Aleutian Tern Colony Abundance

2018 Field Season




6 Nov 2019

Prepared for:
Aleutian Tern Technical Committee
Funded by:
National Fish & Wildlife Foundation
Alaska Department of Fish and Game

Prepared by:
Trent McDonald[1], Jaime Thompson[1], Michael Gerringer[1], Jason Carlisle[1], Robin Corcoran[2], John Skinner[3], Abram Fleishman[4], Matthew McKown[4], and **Kelly Nesvacil[3]**

A collaboration between:

Affiliations:

[1]Western EcoSystems Technology, Inc.; 415 West 17th Street, Suite 200; Cheyenne, WY 82001

[2]U.S. Fish and Wildlife; Kodiak National Wildlife Refuge; 1390 Buskin River Rd.; Kodiak, AK 99615

[3]Alaska Department of Fish and Game; Threatened, Endangered, and Diversity Program; P.O. Box 115526; Juneau, AK 99811

[4]Conservation Metrics; 145 McAllister Way; Santa Cruz, CA 95060

EXECUTIVE SUMMARY

1. At a workshop in January 2018, the Aleutian Tern Technical Committee and other workshop participants recommended pilot testing two direct and one indirect method for estimating Aleutian Tern (ALTE; *Onychoprion aleuticus*) nesting colony abundance. One recommended method involved low-altitude photography obtained by unmanned aerial vehicles (UAV). Another method involved counts in ground-based photographs. The third method involved relating counts to call rates derived from acoustic recorders. In early June 2018, three teams of researchers pilot tested these methods at a total of 16 ALTE colonies in Alaska.

2. *UAVs:* Automated and semi-automated photo recognition routines counted ALTE and Arctic Tern (ARTE, *Sterna paradisaea*) in UAV photography and estimated combined species density. Terns proved relatively easy to identify in UAV photographs, but we could not reliably differentiate ALTE and ARTE from photos alone and hence relied on direct count species ratios. The UAV method estimated densities ranging from 0 to 2.75 ALTE per hectare across six single and mixed-species colonies. We estimated ALTE abundances of 231 (95% CI = 156 to 322) at Black Sand Spit near Yakutat, 2 (95% CI = 0 to 3) at the Kenai Headquarters colony, 0 at the Burton Ranch colony on Kodiak Island, 47 (95% CI = 37 to 57) at the Kalsin Bay colony on Kodiak Island, 14 (95% CI = 9 to 19) at the Middle Bay colony on Kodiak Island, and 17 (95% CI = 12 to 22) at the Women's Bay colony on Kodiak Island. Estimates of abundance at the small colonies derived from UAV photography were lower but generally agreed with contemporaneous direct counts. UAV estimates of abundance at Black Sand Spit, the largest colony we surveyed, were 50% of the direct counts taken the same week, but equaled direct counts taken 3 days before and 9 days after UAV surveys.

3. *Photo counts*: A strong relationship existed between ground-based photo counts and maximum direct counts. We estimated that maximum direct count increased by 0.58 birds for every additional bird counted in photos (R^2 = 0.85). This high correlation makes direct counts and photo counts more or less equivocal in usage because it is possible to convert from one count to the other. The advantages of photo counts relative to direct counts include ease of implementation in the field and creation of a permanent photo record.

4. *Song meters*: We recorded and detected ALTE vocalizations at all 16 deployment sites. Colony specific call rates peaked during different parts of the summer. We estimated that adult direct counts increased by 11% for every one-unit increase in calls per minute when measured during the ± 7 days surrounding the count date. We found positive but statistically insignificant relationships between call rate and nest density.

INTRODUCTION

The Aleutian Tern (ALTE; *Onychoprion aleuticus*) is an uncommon seabird that nests in coastal areas of Alaska and Russia (North 2013). Renner et al. (2015) estimated the number of ALTE at known colonies in Alaska to be declining 8.1% annually since 1960. In 2015, a multi-stakeholder group coalesced to identify information needs and research priorities in an effort to understand the conservation status and needs of the species (summarized in McDonald and Carlisle 2018). The multi-stakeholder group focuses broadly on designing a statewide monitoring program that generates unbiased estimates of ALTE abundance in Alaska. An important part of any future statewide effort will be estimation of the number of individuals at a single nesting colony in a given year. This single-colony estimate has been identified as a critical first hurdle to implementation of future statewide efforts (McDonald and Carlisle 2018).

Certain past ALTE survey methods have centered on human observers flushing birds at nesting colonies and counting birds in flight (Pyare et al. 2013). Other surveys, especially those on Kodiak Island, take advantage of “natural” flush events to count birds in the air. Biologists and others have raised concerns over these types of direct counts, specifically that the proportion of terns that flush is variable, that flying terns are difficult to count accurately, that human-induced flushing may have adverse effects on nesting productivity, and that reliable species identification is difficult where Arctic Terns (ARTE; *Sterna paradisaea*) are present (McDonald and Carlisle 2018).

With a goal toward alleviating some of the concerns over direct counts, the Aleutian Tern Technical Committee and other participants at a 2018 workshop discussed pilot testing three methods for estimating colony abundance. Workshop participants recommended pilot testing low-altitude aerial imagery taken from small unmanned aerial vehicles (UAV, or “drone” method), ground-based photos of flying birds (photo count method), and audio call monitoring (song meter method). In addition, workshop participants emphasized the importance of continuing to apply the direct count method for comparability to previous years. If, for example, one of the pilot methods proves to be highly correlated with direct counts, it may be possible to make past counts comparable to those from the pilot method or the Technical Committee may opt for direct counts only because they are less expensive.

Magness et al. (2019) tested and successfully estimated nest density using a UAV method during summer 2017 at a single mixed-species colony near Kenai AK. Magness et al. (2019) used human observers to count nesting terns in collected photographs and to differentiate between ALTE and ARTE. Importantly, they also found that presence of a drone overhead did not influence nest attendance or tern activity.

Ground-based photo-counts attempted to standardize and reduce variability inherent in traditional visual ground-based direct counts by using multiple photographs in quick succession in an attempt to “freeze” the action and capture flying individuals (McDonald and Carlisle 2018). Photo-counts should reduce the potential for over counting flying birds if multiple photos do not capture the same bird or if duplicates can be identified. Photo-counts still rely on anthropogenic or natural flushes and thus are dependent on the proportion of flushing individuals and estimation of the species ratio at mixed colonies.

The song-meter method deployed acoustic sensors at nesting colonies to record ALTE calling activity over most of the nesting season. During deployment, field technicians intermittently visited the colonies to make visual counts and to assess nest density near the sensors. The song-meter method sought to relate acoustic call rates to concurrent measures of abundance from traditional survey methodologies and hence estimate phenology and relative abundance. Similar song-meter approaches have successfully estimated adult or nest abundance of Forster's Tern (*Sterna forsteri*) (Borker et al. 2014), Leach's Storm-Petrel (*Hydrobates leucorhoa*) (Orben et al. 2019), and Cory's Shearwater (*Calonectris borealis*) (Oppel et al. 2014).

Following recommendations from the Aleutian Tern Technical committee, four teams of researchers conducted pilot-tests of single-colony estimation methods during June 2018 at six mixed-species colonies in southeast and southcentral Alaska. Our objectives were to refine both the in-field and post-field procedures of all four methods. In particular, UAV, photo counts, and song meter methods all require substantial post-field processing in order to produce a colony size estimate. The UAV-based method requires development of a photo-recognition algorithm to identify and count terns in UAV-collected photographs. The photo count method requires counting terns in derived photographs and accounting for duplicates, possibly by stitching photos into a mosaic. The song meter method requires analysis of audio recordings to identify and count ALTE songs and to relate the number of such calls to surrounding nest density or direct counts.

This document is a summary of all four pilot colony-sampling efforts conducted in June 2018. This report contains methodological descriptions and results of the three test methods (UAV, photo count, and song meter), as well as comparisons to the fourth traditional method (direct counts). At the end, we provide a summary of recommendations for future surveys.

STUDY AREAS

For testing of the UAV method, we non-randomly selected single species and mixed ALTE/ARTE colonies in coastal areas of southcentral and southeastern Alaska for study (Figure 1). We tested the photo count method at seven colonies both on and off Kodiak Island's road system (Kalsin Bay, Middle Bay, Alligator Island, Foul Bay, SE Viekoda, Women's Bay, and Three Spruce Island). The song meter team applied their method at sixteen colonies located throughout southern Alaska (Appendix C Figure 3).

METHODS

This section contains one subsection for each of the three primary sampling methods. We provide a summary here and refer to detailed methods in Appendices A, B, and C. A description of the direct count method appears in McDonald and Carlisle (2018).

Un-manned Aerial Vehicle Sampling Methods

We deployed three UAV's at six ALTE colonies in southcentral and southeast Alaska. Four UAV colonies were located on Kodiak Island's road system (Women's Bay, Middle Bay, Kalsin Bay, Burton Ranch), one colony was on the Alaskan road system (Kenai NWR Head Quarters) and one was a remote colony (e.g., Black Sand Spit near Yakutak). The aerial extents of all UAV colonies on Kodiak were approximately the same, while the Black Sand Spit colony was appreciably larger (Table 1). We visited colonies for UAV surveys once between 6 Jun 2018 and 13 Jun 2018 during the anticipated peak of nesting activity.

We conducted all UAV flight missions during daylight hours using high-resolution standard red-green-blue (RGB) imagery. We tested both a census and sample UAV flight patterns over ALTE colonies depending on colony size. At smaller colonies, we employed the census mode wherein UAV photographs overlapped to cover the entire colony. At larger colonies, we employed a spatially balanced sampling method wherein the UAVs flew to pre-determined points and obtained non-overlapping photographs. The spatially balanced method selected 99 points within each colony using the BAS spatially balanced sample algorithm (Robertson et al. 2013, 2017) and applied the travelling salesman algorithm (Kruskal 1956, Garfinkel 1985) to order the points in an efficient survey order (least distance flown, approximately). We photographed colonies at above-ground altitudes of 15 meters, 20 meters, and 30 meters (Appendix A Table 2).

Photo-Recognition

Following photo collection, we trained a customized neural network to search for and detect terns in UAV-derived photographs. During training and afterward, we tolerated a higher than normal number of false positives (not actual terns) due to difficulties distinguishing ALTE and ARTE at the resolutions of our photographs. Later, we verified and determined the species of all putative detections. This allowed us to compare density estimates from a fully-automated method (no verification) to those produced by a computer-assisted method (human verification).

To construct the tern detector's training data set, human observers initially found and marked terns and gulls in a subsample of all images and outlined ("painted") the targets when identified. Those observers initially found terns and gulls in approximately 2,000 of the 11,000 images collected during all surveys at the six colonies. Our initial round of detector training consisted of passing these "painted" images through the detector's neural network which allowed the network to "learn" characteristic differences between the "painted" areas the non-"painted" areas. In other words, by comparing "painted" and non-"painted" areas, the network "learned" what was, and what was not, a tern or gull. Following initial training, we applied the detector to all photos in an effort to find additional terns. Observers again "painted" previously un-detected terns and we completed a second round of training to improve the network's sensitivity and specificity. At the end of the second round of training, we again applied the network to all photos and we labeled the resulting total number of terns the "automated" counts.

Despite two rounds of training and efforts to reduce the number of false-positive detections, we became aware of a substantial number of false-positives in the automated detections. Hence, human observers inspected all detections and classified them as either a tern (ALTE or ARTE), gull, or neither. We labeled this verified count after the second round of detector training as the "computer-assisted" counts because human observers verified all putative detections. During computer-assisted counting, human observers only verified putative detections and did not search photos where the automated detector count was zero. Additional details behind the tern detector, its training, as well as example images, are in Appendix A.

Density and Abundance Estimation

We computed density and abundance of ALTE in a colony from the ground area that photographs covered, the count of terns in each photo, the area of sampled polygons, and the total area of each colony. Using the count and footprint areas from all photographs taken during a mission, we computed average density during the mission as total count over total area sampled by the mission (Appendix A Equation 1). To compute average density at the colony over multiple missions, we multiplied mission density by area of the polygon each mission sampled and divided by total area sampled over all missions (Appendix A Equation 2). We computed colony abundance by multiplying colony density by colony area (Appendix A Equation 3). Finally, we obtained species ratios (ALTE/ARTE) from concurrent direct counts and multiplied it into colony abundance to estimate total ALTE at the colony. We computed standard errors and confidence intervals by bootstrap resampling raw photo counts. Additional details behind the UAV abundance methods appear in Appendix A.

Photo Counts Methods

The general field protocol for photo-counts was largely the same as the protocol for direct counts except for the presence of a photographer and images. During natural flush events, the photographer attempted to cover the entire colony with a single burst of photos lasting 3-5 seconds. The goal was to take multiple photos with overlap so that all in-flight birds appear in the final photo mosaic. Ideally, the photographer captured the entire colony in a single photo or only 2-5 images.

We conducted photo-counts at the following eight Aleutian tern colonies in the Kodiak Archipelago: Kalsin Bay, Middle Bay, Alligator Island, Foul Bay Island 9, Foul Bay Island 9NW, SE Viekoda Bay Islands, Women's Bay Barge, and Three Spruce Island. We visited the first six colonies in the order given between 1 June 2018 and 28 June 2018. We visited the Women's Bay Barge colony on 19 July 2018. We visited Three Spruce Island on 15 August 2018. Some colonies contained nesting ARTE terns alongside nesting ALTE. We could not differentiate ARTE from ARTE in photos and present counts of both tern species together. We surveyed two colonies (Alligator Island and SE Viekoda Bay Islands) from the same location twice during the season for comparison. For analysis, each photo pan contained approximately 1-65 images, and for colonies with multiple images, we examined the degree of photographic overlap and composited only 2-9 of the images to conduct the counts. We manually stitched together some photographs using Microsoft Paint. We digitally stitched together other photographs using Microsoft Image Composite Editor (ICE) and manually counted terns in the composite after importing into Microsoft Paint.

Song Meter Methods

We deployed 16 Song Meter 4 (SM4) sensors and 8 Song Meter 2 (SM2) sensors, both manufactured by Wildlife Acoustics, at 19 ALTE colonies in Southeast Alaska starting 1 May 2018. Field crews retrieved data from 24 deployments at 16 colonies for analysis. We programmed the SM4 units to record 1-minute out of every 6-minutes and the SM2 to record 1 minute out of every 5 minutes 24 hours a day. We collected direct counts at various points in the season at 13 colonies containing one or more acoustic sensors. We also collected counts of nests within 5-, 10-, 15-, and 20-m radii around sensors at 6 colonies (Middle Bay, Women's Bay, Italio, Akhiok, Kalsin Bay, and Black Sand Spit). We computed nest density from these counts as, $Density = \frac{count}{\pi r^2}$, where *r* was the radius of the circle in which each count was conducted.

We detected and counted ALTE calls by training a Deep Neural Network (DNN) to classify 2-second segments of audio recordings as either having an ALTE call or not. The DNN was a software algorithm that attempted to learn the combinations of spectro-temporal features that best differentiated target sounds from other sounds. We trained the detector to find three call types: a long trill, a short trill, and a buzz. Calls we targeted contained peak energies between 2,500 and 5,000 hz. In the end, the detector was not able to distinguish among call types and the metrics of ALTE activity presented here represent all calls regardless of type. The detector also did not effectively reject single-note alarm calls, and some alarm calls are included in results.

We trained the DNN on a dataset of 28,567 manually labeled 2-second sound clips with examples of 59 distinct sounds (classes). We evaluated model performance using ROC curves that plotted the True Positive Rate (# clips with Aleutian Tern calls divided by total number of clips) and False Positive Rate (# clips that did not have Aleutian Tern calls divided by the total clips the DNN flagged as containing Aleutian Tern calls). We built three DNN models from this dataset, using different class combinations for each, and chose the model that best predicted ALTE calls in a hold-out set of audio clips. The hold-out test dataset was manually reviewed and contained 1,185 2-second clips with Aleutian Tern among 7,411 randomly selected clips. The best DNN model (AK_Wetlands_Multi_6Feb19_V8) returned 91.6% of known ALTE calls in the test dataset with a false positive rate of 0.57%. In total, we classified 1,193,893 sounds as ALTE calls across all 22 deployments. We confirmed at least one Aleutian Tern call at each site through manual review.

We computed peak ALTE call rate (#/minute) each day between 90 and 270 minutes after sunrise. We choose this summary period because visual inspection of diel activity patterns indicated peak ALTE activity during this time of day (Appendix C Figure 5). We computed a general activity metric as the average daily peak call rate over 15 May to 15 August (3 months). For comparison to direct counts and nest density, we calculated call rates from a subset of 7 randomly selected days in a 15-day window centered on the date of the direct count. This random subset of days allowed us to use the same number of days of recording for each sensor-count pair and include data from sites that had acoustic data from only before or after direct counts. No recordings occurred on the days immediately prior, or following, the direct count when we moved sensors on the count date. Direct counts with fewer than 7 associated days of data within the 15-day buffer were excluded from the analysis (n=8). Some direct counts covered an area encompassing two acoustic sensors. In these cases, we averaged the call rates for the two sensors.

We investigated the relationship between vocal activity and colony counts using two models. First, we used the entire dataset with observations from all stages of the breeding season to fit a zero-inflated negative binomial model with log-link. Second, we used observations made during incubation to fit a regular negative binomial model with log-link. Data from the incubation period contained fewer zeros and hence a regular, non-zero-inflated, model adequately fitted these data.

We also related call rates to nest density within 5-, 10-, 15-, and 20-m of each recorder measured during each visit to the site. Nest density data were sparse (n=12) and did not lend themselves to complex modeling. We related call rates to nest density using linear regression (ordinary least squares) restricted to the incubation period (n=8). We fitted separate models for each radius.

RESULTS

UAV Results

As anticipated, our automated detector flagged a substantial number of false positives (i.e., objects visually similar to but not terns). Most false positives were wood debris of similar color and general shape as nesting terns. Assuming the computer-assisted counts are correct, 58% of our detector's detections across all missions were ultimately deemed false-positives. Examples of true positive, false positive, and false negative detections appear in Appendix A Addendum B. We rely on the computer-assisted counts below and only plot automated counts for reference.

Colony-wide density from the computer assisted method varied from 0 (Burton Ranch) to 2.75 (95% CI 2.18 to 3.35; Kalsin) individuals per hectare (Table 1; Figure 2). Colony abundance from the computer-assisted counts varied from 0 (Burton Ranch) to 231 individuals (95% CI = 156 to 322; Black Sand Spit) (Table 1; Figure 2).

At smaller colonies, we found that abundance derived from the computer-assisted method generally matched contemporaneous direct counts (Figure 3a). At Women's Bay, UAV estimates and contemporaneous direct counts were virtually identical. At Kalsin and Middle Bay, UAV estimates were 32% (47 vs 62) and 64% (14 vs 23) lower, respectively, than the closest direct count in time.

At Black Sand Spit, the largest colony we sampled, UAV-based estimates were practically identical to a formal direct count conducted 9 days later (230 UAV vs 240 direct count; Figure 3b). Other direct counts at Black Sand Spit, both before and during UAV surveys, were less formal and should be considered rough estimates. Nonetheless, most informal counts prior to the UAV surveys agreed with the general magnitude of UAV estimates.

Photo Count Results

We observed ground-based direct counts ranging from 7 to 129 terns (Appendix B Table 2). The ordinary regression relationship we fitted to these data indicated that maximum direct count increased by an average of 0.58 birds for every additional bird counted in photos (R^2 = 0.85, Figure 4, Appendix B Figure 1). Hence, on average photo counts exceeded direct counts despite the fact that direct counts numerically exceeded the associated photo count in 5 of 11 cases. Photo counts tended to exceed direct counts by substantial numbers at large colonies (counts > 75 approximately) which in turned cause slope of the regression to differ substantially from 1.0.

In four cases, we obtained counts by both manual stitching (Microsoft Paint method) and digital stitching (Microsoft ICE + Paint method). We observed reasonably similar counts among these methods except at Kalsin Bay, one of the larger colonies (Appendix B Table 2). These methodological differences did not appear directional, but this observation is tentative because the number of test cases was low (i.e., 4).

Song Meter Results

We report details on song meter deployments, recording periods, and location in Appendix C – Song Meters. Here, we report song meter results related to phenology, density, and abundance.

Song meters detected ALTE calls at all deployment sites with usable data. In general, we observed two activity peaks; the first from mid-May to mid-June and a second peak from early-July to late-July (Appendix C Figure 7). At Black Sand Spit, the later peak ended slightly sooner than at other sites. Mean daily call rates during 15 May to 15 Aug varied among sites from 0.04 ± 0.08 at Kenai SM4 to 10.43 ± 8.45 calls per minute at Naknek2 (Appendix C Figure 8; Appendix C Table 3; Appendix C Figure 9).

We completed 67 direct ALTE counts during the season. Of these, 66 counts were available for modeling because they occurred during an active acoustic survey window with at least 7 days of acoustic data temporally near the count date. The number of direct counts varied from one (Italio and Three Spruce Island) to 16 (Middle Bay), largely mimicking ease of access. We observed the highest direct counts at Black Sand Spit (80 ± 34.6 sd) and Naknek (78.4 ± 50.9 sd), followed by the single count at Italio (45; Appendix C Figure 10). Average ALTE direct counts at the remaining colonies varied from 0 (Pasagshak River) to 21 (Three Spruce Island).

We obtained nest density at six colonies. We obtained only a single nest count at four of these six (Akhiok, Italio, Kalsin Bay, Womens Bay). We obtained three nest counts at the other two colonies (Black Sand Spit and Middle Bay). Nest searches found few nests within 20-m of the sensors, with a maximum of 6 nests found within 20m of the sensor at Black Sand Spit1. Nest density in the 20m circle ranged from 0.00 to 0.026 nests m^2.

The zero-inflated negative binomial model we fitted to direct counts obtained over the entire season suggested that call rates were inversely related to the probability of detecting a false zero count (i.e. zero count when birds were actually present; odds ratio: 0.26, p = 0.14). This part of the model implied that the probability of counting at least ALTE rose as call rates increased. The structural part of the model indicated that direct colony counts increased by an average of 11.6% ± 3.5 se (95% CI: 4.3-19.5%) for every one unit increase in calls per minute (Figure 5, Appendix C Figure 12). According to predictions from the negative binomial model, one count at Naknek 1 and one count at Black Sand Spit 3 represented extreme outliers (i.e., abnormally high direct counts) among observations in the full dataset.

We conducted 24 of the 66 usable direct counts during the associated colony's incubation stage. The regular negative binomial model we fitted to this incubation stage data indicated that direct counts increased on average by 17.7% ± 3.5 se (95% CI: 9.2-28.7%) for every one unit increase in calls per minute (Figure 6, Appendix C Figure 13). Compared to model results from the full season, the rate of increase observed during incubation was higher and overall the model better fit the observed data.

We conducted 12 nest count surveys during acoustic survey periods. Of these, 8 nest surveys occurred during the incubation period. We did not observe enough nests within 5-m of the sensor to model using ordinary regression. At other survey radii, the ordinary regression model indicated a positive but statistically insignificant relationship between call rates and nest density (R^2=0.006 - 0.10, p=0.44-0.84; Appendix C Figure 14). We hypothesize that the small number of observed nests, inconsistency in nest counting methods, difficulty finding nests near sensors, and the dispersed nature of Aleutian Tern nest sites contributed to the lack of statistical significance in these linear regressions. We also hypothesize that the positive relationships we observed between nest density and call rate will become more apparent after additional data collection.

DISCUSSION

UAV Counts

Sampling tern colonies using UAVs proved efficient and consistent. For colonies on the Kodiak road system, we were able to mobilize the UAVs and associated equipment, conduct multiple missions, de-mobilize, and return to base within three-quarters of a workday. The UAV method required additional mobilization and de-mobolization logistics relative to direct counts at remote colonies (e.g., Black Sand Spit) because the UAV method required more hardware than direct counts (e.g., the UAV, controller, batteries, helipad, etc.) Clearly, most of the efficiency of the UAV method depends on travel distance and survey mode, but when reasonable, the in-field portion of UAV sampling was short and easily managed.

Manually counting terns in photos is prone to errors of omission. Our computer-assisted method directed observers to likely tern targets in photographs and drastically sped up counting as well as increased accuracy. Observers reported that it was much easier to inspect pre-identified objects in photos than to scan entire photos for terns. Our computer- assisted method afforded an approximate 10-fold reduction in photo processing time.

Our UAV-derived abundance estimates were always lower than the associated direct count. For small colonies, the UAV and direct counts were comparable. The UAV produced comparable estimates to the formal direct count taken at Black Sand Spit, the largest colony surveyed, one week after. At small colonies, we surveyed in census mode (overlapping photos). At the single large colony we surveyed, Black Sand Spit, we sampled some areas in census mode and other areas in sample mode. We theorize that the lack of agreement between UAV and direct counts is a function of colony size, over-counting by observers on the ground, or undercounting of birds in the air by the UAV method. Even at small colonies, the UAV abundances were lower than direct counts. Lower UAV counts could be a function of the fact that UAV photography does not capture flying birds or some degree of double-counting by observers. Nonetheless, the general agreement between UAV and direct counts is surprising given that one technique surveys birds on the ground while the other surveys birds in the air. If both UAV and direct counts are accurate, at least at small colonies, this suggests that the natural flushes our observers counted contained a large proportion of the birds associated with the colony and available to be counted. That is, our results suggest either that the proportion of flushing ALTE during those events was nearly 100% of birds in the colony at the time or that observers are over-counting ALTE.

Photo Counts

Two substantial issues presented themselves during processing of the photo count images. First, individual birds appeared in multiple images of the composite image and hence we counted them twice. Second, we lost individual birds in composited images ("ghosting") that caused some undercounting of the actual number of birds at a colony. These two issues had less impact at small colonies where we observed greater agreement between direct and photo counts.

The process of stitching images was time consuming. For small colonies (n < 75 terns), the extra effort required to stitch photos appeared to be un-necessary because maximum direct count and photo counts appeared equivocal and direct counts are faster and easier to collect overall. Direct counts do not require post-processing.

The benefits of the photo-count method relative to direct counts are the creation of a permanent record, a relatively tight relationship between direct- and photo-counts across a range of colony sizes, and the method is relatively quick and easy to accomplish in the field. The downsides of this method include the fact that it relies on anthropogenic or natural flushes and thus is dependent on the proportion of flushing individuals and post processing of images is required. Additionally, double counting and ghosting of birds in composited images affects accuracy and should be quantified in subsequent research.

Song Meters

The relationship between colony counts and call rate, and the ability to detect subtle differences in timing and duration of activity, make passive acoustic monitoring an effective method for detecting and monitoring Aleutian terns in Alaska. However, several questions remain about how best to design large scale acoustic surveys for the species. To fully realize the potential that acoustic monitoring holds for Aleutian terns, future work must address the spatial scale at which an acoustic sensor is monitoring, the most effective arrangement of sensors within a breeding aggregation, and if there is a relationship between nest density and vocal activity at scales larger than 20-meter radius around the sensor. While guidance on some of these questions may be available in the literature, the Aleutian Tern Technical Committee will need to focus thought and resources to develop a robust framework and sampling design for statewide survey efforts.

REFERENCES

Borker, A. L., M. W. McKown, J. T. Ackerman, C. A. Eagles-Smith, B. R. Tershy, and D. A. Croll. 2014. Vocal activity as a low cost and scalable index of seabird colony size. Conservation Biology 00:1–9. <https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.12264>. Accessed 8 Aug 2019.

Garfinkel, R. S. 1985. Motivation and modelling. E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors. The traveling salesman problem - A guided tour of combinatorial optimization. Wiley and Sons.

Kruskal, J. B. 1956. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Proceedings of the American Mathematical Society 7:48–50.

Magness, D. R., T. Eskelin, M. Laker, and H. M. Renner. 2019. Evaluation of small aerial systems as a census tool for Aleutian Tern Onychoprion aleuticus colonies. Marine Ornithology 16:11–16.

McDonald, T. L., and J. D. Carlisle. 2018. Report on the 2018 Aleutian Tern Conservation Planning Meeting. Cheyenne, WY.

North, M. R. 2013. Aleutian Tern (Onychoprion aleuticus), version 2.0. P. G. Rodewald, editor. The Birds of North America. Cornell Lab of Ornithology, Ithaca, NY.

Oppel, S., S. Hervias, N. Oliveira, T. T. Pipa, C. Silva, P. Geraldes, E. I. Goh, and M. W. McKown. 2014. Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. Nature Conservation 7:1. <https://natureconservation.pensoft.net/article/1355/download/pdf/>. Accessed 8 Aug 2019.

Orben, R. A., A. B. Fleishman, A. L. Borker, W. Bridgeland, A. J. Gladics, J. Porquez, P. Sanzenbacher, S. W. Stephensen, R. Swift, M. W. McKown, and R. M. Suryan. 2019. Comparing imaging, acoustics, and radar to monitor Leach’s storm-petrel colonies. PeerJ 7:e6721. <https://peerj.com/articles/6721>. Accessed 8 Aug 2019.

Pyare, S., M. I. Goldstein, D. Duffy, S. Oehlers, N. Catterson, and J. Frederick. 2013. Aleutian Tern (Onychoprion aleutica) research in Alaska: survey methodology, migration, and statewide coordination. Juneau, Alaska.

Renner, H. M., M. D. Romano, M. Renner, S. Pyare, M. I. Goldstein, and Y. Artukhin. 2015. Assessing the breeding distribution and population trends of the Aleutian Tern Onychoprion aleuticus. Marine Ornithology 43:179–187.

Robertson, B. L., J. A. Brown, T. L. McDonald, and P. Jaksons. 2013. BAS: Balanced acceptance sampling of natural resources. Biometrics 69:776–784. <http://dx.doi.org/10.1111/biom.12059>.

Robertson, B. L., T. L. McDonald, C. J. Price, and J. A. Brown. 2017. A modification of balanced acceptance sampling. Statistics & Probability Letters 129:107–112. Elsevier B.V. <http://linkinghub.elsevier.com/retrieve/pii/S0167715217301797>.

TABLES

Table 1: Final estimates of ALTE density and abundance at six nesting colonies sampled by the UAV method in 2018. Estimates presented here derive from the computer assisted count method. Species ratios derive from direct counts during visits temporally close to UAV flights.

