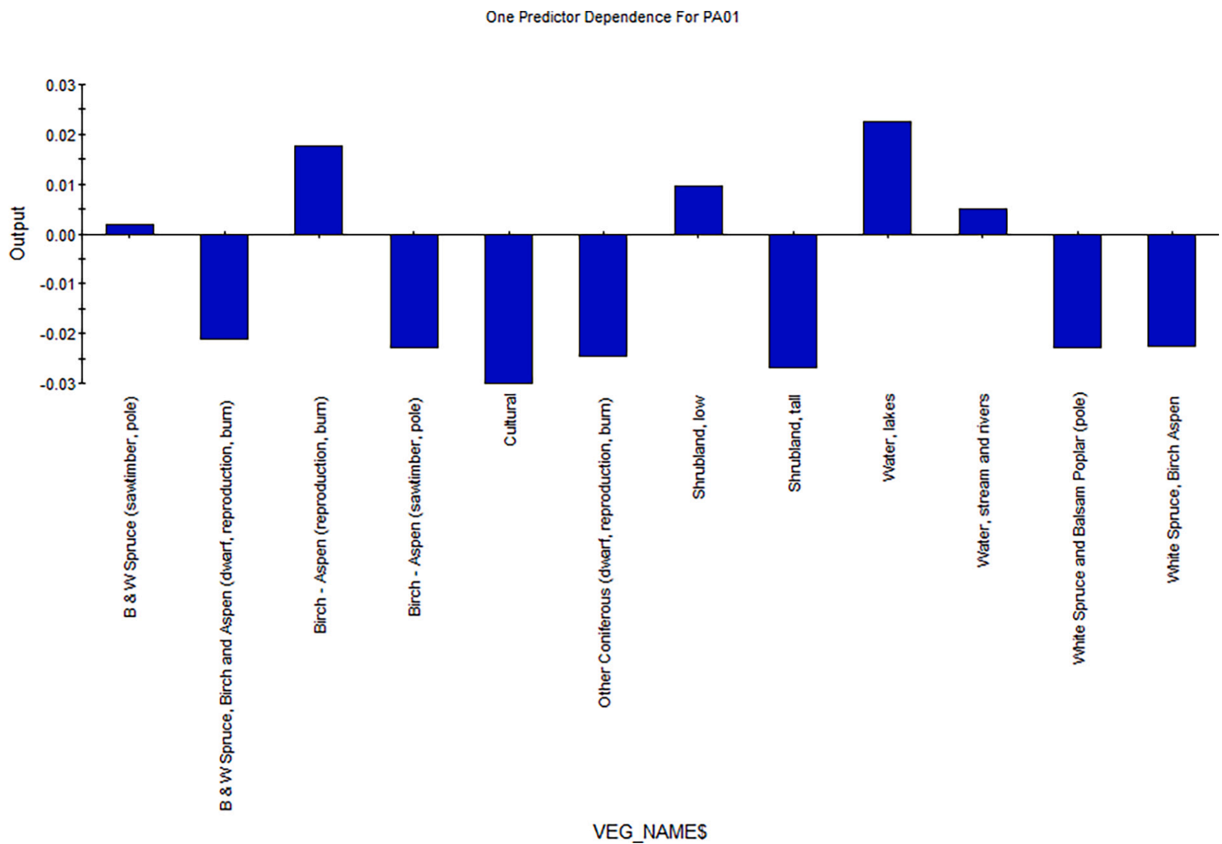



**Fig. 7.** Model Diagnostics Partial Dependence plot of second top-ranked predictor (vegetation class).

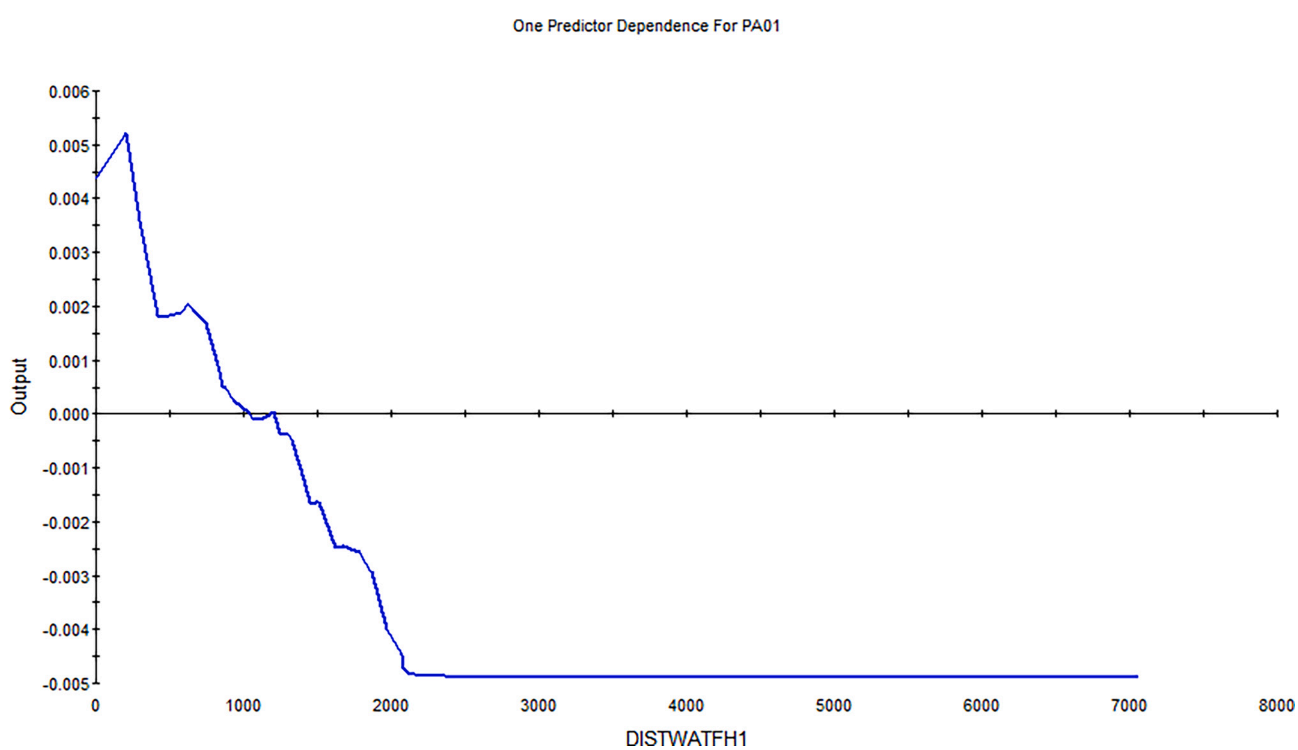



**Fig. 8.** Model Diagnostics Partial Dependence plot of third top-ranked predictor (distance to water).

niche, model assessment data, and model inference for the realized niche distribution.

In machine learning/AI approaches, inference is typically drawn from prediction (Breiman, 2001, Humphries et al., 2018). In ensemble models, one obtains '*a strong learner from many weak learners*' (*sensu* Friedman, 2002). Synergy predictors dominate in real-world ecology

**Table 3**
Model Diagnostics importance rank of predictors from TreeNet machine learning model.

| Predictor name | Importance Rank (percent) | Notes |
|---|---|---|
| MuName | 100 | Categorical predictor (greedy) showing mostly 'water' and 'gravel' as important |
| Veg Name | 61 | Categorical predictor (greedy) showing mostly 'water' and shrubs as important |
| Distance to Water | 44 | |
| Distance to Transfer Station | 36 | |
| Distance to Railways | 33 | |
| Distance to Restaurants | 23 | |
| Age18–21 | 18 | U.S. Census layer (see Table 2 for details of that predictor) |
| HH75TO100K | 18 | U.S. Census layer (see Table 2 for details of that predictor) |
| HH Median | 17 | Peaks around $35,000 to 40,000 household income. U.S. Census layer (see Table 2 for details of that predictor) |
| HH50TO60K | 17 | U.S. Census layer (see Table 2 for details of that predictor) |
| OWNER_OCC | 16 | U.S. Census layer (see Table 2 for details of that predictor) |

