bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extracting socio-spatial networks from photo-ID data using multilevel multinomial models

Tyler R. Bonnell [a,b,c], Robert Michaud [d], Angélique Dupuch [a,b], Véronique Lesage[e], Clément Chion [a,b]

[a] Department of Natural Sciences, Université du Québec en Outaouais, Gatineau, Québec, Canada

[b] Institut des Sciences de la Forêt Tempérée, Université du Québec en Outaouais, Ripon, Québec, Canada

[c] Department of Psychology, University of Lethbridge, 4401 University Drive, Alberta T1K 3M4, Canada

[d] Groupe de Recherche et d'Éducation sur les Mammifères Marins (GREMM), 870 avenue Salaberry, Bureau R24, Québec, QC G1R 2T9, Canada

[e] Maurice Lamontagne Institute, Fisheries and Oceans Canada, P.O. Box 1000, 850 route de la Mer, Mont-Joli, QC G5H 3Z4, Canada

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract:

Photo identification of individuals within a population is a common data source that is becoming more common given technological advances and the use of computer vision and machine learning to re-identify individuals. These data are collected through hand-held cameras, drones, and camera traps, and often come with biases in terms of sampling effort and distribution. In spite of these biases, a common goal of collecting these datasets is to better understand the habitat use pattern of individuals and populations. Here, we examine the potential for multilevel multinomial models to generate socio-spatial networks that capture the similarities in individual users across the spatial distribution of a species. We use this approach with 18 years of photo-ID data to better understand population structuring of beluga whales in the St. Lawrence River. We show using permuted and simulated data that this approach can identify community network structures within populations in a way that accounts for biases in collections methods. Applying this method to the entire 18 years dataset for SLE beluga, we found three spatially distinct communities. These results suggest that within the population's summer range individuals are moving within restricted areas (i.e., home ranges), and have implications for the estimated impacts of localized anthropogenic stressors, such as chemical pollution or acoustic disturbances on animal populations. We conclude that multilevel multinomial models can be effective at estimating socio-spatial networks that describe community structuring within wildlife populations.

Keywords: Multinomial Model, Beluga, Photo ID, Socio-Spatial Network, Bayesian Network, Community Detection

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

An understanding of the spatial and temporal distribution of a species of concern is of central importance to conservation and management (Evans & Hammond, 2004). Increasingly, photo and video are being used to monitor individuals within populations (hereafter photo-ID data), providing a view of within population social mixing and habitat use (Koivuniemi et al., 2016). The increased use of machine learning to identify individuals from these data streams has greatly facilitated the use of these photo-ID data (Schneider et al., 2019). These individual identifications have facilitated the use of novel statistical and computational methods to quantify within population structures, such as social network analysis (Perryman et al., 2019; Schilds et al., 2019).

It is often the case, however, that efforts when collecting photo-ID are not evenly distributed. This differentiation in effort can heavily bias estimates of both habitat usage and population distribution estimates (Hupman et al., 2018). Here we propose the novel use of multilevel multinomial models to account for these biases and to estimate socio-spatial structures within populations.

The existence of social structuring within populations, such as communities, can have important ecological and management implications. If a population as a whole can be considered as highly mixed, i.e., with individuals showing no strong patterns of home range use or sub-structuring within the large population, then all individuals are equally likely to feel the impacts of local changes in the environment. In contrast, if the population cannot be considered to be highly mixed, and shows strong sub-structuring and site-fidelity patterns within the larger population, local stressors might have a disproportionate impact on subsections of the population. For example, if noise pollution increased in only one sector, in a highly mixed population all individuals would be lightly impacted, but in a structured population a subset of the population would be highly impacted. These differences in spatial

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

structuring of populations can lead to biased estimation of the likelihood and magnitude of impacts from local stressors both at the individual and population levels (DeFur et al., 2007).

The multilevel multinomial modeling approach does not have a large body of literature to draw on for use with photo-ID data. However, it does have some unique advantages (Koster & McElreath, 2017). For instance, if sampling effort is biased in different regions, the mean probability of being seen within highly sampled regions will be biased upwards. By taking advantage, however, of the multilevel structure of the model it is possible to extract individual deviations in the probability of being seen within a particular region. Decisively, these individual level deviations from the mean probability are not biased by changes in the sampling effort. That is the mean probability will increase with sampling effort, but the relative difference between individuals within the sector will not. High users of a particular region of a habitat will consistently be higher compared to low users of that habitat, and this difference between high and low users will not be biased by sampling effort. Furthermore, by comparing the individual differences between regions it is possible to see if the high/low users of one region are similarly the high/low users of another region. The similarity, or dissimilarity, between regions can then provide information about which regions share similar user profiles. We suggest that by using the correlations between these individual level deviations in high/low users between regions it is possible to generate socio-spatial networks and identify social structuring within the population. In particular it can help to identify spatial communities, i.e., a set of regions that share similar usage patterns and that differ from other regions.

To evaluate the use of multilevel multinomial models to identify socio-spatial structuring within a population, we make use of a long term photo-ID dataset of beluga whales in the St. Lawrence Estuary, Canada. This population has undergone a drastic decrease from around 10,000 in the late 1800s to less than 1,000 today, and is currently

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

considered as endangered in Canada according to the Species at Risk Act (COSEWIC 2014; Fisheries and Oceans Canada, 2012; Mosnier et al., 2015). The population is part of a larger study on the mitigation of noise pollution due to marine traffic (Chion et al., 2017; Lesage et al., 2014; McQuinn et al., 2011; Parrott et al., 2011).

In this paper we first evaluate the performance of the multilevel multinomial models using simulated data, testing if the method correctly estimates no community structuring when none is present, and identifies the correct structure when it is present. In both cases we use the 18 year beluga photo-ID dataset, randomly permuting uniquely identified individuals to generate unstructured datasets, and randomly placing individuals within pre-specified communities to generate structured datasets. We then apply the method to the observed data and quantify community structures within the population's summer range in the St. Lawrence Estuary. Finally, we discuss some potential extensions to the multilevel multinomial modeling approach.