		Species	Density (#/ha)			Abundance		
Colony	Area (ha)	Ratio (%ALTE)	Est	Low 95%	High 95%	Est	Low 95%	High 95%
Black Sand Spit	226.87	0.8	1.017	0.687	1.418	231	156	322
Burton Ranch	10.51	1	0.000	0.000	0.000	0	0	0
Kalsin Bay	17.04	0.48	2.750	2.183	3.353	47	37	57
Kenai HQ	5.37	1	0.315	0.079	0.632	2	0	3
Middle Bay	31.42	1	0.447	0.301	0.615	14	9	19
Women's Bay	6.62	0.94	2.541	1.831	3.354	17	12	22

FIGURES

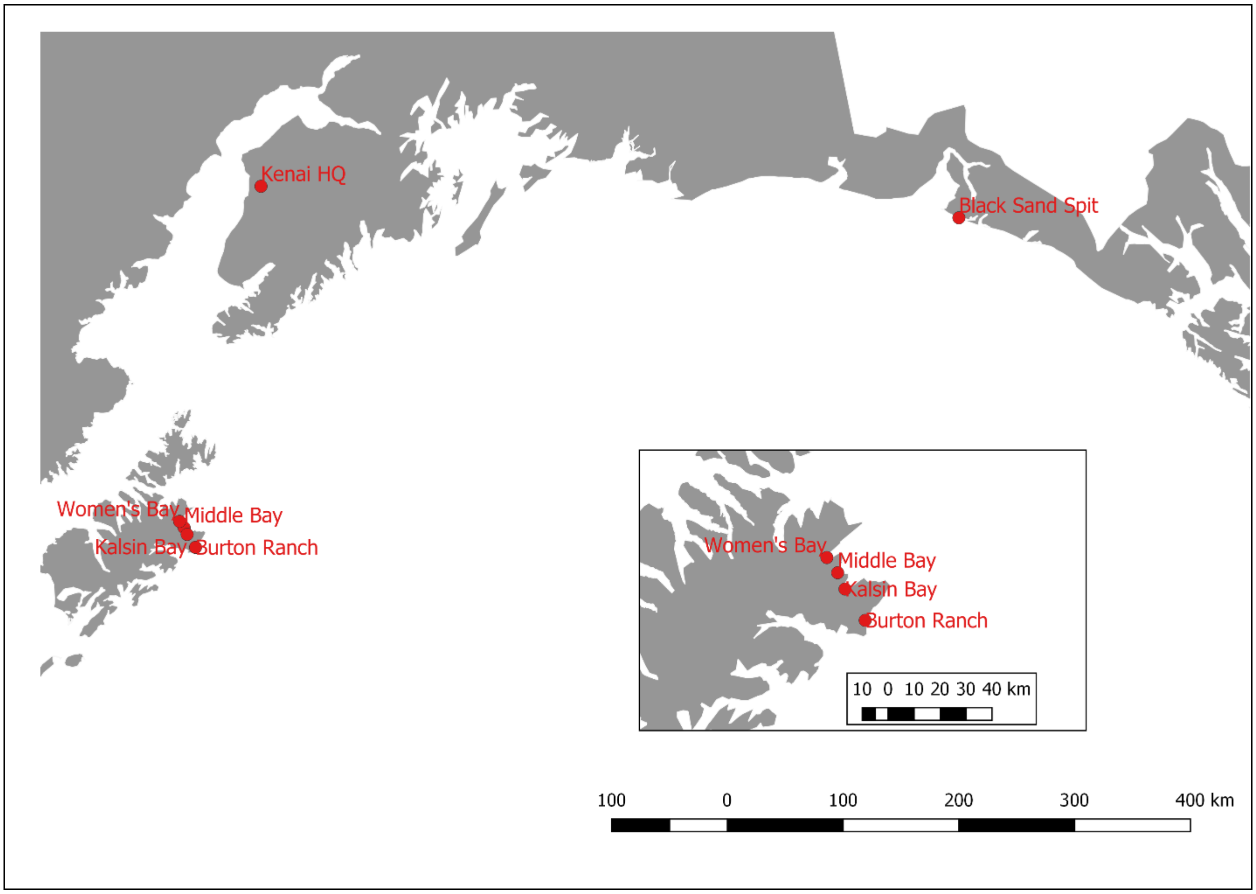


Figure 1: Locations of six tern colonies surveyed by the UAV method in June 2018.

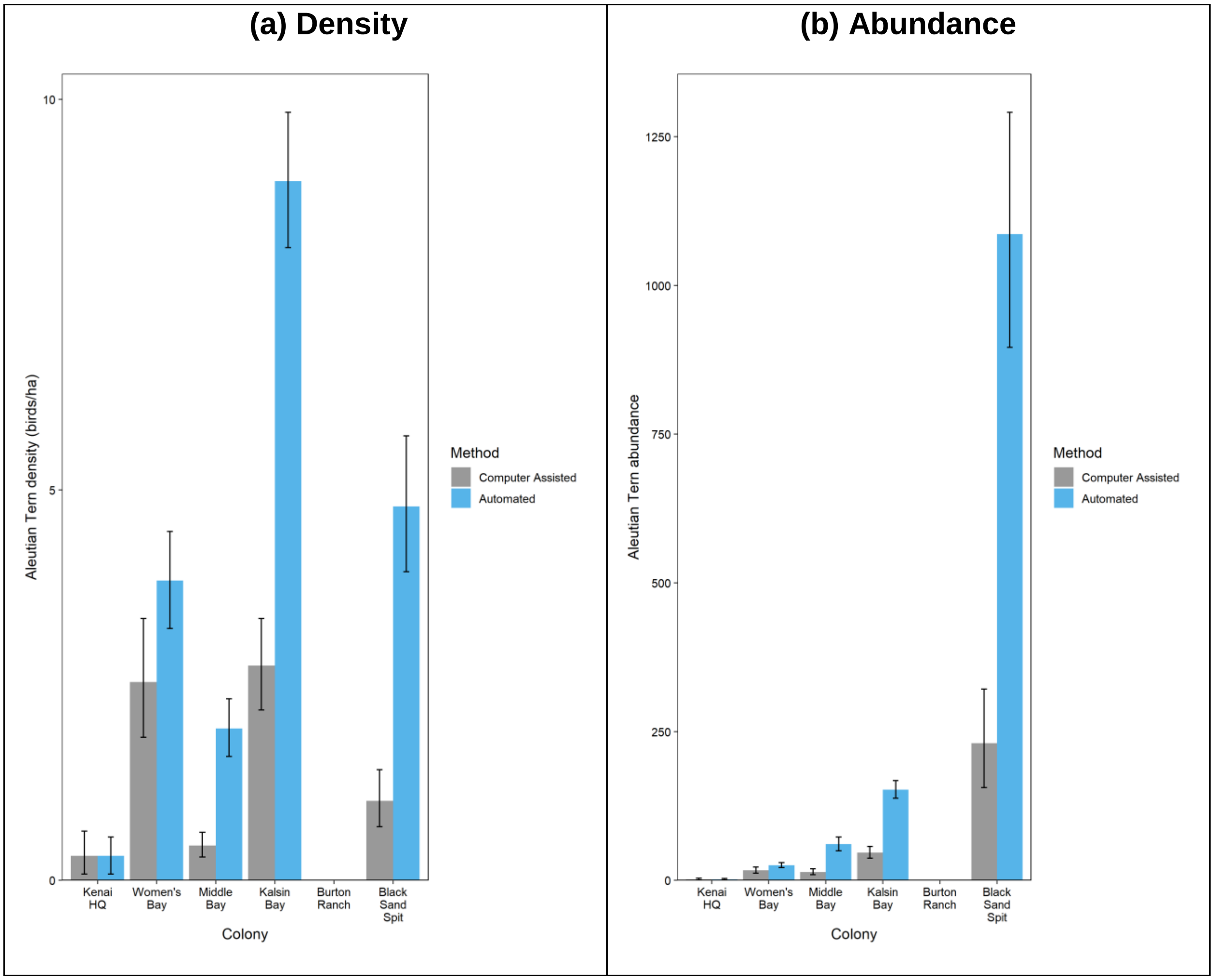


Figure 2: Density and abundance at six colonies sampled by the UAV method in 2018. Additional details in Appendix A.

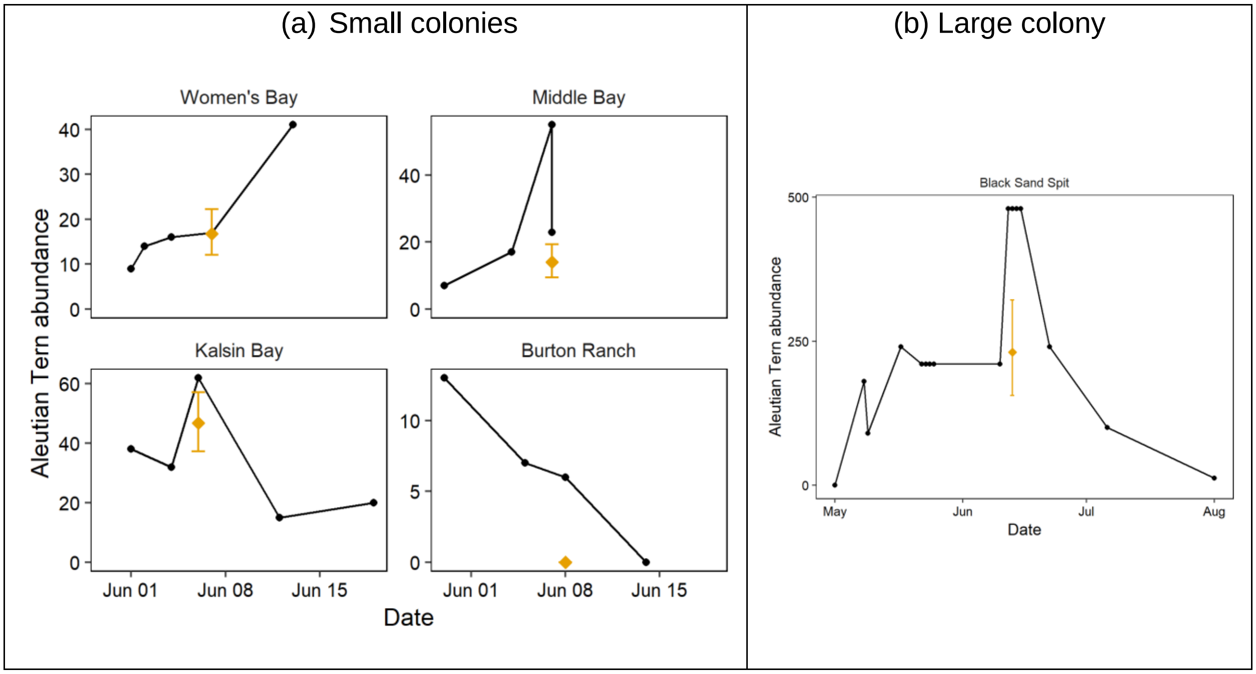


Figure 3: Comparison of abundance estimates derived from the UAV (gold diamonds, with 95% CI bars) method and direct counts (black circles). The only formal direct count at the Large colony (panel b) was conducted on 22 Jun, 9 days after UAV flights. Direct counts at the Large colony, both before and during UAV flights, were less formal and should be considered rough estimates. Additional details in Appendix A.

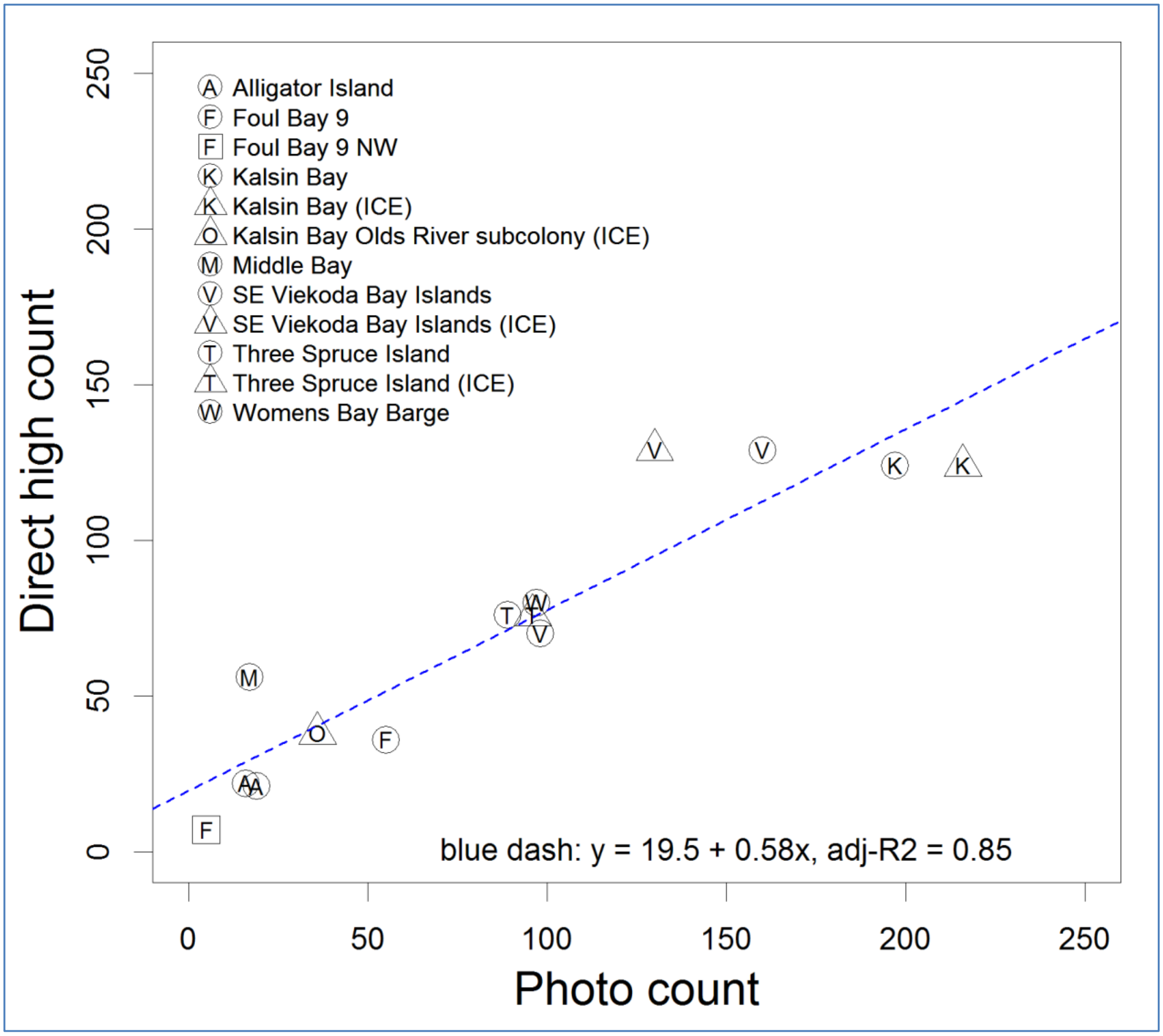


Figure 4: Estimated regression relationship between ground-based photo- and direct- counts for tern colonies surveyed on the Kodiak Archipelago, June-August 2018. Additional details in Appendix B.

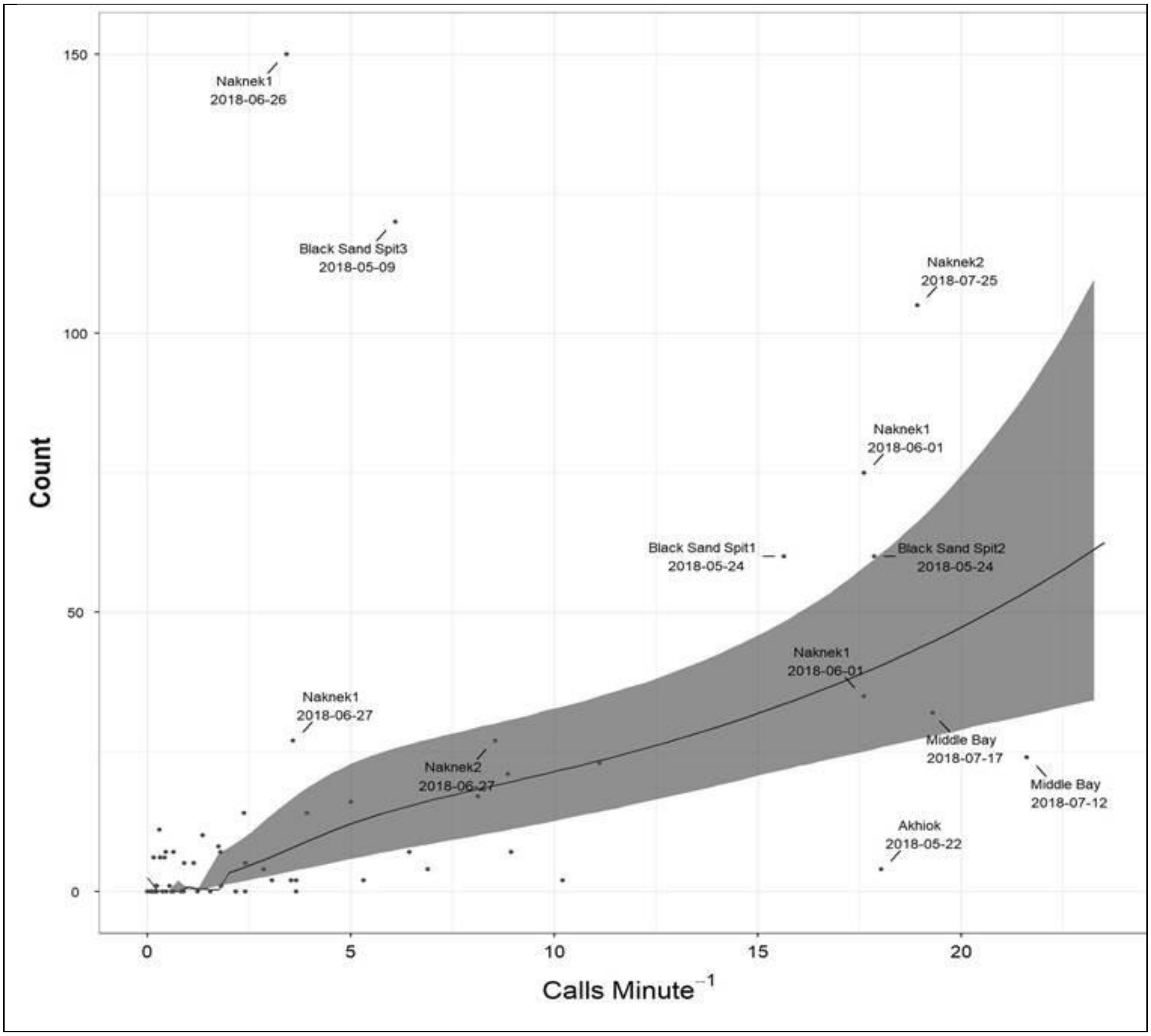


Figure 5: Estimated relationship between ALTE call rate and direct colony counts using the full dataset collected over the entire season. Black line is the mean direct count predicted by a zero-inflated negative binomial with a log link. The grey swath represents 95% confidence interval estimated by bootstrapping 10,000 times. Additional details in Appendix C.

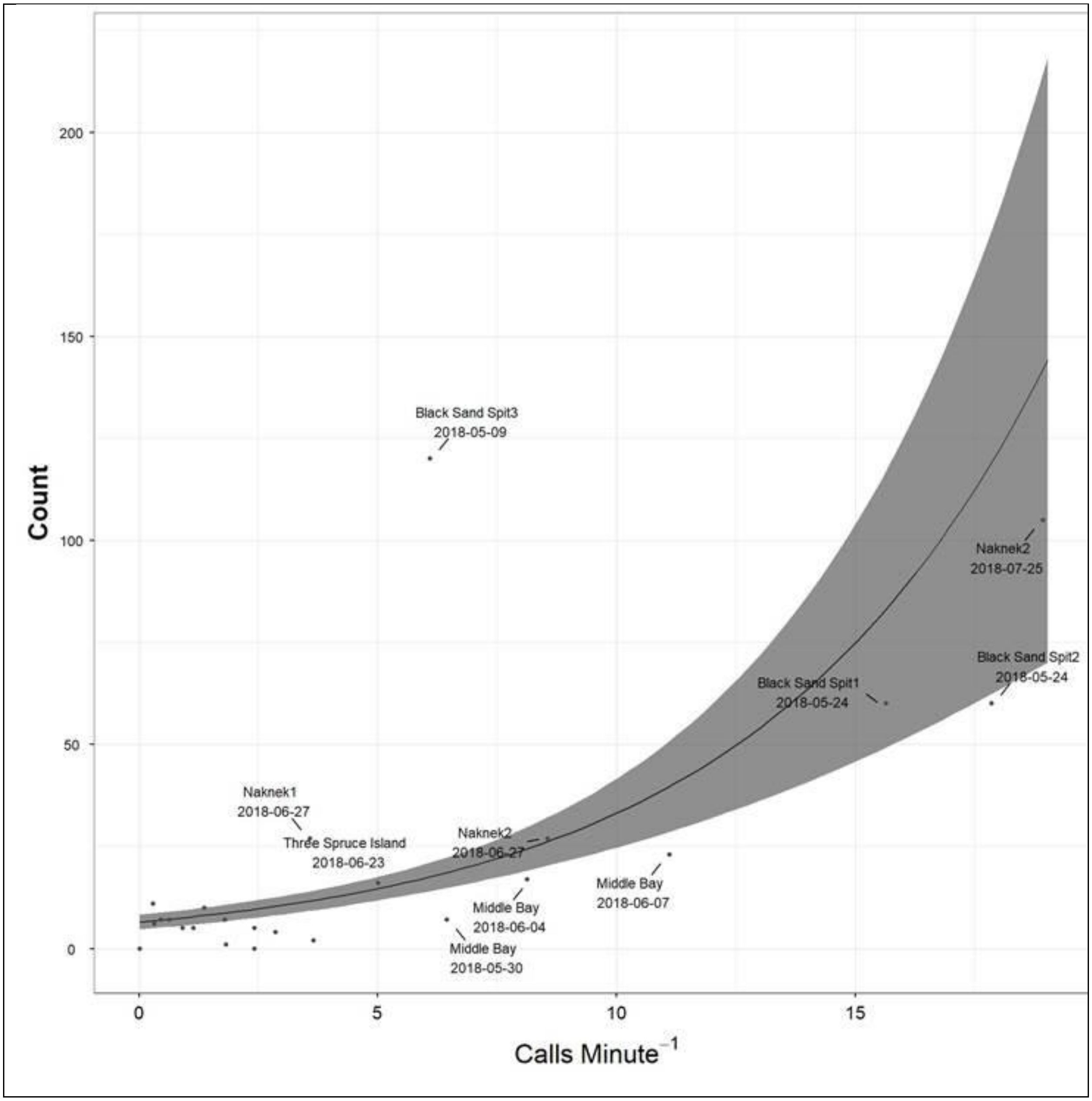


Figure 6: Relationship between ALTE call rate and direct colony counts during the incubation period. Black line is the mean direct count predicted by a regular negative binomial with a log link. The grey swath represents one standard error around predicted mean count. Additional details in Appendix C.

Appendix A – UAVs

2018 Tern Colony Densities from UAV Photos

Trent McDonald, Jason Carlisle, and Jaime Thompson
15 Aug 2019

This appendix contains details and supplementary information on the UAV photo methods applied in 2018.

METHODS

We employed three UAV's at various times to sample six study colonies in south central and south east Alaska (Table 1, Figure 1). At Kodiak Island and Kenai colonies we employed: (1) a DJI Matric 210 quad-copter drone equipped with a 22 Meg sensor, (2) a DJI Phantom 4 quad-copter drone equipped with a 20 Meg sensor, and (3) a 3DR Solo quad-copter drone equipped with a 16 Meg sensor. We conducted all UAV flight missions (hereafter, "missions") during daylight hours using high-resolution standard red-green-blue (RGB) imagery. The pilots of all three UAVs had 507 certification with the FAA that authorization them for commercial UAVs flight. No colonies was within restricted airspace.

We tested two types of UAV-based photographic sampling methods (Figure 2). We employed a census method similar to Magness et al. (2019) in which the UAVs flew a back-and-forth ("lawn mower") transect pattern over the colony. During the census method, UAVs collected overlapping photographs of the entire outlined colony. At larger colonies, we tested a spatially balanced sampling method in which the UAVs flew to pre-determined points and obtained non-overlapping photographs. The spatially balanced method selected 99 points within each colony's polygon using the BAS spatially balanced sample algorithm (Robertson et al. 2013, 2017) and applied the travelling salesman algorithm (Kruskal 1956, Garfinkel 1985) to order the points in an efficient survey order (least distance flown). At each survey point, the UAV hovered for 3 seconds and captured a simple image of the study area directly below. We photographed colonies at above-ground altitudes of 15 meters, 20 meters, and 30 meters (Table 2).

Photo-Recognition

Following photo collection, we trained a customized neural network to search for and detect terns in UAV-derived photographs. During training and afterward, we tolerated a higher than normal number of false positives (not actual terns) due to difficulties distinguishing ALTE and ARTE at the resolutions we used. Later, we verified and determined the species of all putative detections. This allowed us to compare density estimates from a fully-automated method (no verification) to those produced by a computer-assisted method (human verification).

Object detection in photographs is a field of active research, with neural networks currently achieving human-level performance at a variety of recognition tasks. Pre-trained detection networks available from a number of sources (see blog post https://www.analyticsvidhya.com/blog/2018/07/top-10-pretrained-models-get-started-deep-learning-part-1-computer-vision/) are trained on thousands of images containing common items (e.g., cars, people, dogs, etc). Pre-trained networks did not meet our needs because common training data sets do not contain tern examples and the nesting terns in our UAV-derived photographs were quite small in the image. In addition, gulls were common in our aerial photography and we required our network to differentiate terns from gulls. Consequently, we designed and trained a customized neural detection network to be trainable on novel targets (i.e., terns and gulls) when fewer than 1000 example photos are available.

We implemented our tern and gull detector using routines available from the MXnet project (https://mxnet.incubator.apache.org). MXnet provided flexible tools (routines) for construction and training of neural networks on servers that contained graphical processing units (GPUs). We implemented a network architecture capable of generalizing to previously unseen images by using a variety of data augmentation techniques. Data augmentation included randomized image cropping, reflection, rotations and the addition of noise and color balance perturbation. More detail on structure of the tern and gull detector network appears in Addendum A.

To construct a detector training data set, human observers initially found and marked terns and gulls in a subsample of all images and outlined ("painted") the targets when identified (Figure 3a and 3b). Those observers initially found terns and gulls in approximately 2,000 of the 11,000 images collected during all surveys at the six colonies. Our initial round of detector training consisted of passing these tern and "painted" images through the detector's neural network so that the network "learned" to distinguish characteristics of the "painted" areas from those of non-"painted" areas. In other words, by comparing "painted" and non-"painted" areas, the network "learned" what was, and what was not, a tern or gull (Figure 3c). Following initial training, we applied the detector to all photos in an effort to find additional terns. Observers "painted" all new terns which went un-detected during the first round. After this second round of "painting", we completed a second round of training to improve the network's sensitivity and specificity. At the end of the second round of training, we again applied the network to all photos and we labeled the resulting counts the "automated" counts.

Despite two rounds of training and efforts to reduce the number of false-positive detections, we became aware of a substantial number of false-positives in the automated detections. Hence, we inspected all detections and classified them as either an tern (ALTE or ARTE), gull, or neither. We labeled this verified count after the second round of detector training as the "computer-assisted" counts because human observers verified that all putative detections (Figure 3d). This count was "computer-assisted" because human observers only verified putative detections and did not search photos where the automated detector count was zero.

Following training, we applied the network to a small set (100) of hold-out images that included both positive (with tern) and negative (without tern) images. We computed the automated method's positive detection rate as the proportion of known terns in hold-out photos detected by the network. We computed the automated method's false positive rates as the proportion of all detections that were not a tern. Note that these rates use different denominators (i.e., known terns and all detections) to accurately estimate the rates of both under-counting and over-counting. We did not perform hold-out validation of the computer assisted counts.

Density and Abundance Estimation

We recognized that the colony boundaries, and hence the polygons we sampled, were poorly defined for some colonies and that they will likely vary through time. For the purpose of this study, we obtained rough outlines of the colony's extent from biologists familiar with the colony and created colony polygons in GIS software. These rough outlines were later refined in the field, on the day of sampling, based on perceived colony activity and local features (i.e., private land, large trees, etc.).

Given a colony polygon, both the census and sample photographic methods are a type of plot-sampling method (Borchers et al. 2002) designed to estimate tern density on the study areas. In general, plot sampling entails selecting a spatially balanced or random sample of plots from the region of interest, measuring the number of individuals in each plot, determining the area covered by each plot, and computing density from these numbers. Here, we viewed each aerial photograph as one searched plot. Contrary to classical plot-sampling, our plots were searched after field operations, not during, which allowed us to minimize field time and disturbance to the terns.

We computed density and abundance of ALTE in a colony as follows. Assume c_{ij} is the count of terns in photograph i of mission *j*, and a_{ij} is the ground area (ha) covered by photograph i of mission *j*. We computed a_{ij} from meta-data embedded within each photo such as height (AGL), focal length, image sensor size, pitch of the UAV gimbal, yaw of the UAV gimbal, GPS coordinates, UAV heading, etc. Assuming no pitch or yaw of the UAV gimbal, we computed a_{ij} using the well-known relationship,

$$a_{ij} = \frac{h}{f}(xy),$$

where *x* was horizontal dimension of the camera sensor, *y* was vertical dimension of the sensor, *h* was height above ground (AGL) and *f* was focal length (all measurements in mm). When pitch and yaw were present, we applied more complex formula to account for angle of the camera and to accurately adjust are of the photograph on the ground.

We computed the average density of terns in photos during mission j as,

$$\bar{d}_j = \frac{\sum_{i=1}^{n_j} c_{ij}}{\sum_{i=1}^{n_j} a_{ij}}, \tag{1}$$

where n_j was the number of photographs taken during mission *j*. Assuming A_j was the area (ha) of the polygon sampled during mission *j*, we computed the average density of terns for colony k as,

$$\overline{D} = \frac{\sum_{j=1}^{m} \bar{d}_j A_j}{\sum_{j=1}^{m} A_j}, \tag{2}$$

where m is the number of missions in the colony. At smaller colonies (e.g., Womens Bay), the polygon sampled by a single mission covered the entire colony. At larger colonies (e.g., Black Sand Spit), the polygon sampled by a single mission was smaller than the colony and it took several missions to cover the entire colony. At some times, the polygons sampled by different missions on different days overlapped. This method of computing an average density for the colony is equivalent to a weighted average of mission-specific densities, with the weight being proportional to each area sampled by the missions associated with the colony.