and its meaning. Whereas individual model fits and predictor ranks, for instance, are less relevant for defining a real-world synergy relationship between predictor and response variables; it's inference is widely biased (e.g. Humphries et al., 2018 and citations within). Categorical predictors tend to be favored in such ranking exercises, with "greedy" algorithms grabbing most of the variance before considering continuous predictors (Friedman, 2002). In the meantime, the socio-ecological predictors act on the U.S. census block scale and show less spatial resolution and thus, power. This deserves a re-visit with more fine-scaled data.

With this in mind, we show a close association of Short-billed Gull presence with soil type regions, water bodies, 'transfer stations' (garbage dumps), roads, and railroads. The association with restaurants was less strong than when compared with common ravens in winter (Baltensperger et al., 2013). The socio-economic data do not play a large role in this context. But Short-billed Gulls clearly associate with impervious surfaces, specifically in industrial zones and at strip malls (Fig. 5). The niche swap of migratory gulls with common ravens in the spring and summer months occurs from April through July, but is inconsistent across the urbanscape (fewer instances of utilizing fast food restaurants and supermarkets as found with ravens in winter, but a link with water). A core habitat zone for ravens in the winter is consistently occupied by gulls, but with low numbers (Fig. 4). Our models predict a rather patchy distribution of Short-billed Gulls in Fairbanks (Figs. 4, 10 & 11). Presumably, our hotspots are indicative of areas with high detectability, showing the location of nesting colonies, abundant gulls at feeding aggregations and roosting where socializing individuals are easily detected (=heard and seen).

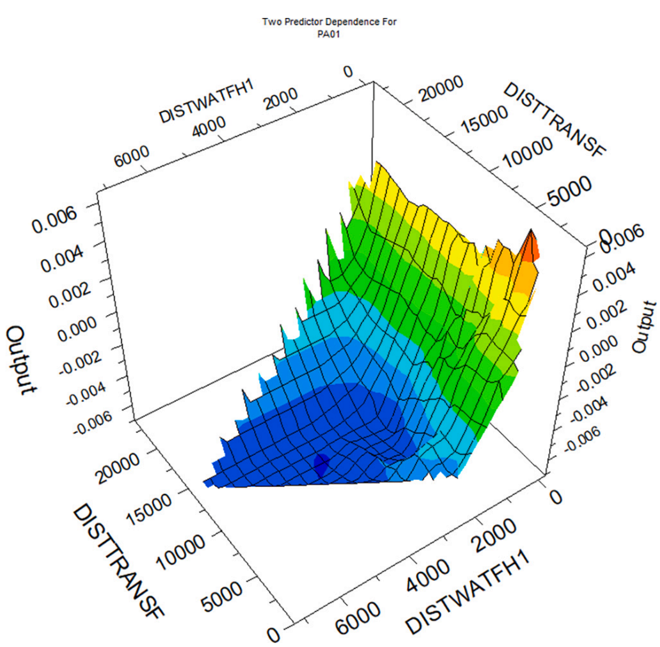



**Fig. 9.** Model Diagnostics 3D Partial Dependence plot of top-ranked predictors (distance to transfer station and distance to water. Highest gull RIOs are found at locations close to water and close to transfer stations).