2. Material and Methods

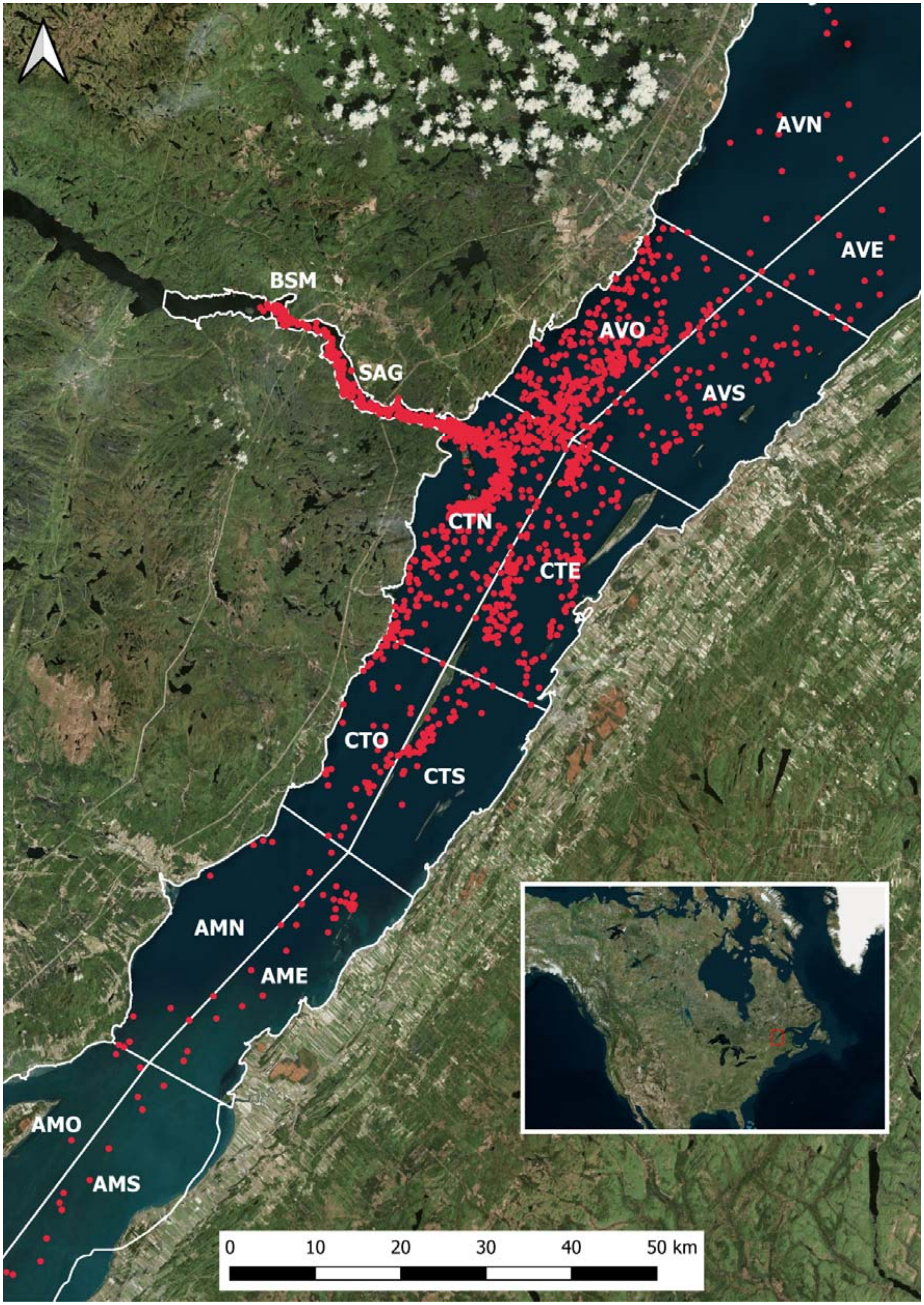
2.1 Data

Photo-ID data were collected using a handheld camera onboard a boat that was able to navigate near to beluga, hereafter referred to as an encounter. Once near beluga a photo ID protocol was then followed to generate photographs used to attempt to identify individuals. Due to the logistical difficulty of covering a large body of water, the sampling effort was unequally distributed across 14 sectors within the St. Lawrence Estuary (Fig. 1). The photo-ID dataset used in this study was collected from 1989-2007 and are part of an ongoing project. The data is stored in a database that facilitates the identification and association between photos to help track the individual identification process. This resulted in a dataset of 7,525 individual encounters where the individual was successfully identified by photo

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sampling (hereafter referred to as a photo-ID), and where a GPS point was taken and the sector recorded. This resulted in 821 unique individuals being successfully identified, with a mean number of photo-IDs per individual of 9 (min = 1, max = 90) (Fig. 1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



123

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1: Spatial distribution of photo-ID data (red points) within the St. Lawrence Estuary, Quebec, Canada (red square in the inset map). The 14 sectors are outlined and labeled in white, and covers the summer habitat of this beluga population.

2.2 Statistical Analysis

Our aim was to use photo-ID data to estimate the probability of seeing an individual in each delineated sector of the St. Lawrence Estuary, and to use these individual probabilities to estimate socio-spatial structures within the beluga population (Fig. 1). To accomplish this aim we used a multilevel multinomial model, where the dependent variable was the number of times an individual was captured photographically (i.e., photo-identified) in each sector. The use of a multilevel model structure allows for the estimation of both the mean probability of photo-identifying an individual in each sector, and the individual level differences in this probability by using individual ID as a random intercept. If we take, as an example, a case where there is only two sectors, then the log-odds of finding individual i in a sector other than the reference sector can be modeled using a multilevel multinomial following (Koster & McElreath, 2017) as:

$$log\left(\frac{p_{1,i}}{p_{r,i}}\right) = \mu_1 + \upsilon_{1i}$$
$$log\left(\frac{p_{2,i}}{p_{r,i}}\right) = \mu_2 + \upsilon_{2i}$$

Where $p_{1,i}$ is the probability of seeing beluga i in sector 1, whereas $p_{r,i}$ is the probability of seeing beluga i in the reference sector. The μ_1 and μ_2 are the intercepts, i.e., the mean probability of seeing a beluga in sectors 1 and 2. This mean probability represents preference/avoidance of the specified sector, however, it is very likely to be biased due to variation in sampling effort. Finally, the υ_{1i} and υ_{2i} are the estimated individual differences (i.e., random intercepts) from the mean probability of capture in sectors 1 and 2, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

These individual differences from the mean probability capture if an individual beluga is a high/low user of that sector, and is not biased by variation in sampling. It is then possible to model the covariance of the individual differences between two sectors using a multivariate normal distribution:

$$\begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix} \sim Multi\ Normal(0, \Omega_v)$$

$$\Omega_v = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix}$$

This multivariate normal distribution has a mean of 0 and a covariance matrix Ω_v. Here the diagonal entries in the covariance matrix (e.g., $\sigma_{1,1}$) represent the magnitude of individual differences within a sector, i.e., are their high and low users in a sector or are all individuals equally likely to be seen? The off-diagonal entries (e.g., $\sigma_{2,1}$) are the covariance estimates between sectors, i.e., do sectors share similar high and low users? By estimating the correlation of individual differences between sectors, this multilevel modeling approach quantifies how much information individual differences in one sector can provide about another sector. Positive correlations suggest that the high/low users in one sector are similarly high/low users in another sector, while negative correlations suggest high/low users in one sector are the low/high users in another sector.

This model can be fit using brms in the R environment using a multivariate syntax: bf(y | trials(n) ~ 1 + (1|q|ID)) + multinomial(). Here, y is a set of column vectors where each column is a sector and each row is an individual. The values in this column vector indicate how many times each individual was seen in each sector. The n is the total number of times an individual was captured, and q is an arbitrary character choice that allows correlations between the estimates of random intercepts for each sector (Bürkner, 2017).

2.2.1 Addition of a common reference sector

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tests using simulated data suggested that adding a preset reference category (i.e., a fifteenth sector used as a reference sector) to the data was required to estimate the correlation between sectors (Fig. S1). To create this reference category, we tally up the observations for each individual and add that number of observations to the new reference sector. This essentially sets the probability of a capture in the reference sector to 0.5 for all individuals, i.e., equal to the probability of being captured outside of this reference sector. This ensures that all individuals have the same baseline probability in the reference category, and as the parameters in the multinomial model measure deviations away from the reference sector, we gain better estimates of the relative deviations between individuals (Fig. S1).