Lastly, assuming A was the area (ha) of the entire colony we estimated total abundance of terns for the colony as,

$$N = \bar{D} A \quad . \tag{3}$$

We calculated a 90% confidence interval (CI) for each mission-specific density ($\bar{d}_j$) and each colony-specific abundance (N) using a non-parametric bootstrap and the percentile method (Borchers et al. 2002, Manly 2006). We applied the bootstrap by resampling the photos within each colony with replacement, and recalculating $\bar{D}$ and N for each of 5,000 bootstrap iterations.

RESULTS

As anticipated, our automated detector flagged a substantial number of false positives (i.e., objects similar to but not terns). Most false positives were wood debris of similar color and general shape as nesting terns. We tolerated the high false-positive rate of the detector in order have higher confidence that the final computer-assisted counts accurate. Assuming the computer-assisted counts are correct, our network produces an average of 58% false-positives across all missions. Examples of true positive, false positive, and false negative detections appear in Addendum B.

Focusing on computer-assisted numbers, ALTE density and abundance varied by as much as 300% to 500% over multiple missions at the same colony depending on the specific area sampled by the mission (Figure 4). Colony-wide density from the computer assisted method varied from 0 (Burton Ranch) to 2.75 (95% CI 2.18 to 3.35; Kalsin) individuals per hectare (Table 3; Figure 5). Colony abundance from the computer-assisted counts varied from 0 (Burton Ranch) to 231 individuals (95% CI = 156 to 322; Black Sand Spit) (Table 3; Figure 5).

We found that abundance derived from the computer-assisted method for smaller colonies (Womens Bay, Kalsin, Middle Bay, Burton Ranch) generally matched direct counts taken by observers near the time of UAV flights (Figure 6a). At Womens Bay, UAV estimates and contemporaneous direct counts were virtually identical. At Kalsin and Middle Bay, UAV estimates were 32% (47 vs 62) and 64% (14 vs 23) lower, respectively, than the nearest direct count. At Black Sand Spit, UAV-based abundance estimates were practically identical to a formal direct count conducted 9 days later (230 UAV vs 240 direct count; Figure 6b). Other direct counts at Black Sand Spit, both before and during UAV surveys, were less formal and should be considered rough estimates. Nonetheless, our UAV-based abundance matched an informal direct count taken three days prior to UAV flights (230 UAV vs 210 direct count).

REFERENCES

Borchers, D. L., S. T. Buckland, and W. Zucchini. 2002. Estimating animal abundance: closed populations. Springer-Verlag, London, UK.

Garfinkel, R. S. 1985. Motivation and modelling. E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors. The traveling salesman problem - A guided tour of combinatorial optimization. Wiley and Sons.

Kruskal, J. B. 1956. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Proceedings of the American Mathematical Society 7:48–50.

Magness, D. R., T. Eskelin, M. Laker, and H. M. Renner. 2019. Evaluation of small aerial systems as a census tool for Aleutian Tern Onychoprion aleuticus colonies. Marine Ornithology 16:11–16.

Manly, B. F. J. 2006. Randomization, bootstrap and Monte Carlo methods in biology. 3rd edition. Chapman & Hall/CRC, London, UK.

Robertson, B. L., J. A. Brown, T. L. McDonald, and P. Jaksons. 2013. BAS: Balanced acceptance sampling of natural resources. Biometrics 69:776–784. <http://dx.doi.org/10.1111/biom.12059>.

Robertson, B. L., T. L. McDonald, C. J. Price, and J. A. Brown. 2017. A modification of balanced acceptance sampling. Statistics & Probability Letters 129:107–112. Elsevier B.V. <http://linkinghub.elsevier.com/retrieve/pii/S0167715217301797>.

TABLES

Table 1: Six Aleutian Tern colonies surveyed using UAV-derived photographs in southern Alaska during June, 2018. Additional flight mission details appear in Table 2.

Colony	Area (ha)	Location	Dominate Landcover	Survey Pilot*	Number of Missions*	Survey Method**
Black Sand Spit	226.9	Yakutat	Sand	W	13	S, C
Burton Ranch	10.5	Kodiak Is.	Grazed grass	W	1	C
Kalsin Bay	17.0	Kodiak Is.	Grass	W	2	C
Middle Bay	31.4	Kodiak Is.	Grass	W, L	2, 1	S, C
Womens Bay	6.6	Kodiak Is.	Grass	W, L	2, 1	C
Kenai HQ	5.4	Kenai	Bog	W	1	C

*W=WEST (Mike Gerringer); L=Mark Laker
**S=Sampled with non-overlapping photographs; C=Census using overlapping photographs

Table 2: Twenty-two UAV survey missions flown at six Aleutian Tern colonies in southern Alaska during June, 2018.

Date	Colony	Mission Name	Pilot*	Flight Height (m)	Survey Method**	Number of Photos	Area Photographed (ha)	Mission Area (ha)
6/4/2018	Kenai HQ	Kenai HQ	W	15	C	571	12.7	5.4
6/6/2018	Kalsin Bay	Kalsin Bay - Inland	W	15	C	656	15.5	10.6
6/6/2018	Kalsin Bay	Kalsin Bay - Coastal	W	15	C	338	7.2	6.4
6/7/2018	Middle Bay	Middle Bay - North	W	15	C	859	21.2	12.6
6/7/2018	Middle Bay	Middle Bay - South	W	15	C	682	16.1	9.7
6/7/2018	Middle Bay	Middle Bay - Laker	L	20	S	519	13	31.4
6/7/2018	Womens Bay	Womens Bay	W	15	C	400	8.8	6.6
6/7/2018	Womens Bay	Womens Bay - Laker	L	20	C	763	34.1	6.6
6/8/2018	Burton Ranch	Burton Ranch	W	20	C	363	14.9	10.5
6/12/2018	Black Sand Spit	BSS - S5 - Small	W	15	C	203	4.5	3.2
6/12/2018	Black Sand Spit	BSS - S3 - V1	W	15	S	88	2	52.6
6/12/2018	Black Sand Spit	BSS - S4 - V1	W	15	S	89	2.1	51.1
6/12/2018	Black Sand Spit	BSS - S4 -Small	W	15	C	216	4.8	3.3
6/13/2018	Black Sand Spit	BSS - S1	W	15	S	88	2	46
6/13/2018	Black Sand Spit	BSS - S2	W	15	S	95	2.2	54.9
6/13/2018	Black Sand Spit	BSS - S3 - V2	W	15	S	92	2.1	52.6
6/14/2018	Black Sand Spit	BSS - S4 - V2	W	15	S	91	2.1	51.1
6/15/2018	Black Sand Spit	BSS - S5 - A	W	15	C	687	16	10
6/15/2018	Black Sand Spit	BSS - S5 - B	W	15	C	995	21.8	14.5
6/15/2018	Black Sand Spit	BSS - S4 - Coastal	W	15	C	277	5.8	4.2
6/15/2018	Black Sand Spit	BSS - S4 - A	W	15	C	933	24	14
6/15/2018	Black Sand Spit	BSS - S4 - B	W	15	C	701	17.1	10.4

*W=WEST (Mike Gerringer); L=Mark Laker

**S=Sampled with non-overlapping photographs; C=Census using overlapping photographs

Table 3: Final estimates of ALTE density and abundance at six nesting colonies sampled by the UAV method in 2018. Estimates derived from the computer assisted count method. Species ratios derived from direct flush counts during visits temporally close to UAV flights.

		Species	Density (#/ha)			Abundance		
	Area	Ratio		Low	High		Low	High
Colony	**(ha)**	**(%ALTE)**	**Est**	**95%**	**95%**	**Est**	**95%**	**95%**
Black Sand Spit	226.87	0.8	1.017	0.687	1.418	231	156	322
Burton Ranch	10.51	1	0.000	0.000	0.000	0	0	0
Kalsin Bay	17.04	0.48	2.750	2.183	3.353	47	37	57
Kenai HQ	5.37	1	0.315	0.079	0.632	2	0	3
Middle Bay	31.42	1	0.447	0.301	0.615	14	9	19
Womens Bay	6.62	0.94	2.541	1.831	3.354	17	12	22

FIGURES

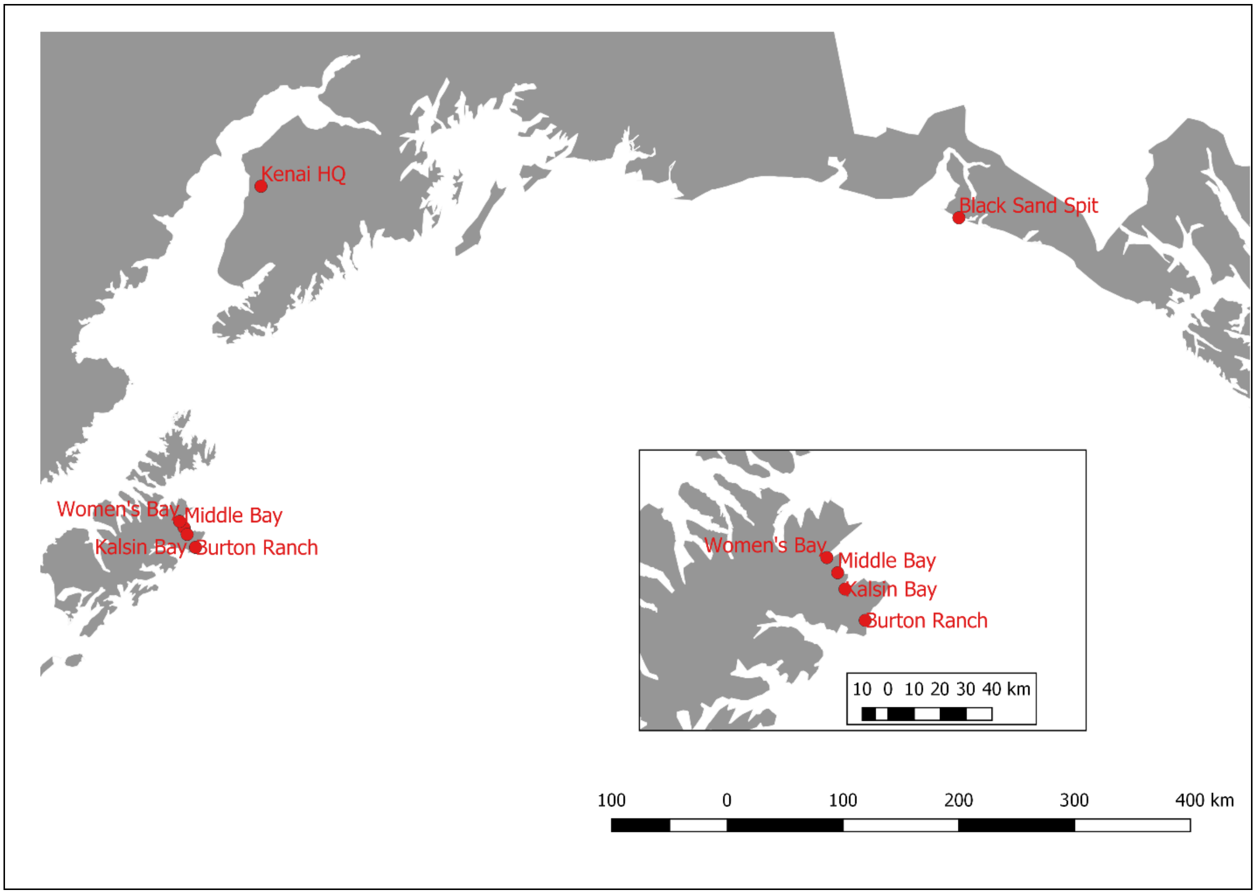


Figure 1: Locations of six tern colonies surveyed by the UAV method in June 2018.

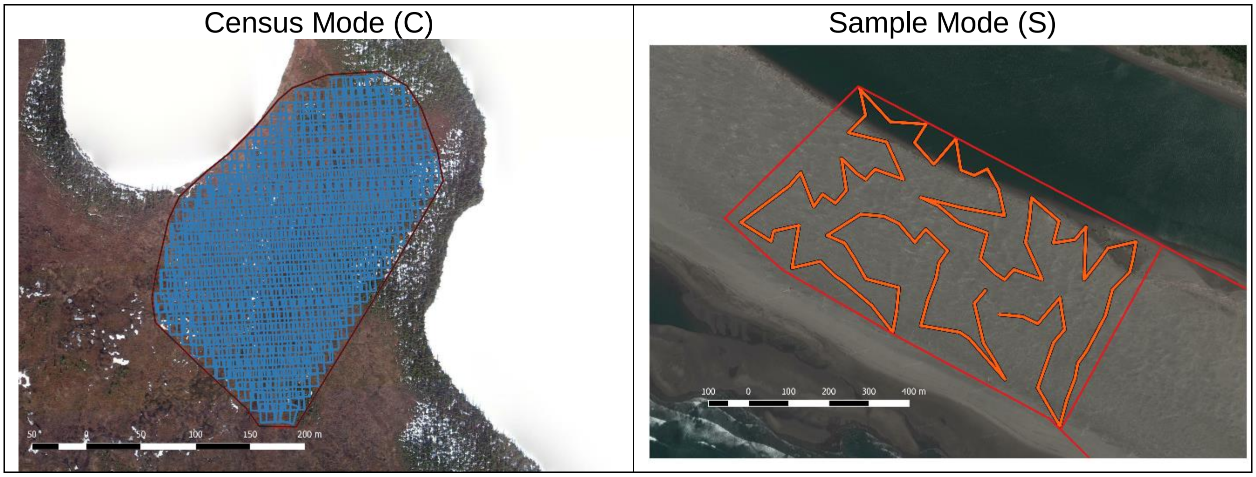


Figure 2: Illustrations of the census and sample modes employed during UAV surveys in June 2018**.**

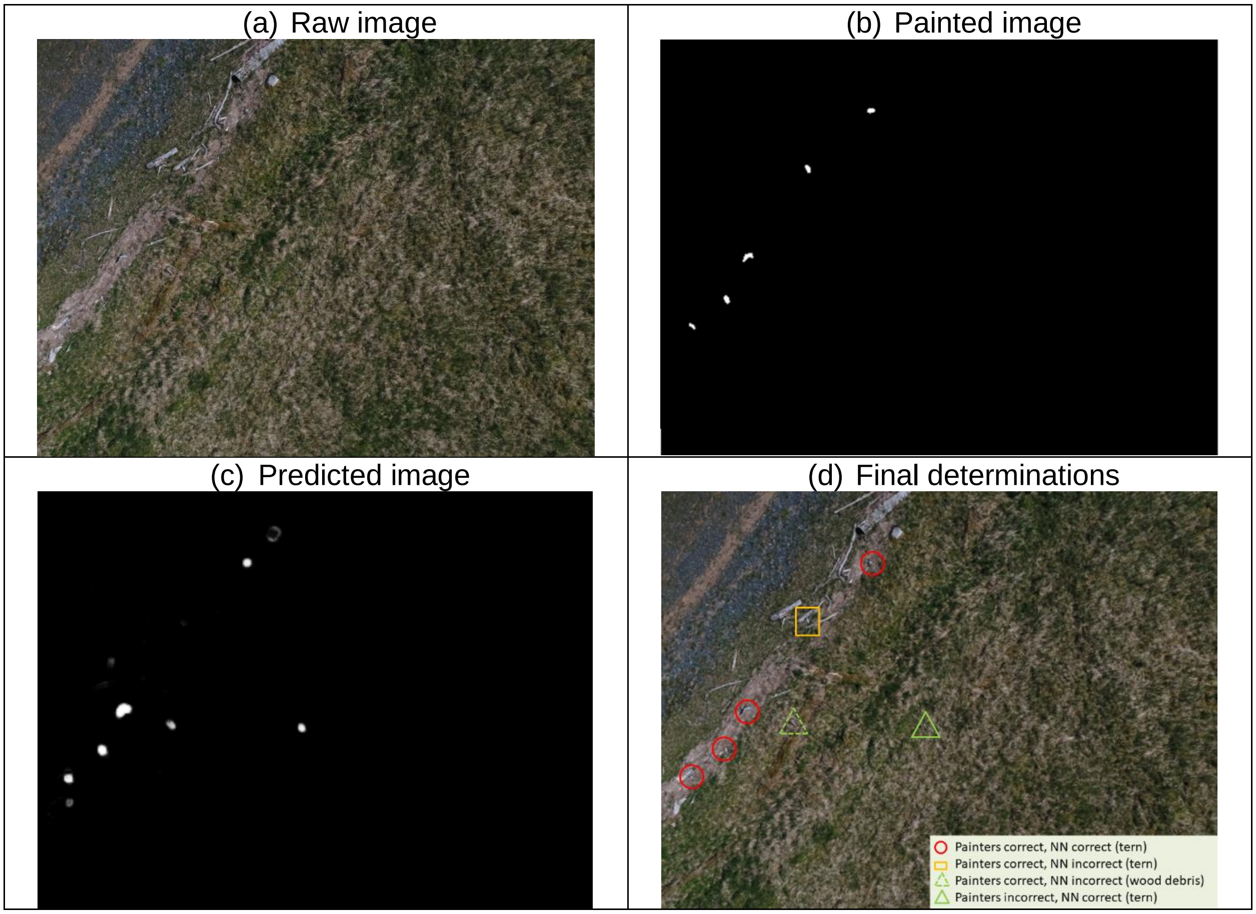


Figure 3: Example images of a raw photo (a), the "painted" version of the raw photo with terns outlined in white (b), the "predicted" image with regions predicted to be terns shown in white (c), and the final determination of terns, false-negatives, and false-positives (d). Human observers detected 5 of 6 terns in this photo. The automated detector found a different set of 5 terns in this photo, and detected one piece of wood debris the same color and general shape as a tern.

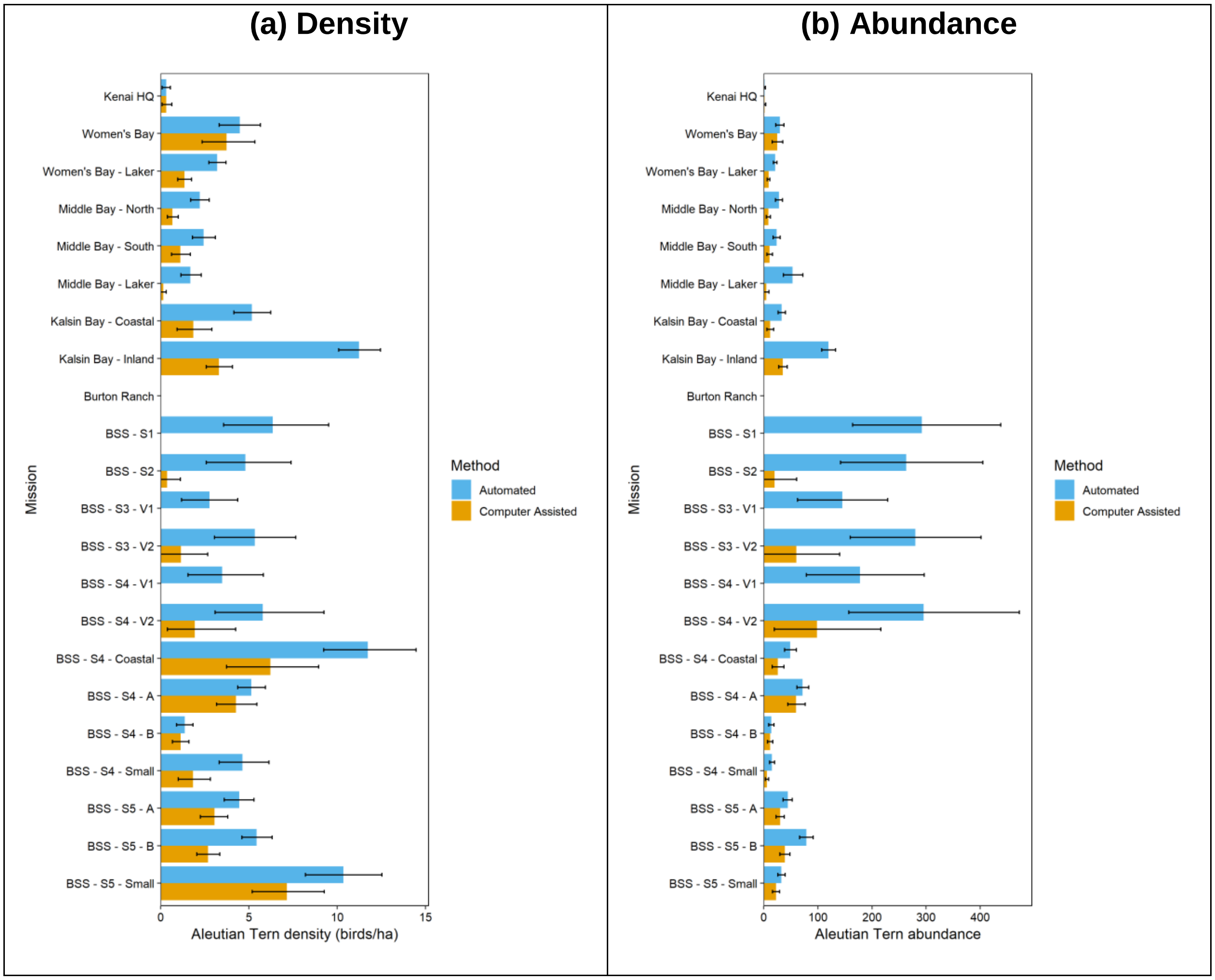


Figure 4: Density and abuncance of ALTE for all UAV missions flown during 2018 field season. “Automated” method uses counts produced by the tern detector, corrected for species ratios. “Computer Assisted” method uses counts verifed as terns by human observers, corrected for species ratios.

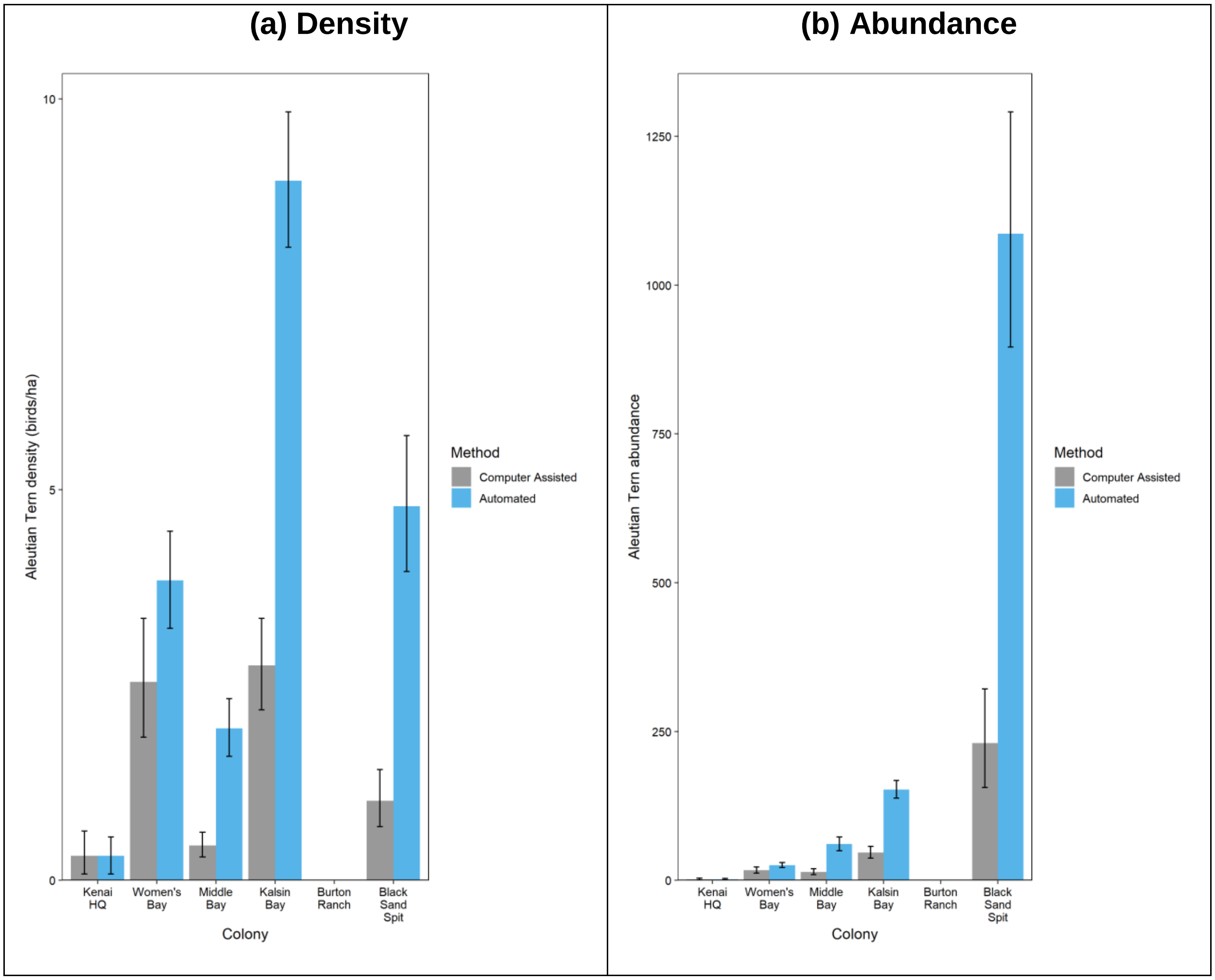


Figure 5: Colony specific density and abundance at six colonies sampled by the UAV method in 2018. Numeric values for the computer assisted method appear in Table 3.

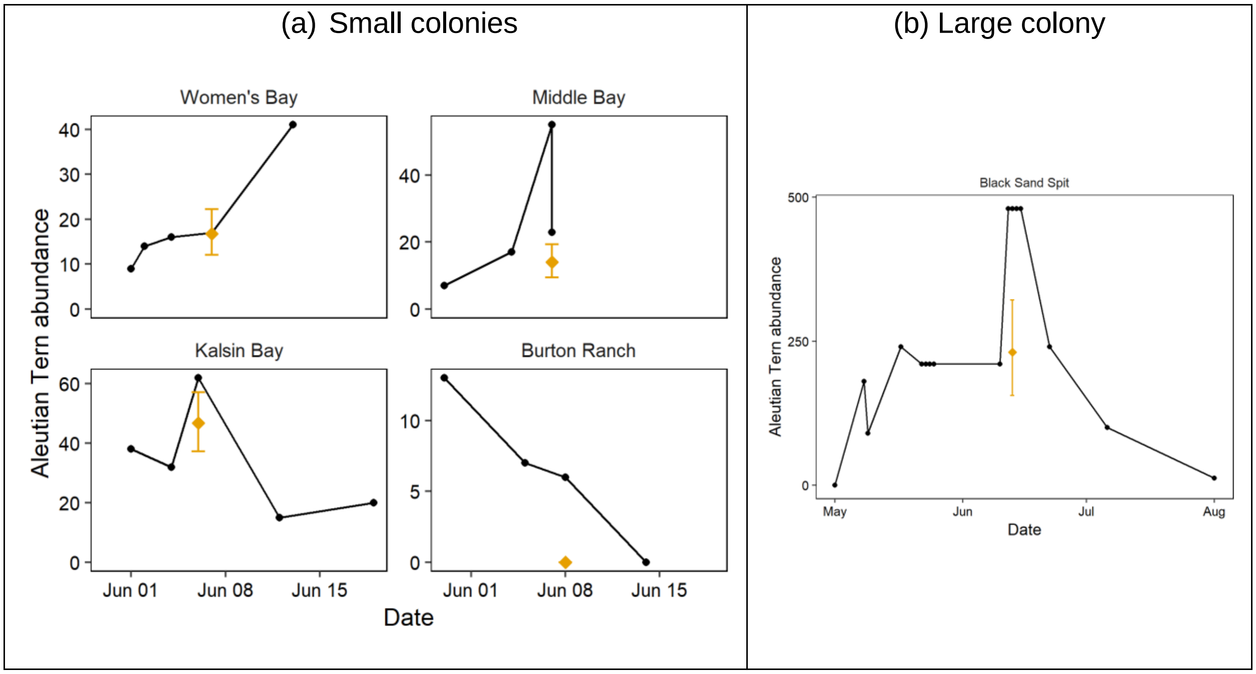


Figure 6: Comparison of abundance estimates derived from the UAV (gold diamonds, with 95% CI bars) method and direct counts (black circles). The only formal direct count at the Large colony (panel b) was conducted on 22 Jun, 9 days after UAV flights. Direct counts at the Large colony, both before and during UAV flights, were less formal and should be considered rough estimates.

ADDENDUM A: Tern Detector Network Details

The architecture (i.e., structure) of the tern and gull detection network we implemented using tools in MXnet was inspired by ResNet. Similarly, we trained the network using common protocols for Data Augmentation. We chose to implement a ResNet-type architecture because it has demonstrated state-of-the-art accuracy on a variety of image recognition tasks and is theoretically appealing due to the identity mappings introduced by simple shortcut connections. Addendum A Figure 1 illustrates the structure of foundational building blocks in the ResNet approach. Under this ResNet approach, "weight layer" contain a collection of convolution filters and the "relu" node is a standard activation function found in many neural networks.

ResNet's use of the identity connection ("x identity" line in Addendum A Figure 1) enables the overall network to learn an efficient model for targets (terns and gulls) by allowing target features to be preserved as they pass through the network. This architecture spares the overall network from learning salient identity mappings explicitly. Instead, this architecture preserves a variety of input features, such as the distribution of colors on input targets, deep in the network where it can be exploited. The architecture we implemented reflects the idea that the most common mapping learned by individual network layers is the identity. By hard coding an identity representation, training is extremely efficient because minor perturbations to the inputs can be used instead of new, whole, identity-preserving representations.

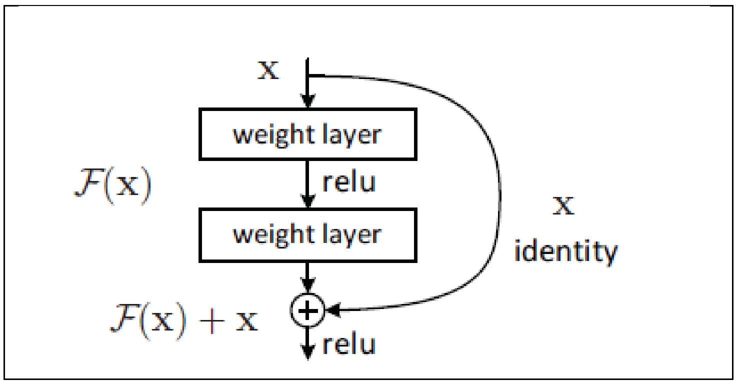
The introduction of an identity connection across multiple layers also gives rise to nth order mappings. This alleviates a typical limitation in the representations generated at each level because individual layers typically produce first order mappings. With first order mappings, saddle points and other stationary extremes in the loss function are difficult to escape during optimization. The identity connection eases this problem by separating the convolutional layers from a copy of the input before adding them back together, thus allowing higher order perturbations to the input before remapping in another set of convolutional layers.