The highest abundances of Short-billed Gulls – up to 200 individuals – are found in the industrial region of Fairbanks, less the downtown zone (Fig. 11). Ecologically, water bodies are the overall driving habitat factor, linked with human impacts like food/waste. Those water bodies associated with gulls are usually not 'natural' nor found in wilderness areas. Instead, they are often centered around gravel-extraction quarries. Gravel is a very precious commodity in the subarctic region. Fairbanks is a former construction hub for the Alaska pipeline, and various other 'development' projects serve as gravel resources, offering an abundance of such water bodies and exposed soils. It is worth noting, such man-made water bodies are known to contain elevated concentrations of contaminants, including PFAS (e.g. Gander, 2022). This topic deserves more study, as this feature clearly defines the "gull-scape".

Alaska, its wilderness, and its wildlife have received a lot of research and survey attention in the past 100 years (Ross, 2006). For avian research efforts, the diffused datasets are poorly documented, rarely coordinated, and largely unavailable or found in good useable formats with metadata. Examples of key efforts include the Alaska Landbird Monitoring Survey (ALMS; Handel and Cady, 2004), and the Alaskan Bird Banding Station (https://aksongbird.org/about-us/our-projects/cfms/); but those are 'punctual' and do not really allow to infer well on a vast array of crucial avian conservation aspects explicit in space and time for Alaska in the Anthropocene. An accessible breeding bird atlas and bird banding recovery Atlas does not exist for Fairbanks, nor for Alaska or U.S. overall (despite a nation with one of the highest GDP in the world; but see Aycrigg et al., 2015). Instead, amateurs, NGOs and

**Table 4**
Internal ROC values for algorithms of the machine learning ensemble.

| Algorithm | ROC value | Detail |
|---|---|---|
| CART | 75 (learn) 64 (test) | A good performance |
| CART Ensemble and Bagger | 83 | One of the highest ROCs in the ensemble |
| TreeNet | 92 (learn) 70 (test) | One of the highest ROCs in the ensemble |
| RandomForest | 63 | An unusual low performance |
| MARS | 74 | A quite high ROC for MARS |
| | Average (test) 71 | |