2.2.2 Dealing with biases in photo-ID datasets

This multilevel multinomial approach accounts for repeated sampling of individuals, and provides an estimate of whether some individuals are found more or less often than the mean probability of captures in each sector. We are particularly interested in the estimates of individual differences from the mean probability of capture (i.e., the random intercepts) as these estimates are not impacted by bias in sampling effort among sectors. This is not the case for estimates of the mean probability of capture for each sector, which are expected to increase in highly sampled sectors. For example, the Saguenay River is over-sampled compared to the other sectors (SAG in Fig. 1), increasing the mean probability of capturing individuals in that sector. However, over-sampling should not affect the individual differences in the probability of being captured, i.e., all individuals' chances of being captured go up or down equally.

Similarly, potential biases due to ease of recognition, e.g., some beluga or age classes might have more distinctive markings, are minimized using a multilevel multinomial approach where the differences in the probability of being seen between sectors is the main

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

focus. For example, if juveniles are less likely to be successfully identified by photo, then they might have a reduced number of photo-IDs compared to other age classes, but the difference in distribution of these fewer photo-IDs across sectors will not be impacted. For example, if both an adult and juvenile spend twice as much time in the SAG sector compared to all other sectors, you might expect a photo-ID distribution (Sag:not-sag) of 10:5 and 2:1, respectively. In both cases the probability of being captured in the SAG sector is twice that of the remaining sector. Due to the adaptive partial pooling properties of multilevel models, individuals with few photo-IDs will, however, be less likely to show differences to the mean probability, i.e., they contain less information. This means that if an age class has very little chance of being identified by photo-ID, they will likely contribute less to the estimated socio-spatial structures estimated by the multilevel multinomial approach.

By using a multilevel modeling approach we also reduce the chance of false positives when making comparisons between many different individuals in many different sectors (i.e., problem of multiple comparisons). For example, if we were to estimate the differences in the probability of each sector separately for each individual, the risk of false positives would be increased. By using a multilevel approach to estimating the differences we can make effective use of partial pooling of information to reduce extreme values, especially where the number of photo IDs is not equal between individuals. Furthermore, by running this in a Bayesian framework we are able to place priors on the individual differences within sectors that start the model assuming that there are no differences between individuals in their use of each sector, e.g., student_t(3,0,1).

2.3 Social Network Analysis

Social networks are often used when visualizing and quantifying social structures within populations, with individuals often represented as nodes and their interactions as edges

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between these nodes. In our case, we use sectors as nodes, and the similarities in user profiles between sectors as edges. The correlations between sectors estimated from the multilevel multinomial model can be used to create a network where the posterior predictions of each correlation parameter corresponds to an edge weight in the network. In this way, each edge has a posterior distribution and can be used to create many networks from which a distribution of network metrics can be generated, e.g., the distribution of node strength values can be calculated for each sector. The advantage of having distributions of network measures is that the measures can be readily compared, e.g., does one sector have a higher node strength than another? It is also possible to use the distribution of edge weights, and a chosen threshold (e.g., 95% credible interval), to highlight only the edges where the sign of the correlation is known with a particular range of certainty. In this paper, we used this latter approach to generate a signed network (i.e., a network with positive and negative edges) and use a simple signed-edge rule to define communities: where a distinct community is a set of nodes that share positive edges but no negative edges. We also made use of signed blockmodeling, an algorithm that can also be used to identify blocks of nodes that maximize within block positive edges and minimize within block negative edges (Doreian & Mrvar, 2015). While the signed-edge rule generally provides relatively intuitive results with simple networks, using the signed blockmodeling is likely to be particularly advantageous when dealing with large networks.

2.4 Testing data

To assess the accuracy of the multilevel multinomial modelling approach, we generated test datasets from the observed photo-ID data. We ensured that the test datasets contained the same number of unique individuals, distribution of sightings (i.e., some individuals are seen more than others), and overall number of photo-IDs as the observed dataset. We, however,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

varied the spatial mixing of the test datasets. To test if the proposed method correctly detected no pattern when none existed, we created a completely random test dataset by permuting the sector associated with each photo-ID in the observed dataset. The expected result was to find no correlations between sectors given that the sectors for each photo-ID had been randomly permuted. To then test whether the proposed method could also correctly identify patterns when a known pattern existed, we generated a structured test dataset by randomly assigning each individual to four equally populated communities with the following and hypothetical home range of adjacent sectors: community 1-BSM, SAG, CTN, community 2- CTN, CTO, AMN, community 3- AVO, AVS, AVN, and community 4-AME, CTS, CTE. Following this, we altered the sector of where the individual photo-IDs were taken so as to fall within sectors associated with an individual's community, i.e., one of their home range sectors. We did this by choosing a sector for each photo-ID based on the individual's assigned community 80% of the time; a random sector was chosen for the other 20% of the time, introducing noise in the assignment of sectors. We then tested whether the model correctly identified the correlations between sectors that defined the home range of each of the communities.

3. Results

3.1 Testing data

When the multinomial multilevel model was fit to the data with sectors randomly permuted between all photo-IDs, the model found no evidence for positive/negative correlations between sectors (Fig. 2a). Similarly, when we simulated data with some known structure, i.e., when we artificially created spatially distinct communities, we found that the model accurately estimated the correlations between sectors that defined these artificial communities (Fig. 2b). The simple signed-edge rule and blockmodelling algorithm applied to

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the simulated datasets both revealed the four artificially generated communities, though the blockmodeling algorithm had difficulty with the multi-membership node as it could not assign a node to two blocks (i.e., the CTN node that was shared between communities 1 and 2).