To deal with the constraints imposed by sparse targets and large images, several important features were introduced to the basic ResNet design. To preserve information in the identity connection while down sampling effectively, average pooling was used in place of subsampling. To improve generalization, a novel form of dropout regularization was implemented where instead of setting neural outputs to 0 with probability .5, values are resampled from an estimation of the generative distribution induced by the neuron with probability .5. This serves to improve robustness by forcing the network as a whole to rely on distributed representations while simultaneously preserving the distribution of layer outputs as a signal moves through the network. This allows for the combination of dropout and batch normalization as regularization methods.

To ensure that the features used to distinguish terns and gulls still shared features to distinguish a "bird" from empty ground, a hierarchical softmax was implemented where the decision hash was hand coded to ensure a grouping between birds in general before distinguishing species.

Data augmentation was an important step during training. Perturbation and distortion helps ensure that the representations learned by the network possess the same invariance properties as a human representation. For instance, many targets look the same when reflected in a mirror. Under minor distortion, most humans can still recognize an object. The perturbations used during training of our tern and gull model included the following:

- Reflections across the x and y axes
- Rotation
- Color balance perturbations
- Random stretches and contractions along the x and y axes
- Gaussian noise



Addendum A Figure 1: Schematic structure of the neural network building blocks used in the ResNet approach to image recognition.

ADDENDUM B: Detection Examples

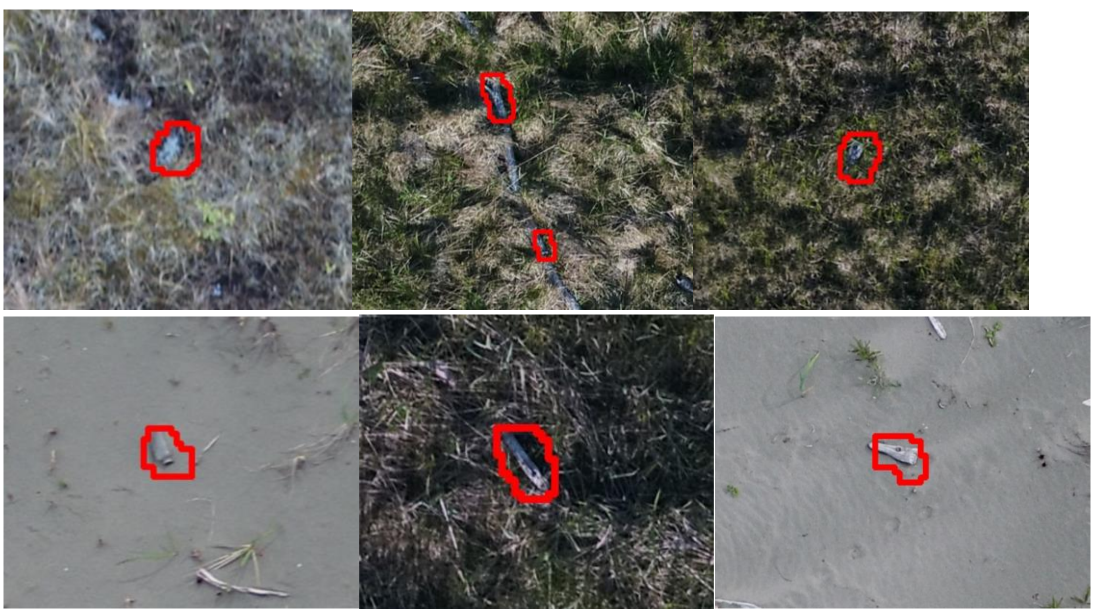
This addendum contains example images of positive, false-positive, and false-negative tern and gull detections. In this addendum, red outlines signify a putative tern detection and gold outlines signify a putative gull detection.

False Negative Examples

Terns undetected by the network

False Positive Examples

Putative tern detections that are not terns.



True Positive Examples

Confirmed tern and gull detections.

Appendix B – Photo Counts

Ground-based photo-counts for 2018 Aleutian tern field season, Kodiak Archipelago Alaska

Robin Corcoran, John Skinner, Kelly Nesvacil

21 June 2019

This appendix contains details and supplementary information on the photo count method applied in 2018.

METHODS

Ground-based photo-counts (photo-counts) were one method recommended from the 2018 Aleutian Tern Conservation Planning meeting. This method attempts to standardize and reduce variability inherent in visual ground-based direct counts by using multiple photographs in quick succession in an attempt to “freeze” the action and capture flying individuals (McDonald and Carlisle 2018). One advantage of photo-counts compared to direct counts is a reduction in the potential for over counting flying birds, but this benefit can be limited if individual birds are captured in multiple frames. Additionally, photo-counts, like direct counts, still rely on anthropogenic or natural flushes and thus are dependent on the proportion of flushing individuals and the species ratio at mixed colonies.

The general protocol for photo-counts is largely the same as the protocol for direct counts and the photographer should attempt to cover the entire statistical colony with a single burst of photos lasting 3-5 seconds (McDonald and Carlisle 2018). The goal is to take multiple photos with overlap so that all in-flight birds appear in the final photo mosaic only once. At small colonies, it is useful to attempt to minimize overlap between images and target only 2-5 images per panto get the entire colony in a single photo. Please see below for a detailed field protocol developed in 2018 for use during the 2019 field season.

Photo-counts were conducted at 8 Aleutian tern colonies on the Kodiak Archipelago from 1 June – 15 August 2018 (Table 1). All counts represent total terns because some of these colonies contained nesting Arctic terns as well which could not typically be distinguished from Aleutian terns in photographs. Two colonies were surveyed at the same location twice during the season using photo-counts. For analysis, each pan had approximately 1-65 images and for colonies with multiple images, photographs were examined for degree of overlap and only 2-9 of the images were composited to conduct the counts. Photographs were manually stitched and counted in Microsoft Paint and/or digitally stitched with Microsoft Image Composite Editor (ICE) and then manually counted in Paint. Results for mixed Aleutian tern and Arctic colonies are presented as total terns.

Table 1. Aleutian tern colony names, dates, and times for ground-based photo-counts during the field season 2018.

Colony Name	Date	Start Time	End Time
Kalsin Bay	6/1/2018	11:20	14:07
Kalsin Bay Olds River subcolony (ALTE only)	6/1/2018	13:32	13:42
Middle Bay	6/7/2018	11:48	13:47
Alligator Island	6/12/2018	8:58	8:59
Alligator Island	6/27/2018	10:36	10:37
Foul Bay 9	6/13/2018	13:05	13:10
Foul Bay 9 NW	6/14/2018	9:35	9:39
SE Viekoda Bay Islands	6/16/2018	13:21	13:40
SE Viekoda Bay Islands	6/28/2018	16:30	17:12
Womens Bay Barge	7/19/2018	8:08	8:16
Three Spruce Island	8/15/2018	10:31	11:02

Photo-count Protocol 2018

The general protocol for photo-counts is largely the same as the protocol for direct-counts. Differences are highlighted in **bold**, below.

- Colonies should be visited at least once at a time when nesting activity is anticipated to be high. Nesting activity and colony residency varies from late April to mid-July.
- Visits can be conducted anytime during daylight hours.
- Colonies with easy access can and should be visited multiple times in a single season. Multiple visits will provide information on nesting phenology and seasonal variation in colony attendance.
- A colony should only be intentionally flushed once per visit. Other "opportunistic" flush counts can be conducted during a visit with the presumed reason for flush recorded (e.g., Bald Eagle flyover, hiker with dog on beach, dreading behavior, etc.).
- Teams of 2-3 people should conduct surveys. Two observers to make independent direct counts and estimate species ratio, and one photographer (If only two observers than photographer can also count and estimate species ratio).
- A total of 3-6 flush counts will be made per colony visit if possible.
- If counting a new colony, observers should pre-scan the colony from a reasonable distance (50 meters) for 10 Minutes to determine whether it is a mixed colony.
- At all mixed species colonies, observers should pre-scan the colony from a reasonable distance (50 meters) for 10 minutes and record independent estimates of the proportion of ALTE at the colony.

- Following pre-scan, the **photographer** should position him/herself at **a suitable vantage point** around the colony.
- When ready, the flusher should approach the colony to flush birds.
- **The photographer should attempt to cover the entire statistical colony with a <u>single photo or a series of overlapping photos lasting 3-5 seconds</u>. That is, multiple photos with overlap should be taken so that all in-flight birds appear in the final photo mosaic once.**
- **The photographer should attempt to minimize overlap between images (10-20% overlap maximum).**
- **"Bookending" by taking a distinctive identifying photograph at the start and end of the photo count sequence is recommended to aid in later photo processing.**
- All people should move away from the colony as quickly as possible after the flushing event and otherwise seek to minimize disturbance.
- **Photographs can be manually stitched and counted in Microsoft Paint and/or digitally stitched with Microsoft Image Composite Editor (ICE) and then manually counted in Paint.**

RESULTS

Ground-based direct-counts of total terns ranged from 7 to 129 terns (Table 2). For 5 out of the 11 surveys, direct-counts did over estimate as compared to photo-counts and these all were at smaller colonies. When comparing the two methods, agreement was good with an adjusted R^2 of 0.85 (Figure 1). For those surveys where it was possible to get Microsoft ICE and Paint estimates, differences appeared to exist between the two analysis techniques, although they did not seem directional. Due to the small sample size, no statistical tests were run and inference for this conclusion is low.

Two substantial issues presented themselves during processing of the colony images. The first was having an individual bird in multiple images that made up the composited image, thus resulting in double counting during the manual counting phase of processing. Additionally, ghosting (loss of individual birds in composited images) was an issue as well and resulted in undercounts of actual birds at the colony.

Table 2. Ground-based direct count and photo-count estimates from the 2018 tern field season, Kodiak Archipelago.

Colony Name	Date	Direct Count High Count	Photo High Count 1 Paint	Photo High Count 1 ICE	Photo High Count 2 Paint	Photo High Count 2 ICE
Kalsin Bay	6/1/2018	124	84		197	216
Kalsin Bay Olds River subcolony (ALTE only)	6/1/2018	38		36		
Middle Bay	6/7/2018	56	25		15	

Alligator Island	6/12/2018	21	19			
Alligator Island	6/27/2018	22	16			
Foul Bay 9	6/13/2018	36	55			
Foul Bay 9 NW	6/14/2018	7	5			
SE Viekoda Bay Islands	6/16/2018	70	98			
SE Viekoda Bay Islands	6/28/2018	129	160	130		
Womens Bay Barge	7/19/2018	80	97			
Three Spruce Island	8/15/2018	76	89	96		

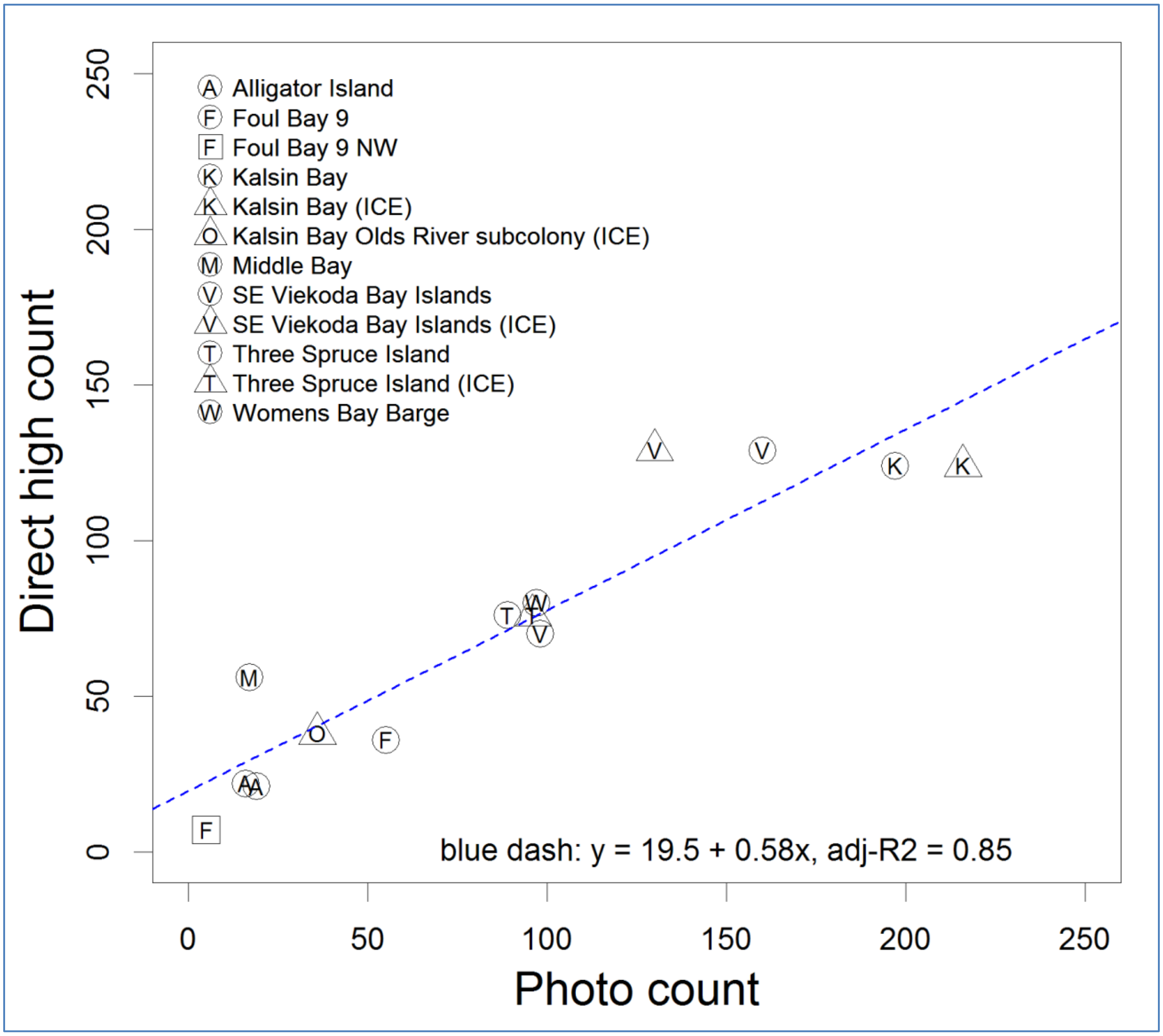


Figure 1. Comparison of ground-based photo- and direct- counts for tern colonies surveyed on the Kodiak Archipelago, June-August 2018.

It was possible to discern Aleutian versus Arctic terns in some photos. Table 3 is a summary of comments for all photo-counts during the 2018 field season.

Table 3. Comments from photo-count surveys during the 2018 field season, Kodiak Archipelago.

Colony Name	Date	Comments
Kalsin Bay	6/1/2018	
Kalsin Bay Olds River subcolony (ALTE only)	6/1/2018	Without flushing observed 15-21 ALTE over Olds River marsh, with flush to nest search & place cameras = 33 (Observer 1) and 38 (Observer 2)
Middle Bay	6/7/2018	
Alligator Island	6/12/2018	Photo-count tern species - 19 at least 5 ALTE
Alligator Island	6/27/2018	Photo-count - all appear to be ALTE
Foul Bay 9	6/13/2018	Photo-count = Tern only, too far to ID species, majority ARTE, counts from above nest habitat & dread
Foul Bay 9 NW	6/14/2018	Photo-count high count = 5 ALTE flying above habitat
SE Viekoda Bay Islands	6/16/2018	Flying high & near constant dreading
SE Viekoda Bay Islands	6/28/2018	Photo count while crew nest searching, minimum count could not get entire flock in single shot
Womens Bay Barge	7/19/2018	High count on skiff survey 67 Adult + 13 HY; high count from 3 separate photos of flying adults = 97,94, & 89 (vs 67)
Three Spruce Island	8/15/2018	

DISCUSSION

The benefits to the photo-count method as compared to direct counts are that there is a permanent record of the count, there seems to be a relatively good relationship between direct- and photo-counts, and the method is relatively quick and easy. The downsides to this method include the fact it relies on anthropogenic or natural flushes and thus is dependent on the proportion of flushing individuals and the species ratio at mixed colonies. Additionally, double counting of birds in composited images is an issue as is ghosting. – loss of individual birds in composited images.

Appendix C – Song Meters

Automated acoustic surveys for Aleutian Tern (*Onychoprion aleuticus*) in Alaska

Abram Fleishman and Matthew McKown

6 June 2019

This appendix contains details and supplementary information on the song meter method applied in 2018.

Summary

Automated acoustic sensors were deployed at twenty-two survey sites within 17 Aleutian Tern (*Onychoprion aleuticus*) colonies in coastal Alaska during the summer of 2018. The goals of these deployments were to evaluate the efficacy of passive acoustic surveys to monitor Aleutian Tern phenology and relative abundance. The study design paired season-long sensor deployments with traditional survey methods, including visual tern counts and nest searches near the sensors by field personnel.

Aleutian Tern vocalizations were detected at all survey sites. Different colonies peaked during different parts of the summer. Adult counts were correlated with mean call rates calculated from ± 7 days around the count date, while there was no relationship with nest density. The nest density result is not unexpected because the low sample size, low nest densities, and low confidence in the quality of the counts around the sensors. There were strong seasonal and diel patterns present at every site monitored, but large differences were seen among colonies. These strong patterns and stark differences likely represent variation in success, synchrony, and/or phenology among colonies.

The relationship between colony counts and call rate, and the clear ability to detect subtle differences in timing and duration of activity, may make passive acoustic monitoring an effective method for detecting and monitoring Aleutian Terns in Alaska. However, several questions remain about how best to design large scale acoustic surveys for the species. To fully realize the potential that acoustic monitoring holds for Aleutian Terns, future work must address the spatial scale at which an acoustic sensor is monitoring, the most effective arrangements of sensors within a breeding aggregation, and if there is a relationship between nest density and vocal activity at large scales than 20-meter radius around the sensor. While guidance on some of these questions may be available in the literature, the Aleutian Tern Technical Committee will need to focus thought and resources to develop a robust framework and sampling design for statewide survey efforts.

Introduction

The Aleutian Tern (*Onychoprion aleuticus*) has a small global population and its breeding range is restricted to Alaska and the Russian Far East. Known breeding populations in Alaska are thought to have declined since the 1960's (Renner et al., 2015). Population trends are difficult to assess, however, due to lack of breeding site fidelity, breeding habitat plasticity, gaps in colony counts, variability in colony attendance within and among years, and potential for high rates of inter-colony movement (Pyare et al., 2013). Moreover, monitoring methods are still being developed and implemented in Alaska. Inconsistent methods among sites have prevented range-wide comparisons and estimates of abundance. Given the unique ecology of the species, and the associated data gaps, Aleutian Tern is listed as a priority Species of Greatest Conservation Need in the 2015 Alaska Wildlife Action Plan (Alaska Department of Fish and Game, 2015).

This report summarizes data collected with acoustic recorders to test their efficacy for monitoring the phenology and relative abundance of Aleutian Tern at colony sites in coastal Alaska. Specifically, acoustic sensors were deployed at sites that have been used by Aleutian Tern for breeding in the past. These sites were also visited by field workers to make visual counts of adult Aleutian Tern and assess nest density near the sensors. The strategy of analyzing acoustic data collected concurrently with measures of abundance from traditional survey methodologies can be used to investigate relationships between the abundance of birds and/or nests and call rate metrics derived from acoustic sensor data. This type of comparison has been used to determine whether or not a functional relationship exists between call rates and adult or nest abundance for several other ground or burrow-nesting seabird species [e.g. Forster's Tern *Sterna forsteri* (Borker et al., 2014), Leach's Storm-Petrel *Hydrobates leucorhoa* (Orben et al., 2019), Cory's Shearwater *Calonectris borealis* (Oppel et al., 2014)]. One important caveat to consider when making these comparisons is that often, traditional methods for making counts are rife with un-quantified error. Poor detection rates due to secretive nesting habits (e.g. burrow nesting, nesting in tall grass) and other counting difficulties (e.g. mixed species flocks, counting birds in flight, difficult weather conditions) are common when surveying seabird breeding aggregations.

Automated acoustic sensors for ecological monitoring

Acoustic cues have long been an important part of bird monitoring projects (Sauer, Peterjohn, & Link, 1994). Technological innovations now make it possible to deploy weatherproof acoustic sensors that can reliably sample the acoustic environment for months at a time without maintenance. Hundreds of hours of field recordings can then be processed with pattern recognition software using deep learning and artificial neural network techniques to derive measures of acoustic activity rates for species of interest. This combination of passive acoustic sensors and automated call detection is especially powerful for monitoring rare/elusive species and species in remote locations (Acevedo & Villanueva-Rivera, 2006; Agranat, 2007; Brandes, 2008a, 2008b).

This survey method takes advantage of the social behavior that occurs at and around breeding aggregations, including frequent vocalizations. Automated acoustic sensors and automated acoustic classification techniques now make it possible to efficiently detect and quantify vocalizations in large datasets. This technology enables researchers to greatly increase the spatial and temporal scale of acoustic surveys - improving detection probabilities for rare and elusive species. The increased survey effort enabled by passive acoustic monitoring is particularly helpful for identifying previously-unknown breeding sites, as well as improving the statistical power of long-term monitoring projects when compared to less intensive monitoring methods (MacKenzie, Nichols, & Lachman, 2002; MacKenzie, Nichols, & Sutton, 2005).

Passive acoustic sensors and automated classification techniques have increasingly been employed to search for rare bird species including many seabirds (McKown 2008; Buxton & Jones 2012; Buxton et al. 2013; Oppel et al. 2014; Borker et al. 2014). In addition, specifically for terns, acoustics have been used as an index for colony size, suggesting that it could be an effective method for monitoring Aleutian Tern (Borker et al., 2014).

Methods

Acoustic sensor hardware and survey design

Sixteen Song Meter 4 (SM4) and 8 Song Meter 2 (SM2) sensors, manufactured by Wildlife Acoustics, were used to collect recordings for this survey. SM4s were programmed to record 1-minute out of every 6-minutes and SM2s were programmed to record 1 minute out of every 5 minutes 24 hours a day.

Traditional colony monitoring methods were paired with acoustic sensors at 13 colonies. Adults were counted using two methods, flush counts and direct counts. Direct counts only occurred at Burton ranch, Pasagshak River, and Naknek. These counts were considered equivalent in this report. Nest counts within 5-, 10-, 15-, and 20-m radii around sensors were undertaken at 6 colonies (Middle Bay, Womens Bay, Italio, Akhiok, Kalsin Bay, and Black Sand Spit). Nest density was calculated from these counts as: $Density = \frac{count}{\pi r^2}$ where r is the radius of the circle in which each count was conducted.

Automated call detection

Automated acoustic analysis of all field recordings was carried out with custom detection and classification software created by Conservation Metrics (CMI). We apply a machine learning technique known as a Deep Neural Network, which detects sounds on field recordings that have spectro-temporal properties similar to those measured from signals produced by target species. Deep Neural Networks (DNNs) are a powerful classification tool used in many fields to perform speech recognition, image recognition, and computer vision tasks (Cichy, Khosla, Pantazis, & Torralba, 2016; Deng, Hinton, & Kingsbury, 2013; Min, Lee, & Yoon, 2016; Schmidhuber, 2015).

Our approach to acoustic analysis splits field recordings into 2-second clips and extracts measurements of 10 spectro-temporal features typically found in animal sounds. A DNN classification model is then trained for each species of interest using training and cross-validation datasets containing examples of vocalizations from target species and a representative sample of other sounds from the study region. The DNN learns which combination of spectro-temporal features best differentiates target sounds from other sounds in the environment. The trained DNN can then be applied to new acoustic data from survey sites, returning a probability that a given 2-second window of field recordings contains a sound produced by the target species (Figure 1).

Aleutian Tern vocalizations are diverse, and we chose to train the detector to identify three common flight calls: a long trill, and short trill and a buzzy call all with peak energy between 2,500 and 5,000 hz (Figure 1). These calls were not distinguished from each other by the detector, and the metrics of Aleutian Tern activity presented in this report represent the sum of all three call types together. In addition to the three focal calls, the detector did not effectively reject single-note alarm calls (Figure 1), and some of the calls detected were this call.

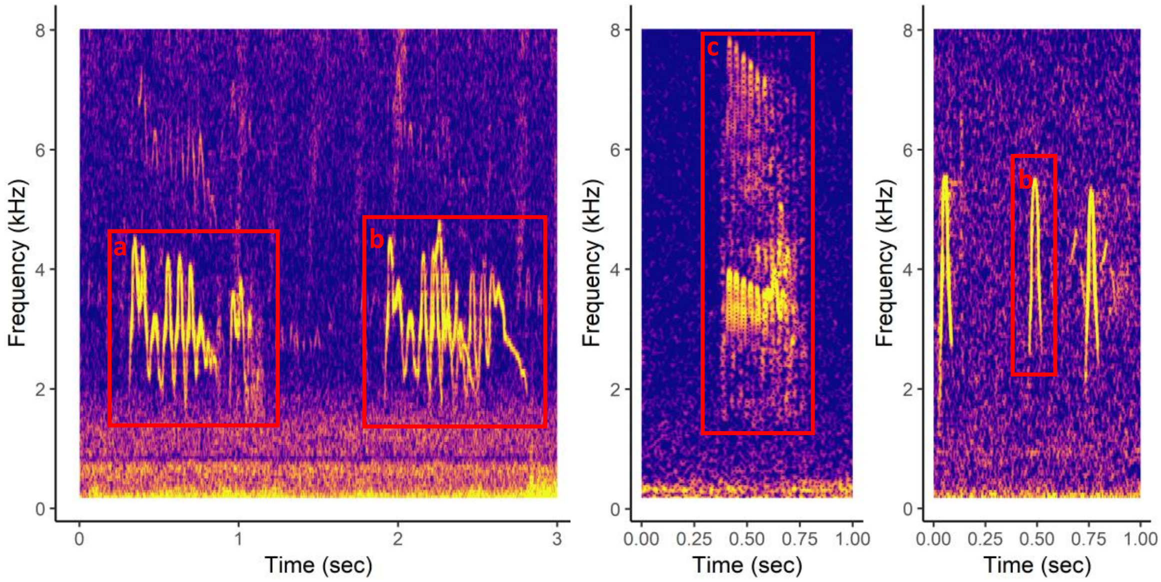


Figure 1. Long (a) and short (b) Aleutian Tern trill calls, (c) Aleutian tern buzz call, and (d) Aleutian Tern alarm call.

We iteratively trained a multiclass DNN detection model using an “active learning” workflow. Active learning uses successive rounds of data labeling and model training to develop neural network models. Each round of labeling informs the following model training, which then enhances the next round of labeling. This workflow accelerates the creation of training datasets by efficiently identifying and labeling examples of target and non-target sounds from randomly selected recordings. After five active learning cycles we had compiled a training dataset with 28,567 manually labeled 2-second sound clips with examples of 59 distinct sounds (classes). We evaluated model performance with ROC curves using True Positive Rate (# events with Aleutian Tern calls above a classification threshold divided by total events predicted to have Aleutian Tern calls) and False Positive Rate (# events that did not have Aleutian Tern calls above a classification threshold divided by the total events without Aleutian Tern calls in the dataset). We built three models from this dataset, using different class combinations for each, and chose the model that performed the best on the test dataset with a true positive rate close to 50% on a manually reviewed independent test dataset containing 1,185 2-second clips with Aleutian Tern and 7,411 other randomly selected clips from the soundscape. The selected model (AK_Wetlands_Multi_6Feb19_V8) returned 49.8% of Aleutian Tern flight vocalizations above the chosen DNN classification score threshold (0.75) with a False Positive Rate of 0.57% (Figure 2).

The high accuracy (91.6%) of our model, and the abundance of Aleutian Tern signal in the acoustic dataset allowed us to use the raw model output to inform patterns of seasonal, diel and relative abundance among sites. While we acknowledge that there are false positive detections in the data, they are too infrequent to substantially influence our analysis results. In total, there were 1,193,893 events classified as Aleutian Tern across the entire acoustic dataset. We confirmed at least one Aleutian Tern call at each site through manual review.