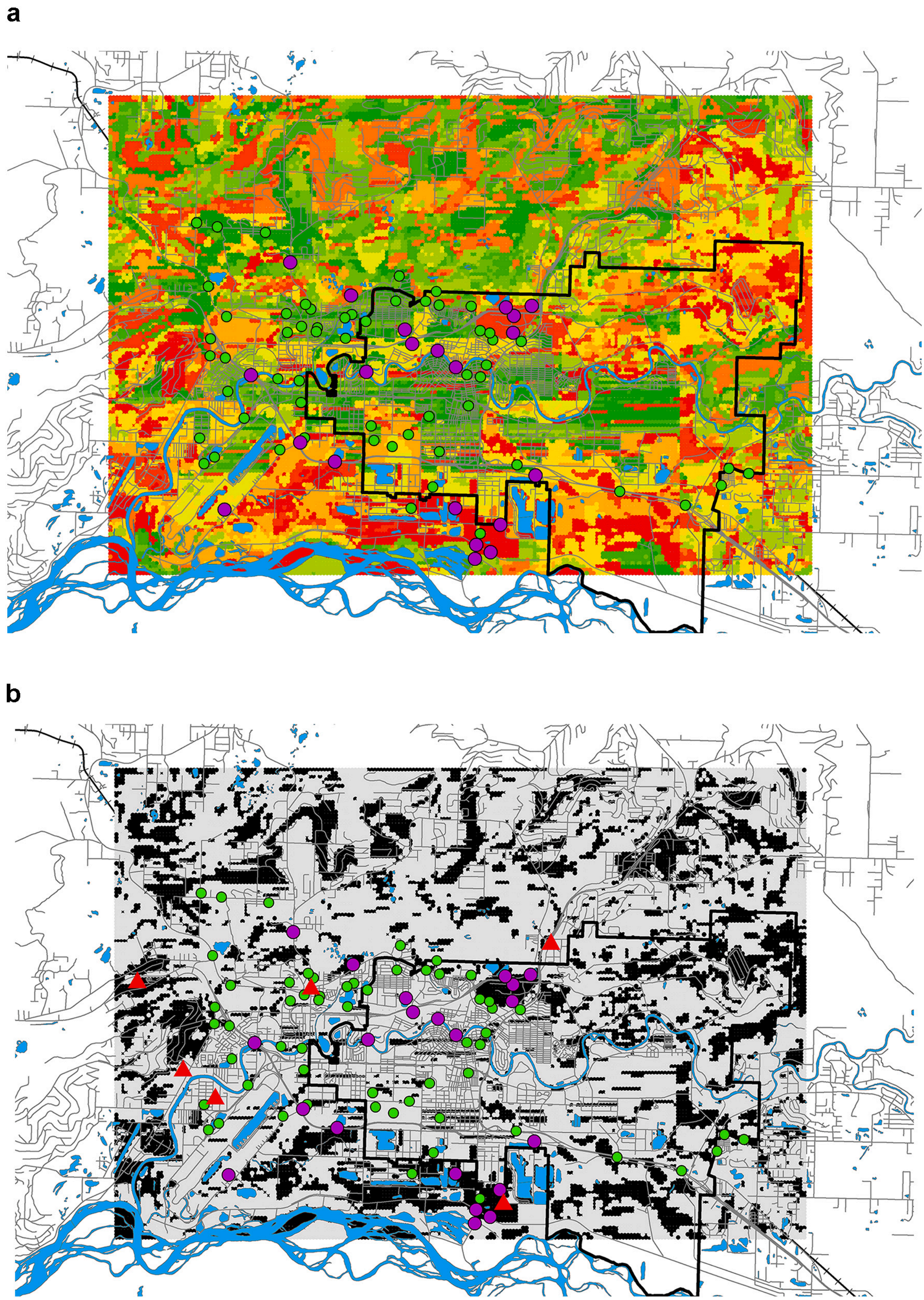



**Fig. 10.** a) Model assessment data of alternative 80 locations for spring 2013 (Fig. 3) overlaid over the predicted RIO for Short-billed Gulls based on a machine learning ensemble (Fig. 4). b) Map as above but binary classification (black is presence, grey is absence).

industrial contractors dominate most of the ornithiology and its information. Thus, the whereabouts of Alaskan gulls and ravens throughout the year remain poorly poorly quantified and known, if at all.

This study is part of a wider ongoing global research effort analyzing avian use of urban areas with GIS and open access 'Big Data' techniques for testable, repeatable and quantified efforts (Baltensperger et al., 2013; Kövér et al., 2015; Hansen and Huettmann, 2020, *sensu* Huettmann and Arhonditsis, 2023). Here we are able to provide substantial

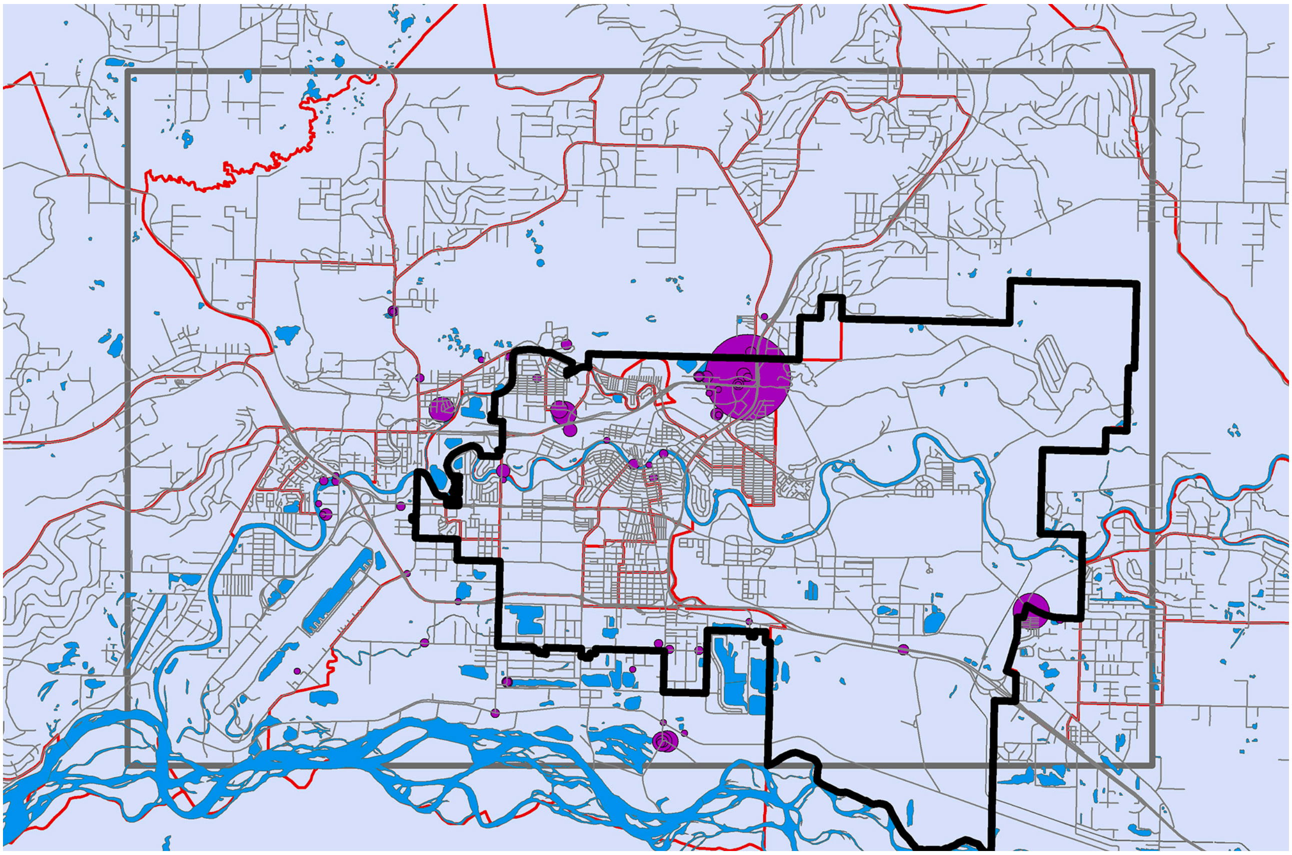


**Fig. 11.** Proportional abundance plot of data 2013, 2014 and 2015 for Short-billed Gulls overlaid with a binary RIO prediction (RIO map see Fig. 4; for survey locations see Figs. 2 and 3).