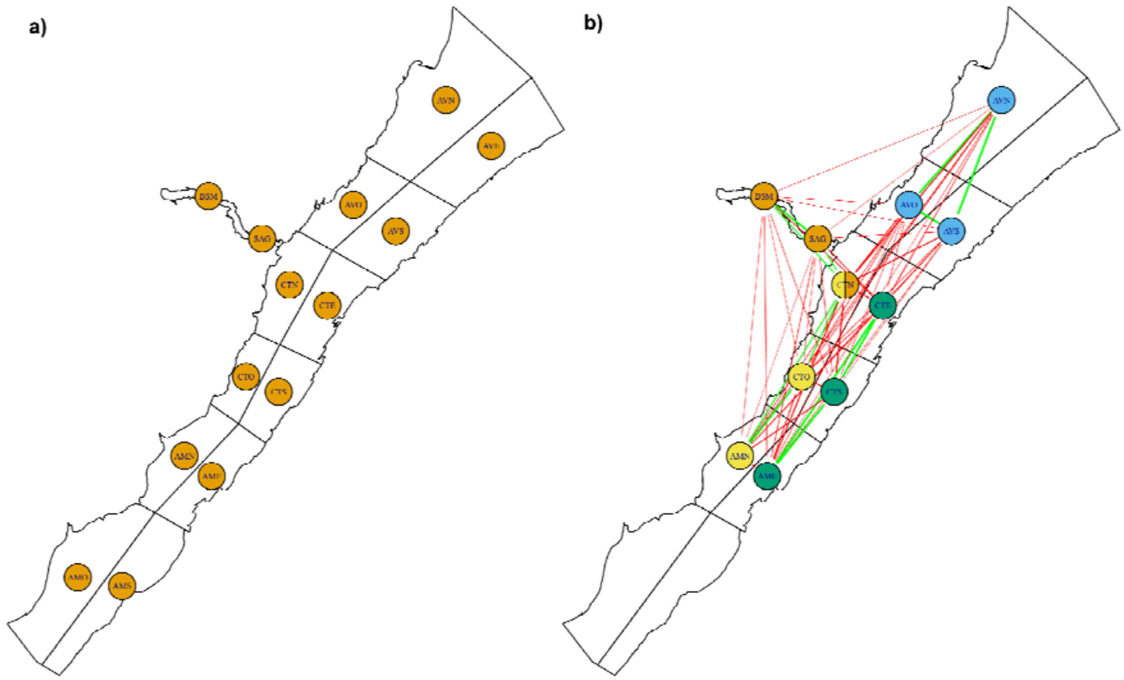


Figure 2: Similarity and dissimilarity between sectors in the simulated datasets: a) randomly permuted data, where there are no spatial communities, and b) structured data, where there are four distinct communities. In b) the simulated communities are represented by color codes for each of their sectors (Note: CTN is part of the orange and yellow communities). The green edges between two sectors signify that the sectors share high/low users, while red edges signify that they have dissimilar high/low users. The lack of an edge signifies that the high/low users of one sector does not provide information about the high/low users of other sectors. Nodes represent sectors, and are coloured based on the communities imposed when simulating the data.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.2 Observed data

The results from our multilevel multinomial model found that within sectors there were consistent individual differences in how likely it was to see individual beluga (Table 1). That is, within sectors, there were some beluga that used the sector heavily, while others did not. The model also found that between sectors these individual differences were correlated (Table S1). These correlations quantify the magnitude of similarity/dissimilarity between sectors in terms of which beluga are using those sectors heavily or rarely. If we take two sectors as examples, e.g., the SAG and CTE sectors, representing, respectively, a tributary to the St. Lawrence Estuary and a sector on the opposite side close to the South shore of the Estuary, and we look at the top 10 estimated high users (i.e., relatively high probability of being found there) within the SAG, we find that they are found to be low users in the CTE sector (see blue dots in Fig 3 a) and b)).

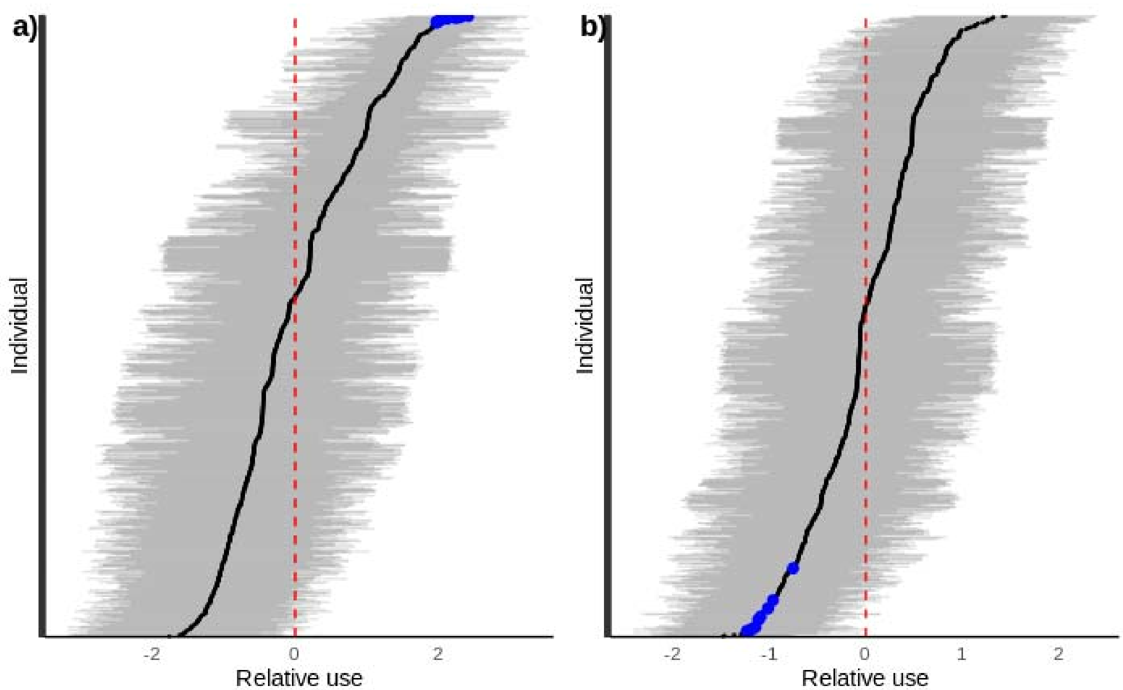


Figure 3: Estimate of the relative use (i.e., deviation from mean use) for each individual within the SAG (a) and CTE sectors (b) of the St. Lawrence beluga summer habitat. The values are deviations from the mean probability of observing individuals within a sector and

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are on the logit scale. The red dashed line represents the mean use, black points represent the estimated deviation from the mean, while the horizontal grey lines represent the 95% credible interval. To highlight how correlations are estimated between sectors, the estimated top 10 users of the SAG sector are represented by blue dots (panel a), and those same individuals are also highlighted in blue in the CTE sector (panel b).

The use of a multilevel model also allowed us to estimate the magnitude of individual differences in each sector, i.e., the extent to which there are high/low beluga users in a sector. Our model found that the CTN sector showed very little individual differences in use (Table 1, i.e., low “sd” value) compared to other sectors, suggesting very little differences in high and low users of that sector. While the BSM sector showed large individual differences, with some very high/low users of that sector (Table 1).

Table 1: Parameter estimates from the multilevel multinomial model predicting the probability of capturing a photo-ID by sector. Estimated magnitudes of individual differences (sd) are presented for each sector. Higher ‘sd’ estimates indicate more individual differences in individual use of that sector, whereas lower estimates indicate individuals are using the sector at very similar levels. To facilitate interpretation we have ordered the table by lowest to highest estimates of individual differences. As the number of parameters in the model is large, the overall mean by sector, and estimated correlations between individual differences, are presented in the supplementary section (Table S1).