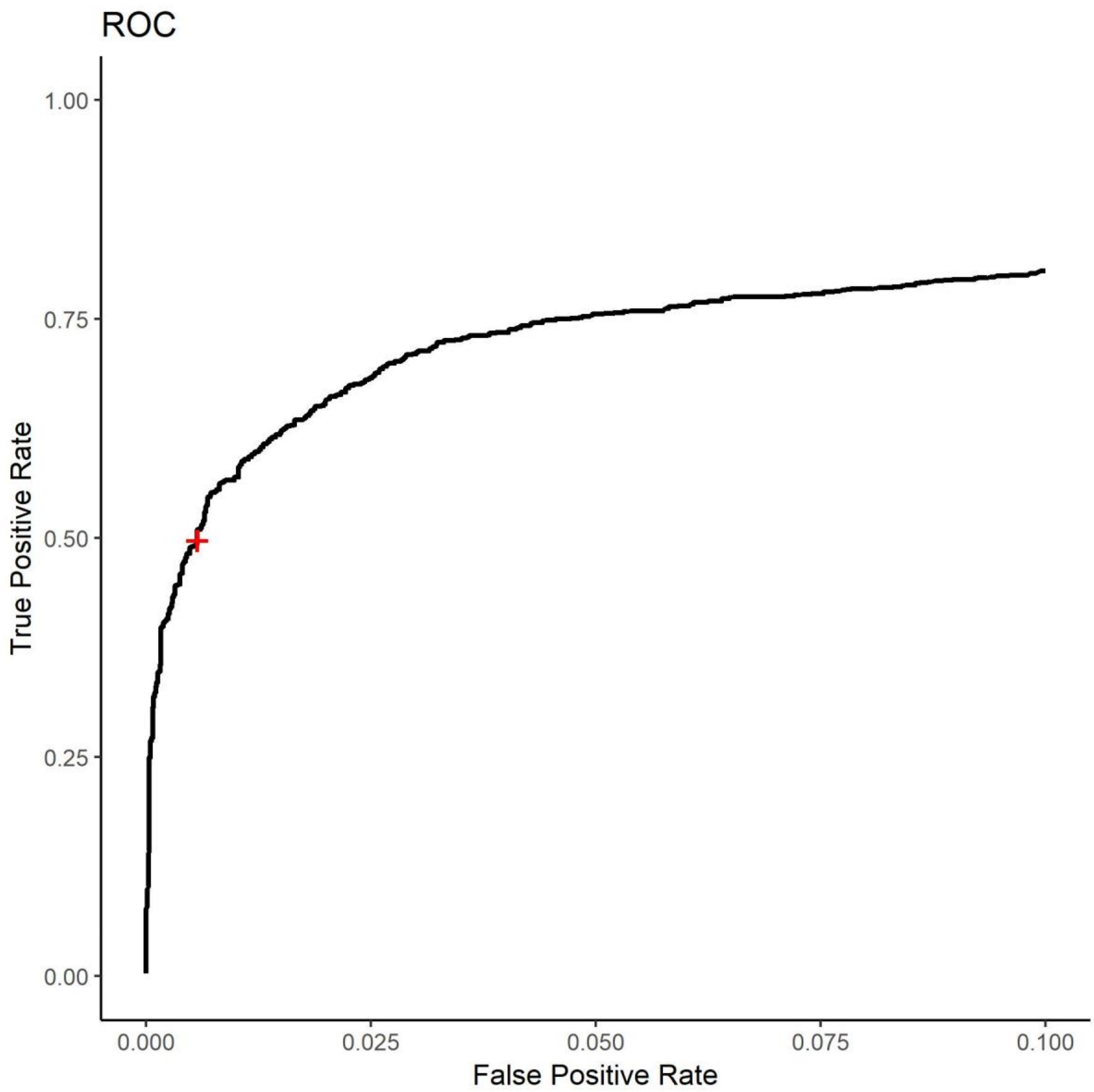


Figure 2: Receiver Operator Curve for the Aleutian Tern class in model version AK_Wetlands_Multi_6Feb19_V8. The red cross indicates the chosen DNN classification score threshold with performance characteristics of a false positive rate of 0.57% and a true positive rate of 49.8%.

Statistical Analysis

Call rates were summarized by day during a peak activity time period based on an offset from sunrise. This period, 90 - 270 minutes after sunrise, was chosen based on visual inspection of the diel activity patterns shown in *Figure 5*. Daily peak period call rates were then averaged across a 3-month date range from 15 May to 15 August to calculate an average calls per minute as an activity metric. The seasonal peak date range was selected based on the general activity period from initial data exploration. Most sensors did not collect data for this entire survey period. Call rates are presented as mean calls per minute ± 1 sd.

For comparisons of call rates with traditional tern counts and nest density metrics, we calculated rates using a subset of the recordings collected before and/or after the date of the traditional surveys. We used recordings from 7 randomly selected days from a 15-day window centered on the date that each count occurred at each site. The randomization allowed us to select the same number of days of recording from each sensor-count pair and include data from sites that had acoustic data from only before or after traditional counts were conducted. This was required when sensors were moved on the count date; and therefore, had no recordings on the days immediately prior, or following, the count. Counts that had fewer than 7 days of data within the 15-day buffer were excluded from the analysis (n=8). Some traditional colony counts covered an area encompassing two acoustic sensors. In these cases, we averaged the call rates for the two sensors.

We investigated the relationship of vocal activity to colony counts with two models. First, we modeled the entire dataset with counts from all stages of the breeding season. These data were over dispersed with a large number of zeros in the counts, and we used a zero-inflated negative binomial GLM with a log-link to model the full dataset. We also split out the counts that were made during incubation (n=8). Although the data during incubation were still over dispersed, there were fewer zeros and we fit a negative binomial model with a log-link.

We also modeled the relationship of call rates to nest density. These data were sparse (n=12) and did not lend themselves to complex modeling. On each survey the number of nests within 5-, 10-, 15-, and 20-m radius were counted. We used linear regression (ordinary least squares) restricted to the incubation period (n=8) to check for a relationship between nest density within each radius and call rate. Each radius was modeled separately.

Results

Survey Effort

Acoustic sensors were deployed at 25 sites in Alaska between 1 May and 4 October (for all recordings and did not identify any microphone failures. Three sensors were deployed for which we did not receive any data: ALTE13 at Kalsin1 from 16 May to 3 July, ALTE 15 at Womens Bay2 from 13 June to 3 July ALTE 20 at Foul Bay NC from 28 June to 11 August. Additionally, we received acoustic recordings from a sensor named ALTE1 with dates from 2017 that did not have deployment information and was therefore removed from analysis.

The sensors at Naknek1 and Naknek2 had a different recording schedule, recording up to 24 hours of data per day (Figure 4). We subsampled these data to match the 1 minute every 6-minute recording schedule for reporting, including the effort reporting.

Table 1, Figure 3, and Figure 4). They recorded 5,542.17 hours of audio across 1,412 sensor-days (for all recordings and did not identify any microphone failures. Three sensors were deployed for which we did not receive any data: ALTE13 at Kalsin1 from 16 May to 3 July, ALTE 15 at Womens Bay2 from 13 June to 3 July ALTE 20 at Foul Bay NC from 28 June to 11 August. Additionally, we received acoustic recordings from a sensor named ALTE1 with dates from 2017 that did not have deployment information and was therefore removed from analysis.

The sensors at Naknek1 and Naknek2 had a different recording schedule, recording up to 24 hours of data per day (Figure 4). We subsampled these data to match the 1 minute every 6-minute recording schedule for reporting, including the effort reporting.

Table 1 and Table 2).

Over the course of an acoustic sensor deployment, exposure to the elements can degrade the sensitivity of microphones. We evaluated sound quality for all recordings and did not identify any microphone failures. Three sensors were deployed for which we did not receive any data: ALTE13 at Kalsin1 from 16 May to 3 July, ALTE 15 at Womens Bay2 from 13 June to 3 July ALTE 20 at Foul Bay NC from 28 June to 11 August. Additionally, we received acoustic recordings from a sensor named ALTE1 with dates from 2017 that did not have deployment information and was therefore removed from analysis.

The sensors at Naknek1 and Naknek2 had a different recording schedule, recording up to 24 hours of data per day (Figure 4). We subsampled these data to match the 1 minute every 6-minute recording schedule for reporting, including the effort reporting.

Table 1. Deployment Table

SPID	Recording Unit	Latitude	Longitude	First Recording	Last Recording
Akhiok	ALTE3	56.94	-154.14	2018-05-22 10:22:00	2018-07-20 11:18:00
Black Sand Spit1	ALTE6	59.45	-139.61	2018-05-24 13:44:00	2018-08-17 10:04:00
Black Sand Spit2	ALTE7	59.45	-139.61	2018-05-24 11:50:00	2018-08-17 11:04:00
Black Sand Spit3	ALTE9	59.44	-139.58	2018-05-09 11:22:00	2018-08-17 10:46:00

SPID	Recording Unit	Latitude	Longitude	First Recording	Last Recording
Burton Ranch	ALTE2	57.48	-152.33	2018-06-05 13:05:00	2018-07-08 14:07:00
Foul Bay 9	ALTE16	57.76	-152.50	2018-06-01 17:37:00	2018-06-13 14:15:00
Foul Bay 9	ALTE18	57.76	-152.50	2018-06-13 13:21:00	2018-08-11 17:38:00
Grassy Island	ALTE20	58.45	-152.78	2018-06-13 05:09:00	2018-06-27 11:59:00
Italio	ALTE8 SM4	59.32	-139.18	2018-05-26 16:09:00	2018-09-01 10:26:00
Kalsin2	ALTE8 SM2	57.59	-152.46	2018-07-03 11:24:00	2018-08-23 11:05:00
Kenai SM2	KENAISM2	60.46	-151.07	2018-05-11 15:48:00	2018-05-30 12:56:00
Kenai SM4	S4KENAI	60.46	-151.07	2018-06-08 11:46:00	2018-08-13 05:12:00
Lost River	ALTE10	59.46	-139.61	2018-05-08 15:13:00	2018-08-17 13:10:00
Middle Bay1	ALTE4	57.65	-152.50	2018-05-01 12:31:00	2018-07-11 05:57:00
Middle Bay2	ALTE4	57.65	-152.51	2018-07-12 11:40:00	2018-08-23 09:58:00
Middle Bay3	ALTE5	57.65	-152.50	2018-05-01 13:07:00	2018-07-09 00:05:00
Naknek1	ALTE12	58.74	-157.06	2018-06-01 11:23:00	2018-06-27 13:56:00
Naknek2	ALTE12	58.73	-157.06	2018-06-27 14:20:00	2018-09-30 09:07:00
Pasagshak1	ALTE1	57.48	-152.47	2018-05-04 12:30:00	2018-07-09 01:55:00
Pasagshak2	ALTE2	57.47	-152.47	2018-05-04 12:54:00	2018-06-05 11:59:00
Sheep	ALTE14	57.22	-153.25	2018-05-22 09:30:00	2018-06-09 23:34:00
Sheep	ALTE19	57.22	-153.25	2018-06-11 09:40:00	2018-08-01 09:35:00
Stonestep Lake SM2	STPSTN2018	59.69	-151.35	2018-05-02 09:57:00	2018-05-16 07:14:00
Stonestep Lake SM4	ALTE18-HOMER	59.69	-151.35	2018-05-26 11:07:00	2018-07-19 14:44:00
Three Spruce Island	ALTE14	58.59	-152.51	2018-06-24 04:10:00	2018-07-10 02:39:00
Womens Bay1	ALTE11	57.71	-152.57	2018-07-03 16:35:00	2018-08-23 09:04:00
Foul Bay NC	ALTE20	58.34	-152.89	No Data	
Kalsin1	ALTE13	57.59	-152.46	No Data	
Womens Bay2	ALTE15	57.70	-152.57	No Data	

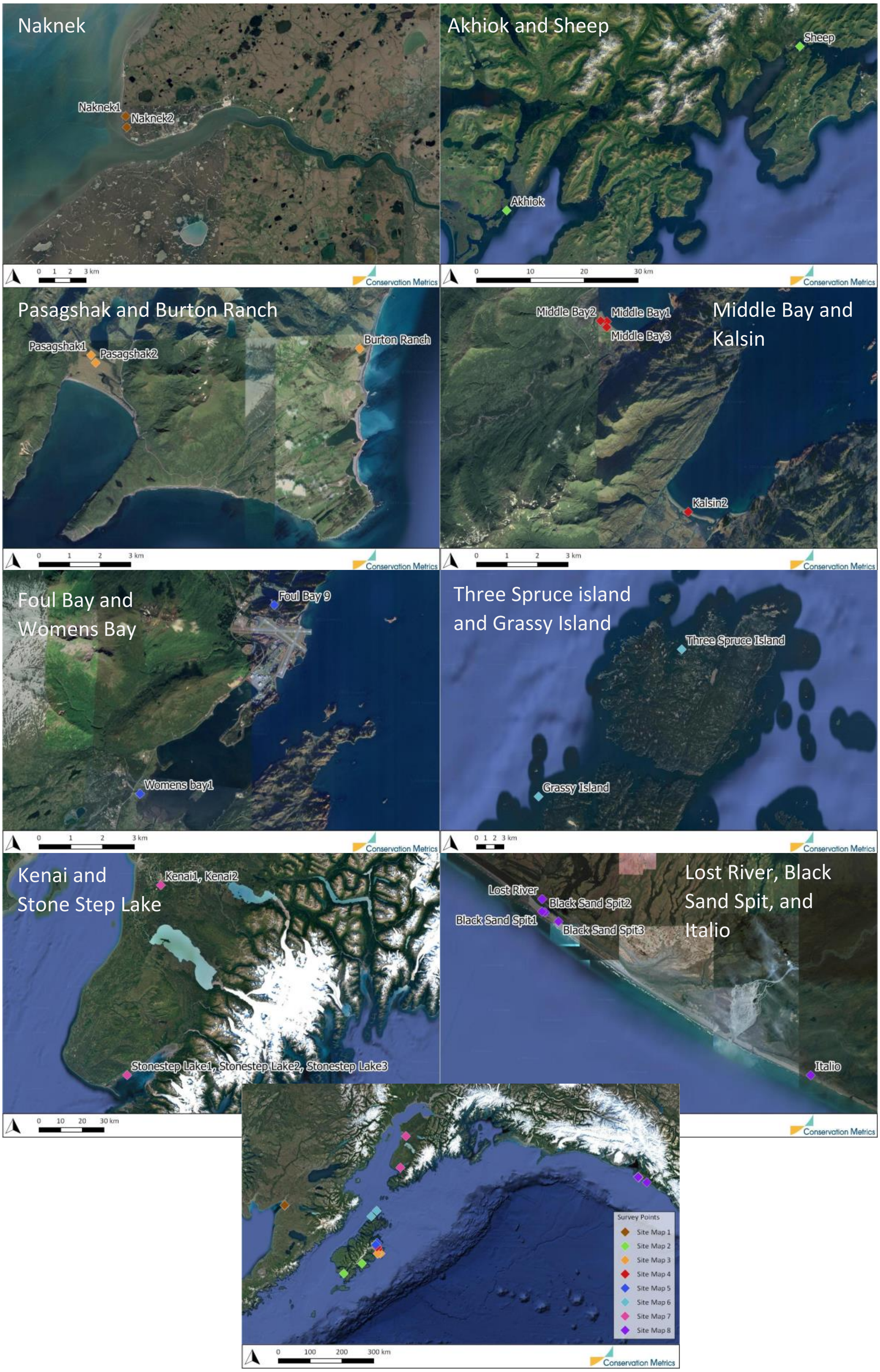


Figure 3. Survey location maps sites (1-8). See Appendix A for higher resolution maps.

Table 2. Survey effort

SPID	Total Nights	Total Hours
Akhiok	60	236.13
Black Sand Spit1	86	339.18
Black Sand Spit2	86	339.12
Black Sand Spit3	101	399.78
Burton Ranch	34	132.08
Foul Bay 9	72	278.83
Grassy Island	15	68.58
Italio	99	362.37
Kalsin2	52	218.55
Kenai SM2	20	74.80
Kenai SM4	67	258.72
Lost River	102	403.53
Middle Bay1	72	283.00
Middle Bay2	43	167.50
Middle Bay3	70	273.95
Naknek1	27	104.47
Naknek2	96	378.58
Pasagshak1	67	262.33
Pasagshak2	33	128.02
Sheep	71	287.40
Stonestep Lake SM2	15	53.37
Stonestep Lake SM4	55	216.58
Three Spruce Island	17	72.92
Womens Bay1	52	202.37
Total	**1,412**	**5,542.17**

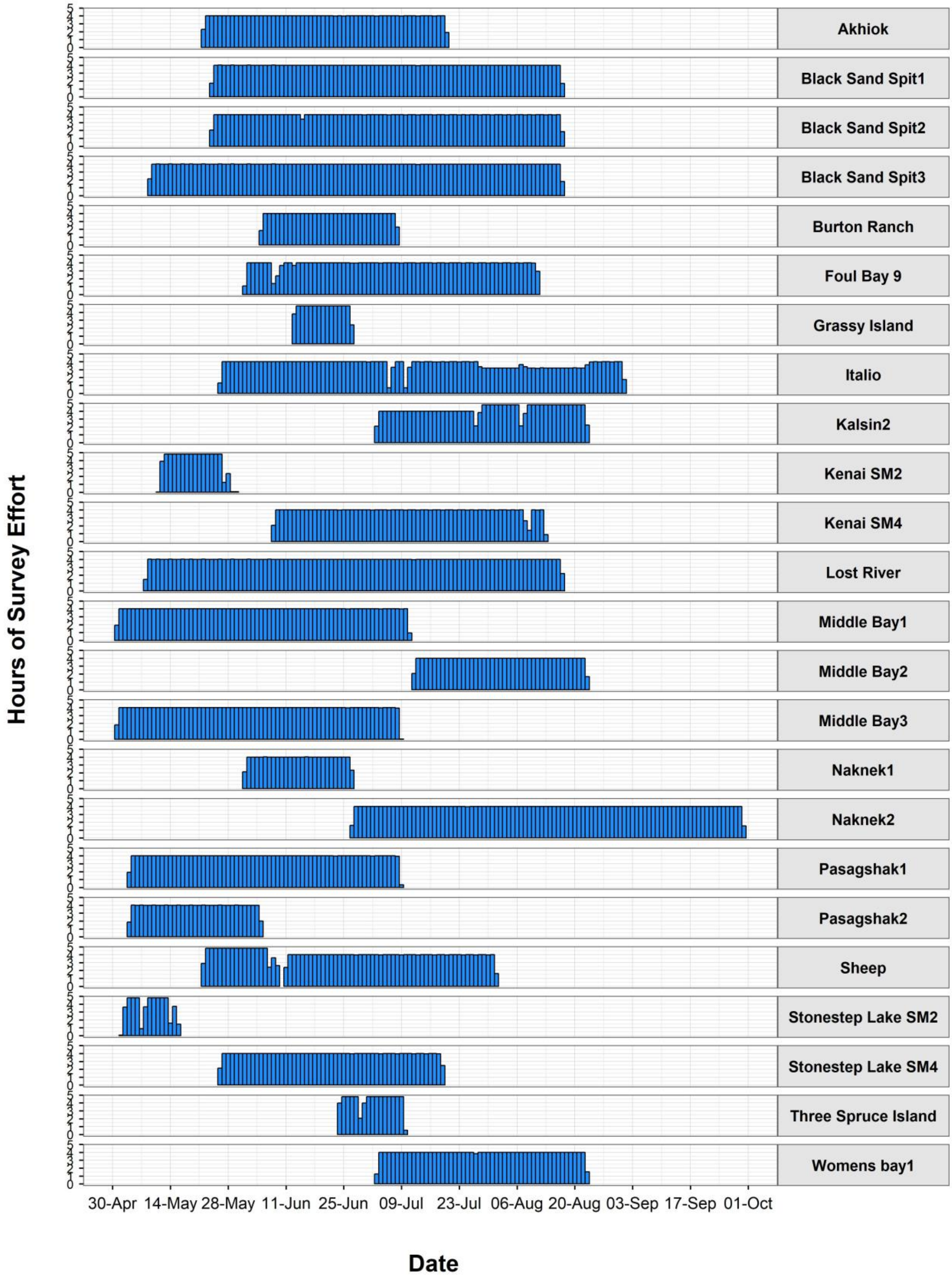


Figure 4. Survey effort. Blue bars show the number of hours of data received by day.

Aleutian Tern Phenology and Relative Abundance

Aleutian Tern calls were detected at all 25 survey sites (Figure 7). Daily activity patterns generally showed a peak in vocal activity in the morning from 90 to 360 minutes after sunrise, with a steady decline throughout the day and very low level of calling at night (Figure 5). This pattern was generally consistent, with some variation within each site as the season progressed and some inter-site variation (Figure 6, See Appendix B for larger scale figures by site). We chose a 2-hour morning period between 90 to 270 minutes after sunrise to compare mean activity patterns over the season. In general, there were two peaks in activity; the first from mid-May to mid-June and a second later peak from early-July to late-July (Figure 7). At Black Sand Spit sites, the later peak ended a little sooner than other sites (Figure 7).

Mean call rates during the comparison period (15 May to 15 Aug) varied among sites from 0.04 ± 0.08 at Kenai SM4 to 10.43 ± 8.45 calls per minute at Naknek2 (Figure 8, Table 3, Figure 9, See *Appendix A* for larger maps).

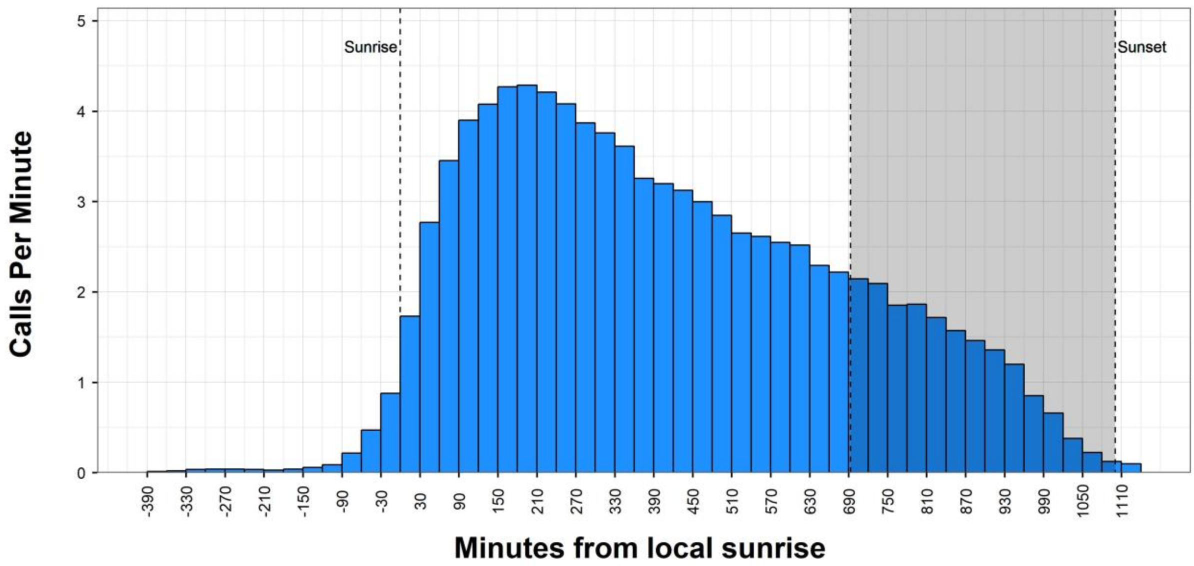


Figure 5. Species activity as a function of time from sunrise. Each bar is a 30-minute bin.

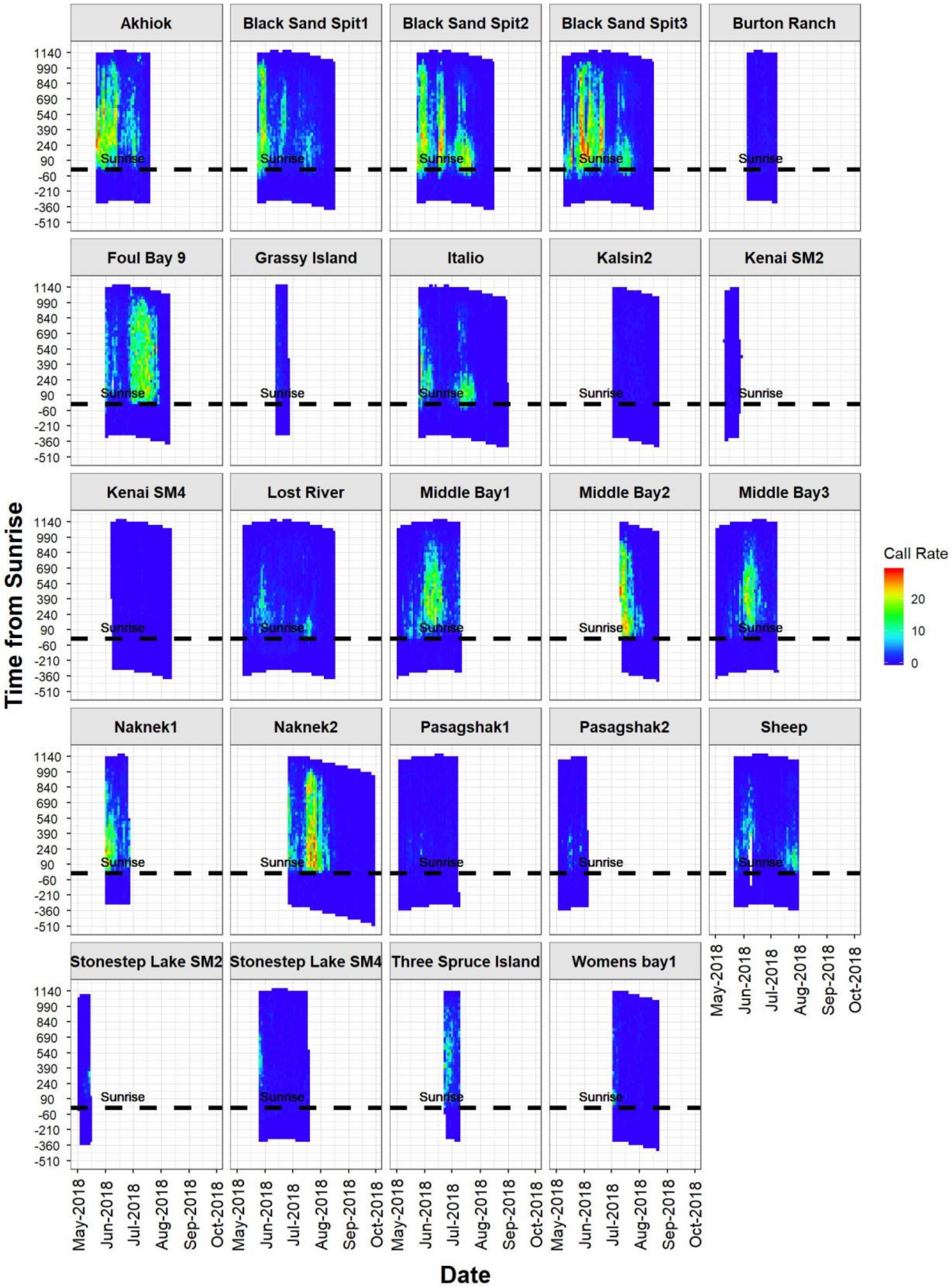


Figure 6: Phenological patterns of Aleutian Tern activity in relation to sunrise at each survey site. Each cell in these raster graphs represents the mean call rate for non-overlapping 30-minute time bins (y-axis) and 2-day date bins (x-axis). Warmer colors indicate higher rates of acoustic activity. NOTE: Survey effort varied considerably across sites.

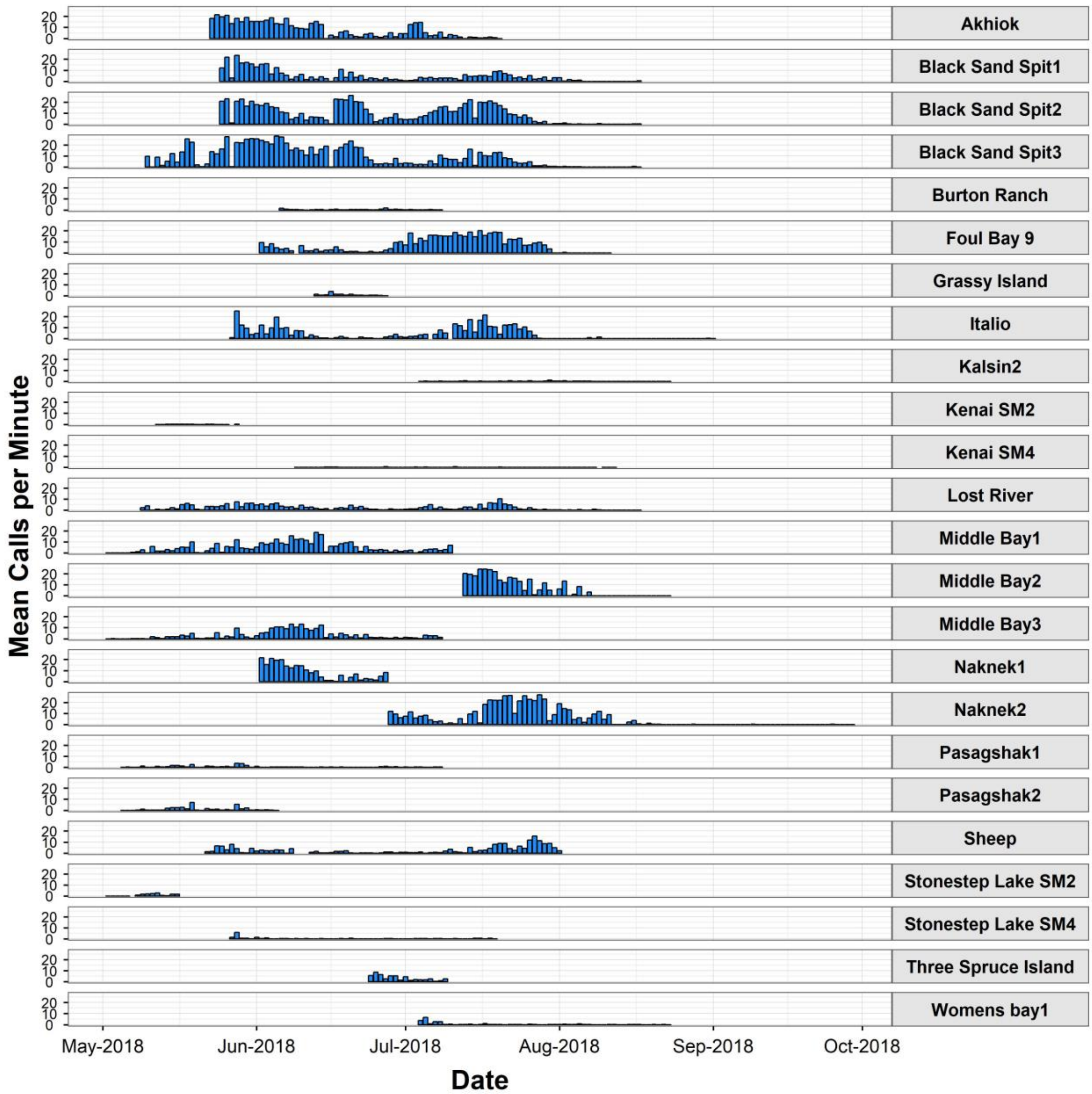


Figure 7. Aleutian Tern acoustic activity rates by date and site during the peak calling hour (90-270 minutes after sunrise).