**Fig. 12.** Photo of a field plot at an industrial gravel pit with high abundances (credit: field team).

progress in this effort, for Fairbanks and for interior Alaska. We measure this progress via Open Access data compilations. We provide a first known, multi-year survey dataset for Short-billed Gulls in Alaska, explicit in space and time with ISO-compliant metadata. This dataset allows for presence/absence data analysis, as well as some limited abundance estimates (Figs. 2, 3 and 4). Further, we compiled seven GIS layers, combined with 61 US Census data predictors, for the Fairbanks municipality, also with ISO-compliant metadata (see Appendix). This combined effort provides a baseline infrastructure for future studies and allows for transparent and replicable research and inference to work from fo updates. We also provide a first and assessed machine learning ensemble prediction for Short-billed Gull presence/absence in Fairbanks, resulting in a first-known high resolution (100-m) map for urban gulls in the boreal forest.

Put Fig. 11 app here: Proportional abundance plot of data 2013, 2014 and 2015 for Short-billed Gulls overlaid with a binary RIO prediction (RIO map see Fig. 4; for survey locations see Figs. 2 and 3).

Put Fig. 12 app here: Photo of short-billed gulls and their urban habitat.

Additional efforts will help to expand knowledge provided in this baseline study. For example, the surveys we conducted for Short-billed Gulls include co-occurring gull and other avian species. Short-billed Gulls are likely embedded in the wider avian and urban ecology (see U.S. census data). Analyses of species diversity, co-occurrence, behavior, and detectability are planned. This study does not survey nest sites nor use nest locations as a predictor variable. Such data do not exist yet, and should be a focus of future investigations. A citizen science data source, e.g., eBird, was not used in this study, but is planned for a wider quantitative assessment of gulls in the region over time. The models we developed herein provide spatial outputs on a 100-m scale. While this is high resolution for the study area, 1-m resolution is achievable where existing datasets and frameworks allow (Boulanger-Lapointe et al., 2022; Robold and Huettmann, 2021).

The implications for this work in a larger context open up many new and exciting research opportunities. Specifically, underlying questions of great relevance include disease risk, population structure, socio-economics and urban planning on a landscape-scale (*sensu* O'Connor et al., 1996, Allen and O'connor, 2000). Such considerations expand on the current perspective of gulls and, broadly, avian research in urban and Arctic landscapes (Auman et al., 2008, 2011). The creation and use of city- and gull-specific predictors should be encouraged and utilized to produce novel outputs, e.g., *E. coli* map, contaminants and disease transmission risk maps. Zoonotic diseases, and the role that gulls and cities play as reservoirs and spill-over areas for avian influenza (Gulyaeva et al., 2020) should also be considered, as well as OneHealth approaches (Huettmann and Hueffer, 2021; Krishna et al., 2022). From an urban planning perspective, analyses can be expanded to issues of contamination, power plant associated heavy metal loads, water quality, heat islands, and man-made climate change. All of these topics are of major relevance for the Anthropocene, and their effects will extend even to remote areas of interior Alaska. Finally, we offer that this research provides a path for addressing questions of telecoupled spill-overs (Liu et al., 2018), specifically for polar regions (Raya Rey et al., 2017). This should be investigated in the future for more advanced inference, policy, and progress.

## Data availability

Data are shared Open Access, as per Methods and Appendix https://drive.google.com/file/d/1VaF7LOBTwUdLZuSs1qYZKV9MZ3j18SoN/view?usp=drive_link

## Acknowledgements

We would like to express our gratitude to all GIS data providers and open access sources, namely the Fairbanks Municipality. FH appreciates the help by the H. Berrios and E.J. Huettmann team, including Chrome et al. For LK, project no. TKP2021-NKTA-32 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NKTA funding scheme. This is EWHALE lab publication # 303.

## Appendix A. Appendices

*Appendix 1: Project GIS files and ISO-compliant metadata*

Study area (shapefile).
Lattice (shapefile).
Fairbanks City boundary (shapefile).

*Appendix 2: Gull survey data and individual ISO-compliant metadata per survey*

2013 80 surveys locations (alternative model assessment data).
2014 50 survey locations (model training data).
2015 50 survey locations (model training data).
2016 time series of summer 2016.

*Appendix 3: GIS predictors (ISO-compliant metadata as in Appendix 1)*

Transfer Stations (points) and proximity.
Restaurants and proximity.
Water and proximity.
Railroad and proximity.
Road and proximity.
Forest cover.
Soil map.
Census data.

*Appendix 4: GIS prediction (ISO-compliant metadata as in Appendix 1)*

Data Cube of Lattice for Ensemble Model prediction.

## Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2023.102364

## References

Allen, A.P., O'connor, R.J., 2000. Hierarchical correlates of bird assemblage structure on northeastern USA lakes. Environ. Monit. Assess. 62, 15–37.
Auman, H.J., Meathrel, C.E., Richardson, A., 2008. Supersize me: does anthropogenic food change the body condition of silver gulls? A comparison between urbanized and remote, non-urbanized areas. Waterbirds 31 (1), 122–126 (1 March 2008). https://doi.org/10.1675/1524-4695(2008)31[122:SMDAFC]2.0.CO;2.
Auman, H.J., Bond, A.L., Meathrel, C.E., Richardson, A.M.M., 2011. Urbanization of the silver Gull: evidence of anthropogenic feeding regimes from stable isotope analyses. Waterbirds 34 (1), 70–76 (1 March 2011). https://doi.org/10.1675/063.034.0108.
Aycrigg, J., Beauvais, G., Gotthardt, T., Huettmann, F., Pyare, S., Andersen, M., Keinath, D., Lonneker, J., Spathelf, M., Walton, K., 2015. 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.
Baltensperger, A.P., Mullet, T.C., Schmid, M.S., Humphries, G.R.W., Kövér, L., Huettmann, F., 2013. Summer and winter observations and machine-learning-based spatial model predictions for the common raven (*Corvus corax*) in the urban, sub-arctic environment of Fairbanks, Alaska. Polar Biol. 36, 1587–1599.
Barua, M., 2022. Feral ecologies: the making of postcolonial nature in London. J. R. Anthropol. Inst. 28 (3), 896–919.
Beaulieu, A., 2022. Generalized reciprocity in common ravens, Corvus Corax, In Interior Alaska During Winter: A Digital Data Mining Approach. Unpublished, M.Sc. thesis. University of Alaska Fairbanks.
Belant, J.L., 1997. Gulls in urban environments: landscape-level management to reduce conflict. Landsc. Urban Plan. 38 (3–4), 245–258.
Benmazouz, I., Jokimäki, J., Juhász, L., Kaisanlahti-Jokimäki, M.L., Paládi, P., Kardos, G., Kövér, L., 2023. Morphological changes in Hooded Crows (*Corvus cornix*) related to urbanization. Front. Ecol. Evol. 11, 1196075.
Betts, M.G., Yang, Z., Hadley, A.S., Smith, A.C., Rousseau, J.S., Northrup, J.M., Gerber, B. D., 2022. Forest degradation drives widespread avian habitat and population declines. Nat. Ecol. Evol. 6 (6), 709–719.
Bocharnikov, V., Huettmann, F., 2019. Wilderness condition as a status Indicator of Russian Flora and Fauna: implications for future protection initiatives. Int. J. Wilderness 25, 26–39.
Bonnefoy, X., Kampen, H., Sweeney, K., 2008. Public Health Significance of Urban Pests. World Health Organization.
Booms, T., Huettmann, F., Schempf, P., 2010. Gyrfalcon nest distribution in Alaska based on a predictive GIS model. Polar Biology 33, 1601–1612.
Booms, T.R.A.V.I.S., Lindgren, M.I.C.H.A.E.L., Huettmann, F.A.L.K., 2011. Linking Alaska's Predicted climate, Gyrfalcon, and ptarmigan distributions in space and time: A unique 200-year perspective. In: Watson, R.T., Cade, T.J., Fuller, M., Hunt, G., Potapov, E. (Eds.), Gyrfalcons and Ptarmigan in a Changing World. the Peregrine Fund, pp. 1–14. Boise, Idaho, USA. https://doi.org/10.4080/gpcw.2011.0116.
Boulanger-Lapointe, N., Ágústsdóttir, K., Barrio, I.C., Defourneaux, M., Finnsdóttir, R., Jónsdóttir, I.S., Marteinsdóttir, B., Mitchell, C., Möller, M., Nielsen, Ó.K., Sigfússon, A.Þ., Þórisson, S.G., Huettmann, F., 2022. Understanding herbivore species coexistence in changing rangeland ecosystems: first high resolution national open-source and open-access ensemble models for Iceland. Sci. Total Environ. 845, 157140.
Breiman, L., 2001. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat. Sci. 16 (3), 199–231.
Burger, J., 1981. Feeding competition between laughing gulls and herring gulls at a sanitary landfill. Condor 83 (4), 328–335.
Burger, J., Gochfeld, M., 1983. Behavior of nine avian species at a Florida garbage dump. Colon. Waterbirds 54–63.
Carlen, E., Munshi-South, J., 2021. Widespread genetic connectivity of feral pigeons across the northeastern megacity. Evol. Appl. 14 (1), 150–162.
Durant, J.M., Hjermann, D.Ø., Frederiksen, M., Charrassin, J.B., Le Maho, Y., Sabarros, P. S., Stenseth, N.C., 2009. Pros and cons of using seabirds as ecological indicators. Clim. Res. 39 (2), 115–129.
Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15 (1), 3133–3181.
Fox, C.H.F., Huettmann, G.K.A., Harvey, K.H., Morgan, J., Robinson, R. Williams, Paquet, P.C., 2017. Predictions from machine learning ensembles: marine bird distribution and density on Canada's Pacific coast. Mar. Ecol. Prog. Ser. 566, 199–216.
Friedman, J.H., 2002. Stochastic gradient boosting. Comp. Stat. Data Analys. 38 (4), 367–378.
Fuirst, M., Veit, R.R., Hahn, M., Dheilly, N., Thorne, L.H., 2018. Effects of urbanization on the foraging ecology and microbiota of the generalist seabird *Larus argentatus*. PLoS One 13 (12), e0209200. https://doi.org/10.1371/journal.pone.0209200.
Gabrey, S.W., 1997. Bird and small mammal abundance at four types of waste-management facilities in Northeast Ohio. Landsc. Urban Plan. 37 (3–4), 223–233.
Gander, M.J., 2022. Climate change and the water quality threats posed by the emerging contaminants per-and polyfluoroalkyl substances (PFAS) and microplastics. Water Int. 1–23.
Goumas, M., Collins, T.R., Fordham, L., Kelley, L.A., Boogert, N.J., 2020. Herring gull aversion to gaze in urban and rural human settlements. Anim. Behav. 168 (2020), 83–88. https://doi.org/10.1016/j.anbehav.2020.08.008.