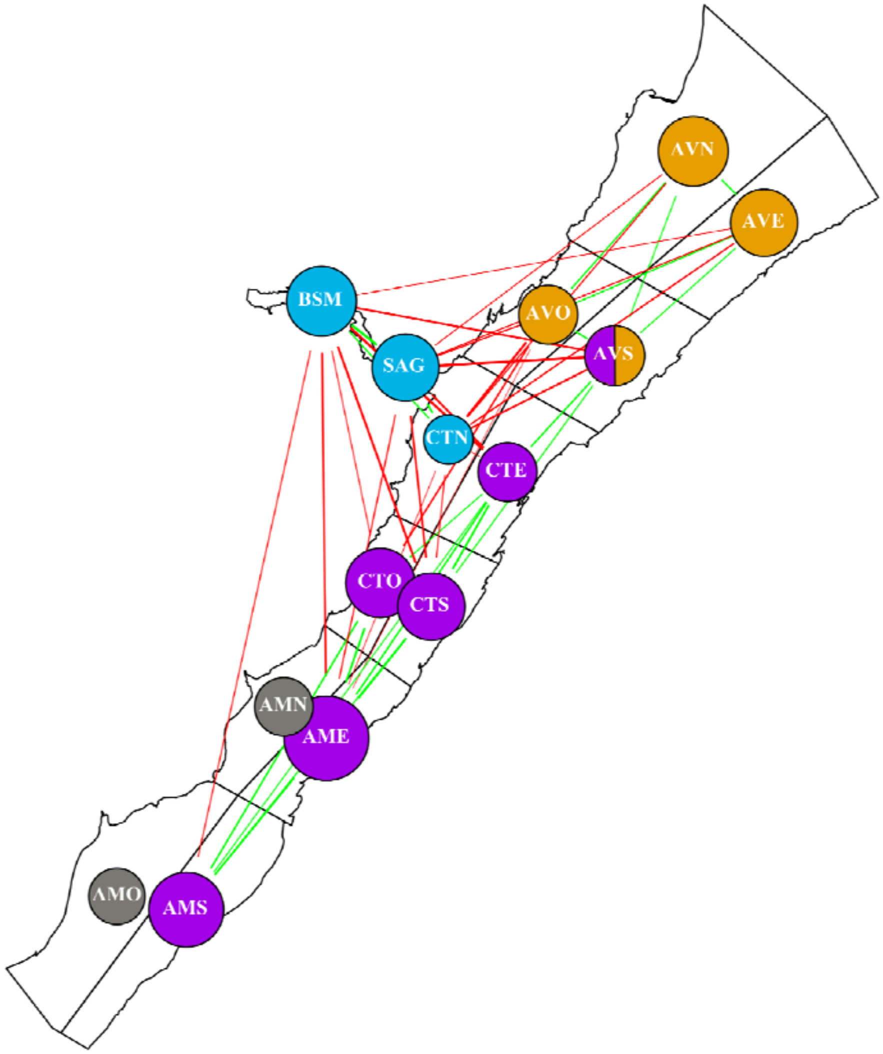
Parameter	**Estimate**	**SD**	**l-95% CI**	**u-95% CI**
sd(mu_CTN)	0.33	0.03	0.27	0.39
sd(mu_AMO)	0.65	0.38	0.05	1.41
sd(mu_AMN)	0.76	0.29	0.14	1.31

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sd(mu_AVO)	0.78	0.05	0.68	0.88
sd(mu_CTE)	0.78	0.05	0.68	0.89
sd(mu_AVS)	0.83	0.07	0.70	0.97
sd(mu_AVE)	1.16	0.20	0.78	1.57
sd(muy_SAG)	1.16	0.08	1.01	1.31
sd(mu_CTS)	1.18	0.11	0.97	1.40
sd(mu_CTO)	1.27	0.14	1.00	1.55
sd(mu_AVN)	1.30	0.17	0.98	1.65
sd(mu_BSM)	1.30	0.11	1.09	1.52
sd(mu_AMS)	1.53	0.22	1.10	1.98
sd(mu_AME)	2.00	0.19	1.64	2.38

Using the between sector correlations to generate a signed network overlaid on top of the sectors in the St. Lawrence, suggests, for example, that individuals that are seen in the SAG sector a lot, are also seen in the BSM and CTN sectors a lot, but are seen very little in the CTE and CTS sectors (Fig. 4). Applying the simple-signed rule and the blockmodeling algorithm to delineate communities, both find that there are three distinct communities (Fig. 4). Though, in the case of AVS the simple sign-rule suggested multi-membership, while the blockmodeling algorithm found AVS to be part of the cluster containing (AVO, AVN, AVE) or that the two clusters (orange and purple in fig. 4) merged into one depending on the choice of weighting parameter (i.e., emphasizing positive or negative edges).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



331

332 Figure 4: Similarity and dissimilarity between sectors in the beluga whale population of the
333 St. Lawrence. The green edges between two sectors signify that the sectors share high/low
334 users, while red edges signify that they have dissimilar high/low users. The lack of an edge
335 signifies that the high/low users of one sector does not provide information about the
336 high/low users of other sectors. Nodes represent sectors, and are coloured based on shared
337 communities: i.e., shared green edges, and no red edges. Node sizes represent the magnitudes
338 of individual differences in use within the sector, i.e., larger nodes suggest larger differences
339 between high and low users.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Discussion

Here we've shown that using photo-ID data with a multilevel multinomial model it is possible to estimate socio-spatial networks, identifying spatial communities while controlling for sampling biases. Our results suggest that the beluga population shows a non-random social-spatial structuring within the summer range.

Within sectors of the St. Lawrence, our model suggests that subsets of belugas are heavily using some sectors, while other sectors show little evidence of differences in use. The magnitude of individual differences in each sector, i.e., how much individuals differ in their probability of being observed in a particular sector, shows that the CTN sector, in particular, has very little in the way of individual differences in the probability of being seen in that sector (Table 1). This result suggests that the CTN sector is used similarly by most individuals, and represents a potential high mixing zone for the population. In contrast, the AME sector shows the highest level of individual differences, suggesting that there are large differences in how beluga are using this sector. These results suggest that the population is not randomly mixing with the St. Lawrence, and that there are belugas that make use of some sectors more than other belugas.

Between sectors of the St. Lawrence, our results add to the evidence that the beluga population cannot be assumed to be randomly mixing within its summer habitat. Rather, comparing the individual differences in beluga usage patterns within sectors suggests similar/dissimilar user populations across sectors (Fig. 4). By using correlations between sector usage patterns to create a socio-spatial network, and running community detection algorithms, our results found that there are spatially distinct communities that make use of particular regions of the St. Lawrence and the Saguenay River (Fig. 4). We found that the beluga population in the St. Lawrence could be separated into three distinct communities: 1)

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The lower St. Lawrence (AVO, AVS, AVE, AVN), the Saguenay River and mouth (BSM, SAG, CTN), and the upper and eastern portion of the St. Lawrence (CTE, CTS, CTO, AME, AMS) (Fig. 4).

Our findings have direct implications for estimating the impacts of anthropogenic disturbances on this population. As the population shows evidence of spatially restricted habitat use, disturbances to particular regions can have a disproportionate impact on particular segments of the larger population. In particular, the cumulative impacts over time are likely to be greatly increased in some segments while reduced in other segments of the population, altering estimations of the distribution of impacts. If cumulative impacts, such as noise, or environmental contaminants, have a threshold beyond which individual survival is greatly reduced, properly estimating the distribution of cumulative impacts can have large implications for conservation management. Our results add to the current understanding of socio-spatial structuring within this population (Michaud, 1993, 2005), and suggest that more empirical data, e.g., photo-ID data, movement data, aerial surveillance, should be collected to better refine socio-spatial mixing in this population.