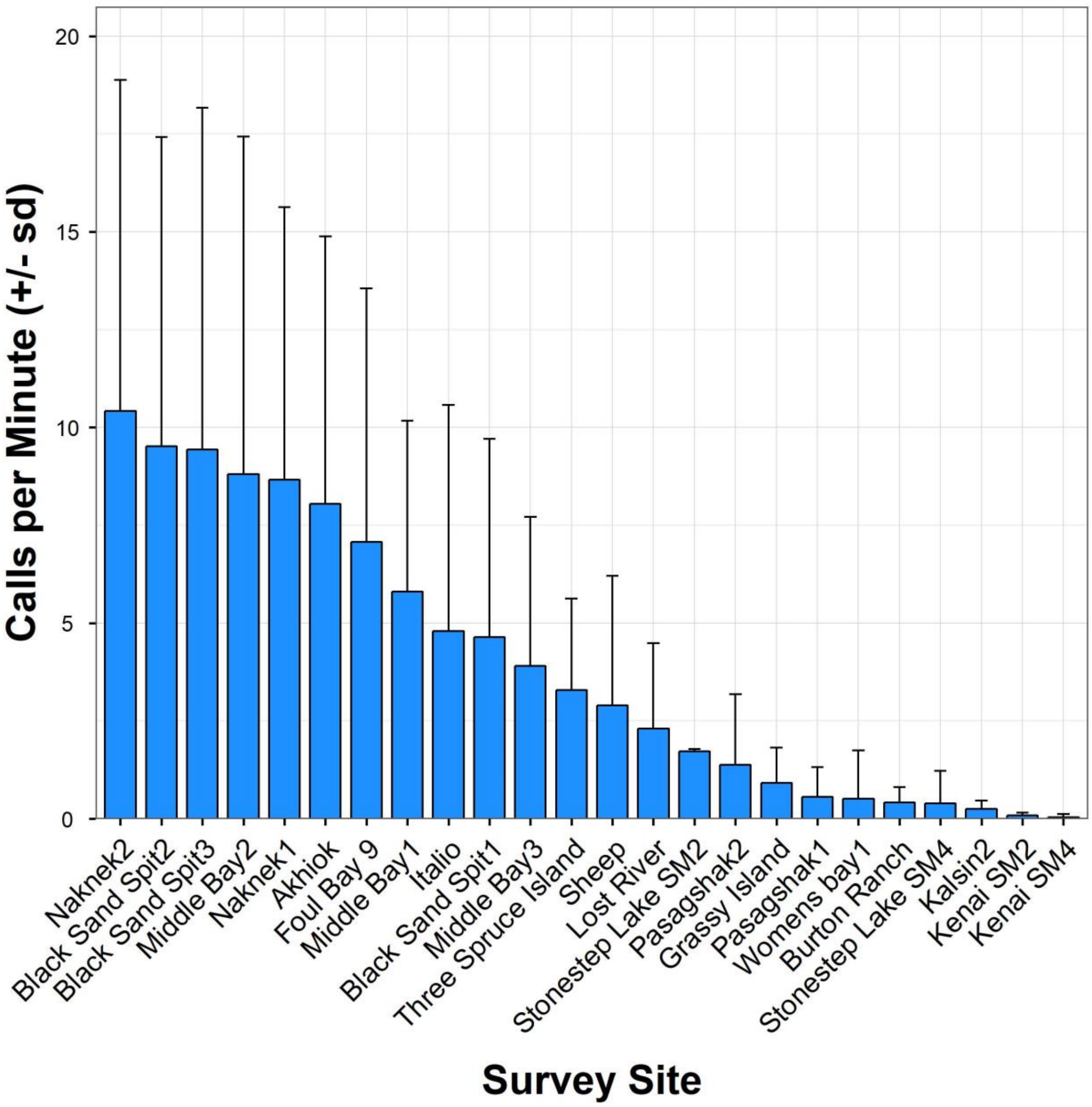


Figure 8. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August).

Table 3. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August).

SPID	Rate Min^{-1}	N	sd	se
Akhiok	8.05	59	6.84	0.89
Black Sand Spit1	4.65	83	5.07	0.56
Black Sand Spit2	9.51	83	7.90	0.87
Black Sand Spit3	9.43	93	8.73	0.91
Burton Ranch	0.41	33	0.39	0.07
Foul Bay 9	7.07	70	6.48	0.77
Grassy Island	0.91	15	0.91	0.24
Italio	4.80	79	5.77	0.65
Kalsin2	0.25	43	0.21	0.03
Kenai SM2	0.08	13	0.07	0.02
Kenai SM4	0.04	64	0.08	0.01
Lost River	2.31	93	2.18	0.23
Middle Bay1	5.80	57	4.37	0.58
Middle Bay2	8.81	34	8.62	1.48
Middle Bay3	3.90	55	3.82	0.51
Naknek1	8.66	26	6.97	1.37
Naknek2	10.43	49	8.45	1.21
Pasagshak1	0.56	55	0.76	0.10
Pasagshak2	1.37	22	1.81	0.38
Sheep	2.90	69	3.30	0.40
Stonestep Lake SM2	1.72	2	0.07	0.05
Stonestep Lake SM4	0.39	54	0.83	0.11
Three Spruce Island	3.28	16	2.34	0.58
Womens bay1	0.52	43	1.23	0.19

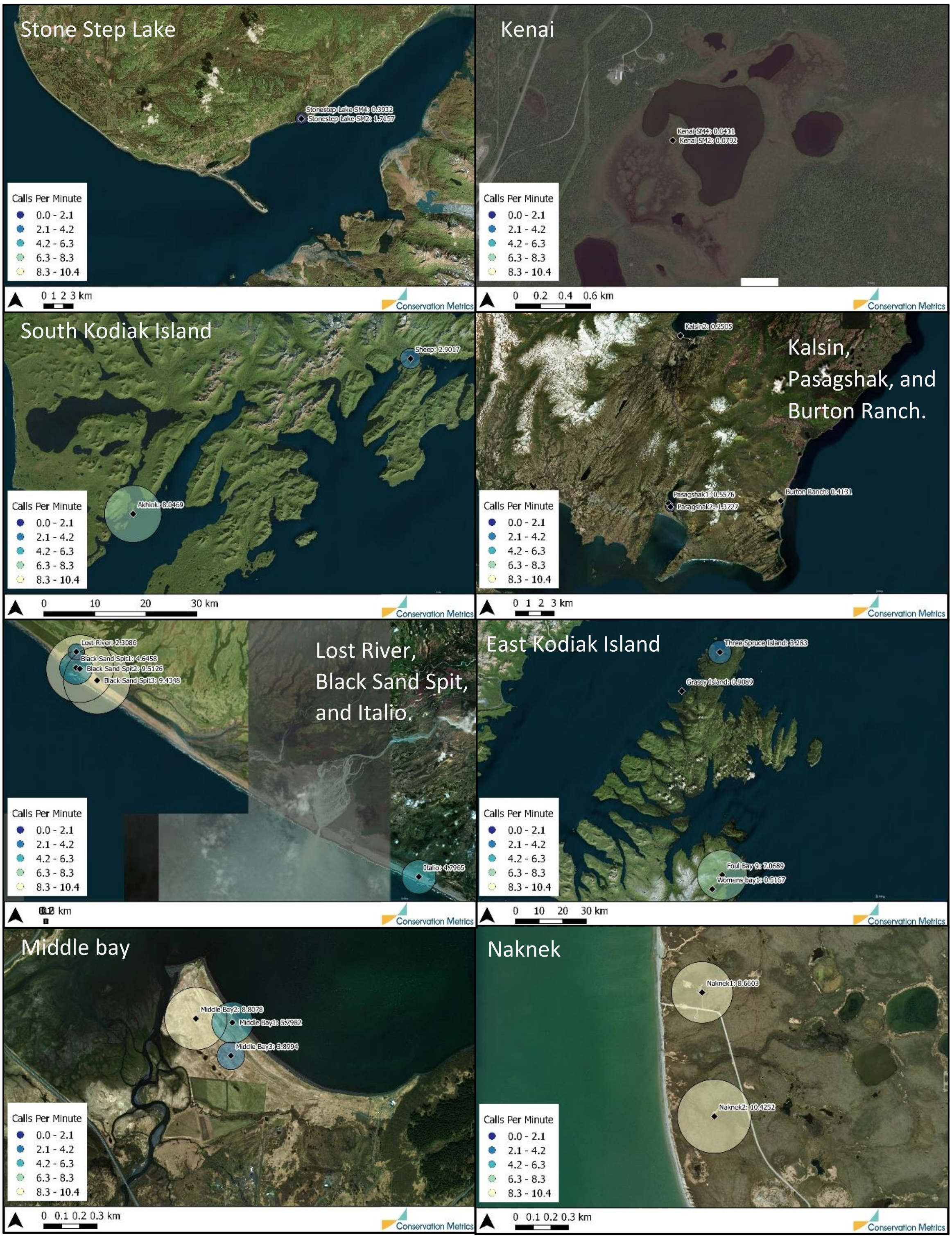


Figure 9. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August). See Appendix A for colony specific higher resolution maps.

Traditional Count Metrics

There were 67 traditional counts completed during this survey. Some colonies had a single count (Italio, Three Spruce Island) while other sites had up to 16 counts (Middle Bay). Black Sand Spit (80 ± 34.6) and Naknek (78.4 ± 50.9) had the highest mean counts, followed by the single count at Italio (45; Figure 10).

Nest density was monitored at six colonies; four of which were only monitored on a single date (Akhiok, Italio, Kalsin Bay, Womens Bay). The other two colonies (Black Sand Spit and Middle Bay) were monitored three times each. Nest searches found few nests within 20-m of the sensors monitored. The maximum number of nests found within 20m of a sensor was 6 nests (Black Sand Spit1). Density ranged from 0.00 to 0.026 nests m^{-2}, the maximum nest density was recorded at Black Sand Spit2 at a 5-m radius (2 nests within 5 meters).

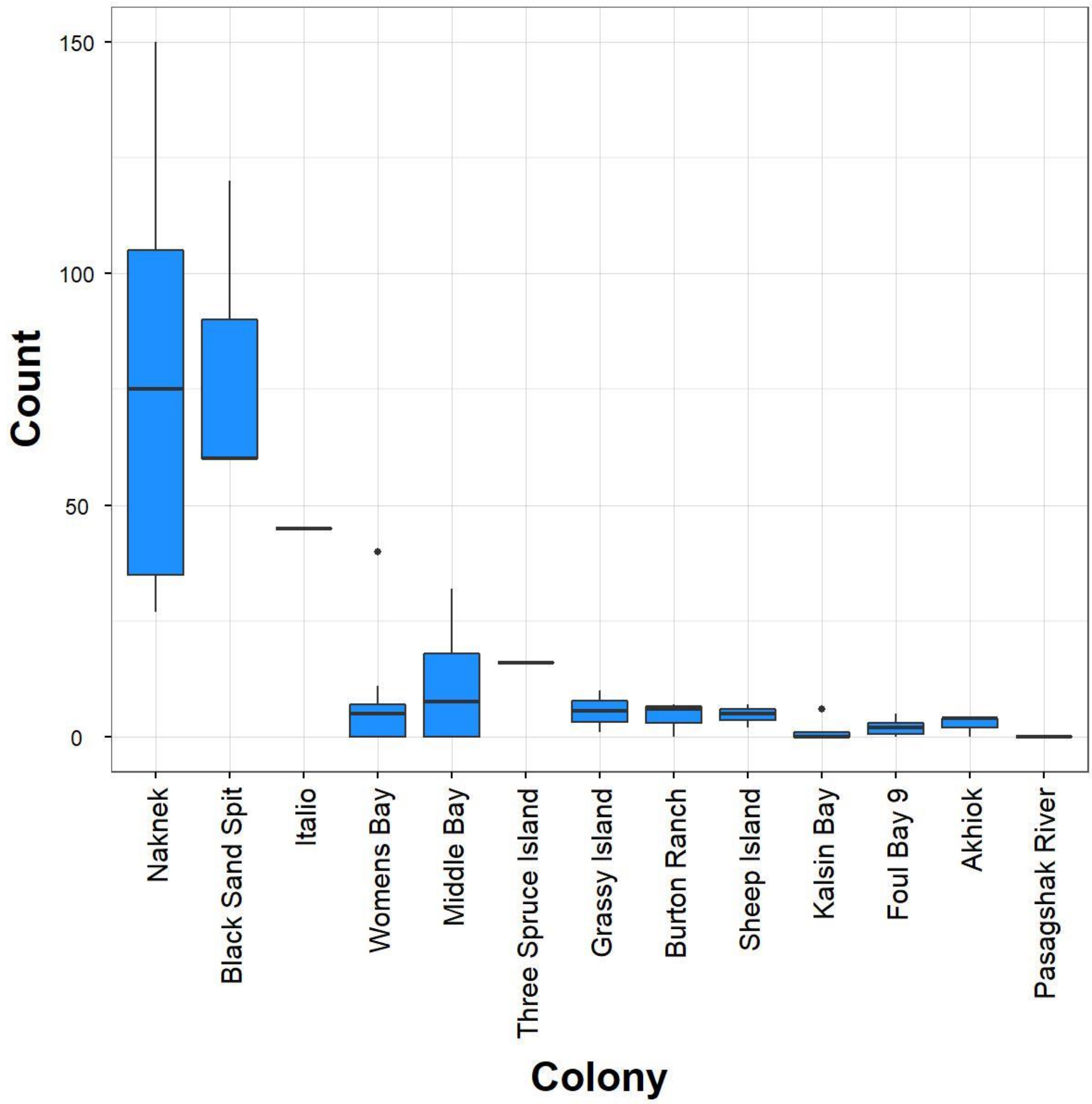


Figure 10: Summary of traditional adult Aleutian Tern colony counts. The black center line represents the median, the box represents 25%-75% quartiles, and the whiskers represent 95% quantiles. Data points outside the 95% quantile are represented as black dots.

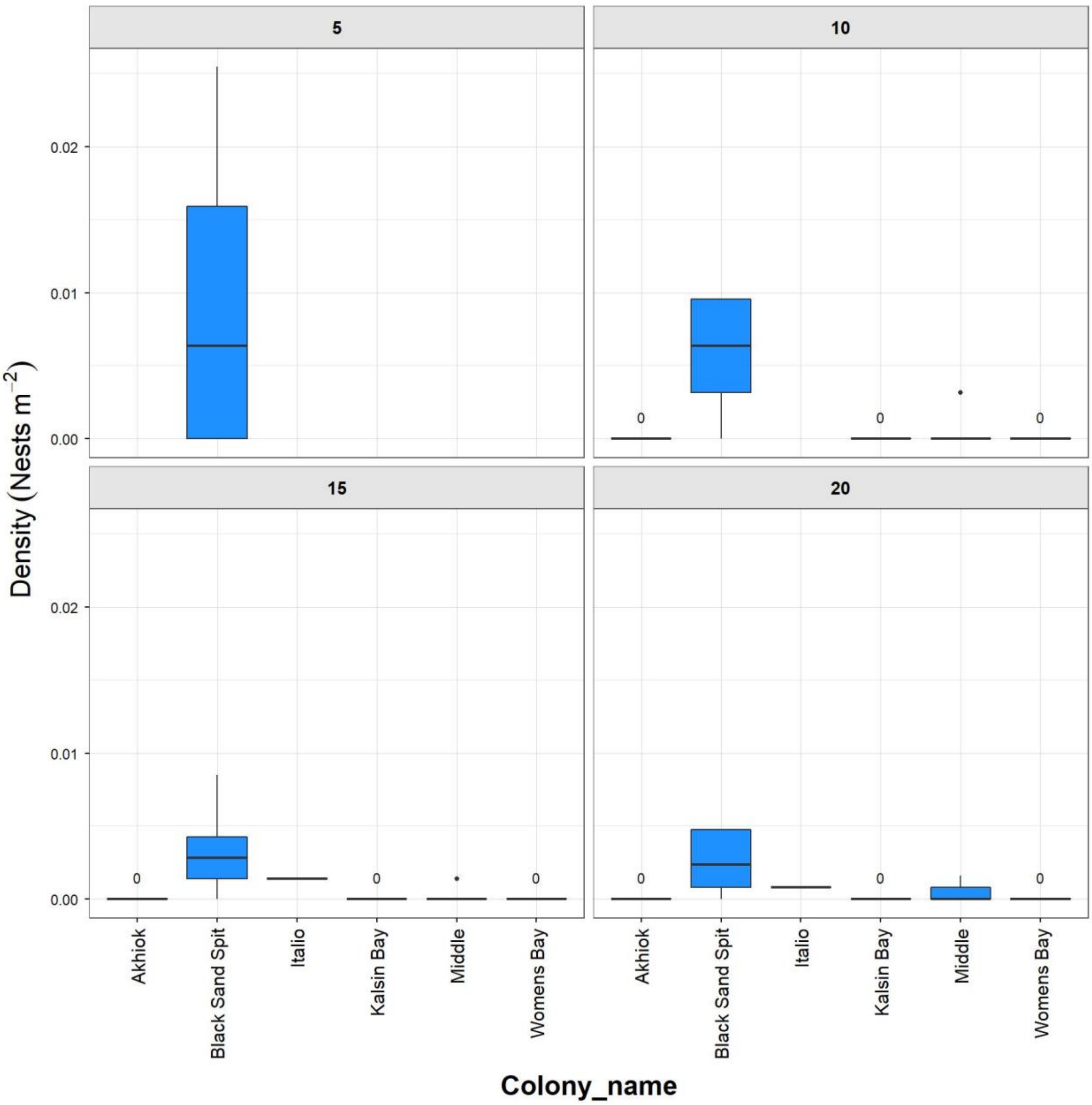


Figure 11: Aleutian Tern nest density at 5-,10-,15-,and 20-meter radii around acoustic sensors at 6 colonies in coastal Alaska. The black center line represents the median, the box represents 25%-75% quartiles, and the whiskers represent 95% quantiles. Data points outside the 95% quantile are represented as black dots.

Comparing Call Rates to Traditional Monitoring Metrics

Counts - full data set

There were 66 observations where colony counts occurred during an active acoustic survey window with at least 7 days of data near the count date. The traditional colony count data in this dataset were over-dispersed and there were many counts of zero terns. Thus we chose a Generalized Linear Modeling approach with a zero-inflated negative binomial distribution and a log link function to compare colony count data to call rates. This two-part mixture model suggests that call rate is inversely related to the probability of detecting a false zero count (i.e. a zero generated when birds were missed by counters; odds ratio: 0.26, p = 0.14). This means that the probability of counting at least one tern rose as call rates increased. The second part of the model predicts the relationship of calls to counts, and the results indicate that colony counts increased by 11.6% ± 3.5 se (95% CI: 4.3-19.5%) for each increase of 1 call minute^{-1} (df = 5, p = 0.002; Figure 12).

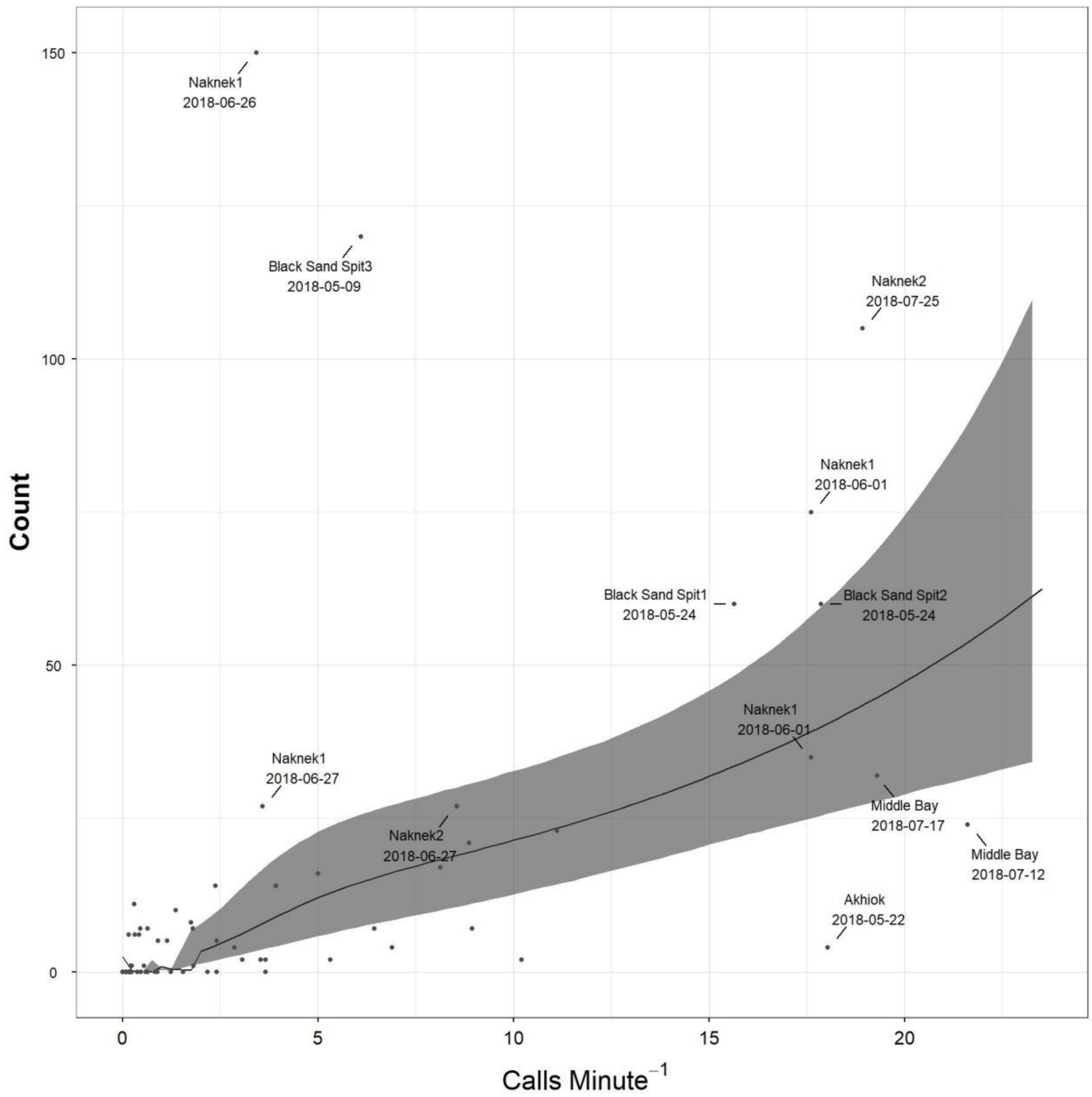


Figure 12: Relationship between call rate and colony counts using the full dataset. Line is a model fit from a zero-inflated negative binomial GLM with a log link. The grey swath represents 95% confidence interval estimated by bootstrapping 10,000 times.

Counts - Incubation

There were 24 counts conducted during the incubation stage (colony status "nesting" or "active") and there were at least 7 days of acoustic survey data near the count date. The count data in this dataset were over-dispersed and we used a GLM with a negative binomial distribution and a log link-function to compare counts to call rates. Counts were positively correlated with call rates, and showed a 17.7% ± 3.5 se (95% CI: 9.2-28.7%) increase for every increase of 1 call minute^{-1} (deviance: 27.6%, p < 0.0001; Figure 13).

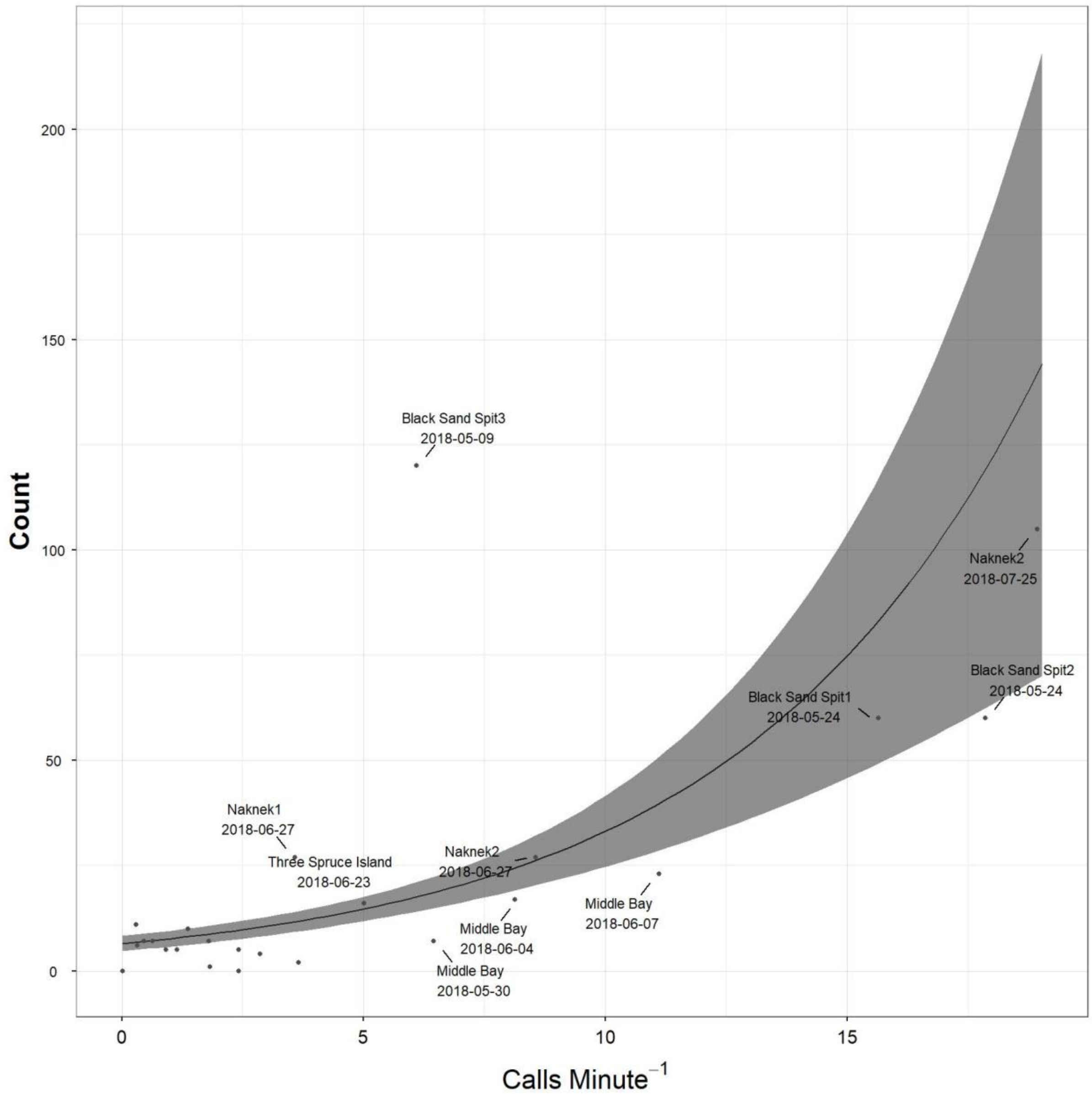


Figure 13: Relationship between call rate and colony counts during the incubation period. Line is a model fit from a negative binomial GLM with a log link. The grey swath represents the standard error around the prediction.

Nest density

There were 12 surveys for nest density conducted during acoustic survey periods. Of these, 8 nest surveys occurred during the incubation period. We used linear regression (ordinary least squares) to investigate the relationship between call rate and nest density. There were not enough samples within 5-m of the sensor to include in our analysis. At other survey radii, call rates were not related to nest density within 10-, 15-, or 20-m radius of the acoustic sensor (i.e. at 15m; R^2=0.006 - 0.10, p=0.44-0.84; Figure 14). The small sample size, inconsistency in counting methods, difficulty of finding and counting nests around each sensor, and dispersed nature of Aleutian Tern nest sites likely help to explain why there was not a significant relationship between these two metrics.

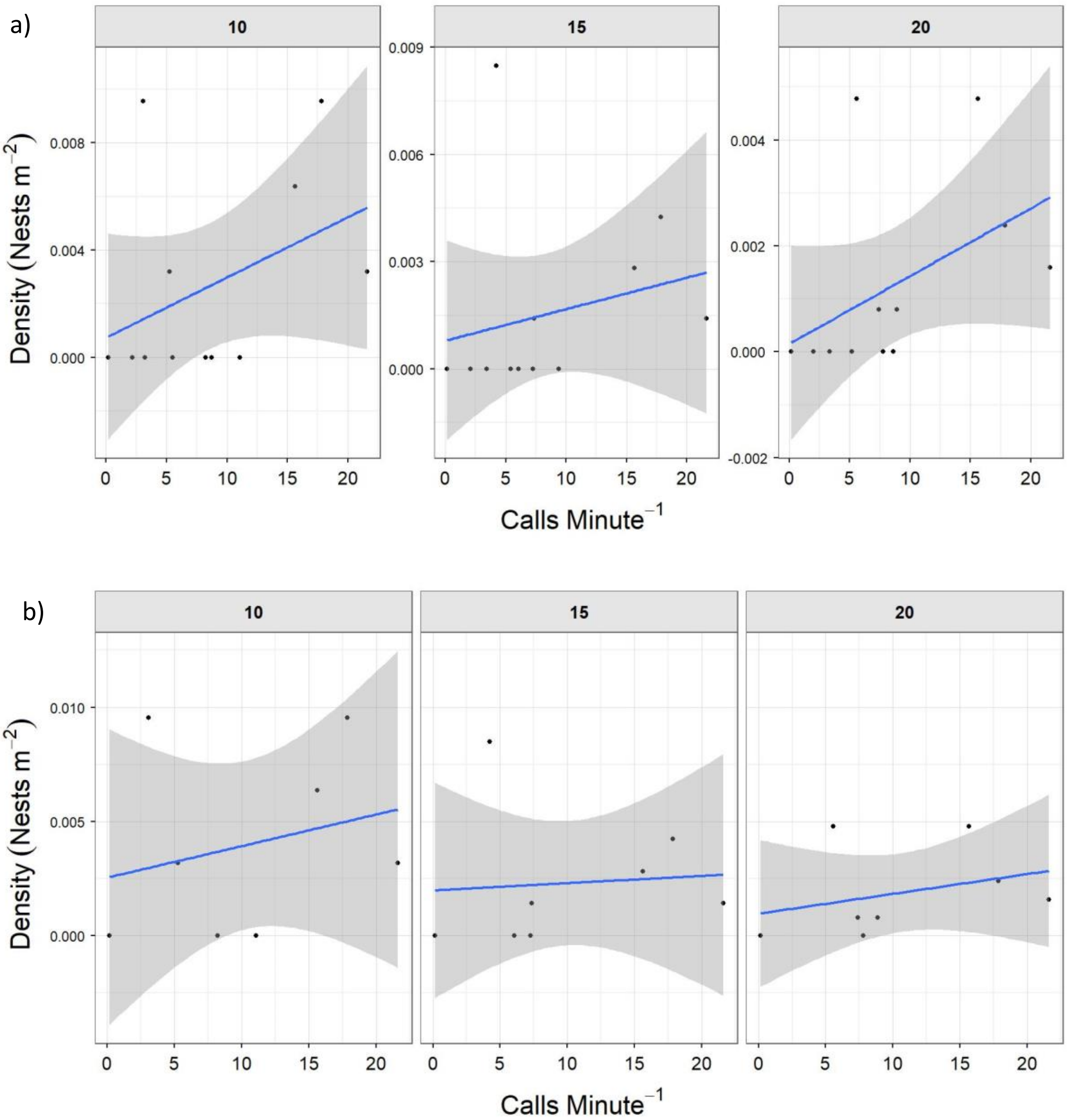


Figure 14: Nest density call rate relationship for a) full dataset and b) incubation with 95% confidence intervals for 10-, 15-, and 20-m radius circles. Only four counts were done at a 5-m radius and it was removed from further analysis.