Gulyaeva, M., Huettmann, F., Shestopalov, A., Okamatsu, M., Matsuno, K., Chu, D.-H., Sakoda, Y., Glushchenko, A., Milton, E., Bortz, E., 2020. 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.
Handel, C.M., Cady, M.N., 2004. Alaska Landbird Monitoring Survey: Protocol for Setting up and Conducting Point Count Surveys U.S. Geological Survey. Alaska Science Center Anchorage, AK, USA. https://alaska.usgs.gov/science/biology/bpif/monitor/alms/ALMSprotocol_2004.pdf.
Hansen, L., Huettmann, F., 2020. Chapter 18. Swallows and sparrows in the human street-market Interface of urban Nepal: Towards a first open access GIS data and model inference on the role of religion and culture in bird distribution. In: Huettmann, F., Regmi, G.R., Huettmann, F. (Eds.), Hindu Kush-Himalaya Watersheds Downhill: Landscape Ecology and Conservation Perspectives. Switzerland, Springer Gland, pp. 361–399.
Hardy, S.M., Lindgren, M., Konakanchi, H., Huettmann, F., 2011. Predicting the distribution and ecological niche of unexploited snow crab (*Chionoecetes opilio*) populations in Alaskan waters: a first open-access ensemble model. Integr. Comp. Biol. 51 (4), 608–622. https://doi.org/10.1093/icb/icr102.
Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. Springer, New York, pp. 1–758.
Hedblom, M., Murgui, E., 2017. Urban bird research in a global perspective. In: Ecology and Conservation of Birds in Urban Environments. Springer, Cham, pp. 3–10.
Hegel, T.S., Cushman, A., Evans, J., Huettmann, F., 2010. Current state of the art for statistical modelling of species distributions. Chapter 16, pp. 273–312. In: Cushman, S., Huettmann, F. (Eds.), Spatial Complexity, Informatics and Wildlife Conservation. Springer, Tokyo, Japan, pp. 273–312.
Huettmann, F., 2011. Serving the Global Village through public data sharing as a mandatory paradigm for seabird biologists and managers: Why, What, How, and a call for an efficient action plan. Open Ornithol. J. 4, 1–11.
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: F. Huettmann F. (Ed.), Central American Biodiversity: Conservation, Ecology, and a Sustainable Future. Springer, New York, pp. 391–418.
Huettmann, F., Arhonditsis, G., 2023. Editorial: towards an ecological informatics scholarship that is reflective, repeatable, transparent, and sharable! Eco. Inform. https://doi.org/10.1016/j.ecoinf.2023.102132.
Huettmann, F., Hueffer, K., 2021. The ecological niche of reported rabies cases in Canada is similar to Alaska. Zoonoses Public Health 68, 677–683. https://doi.org/10.1111/zph.12835.
Huettmann, F., Young, B., 2022. The so-called modern 'sustainable forestry' destroys wilderness, old-growth forest landscapes and ecological services worldwide: A short first-hand review and global narrative on the use of 'growth-and-yield' as a destructive and even impossible goal. In: Forest Dynamics and Conservation. Springer, Singapore, pp. 53–82.
Huettmann, F., MacIntosh, K., Stevens, C., Dean, T., Diamond, A.W., 2000. A large mid-winter observation of Bonaparte's gulls, *Larus philadelphia*, in head harbour passage, New Brunswick. Canad. Field-Natural. 114 (2), 327–330.
Huettmann, F., Mi, C., Yu, Guo, 2018. '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: Humphries, G., Magness, D.R., Huettmann, F. (Eds.), Machine Learning for Ecology and Sustainable Natural Resource Management, pp. 163–184.
Huig, N., Buijs, R.J., Kleyheeg, E., 2016. Summer in the city: behaviour of large gulls visiting an urban area during the breeding season. Bird Study 63 (2), 214–222.
Humphries, G., Magness, D.R., Huettmann, F., 2018. Machine Learning for Ecology and Sustainable Natural Resource Management. Springer, Switzerland.
Jiménez, A., Brush, J., Zambrano, R., Warraich, N., Korosy, M., 2023. Using double sampling to estimate the population of least terns (*Sternula antillarum*) nesting on Florida's rooftops. Waterbirds 45 (4), 440–449.
Johnson, C.J., Ehlers, L.P., Seip, D.R., 2015. Witnessing extinction–cumulative impacts across landscapes and the future loss of an evolutionarily significant unit of woodland caribou in Canada. Biol. Conserv. 186, 176–186.
Karmacharya, D.K.F., Huettmann, C., Mi, X., Han, R., Duwal, S.K. Yadav, Guo, Yu, 2020. Chapter 28. A first high-resolution open access data and open source GIS model-prediction for the globally threatened Sarus crane (Antigone antigone) in Nepal: Data mining of 81 predictors support evidence for ongoing declines in distribution and abundance. In: Regmi, G.R., Huettmann, F. (Eds.), Hindu Kush-Himalaya Watersheds Downhill: Landscape Ecology and Conservation Perspectives. Springer Gland, Switzerland, pp. 577–591.
Kirk, D.A., Bellerby, G., Brook, R.W., Weseloh, D.C., Ewins, P.J., 2008. Assessing seasonal variation in counts and movements of Bonaparte's gulls *Larus philadelphia* on the Niagara River, Ontario. Waterbirds 31 (2), 193–202.