The modeling approach presented in this paper relies on defined sectors within a particular spatial range, e.g., SAG sector, CTN sector… etc (Fig. 1). In some cases, these delineations can be justified as they identify management zones, but in other cases, the delineation and scale of these sectors can be delineated somewhat arbitrarily. Future work could assess the use of continuous random effects (as opposed to categorical) where individual differences in the probability of being seen could be on a continuous surface. Point estimates of individual differences could then be estimated at any location, and correlations between individual differences obtained between any two points in continuous space. This approach could avoid reliance on user-defined sectors and facilitate a means of looking at the results at different scales (e.g., grids of points at various scales could be used when estimating

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

correlations). Similarly, given that the method requires repeated sampling of individuals to obtain probability of being seen in any one sector, there is a reliance on longitudinal data. In well sampled populations, it could be feasible to estimate the change in time of individuals being seen in particular sectors. Here differences in being seen in any particular area could be explicitly modeled to capture any temporal changes in community substructures, or developmental trajectories related to habitat use.

In terms of implementing the multilevel multinomial model on other photo-ID datasets, the use of test datasets should hold a prominent role in the analysis. The use of permutation/randomization methods to both generate structured and unstructured datasets, while maintaining the sample size distribution of the original datasets, can be very valuable in helping to set model priors and to interpret the final model results. The use of permutation approaches is common in social network analysis (Croft et al., 2011; Farine, 2017), and is becoming more common in statistical workflows more generally (Gelman et al., 2013; McElreath, 2020).

5. Conclusions

We have introduced the use of multilevel multinomial modeling to estimate socio-spatial networks from photo-ID data. We've shown, using testing datasets, that the proposed method is effective at detecting socio-spatial structures. When applied to 18 years of photo-ID data from an endangered population of beluga whales in the St. Lawrence, our results suggest strong evidence that the population has three distinct spatial communities. We suggest that multilevel multinomial models can be effective in extracting socio-spatial structuring within animal populations monitored by photo-ID, and can have direct implications for conservation management.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6. Acknowledgment

We would like to thank all those who worked with and supported the GREMM in collecting these long term data.

7. Funding

Funding was provided by Ministère des Forêts, de la Faune et des Parcs du Québec and Secrétariat à la stratégie maritime du Québec.

8. Data accessibility

The permuted and simulated photo ID datasets are available on github (github.com/tbonne/photoID_multinomial), along with code used in the analysis.

9. Author Contributions

RM collected the data; TRB conceived the analytical methodology and performed the analysis; all authors contributed critically to the drafts and gave final approval for publication.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References:

Chion, C., Lagrois, D., Dupras, J., Turgeon, S., McQuinn, I. H., Michaud, R., Ménard, N., & Parrott, L. (2017). Underwater acoustic impacts of shipping management measures: Results from a social-ecological model of boat and whale movements in the St. Lawrence River Estuary (Canada). *Ecological Modelling*, *354*, 72–87. https://doi.org/10.1016/j.ecolmodel.2017.03.014

Croft, D. P., Madden, J. R., Franks, D. W., & James, R. (2011). Hypothesis testing in animal social networks. *Trends in Ecology & Evolution*, *26*(10), 502–507.

DeFur, P. L., Evans, G. W., Hubal, E. A. C., Kyle, A. D., Morello-Frosch, R. A., & Williams, D. R. (2007). Vulnerability as a function of individual and group resources in cumulative risk assessment. *Environmental Health Perspectives*, *115*(5), 817–824.

Doreian, P., & Mrvar, A. (2015). Structural balance and signed international relations. *Journal of Social Structure*, *16*, 1.

Evans, P. G., & Hammond, P. S. (2004). Monitoring cetaceans in European waters. *Mammal Review*, *34*(1–2), 131–156.

Farine, D. R. (2017). A guide to null models for animal social network analysis. *Methods in Ecology and Evolution*, *8*(10), 1309–1320.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Hupman, K., Stockin, K. A., Pollock, K., Pawley, M. D. M., Dwyer, S. L., Lea, C., & Tezanos-Pinto, G. (2018). Challenges of implementing Mark-recapture studies on poorly marked gregarious delphinids. *PLOS ONE*, *13*(7), e0198167. https://doi.org/10.1371/journal.pone.0198167

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Koivuniemi, M., Auttila, M., Niemi, M., Levänen, R., & Kunnasranta, M. (2016). Photo-ID as a tool for studying and monitoring the endangered Saimaa ringed seal. *Endangered Species Research*, *30*, 29–36.

Koster, J., & McElreath, R. (2017). Multinomial analysis of behavior: Statistical methods. *Behavioral Ecology and Sociobiology*, *71*(9), 138.

Lesage, V., McQuinn, I. H., Carrier, D., Gosselin, J.-F., & Mosnier, A. (2014). Exposure of the beluga (Delphinapterus leucas) to marine traffic under various scenarios of transit route diversion in the St. Lawrence Estuary. *DFO Can. Sci. Advis. Sec., Res. Doc. 2013/125. Iv*, 28.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

McQuinn, I. H., Lesage, V., Carrier, D., Larrivée, G., Samson, Y., Chartrand, S., Michaud, R., & Theriault, J. (2011). A threatened beluga (Delphinapterus leucas) population in the traffic lane: Vessel-generated noise characteristics of the Saguenay-St. Lawrence Marine Park, Canada. *The Journal of the Acoustical Society of America*, *130*(6), 3661–3673. https://doi.org/10.1121/1.3658449

Michaud, R. (1993). *Distribution estivale du béluga du Saint-Laurent: Synthèse 1986 à 1992*. Ministère des pêches et des océans, Direction de la gestion des pêches et de ….

Michaud, R. (2005). Sociality and ecology of the odontocetes. *Sexual Segregation in Vertebrates*, 303–326.

Parrott, L., Chion, C., Martins, C. C. A., Lamontagne, P., Turgeon, S., Landry, J. A., Zhens, B., Marceau, D. J., Michaud, R., Cantin, G., Ménard, N., & Dionne, S. (2011). A decision support system to assist the sustainable management of navigation activities in the St. Lawrence River Estuary, Canada. *Environmental Modelling & Software*, *26*(12), 1403–1418. https://doi.org/10.1016/j.envsoft.2011.08.009

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Perryman, R. J. Y., Venables, S. K., Tapilatu, R. F., Marshall, A. D., Brown, C., & Franks, D. W. (2019). Social preferences and network structure in a population of reef manta rays. *Behavioral Ecology and Sociobiology*, *73*(8), 114. https://doi.org/10.1007/s00265-019-2720-x

Schilds, A., Mourier, J., Huveneers, C., Nazimi, L., Fox, A., & Leu, S. T. (2019). Evidence for non-random co-occurrences in a white shark aggregation. *Behavioral Ecology and Sociobiology*, *73*(10), 138. https://doi.org/10.1007/s00265-019-2745-1

Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, *10*(4), 461–470. https://doi.org/10.1111/2041-210X.13133

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary material:

Use of a preset reference sector:

We use the simulated dataset with structure, presented in the main text, to fit a multilevel multinomial model with and without a preset reference sector. That is, in the model with the preset reference sector we duplicated each individual's photo-IDs and placed them in the reference sector. This results in a probability of 0.5 for being seen in the reference sector for all individuals. The results suggest that using a preset reference sector where all individuals have the same probability of being seen is required to estimate correlations between sectors and produces appropriate socio-spatial networks (Fig S1).