Discussion

Automated acoustic surveys show promise as a tool to monitor Aleutian Terns at breeding aggregations in Alaska. First, results from the pilot surveys demonstrated that it was possible to develop an effective acoustic classification model to automate the analysis of data from passive acoustic surveys. Our classification model was able to process >5,000 hours of recordings quickly, and with a reasonable levels of classification error. Second, acoustic detection data showed diel and seasonal patterns that could help document general attendance patterns at each Aleutian Tern breeding aggregation across an entire breeding season. Finally, although no clear correlation was found between Aleutian Tern nest densities and acoustic activity rates within a 20-m radius of the acoustic sensors, our analysis did find a statistical relationship between call rates and the number of birds counted during traditional visual colony counts. This relationship, with further ground truthing, could make passive acoustic monitoring an effective tool for monitoring relative abundance of Aleutian Terns at scale.

Detection Model

Our analysis was based on the raw output of our Deep Neural Network classification algorithm for Aleutian Terns. We chose a classification threshold that resulted in a low level of false positives. We estimate that ~50,000 (~4.2%) of the 1,193,893 events classified as Aleutian Tern calls are false detections. False positive classifications were generally calls from other bird species that were mis-classified as Aleutian Tern calls. It is likely that model performance could be improved in the future by increasing the number of training examples for the songbirds and other signal classes included in the model (Total signal classes = 59). Also, separation and reclassification of the Aleutian Tern training data to individual call types (e.g. long trill, short trill, alarm, buzz) may boost classification performance. Creating more specific tern call classes may also help improve the fit of statistical models of the relationship between visual colony counts/nest density counts and call rates by focusing on calls that may be more likely near nest sites (if any). We plan to explore some of these ideas once we receive the 2019 acoustic recordings.

Call Rates as an Index of Abundance

Aleutian Tern call rates were not correlated with nest counts within a 20m radius of the acoustic sensors. Several factors may have influenced these results, including:

1. Variation in count methodology among study locations;
2. Small sample size of traditional nest counts vs. acoustic counts; and
3. The 20m radius limit to the spatial area surveyed for nests.

In future years, nest counts quantified from drone imagery could be used to compare call rates to nest densities over larger spatial scales (i.e. > 20m radius around each sensor). This approach could also help to standardize count methodologies and increase samples sizes, especially with replicate drone flights during the breeding season.

Mean call rates were correlated with visual colony counts. This result is similar to previous research on Forster's Tern (*Sterna forsteri*) (Borker et al., 2014), which found a statistically significant correlation between acoustic metrics and total nests in their study colonies.

The lack of a relationship between the density metrics for Aleutian Terns at the 20m radius scale, but a relationship to colony wide counts suggest that acoustic sensors may be integrating information over larger spatial scales. This would occur when local call rates at a monitoring site are influenced by individuals from across the colony (spatial autocorrelation). The relationship we report is likely a result of individuals (breeders and non-breeders) calling in flight while leaving or returning to the colony, general aerial social behavior over the colony, and periodic bursts of aerial activity from disturbance events.

While this relationship between Aleutian Tern call rates and traditional visual counts of individuals is an encouraging result, it would be preferable to establish a relationship between call rates and numbers of nests (breeding pairs) at a known spatial scale. This would preclude the need to estimate the percentage of breeders vs. non-breeders contributing to the acoustic activity through some other means, and would offer guidance on how many sensors to deploy, and where to place them for large-scale surveys. Thus, we think it would be valuable to explore ways to compare call rates and nest counts over larger radii around each sensor.

In the absence of such a relationship, the fact that acoustic activity rates track colony abundance is still a potential valuable metric, especially when coupled with the ability to use acoustic survey data to track seasonal patterns at each site (*See below*). Large-scale passive acoustic surveys could therefore track presence of Aleutian Terns at survey sites, provide information on the duration of activity at each site, compare relative abundance across sites, and track changes in relative abundance through time (statistical power still to be determined). A large-scale survey effort would likely involve deployment of replicate acoustic sensors (N=3?) per survey area.

Predictive modeling

Modeling for both count datasets (full season and incubation) could likely be improved with the addition of other explanatory variables. There are many covariates that may influence both call rates and human counts that were not included here. These include:

1) *weather* - rain and wind can produce masking noise that will reduce call rates and potentially change calling behavior. These same two variables can influence human counters as well;
2) *co-occurring species* - presence and abundance of other species such as Arctic Terns that may be confused for Aleutian Terns in traditional counts and may mask Aleutian Tern acoustic signals,
3) *factor correlated with nest-count errors* - detection probabilities of traditional counts (nests obscured by grass, areas out of view of the counter),
4) *timing of traditional counts* - diel patterns in calling activity showed a peak of Aleutian Tern calling activity from 90 to 270 minutes after sunrise. This peak may be related to colony attendance, and not all the traditional counts were conducted during the peaks in acoustic activity, so including a time of day of the counts as a covariate may be important for improving model fit.

Phenological Patterns

Acoustic surveys can produce an extremely rich dataset for exploring questions about colony attendance and breeding phenology at survey sites. Figure 6 shows some of the patterns available for exploration. For example, many sites showed clear differences in the time of day when Aleutian Terns called (e.g. Foul Bay 9 vs. Black Sand Spit sites). At some sites, vocal activity shifted within the season. For example, Italio, had activity throughout the day early in the season, but only high activity during the morning towards the end of the season. Similar pulses of activity late in the season were documented at Black Sand Spit, Lost River, Akhiok, and Sheep. Could this be indicative of staging/flocking behavior associated with post-breeding?

Alternatively, could the early season periods of all-day vocal activity (spanning most day-light hours) be related to courtship activity? This pattern was visible at many sites for some part of the summer as exemplified by Akhiok. Finally, are truncated (<3 week) periods of all-day vocal activity indicative of failed breeding attempts? For example, might the extended duration of activity at Foul Bay 9 or Naknek2 be indicative of asynchronous courtship behavior leading in to a successful breeding season, while the shorter pulses at Black Sand Spit, indicate more synchronous courtship or a failed breeding attempt? At the moment these are speculative questions, but additional data from camera traps, drones, and traditional surveys could aid with interpretation of these acoustic patterns and could provide a useful method for monitoring Aleutian Tern breeding phenology at scale.

Considerations for future surveys

The acoustic survey results from the pilot season (2018) were encouraging. However, several questions remain about how best to design large scale acoustic surveys for the species. Within breeding aggregations, Aleutian Tern nests appears to be patchily distributed and in low densities. However, communication behavior of Aleutian Terns and individuals vocalizing over the colony appear to allow acoustic sensors to gather meaningful information about tern abundance at breeding aggregations. The question remains, what is the spatial scale of an acoustic sensor for this species, and how can that be used to plan sampling strategies for Aleutian Terns, especially when surveys will need to be carried out before the initiation of the breeding season, and in areas with no prior knowledge of nest sites. In the absence of an observed correlation between nest densities and call rates at larger spatial scales, the best strategy may be the deployment of replicate sensors in potential breeding habitat at fixed distances. But some assumptions will need to be made about what that distance should be, and how the sensors should be spatially distributed. It would be useful to try and quantify the total available breeding habitat at each of the breeding aggregations monitored in 2018 to explore this further. Finally, if the best acoustic index of abundance is tied to total birds, we will need to estimate the ratio of breeding birds to non-breeding birds contributing to the acoustic call rate in order to estimate breeding abundance at survey sites. While guidance on some of these questions can be gleaned from the literature, the Aleutian Tern Technical Committee will need to focus some thought and resources to develop a robust framework and sampling design for statewide survey efforts.

Data Issues from the 2018 survey

Several changes could improve future survey results. Below is a list of some of the challenges we encountered while analyzing these data.

- Some sensors were programed with the wrong time. Aleutian Tern18 was programed to Eastern Time. It is important to check the date and time on every visit to the acoustic sensor.
- Some sensors were programed with the wrong recording schedule. The sensor at Womens Bay was initially programed to record hour long files, 24 hours per day. Additionally, the rest of units were set to a 1 minute every 5- or 6-minute schedule. This was due to a programing mistake that Conservation Metrics made when creating the Song Meter program for the SM4s. We have corrected the schedule and for the 2019 season all units should be deployed on a 1 minute every 5 minutes 24 hours a day schedule.
- Count and density data were not standardized in the data entry spreadsheets. For instance, if the flock composition was estimated (80% Aleutian Tern, 20% Arctic Tern) with a total number of terns, make a note that it was estimated with the composition, and also calculate the number for each species. This season the counts were left blank.
- Two sensors were programed with the same prefix. This is the most difficult issue for us to deal with at Conservation Metrics. Our data pipeline is built around each sensor having a unique name and never having two sensors in the field at the same time with the same prefix. The prefix ALTE8 was given to two sensors, one a Kalsin and one at Italio. This was not caught

when we ingested the data and files became mixed. We were able to recover most of the data (but there were some files that were overwritten and lost. If data from both ALTE8 sensors could be sent with data from the 2019 survey season, we can recover the missing files. Please avoid naming sensors with the same prefix in the future (even if they are SM2 vs SM4).

Literature Cited

Acevedo, M. A., & Villanueva-Rivera, L. J. (2006). Using Automated Digital Recording Systems as Effective Tools for the Monitoring of Birds and Amphibians. *Wildlife Society Bulletin*, *34*(1), 211–214. https://doi.org/10.2193/0091-7648(2006)34[211:uadrsa]2.0.co;2

Agranat, I. (2007). *Automatic detection of cerulean warblers using autonomous recording units and song scope bioacoustics software*. Wildlife Acoustics, Inc.

Alaska Department of Fish and Game. (2015). *2015 Alaska wildlife action plan.* Juneau, Alska.

Borker, A. L., McKown, M. W., Ackerman, J. T., Eagles-Smith, C. a, Tershy, B. R., & Croll, D. a. (2014). Vocal activity as a low cost and scalable index of seabird colony size. *Conservation Biology : The Journal of the Society for Conservation Biology*, *00*(4), 1–9. https://doi.org/10.1111/cobi.12264

Brandes, T. S. (2008a). Automated sound recording and analysistechniques for bird surveys and conservation. *Bird Conservation International*, *18*, S163–S173.

Brandes, T. S. (2008b). Feature vector selection and use with hidden Markov models to identify frequency-modulated bioacoustic signals amidst noise. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(6), 1173–1180.

Buxton, R. T., & Jones, I. L. (2012). Measuring nocturnal seabird activity and status using acoustic recording devices: applications for island restoration. *Journal of Field Ornithology*, *83*(1), 47–60.

Buxton, R. T., Major, H. L., Jones, I. L., & Williams, J. C. (2013). Examining patterns in nocturnal seabird activity and recovery across the Western Aleutian Islands, Alaska, using automated acoustic recording. *The Auk*, *130*(2), 331–341. https://doi.org/10.1525/auk.2013.12134

Cichy, R. M., Khosla, A., Pantazis, D., & Torralba, A. (2016). Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition. *Arxiv*, (January), 1–35. https://doi.org/10.1038/srep27755

Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: an overview. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 8599–8603). Vancouver, BC: IEEE. https://doi.org/10.1109/ICASSP.2013.6639344

MacKenzie, D., Nichols, J., & Lachman, G. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255.

MacKenzie, D., Nichols, J., & Sutton, N. (2005). Improving inferences in population studies of rare species that are detected imperfectly. *Ecology*, *86*(5), 1101–1113.

McKown, M. W. (2008). *Acoustic communication in colonial seabirds: individual, sexual, and species-specific variation in acoustic signals of Pterodroma petrels*. *Curriculum in Ecology*. University of North Carolina, Chapel Hill.

Min, S., Lee, B., & Yoon, S. (2016). Deep learning in bioinformatics. *Briefings in Bioinformatics*, (March), 1–19. https://doi.org/10.1093/bib/bbw068

Oppel, S., Hervías, S., Oliveira, N., Pipa, T. T., Silva, C., Geraldes, P., … Hervias, S. (2014). Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. *Nature Conservation*, *7*, 1–13. https://doi.org/10.3897/natureconservation.7.6890

Orben, R. A., Fleishman, A. B., Borker, A. L., Bridgeland, W., Gladics, A. J., Porquez, J., … Suryan, R. M. (2019). Comparing imaging, acoustics, and radar to monitor Leach's storm-petrel colonies. *PeerJ*, *7*, e6721. https://doi.org/10.7717/peerj.6721

Pyare, S., Goldstein, M. I., Duffy, D., Oehlers, S., Catterson, N., & Frederick, J. (2013). Aleutian Tern (Onychoprion aleuticus) research in Alaska: survey methodology, migration, and statewide coordination. *Final Report to the Alaska Department of Fish and Game. Juneau, AK: Alaska Department of Fish and Game*.

Renner, H. M., Romano, M. D., Renner, M., Pyare, S., Goldstein, M. I., & Artukhin, Y. (2015). Assessing the breeding distribution and population trends of the Aleutian Tern Onychoprion aleuticus. *Marine Ornithology*, *43*(2), 179–187.

Sauer, J. R., Peterjohn, B. G., & Link, W. A. (1994). Observer differences in the North American breeding bird survey. *The Auk*, 50–62.

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Appendix A: Colony Specific Call Rate Maps

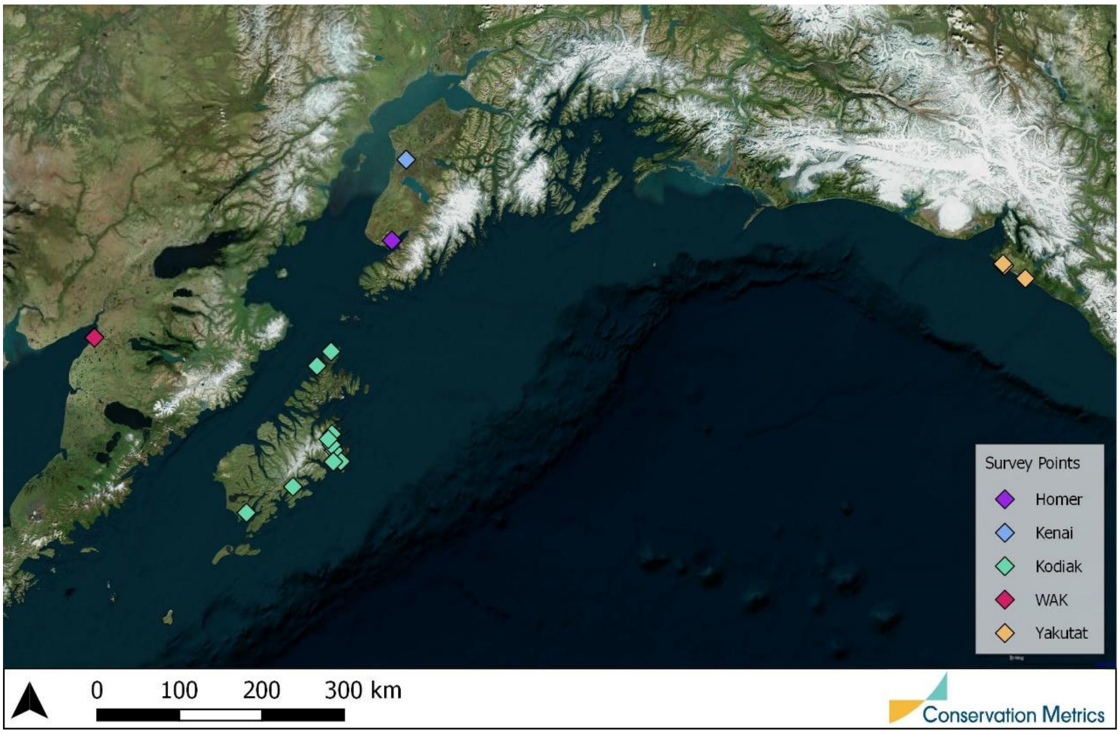


Figure 15: Overview site map showing the survey points colored by region.

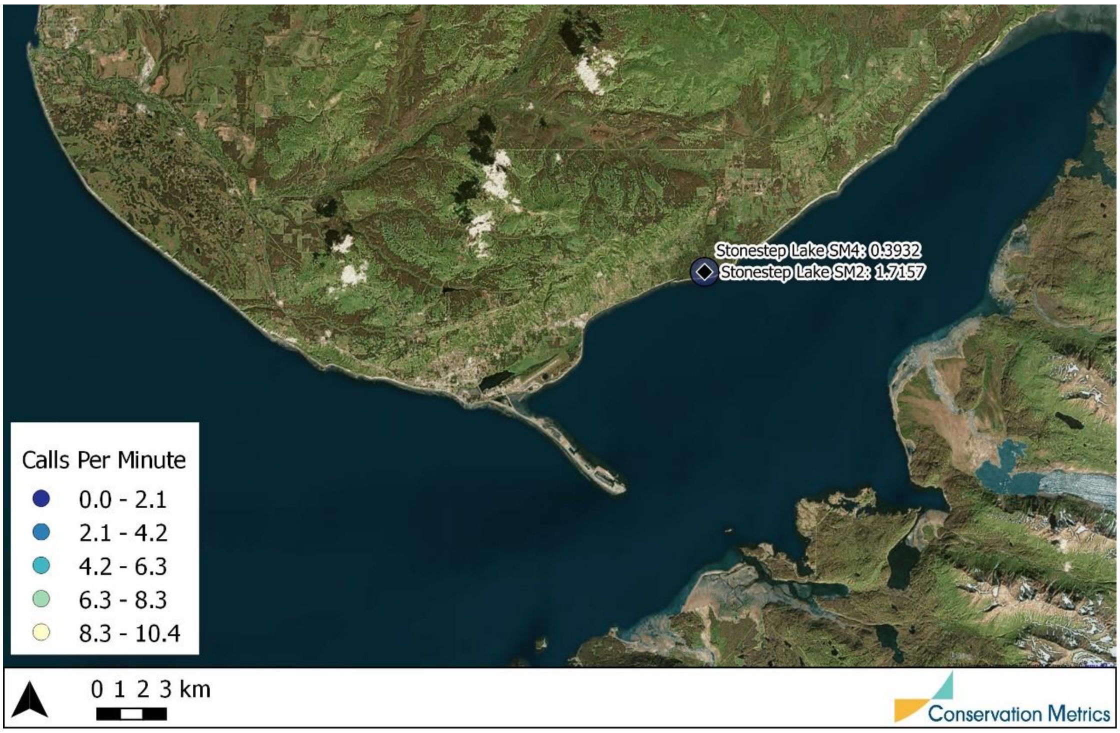


Figure 16. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at Stone Step Lake.

Figure 17. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at Kenai.

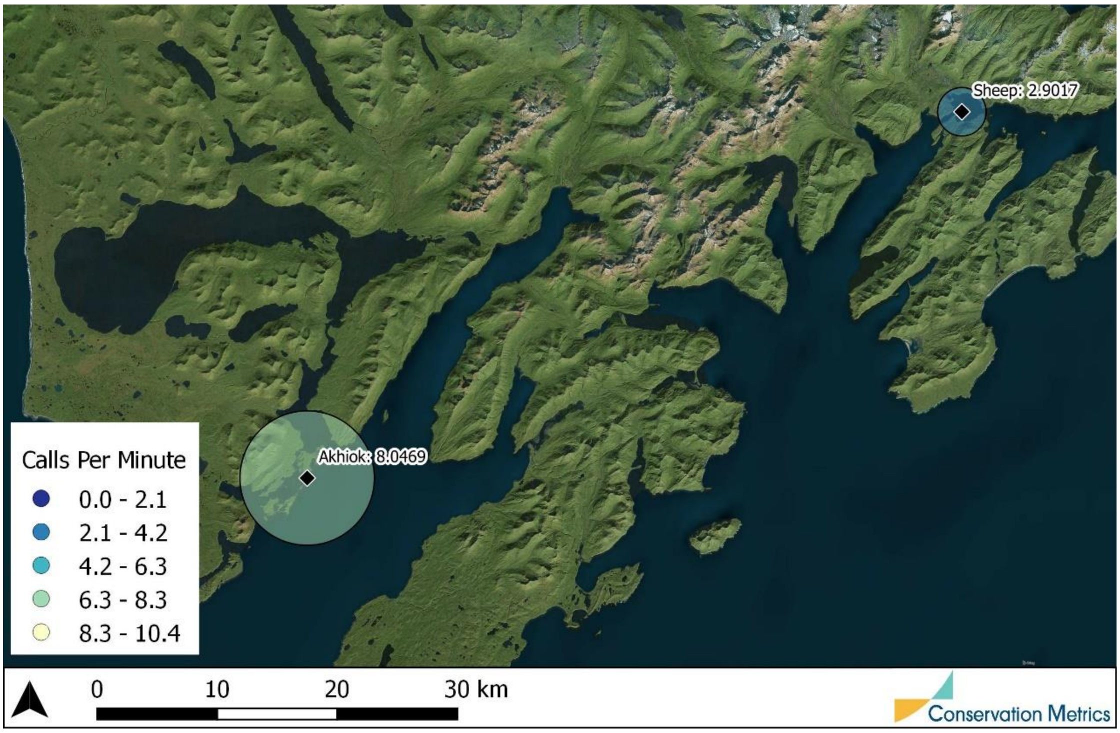


Figure 18. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at Sheep and Akhiok.

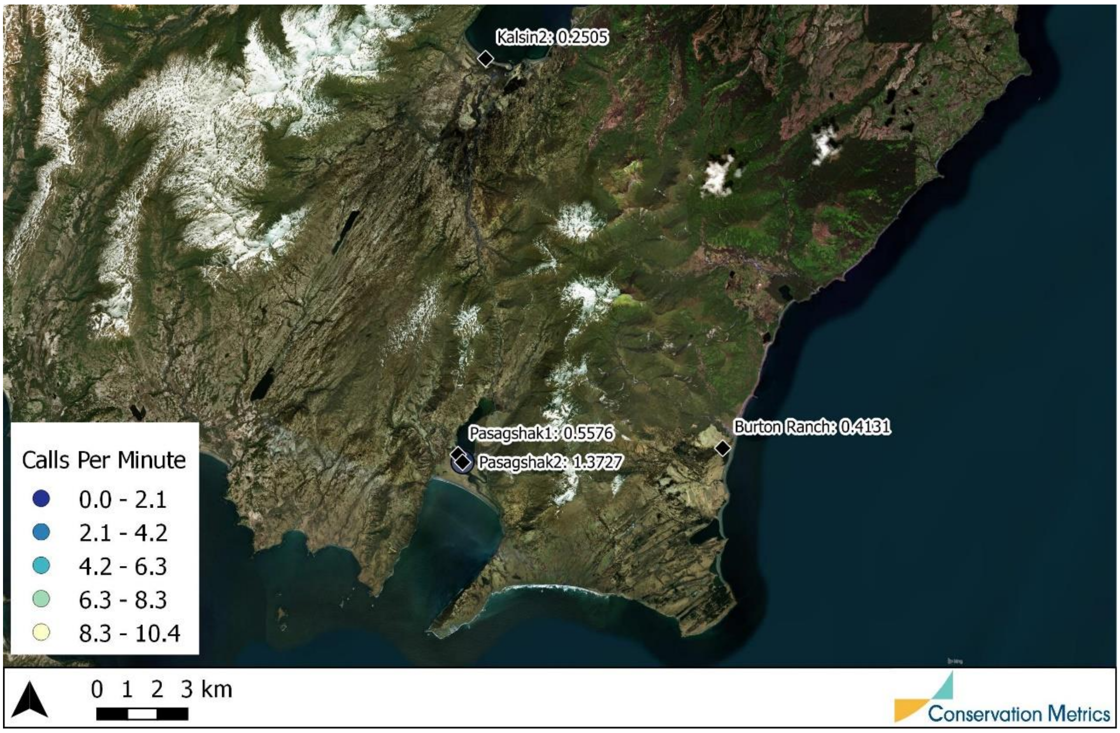


Figure 19.Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at Kalsin, Pasagshak, and Burton Ranch.

Figure 20. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at Middle Bay.

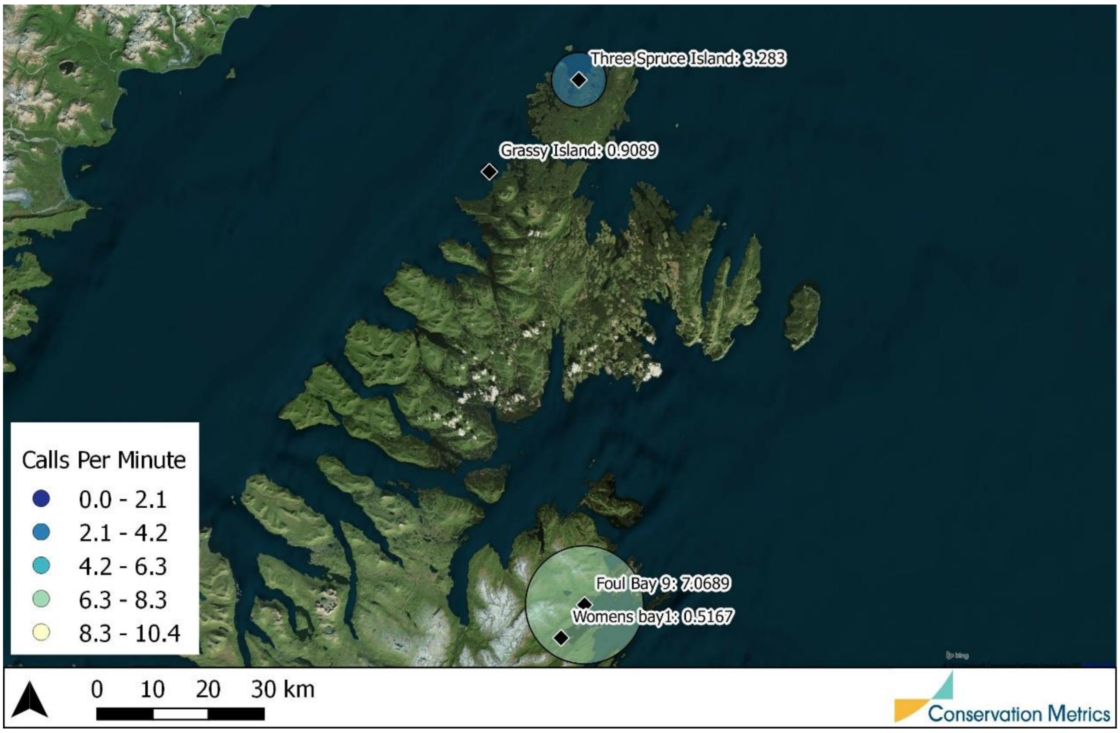


Figure 21. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at North Kodiak colonies: Grassy Island, Three Spruce Island, Foul Bay, and Womens Bay.