Kövér, L., Gyüre, P., Balogh, P., Huettmann, F., Lengyeld, S., Juhász, L., 2015. Recent colonization and nest site selection of the hooded crow (*Corvus corone cornix* L.) in an urban environment. Landsc. Urban Plan. 133, 78–86.
Krishna, P.A., Chand, R., Huettmann, F., Ghimire, T.R., 2022. Rabies elimination: Is it feasible without considering wildlife? J. Trop. Med. 2022, 5942693 https://doi.org/10.1155/2022/5942693.
Langley, L., 2021. The Ecology of Lesser Black-Backed Gulls (*Larus fuscus*) in the Anthropocene: Implications for Conservation and Management. Doctoral dissertation. University of Exeter.
Liu, J., Dou, Y., Batistella, M., Challies, E., Conno, T., Friis, C., Millington, J.D.A., Parish, E., Romulo, C.L., Bicudo Silva, R.F., Triezenberg, H., Yang, H., Zhao, Z., Zimmerer, K.S., Huettmann, F., Treglia, M.L., Basher, Z., Chung, M.G., Herzberger, A., Lenschow, A., Mechiche-Alami, A., Newig, J., Roch, J., Sun, J., 2018. Spillover systems in a telecoupled Anthropocene: typology, methods, and governance for global sustainability. Environ. Sustain. 33, 58–69. https://doi.org/10.1016/j.cosust.2018.04.009.
Louise, H.G., 2020. Thinking with gulls: multi-species interactions in the Anthropocene. The Elphinstone Review 80.
McArdle, B.H., 1988. The structural relationship: regression in biology. Can. J. Zool. 66 (11), 2329–2339.
Mi, C., Huettmann, F., Yu Guo, X., Han, L. Wen, 2017. Why to choose RandomForest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ. https://doi.org/10.7717/peerj.2849.
Naess, A., Jickling, B., 2000. Deep ecology and education: a conversation with Arne Næss. Can. J. Environ. Educ. 5, 48–62.
O'Connor, R.J., Jones, M.T., White, D., Hunsaker, C., Loveland, T.O.M., Jones, B., Preston, E., 1996. Spatial partitioning of environmental correlates of avian biodiversity in the conterminous United States. Biodivers. Lett. 97–110.
Ohse, B., Huettmann, F., Ickert-Bond, S., Juday, G., 2009. 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.
Ouled-Cheikh, J., Morera-Pujol, V., Bahillo, Á., Ramírez, F., Cerdà-Cuéllar, M., Ramos, R., 2021. Foraging in the Anthropocene: feeding plasticity of an opportunistic predator revealed by long term monitoring. Ecol. Indic. 129, 107943.
Pais de Faria, J., Paiva, V.H., Veríssimo, S., Gonçalve, A.M.M., Ramos, J.A., 2021. Seasonal variation in habitat use, daily routines and interactions with humans by urban-dwelling gulls. Urban Ecosyst. 24, 1101–1115. https://doi.org/10.1007/s11252-021-01101-x.
Piatt, J.F., Sydeman, W.J., Wiese, F., 2007. Introduction: a modern role for seabirds as indicators. Mar. Ecol. Prog. Ser. 352, 199–204.
Raya Rey, A.N., Pizarro, J.C., Anderson, C.B., Huettmann, F., 2017. Even at the uttermost ends of the Earth: How seabirds telecouple the Beagle Channel with regional and global processes that affect environmental conservation and social-ecological sustainability. Ecol. Soc. 22 (4), 31 [online] URL: https://www.ecologyandsociety.org/vol22/iss4/art31/.
Robold, R., Huettmann, F., 2021. 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 9, e11830. https://peerj.com/articles/11830/.
Rock, P., 2005. Urban gulls. Br. Birds 98, 338–355.
Ross, K., 2006. Pioneering Conservation in Alaska. University Press of Colorado, Bolder.
Russo, F., 2011. Correlational data, causal hypotheses, and validity. J. Gen. Philos. Sci. 42, 85–107.
Schilthuizen, M., 2018. Darwin Comes to Town: How the Urban Jungle Drives Evolution. Picador, New York, p. 293.
Sinclair, Pamela H., Nixon, Wendy A., Eckert, Cameron D., Hughes, Nancy L. (Eds.), 2011. Birds of the Yukon Territory. UBC Press, Vancouver, Canada.
Steiner, M., Huettmann, F., 2023. Chapter 8: Tree squirrels in old-growth forests? Landscape Metrics, Open Access Field Data, Machine Learning, and GIS models from Remotely-Sensed Imagery in Tanana State Forest Wilderness of Alaska. In: Steiner, M., Huettmann, F. (Eds.), Sustainable Squirrel Conservation A Modern Reassessment of Family Sciuridae. Springer, New York, pp. 251–262.
Weiser, E.L., Powell, A.N., 2011. Reduction of garbage in the diet of nonbreeding glaucous gulls corresponding to a change in waste management. Arctic 220–226.
Zelenskaya, L.A., 2019. Ecology of an urban population of the Slaty-backed Gull (Larus schistisagus) in comparison with natural colonies: features of nest location and productivity. Biol. Bull. 46 (9), 1108–1123. https://doi.org/10.1134/S106235901909019X.
Zelenskaya, L.A., 2021. (2021): ecology of an urban population of the Slaty-backed Gull (*Larus schistisagus*) in comparison with natural colonies, Feeding and foraging flights. Biol. Bull. 48 (Suppl. 1), S85–S102. https://doi.org/10.1134/S1062359021140223.