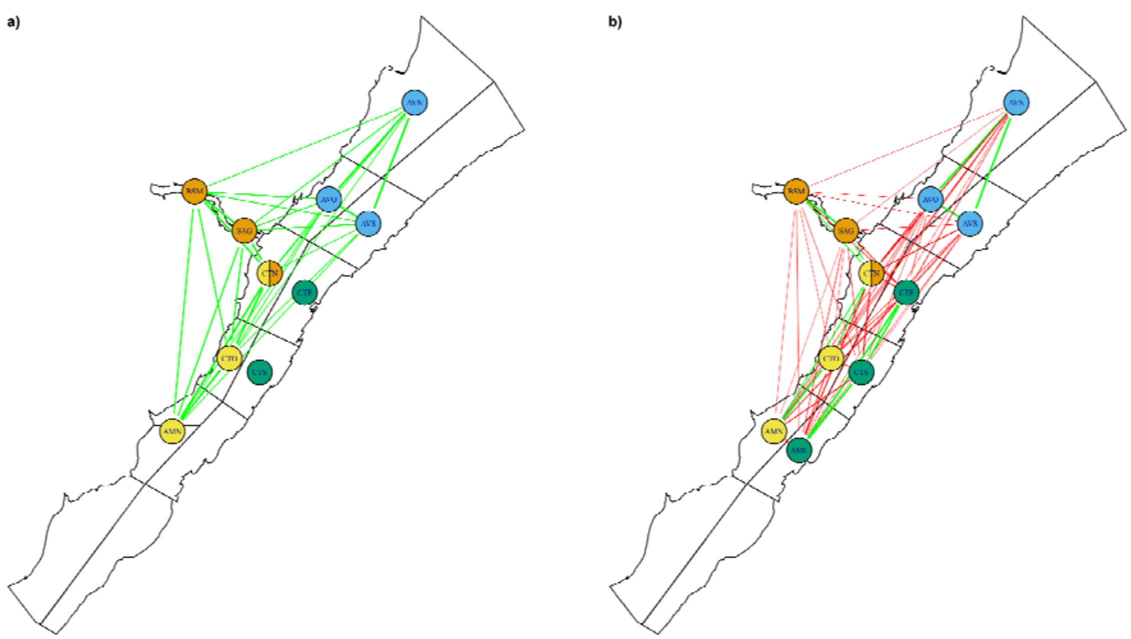


Fig S1: The estimate socio-spatial networks for a multilevel multinormial model a) without a preset reference sector, i.e., AME is used as the reference, and b) with a preset reference sector. The green edges between two sectors signify that the sectors share high/low users, while red edges signify that they have dissimilar high/low users. The lack of an edge signifies that the high/low users of one sector does not provide information about the high/low users of

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

other sectors. Nodes represent sectors, and are coloured based on the communities imposed when simulating the data.

Full model table:

Table S1: Parameter estimates from the multilevel multinomial model predicting probability of capturing a photo-ID by sector. Estimated mean probability (logit scale), individual differences in mean probability ('sd'), and correlation of individual differences between the different sectors are presented along with an estimate of their 95% credible intervals. Positive correlations suggest that the high/low users in one sector are similarly high/low users in another sector, while negative correlations suggest high/low users in one sector are the low/high users in another sector.

Type	**Parameter**	**Estimate**	**SD**	**l-95% CI**	**u-95% CI**
Mean probability (logit scale)					
	mu_AME	-5.85	0.28	-6.43	-5.33
	mu_AMN	-5.60	0.24	-6.15	-5.17
	mu_AMO	-6.68	0.34	-7.41	-6.08
	mu_AMS	-6.33	0.33	-7.00	-5.73
	mu_AVE	-5.47	0.24	-5.99	-5.02
	mu_AVN	-5.06	0.22	-5.52	-4.65
	mu_AVO	-1.82	0.05	-1.93	-1.72
	mu_AVS	-2.90	0.07	-3.04	-2.76
	mu_BSM	-3.78	0.13	-4.04	-3.52
	mu_CTE	-2.20	0.05	-2.30	-2.10
	mu_CTN	-1.24	0.03	-1.30	-1.18
	mu_CTO	-4.58	0.16	-4.92	-4.28

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	mu_CTS	-4.09	0.13	-4.35	-3.85
	mu_SAG	-2.78	0.09	-2.96	-2.61
Individual Differences					
	sd(mu_AME)	2.00	0.19	1.64	2.38
	sd(mu_AMN)	0.76	0.29	0.14	1.31
	sd(mu_AMO)	0.65	0.38	0.05	1.41
	sd(mu_AMS)	1.53	0.22	1.10	1.98
	sd(mu_AVE)	1.16	0.20	0.78	1.57
	sd(mu_AVN)	1.30	0.17	0.98	1.65
	sd(mu_AVO)	0.78	0.05	0.68	0.88
	sd(mu_AVS)	0.83	0.07	0.70	0.97
	sd(mu_BSM)	1.30	0.11	1.09	1.52
	sd(mu_CTE)	0.78	0.05	0.68	0.89
	sd(mu_CTN)	0.33	0.03	0.27	0.39
	sd(mu_CTO)	1.27	0.14	1.00	1.55
	sd(mu_CTS)	1.18	0.11	0.97	1.40
	sd(mu_SAG)	1.16	0.08	1.01	1.31
Correlations between individual differences					
	cor(mu_AME,mu_AMN)	0.39	0.19	-0.03	0.71
	cor(mu_AME,mu_AMO)	0.27	0.24	-0.26	0.67
	cor(mu_AMN,mu_AMO)	0.14	0.24	-0.36	0.57
	cor(mu_AME,mu_AMS)	0.63	0.11	0.39	0.82
	cor(mu_AMN,mu_AMS)	0.25	0.21	-0.19	0.62
	cor(mu_AMO,mu_AMS)	0.27	0.24	-0.26	0.67
	cor(mu_AME,mu_AVE)	0.18	0.16	-0.14	0.46
	cor(mu_AMN,mu_AVE)	0.15	0.21	-0.28	0.56

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cor(mu_AMO,mu_AVE)	0.10	0.23	-0.36	0.52
cor(mu_AMS,mu_AVE)	0.05	0.19	-0.32	0.41
cor(mu_AME,mu_AVN)	-0.04	0.15	-0.33	0.25
cor(mu_AMN,mu_AVN)	-0.02	0.21	-0.42	0.39
cor(mu_AMO,mu_AVN)	0.00	0.22	-0.42	0.44
cor(mu_AMS,mu_AVN)	-0.07	0.18	-0.41	0.28
cor(mu_AVE,mu_AVN)	0.50	0.14	0.19	0.75
cor(mu_AME,mu_AVO)	-0.21	0.10	-0.40	0.00
cor(mu_AMN,mu_AVO)	0.03	0.19	-0.35	0.40
cor(mu_AMO,mu_AVO)	-0.06	0.21	-0.45	0.36
cor(mu_AMS,mu_AVO)	-0.28	0.14	-0.54	0.01
cor(mu_AVE,mu_AVO)	0.48	0.13	0.22	0.71
cor(mu_AVN,mu_AVO)	0.50	0.10	0.28	0.69
cor(mu_AME,mu_AVS)	0.13	0.12	-0.10	0.36
cor(mu_AMN,mu_AVS)	0.21	0.18	-0.16	0.53
cor(mu_AMO,mu_AVS)	0.13	0.21	-0.31	0.50
cor(mu_AMS,mu_AVS)	0.03	0.15	-0.26	0.32
cor(mu_AVE,mu_AVS)	0.40	0.14	0.12	0.66
cor(mu_AVN,mu_AVS)	0.38	0.12	0.14	0.60
cor(mu_AVO,mu_AVS)	0.54	0.08	0.38	0.69
cor(mu_AME,mu_BSM)	-0.57	0.09	-0.75	-0.38
cor(mu_AMN,mu_BSM)	-0.37	0.17	-0.66	0.00
cor(mu_AMO,mu_BSM)	-0.22	0.23	-0.62	0.26
cor(mu_AMS,mu_BSM)	-0.37	0.13	-0.62	-0.10
cor(mu_AVE,mu_BSM)	-0.34	0.13	-0.59	-0.06
cor(mu_AVN,mu_BSM)	-0.22	0.12	-0.44	0.02

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cor(mu_AVO,mu_BSM)	-0.16	0.08	-0.32	0.01
cor(mu_AVS,mu_BSM)	-0.66	0.07	-0.80	-0.51
cor(mu_AME,mu_CTE)	0.54	0.10	0.34	0.72
cor(mu_AMN,mu_CTE)	0.34	0.18	-0.04	0.65
cor(mu_AMO,mu_CTE)	0.21	0.23	-0.29	0.60
cor(mu_AMS,mu_CTE)	0.38	0.14	0.10	0.62
cor(mu_AVE,mu_CTE)	0.07	0.14	-0.21	0.36
cor(mu_AVN,mu_CTE)	-0.11	0.12	-0.35	0.14
cor(mu_AVO,mu_CTE)	-0.10	0.08	-0.26	0.07
cor(mu_AVS,mu_CTE)	0.52	0.08	0.34	0.67
cor(mu_BSM,mu_CTE)	-0.80	0.06	-0.89	-0.68
cor(mu_AME,mu_CTN)	-0.26	0.12	-0.49	-0.03
cor(mu_AMN,mu_CTN)	-0.24	0.19	-0.59	0.16
cor(mu_AMO,mu_CTN)	-0.12	0.21	-0.50	0.32
cor(mu_AMS,mu_CTN)	-0.09	0.15	-0.39	0.21
cor(mu_AVE,mu_CTN)	-0.53	0.13	-0.77	-0.26
cor(mu_AVN,mu_CTN)	-0.44	0.12	-0.67	-0.18
cor(mu_AVO,mu_CTN)	-0.75	0.07	-0.86	-0.60
cor(mu_AVS,mu_CTN)	-0.63	0.09	-0.79	-0.44
cor(mu_BSM,mu_CTN)	0.47	0.09	0.28	0.64
cor(mu_CTE,mu_CTN)	-0.25	0.10	-0.46	-0.05
cor(mu_AME,mu_CTO)	0.63	0.09	0.43	0.79
cor(mu_AMN,mu_CTO)	0.28	0.20	-0.15	0.63
cor(mu_AMO,mu_CTO)	0.16	0.22	-0.30	0.56
cor(mu_AMS,mu_CTO)	0.55	0.13	0.27	0.78
cor(mu_AVE,mu_CTO)	-0.12	0.17	-0.45	0.22

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cor(mu_AVN,mu_CTO)	-0.24	0.15	-0.54	0.07
cor(mu_AVO,mu_CTO)	-0.55	0.09	-0.72	-0.36
cor(mu_AVS,mu_CTO)	-0.16	0.12	-0.39	0.07
cor(mu_BSM,mu_CTO)	-0.33	0.11	-0.53	-0.12
cor(mu_CTE,mu_CTO)	0.39	0.11	0.17	0.60
cor(mu_CTN,mu_CTO)	0.19	0.12	-0.06	0.42
cor(mu_AME,mu_CTS)	0.70	0.08	0.53	0.85
cor(mu_AMN,mu_CTS)	0.42	0.19	0.00	0.73
cor(mu_AMO,mu_CTS)	0.21	0.23	-0.27	0.61
cor(mu_AMS,mu_CTS)	0.42	0.14	0.14	0.67
cor(mu_AVE,mu_CTS)	0.16	0.16	-0.15	0.45
cor(mu_AVN,mu_CTS)	0.06	0.14	-0.23	0.33
cor(mu_AVO,mu_CTS)	-0.04	0.10	-0.23	0.15
cor(mu_AVS,mu_CTS)	0.42	0.11	0.20	0.61
cor(mu_BSM,mu_CTS)	-0.72	0.08	-0.85	-0.55
cor(mu_CTE,mu_CTS)	0.64	0.09	0.46	0.80
cor(mu_CTN,mu_CTS)	-0.37	0.11	-0.57	-0.14
cor(mu_CTO,mu_CTS)	0.48	0.11	0.25	0.68
cor(mu_AME,mu_SAG)	-0.43	0.09	-0.61	-0.24
cor(mu_AMN,mu_SAG)	-0.36	0.17	-0.66	0.00
cor(mu_AMO,mu_SAG)	-0.17	0.22	-0.56	0.29
cor(mu_AMS,mu_SAG)	-0.22	0.13	-0.47	0.03
cor(mu_AVE,mu_SAG)	-0.44	0.12	-0.68	-0.19
cor(mu_AVN,mu_SAG)	-0.37	0.10	-0.57	-0.16
cor(mu_AVO,mu_SAG)	-0.48	0.06	-0.60	-0.35
cor(mu_AVS,mu_SAG)	-0.80	0.06	-0.90	-0.68

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.06.442957; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cor(mu_BSM,mu_SAG)	0.83	0.04	0.74	0.91
cor(mu_CTE,mu_SAG)	-0.68	0.06	-0.79	-0.55
cor(mu_CTN,mu_SAG)	0.67	0.07	0.52	0.80
cor(mu_CTO,mu_SAG)	-0.12	0.11	-0.32	0.09
cor(mu_CTS,mu_SAG)	-0.62	0.08	-0.77	-0.45

Note on interpreting low sd within sectors:

Similar to the CTN sector, the sector AMO is also estimated to have a low magnitude of individual differences. However, it has a large uncertainty in this estimate. This highlights that it is possible to have little individual differences in a sector due to either: 1) limited data, resulting in all individuals being pooled to the mean value, and 2) limited data is not a factor, but individuals are using this sector relatively equally. Care should therefore be taken when interpreting the magnitude of individual differences, nevertheless, the estimated uncertainty around magnitude estimates is one way to help identify sectors with limited data.