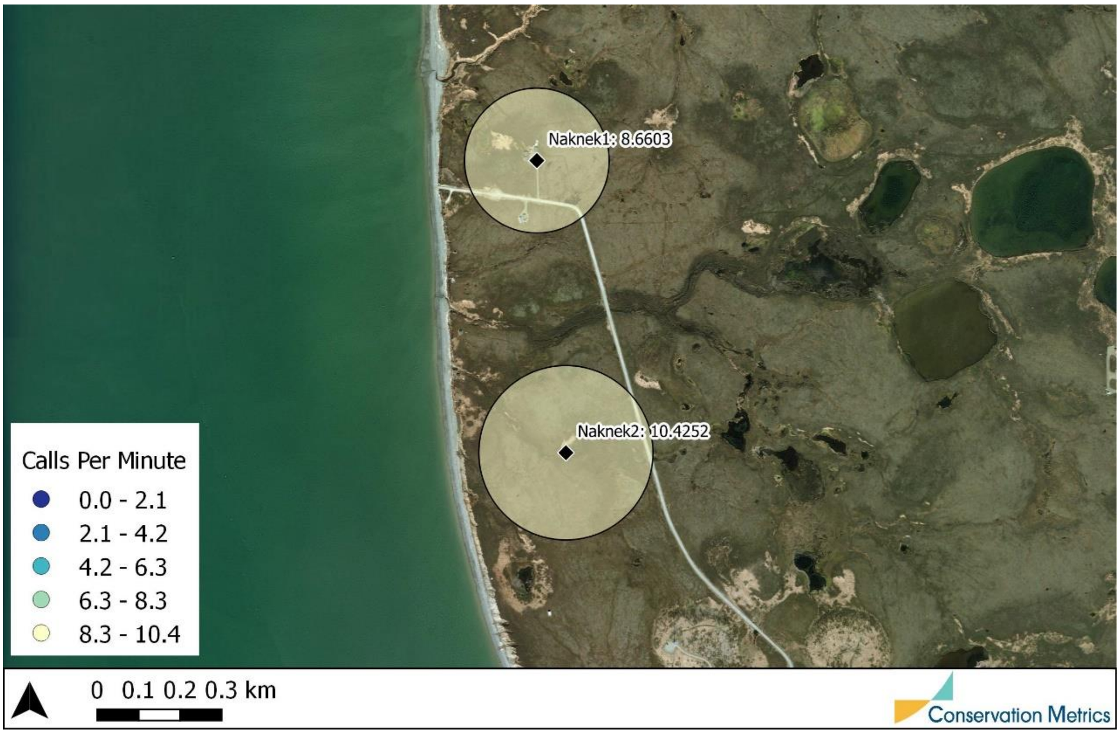
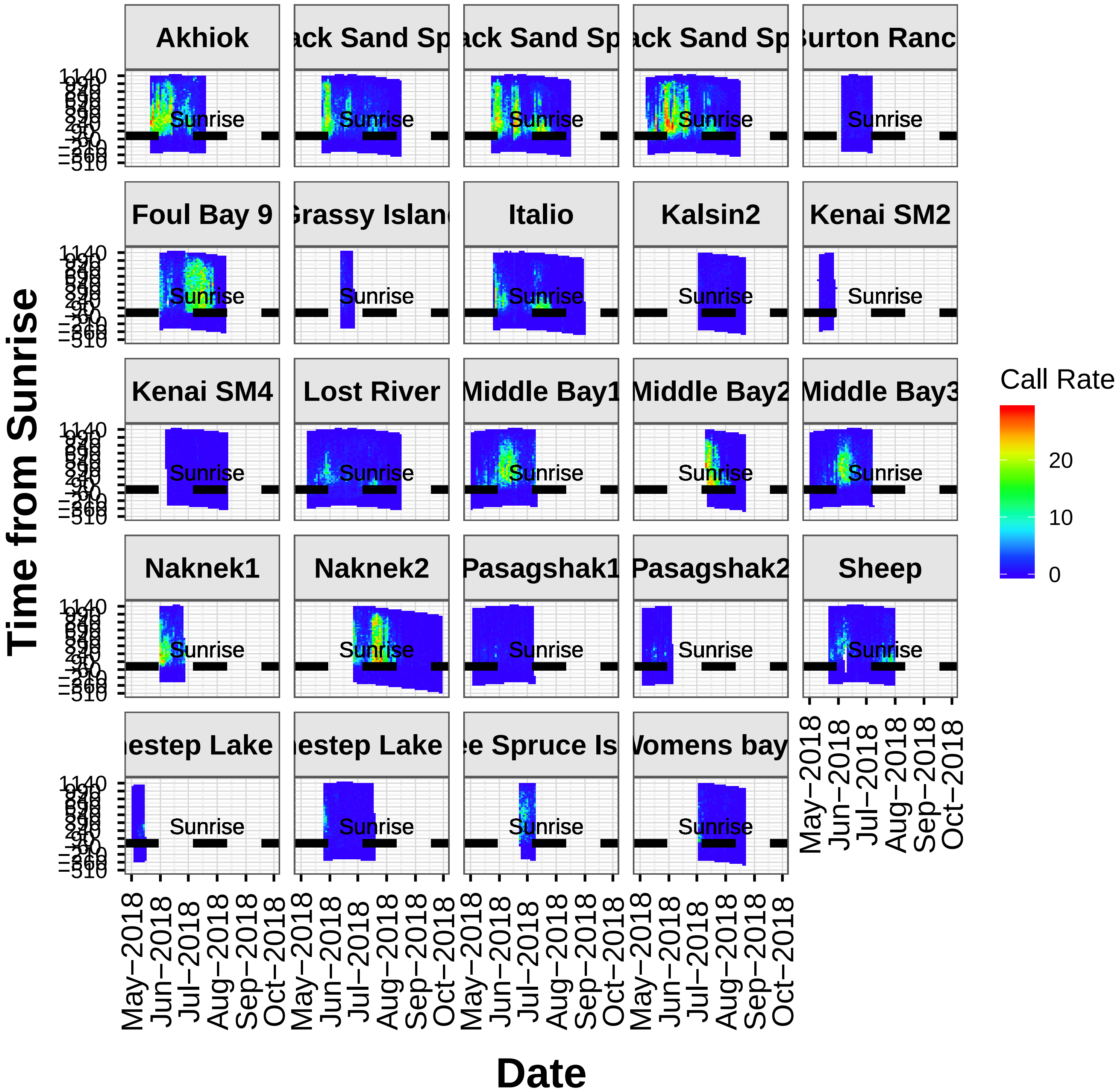


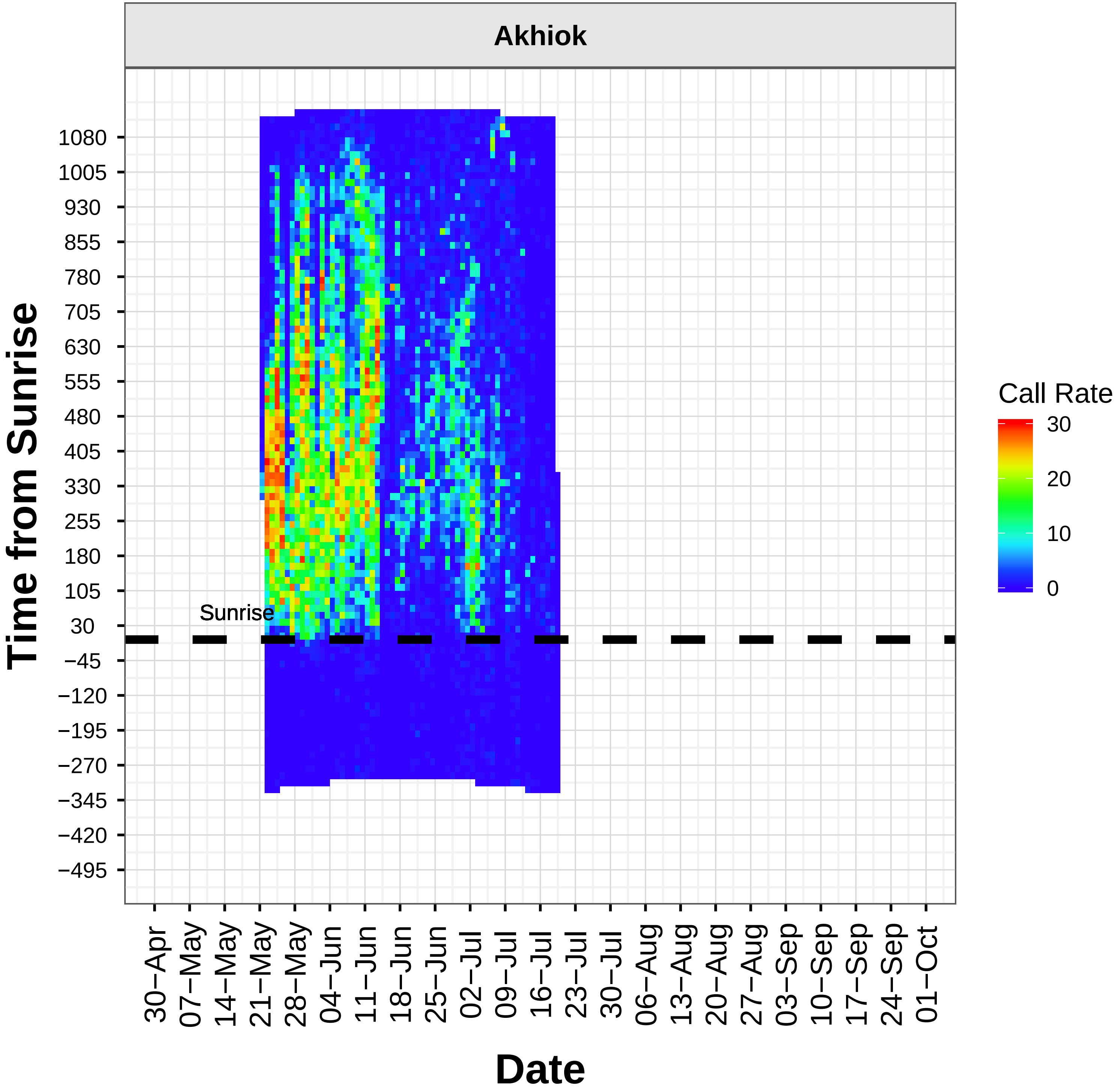
Figure 22. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at Naknek.

Figure 23. Aleutian Tern call rate estimates during peak calling hour (90-270 minutes after sunrise) and acoustic comparison period (15 May - 15 August) at Lost River, Black Sand Spit, and Italio.

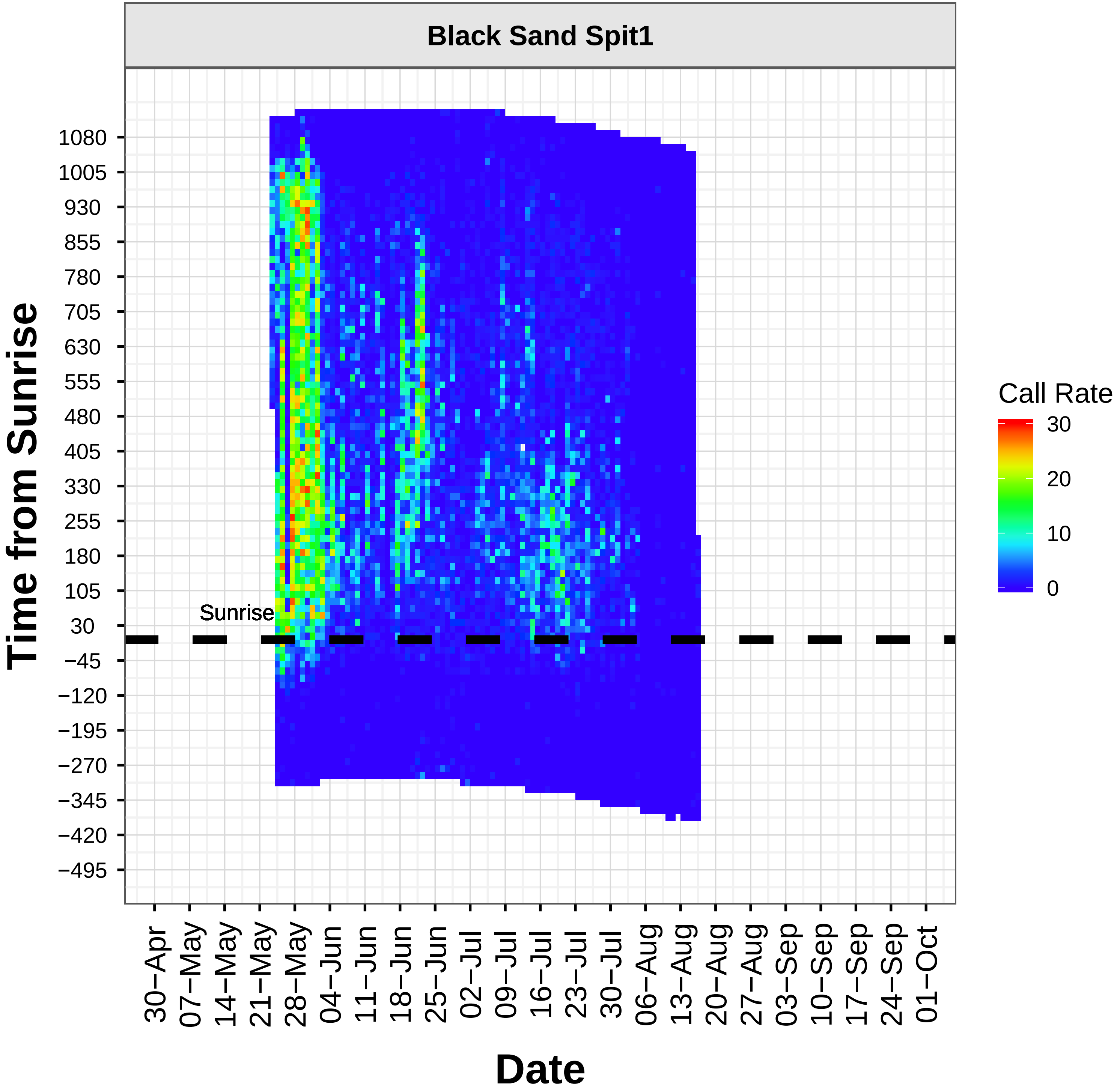
Appendix B


Akhiok
ack Sand Sp
ack Sand Sp
ack Sand Sp
urton Ranc
Foul Bay 9
Grassy Islan
Italio
Kalsin2
Kenai SM2
Kenai SM4
Lost River
Middle Bay1
Middle Bay2
Middle Bay3
Naknek1
Naknek2
Pasagshak1
Pasagshak2
Sheep
estep Lake
estep Lake
e Spruce Is
Womens bay
Sunrise
Time from Sunrise
Date
May−2018
Jun−2018
Jul−2018
Aug−2018
Sep−2018
Oct−2018
Call Rate
20
10
0

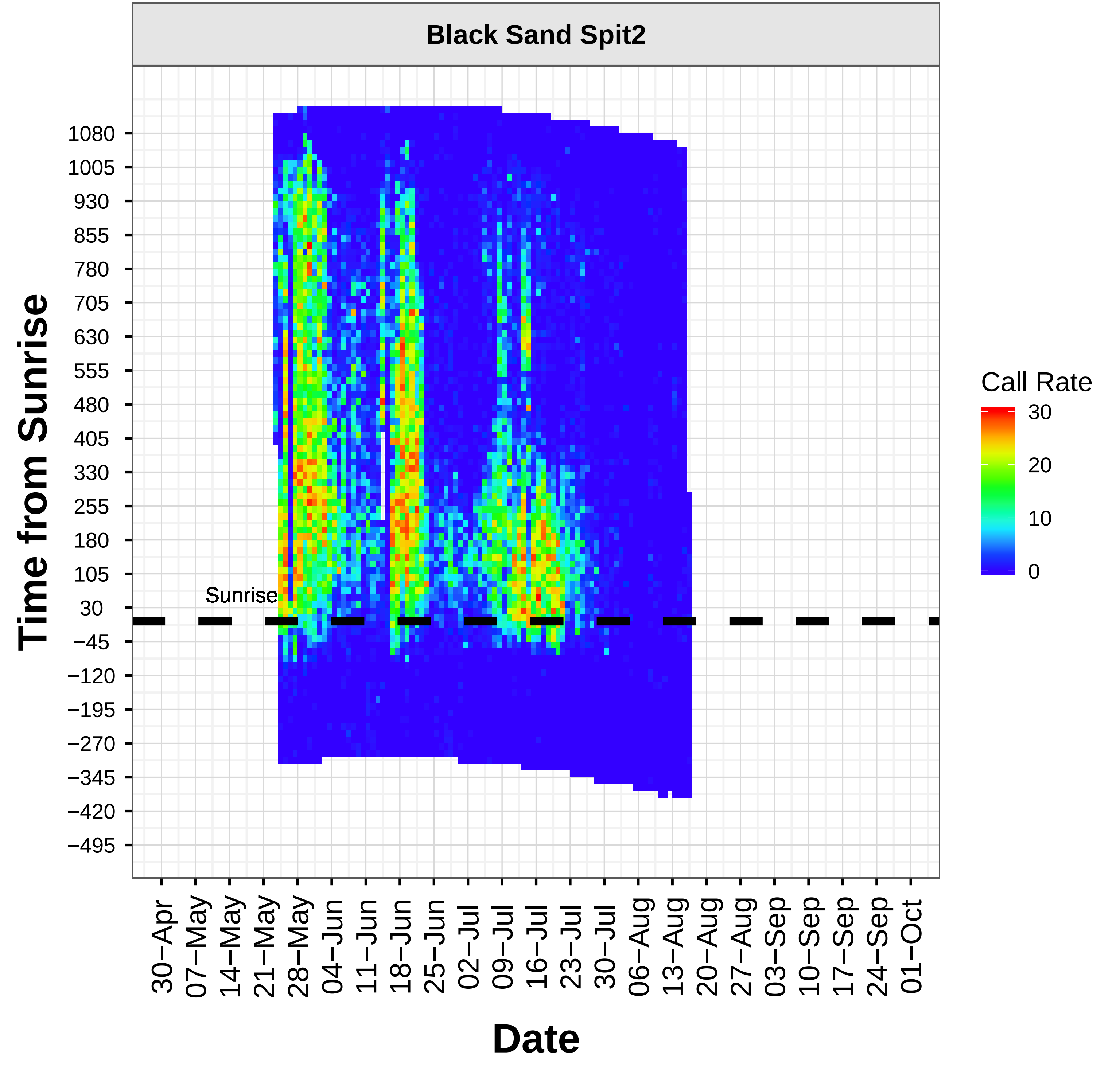
Appendix B


Akhiok
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B

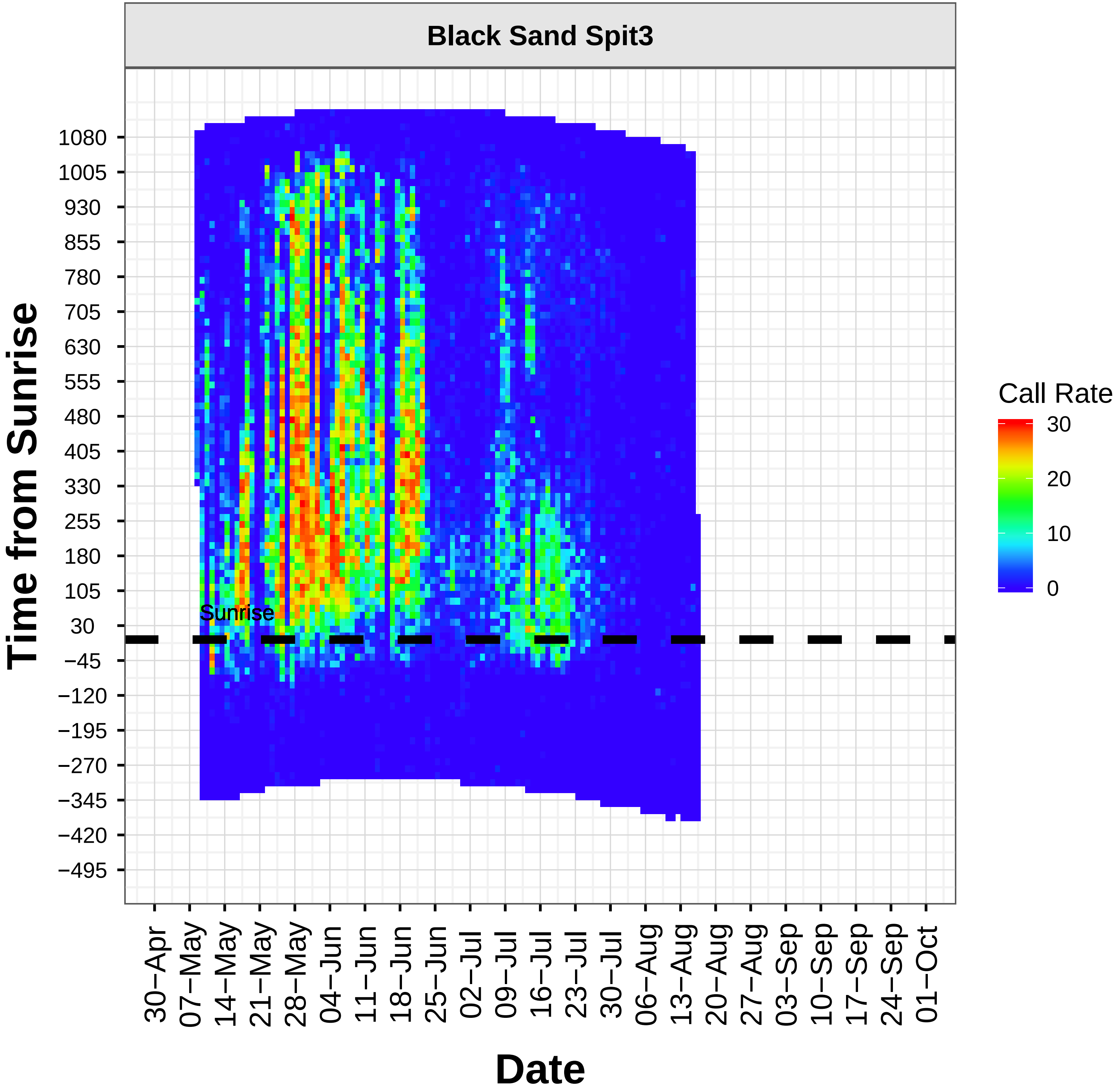

Black Sand Spit1
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B



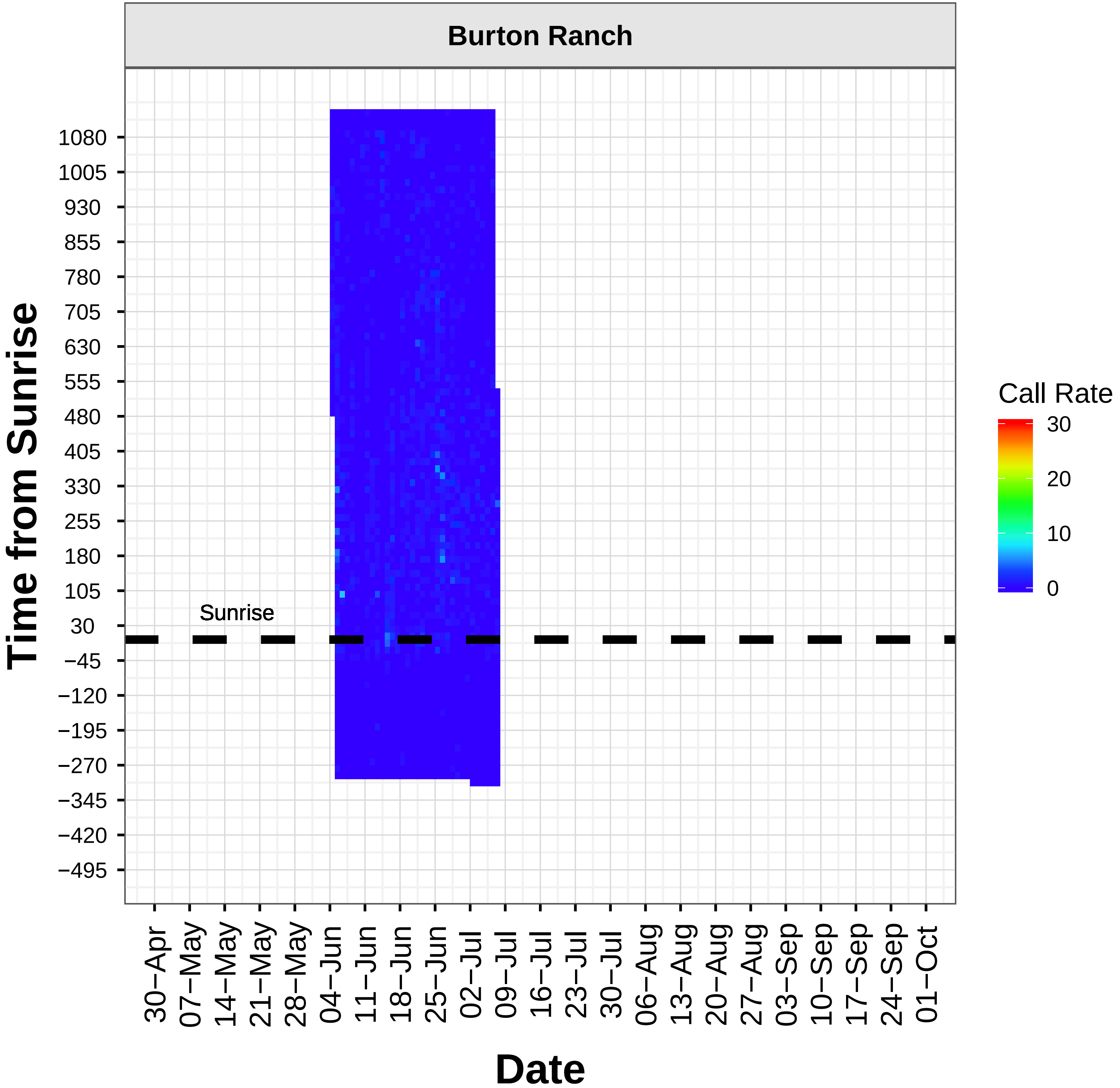
Black Sand Spit2
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B



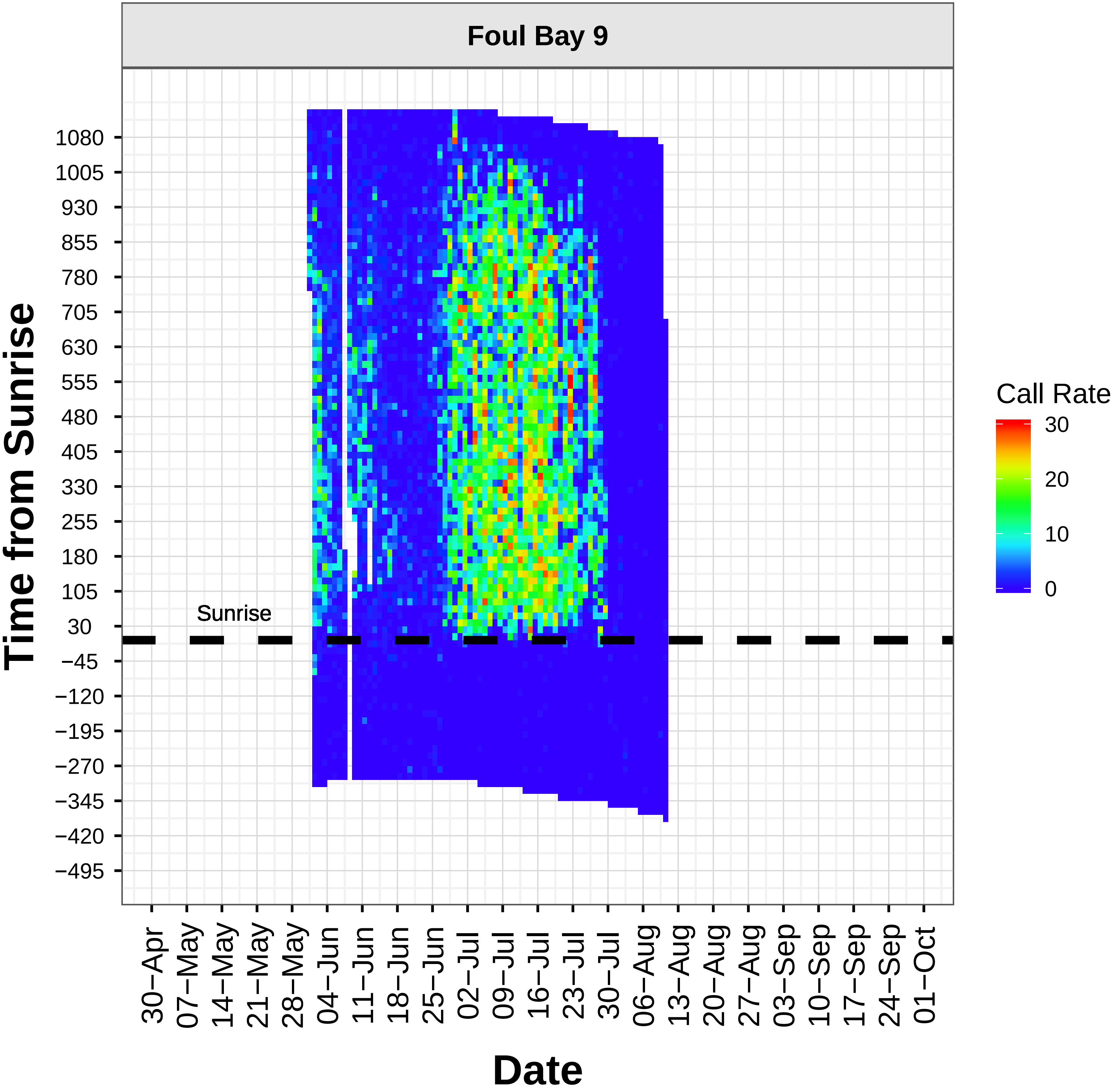
Black Sand Spit3
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B



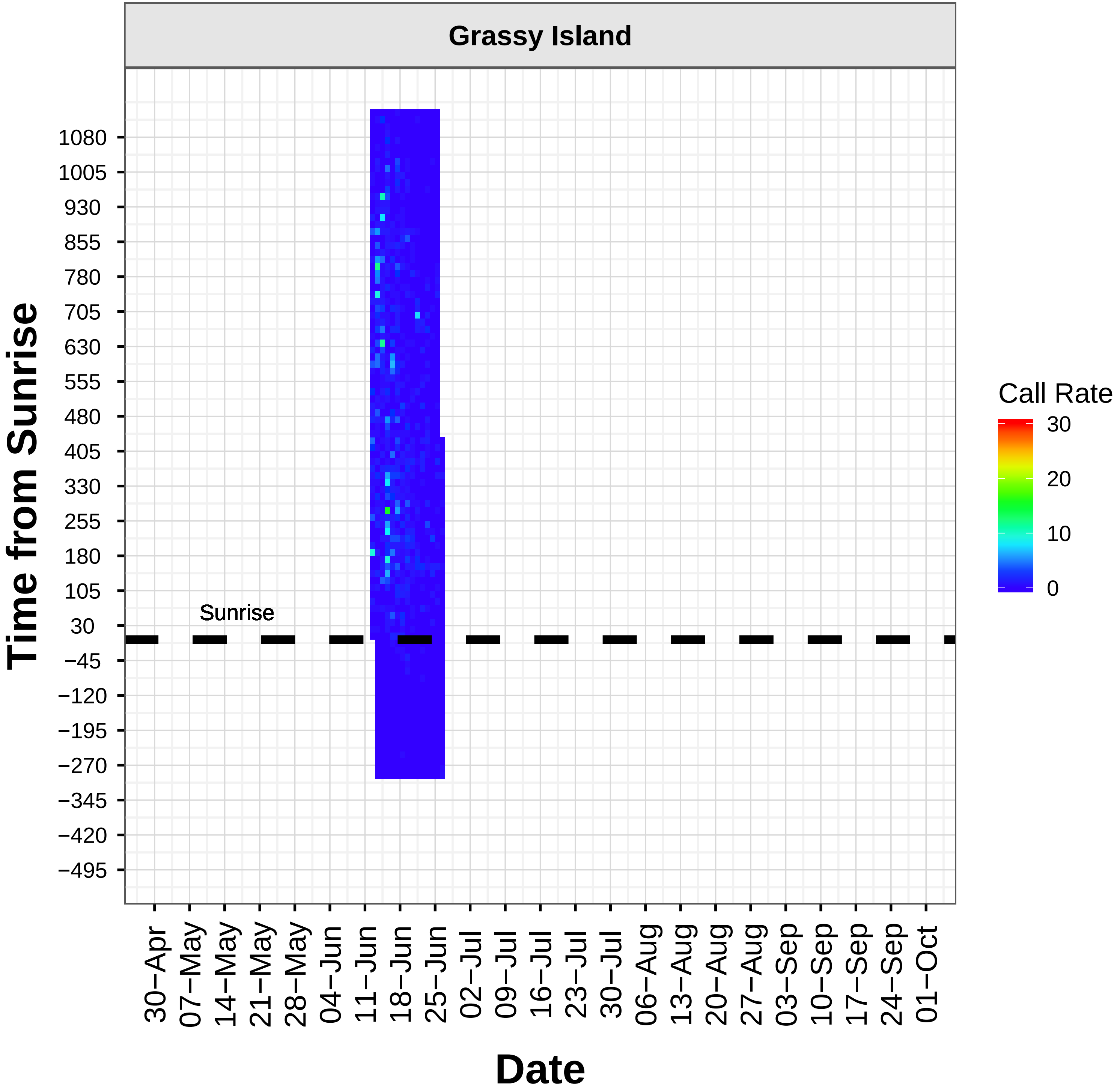
Burton Ranch
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B

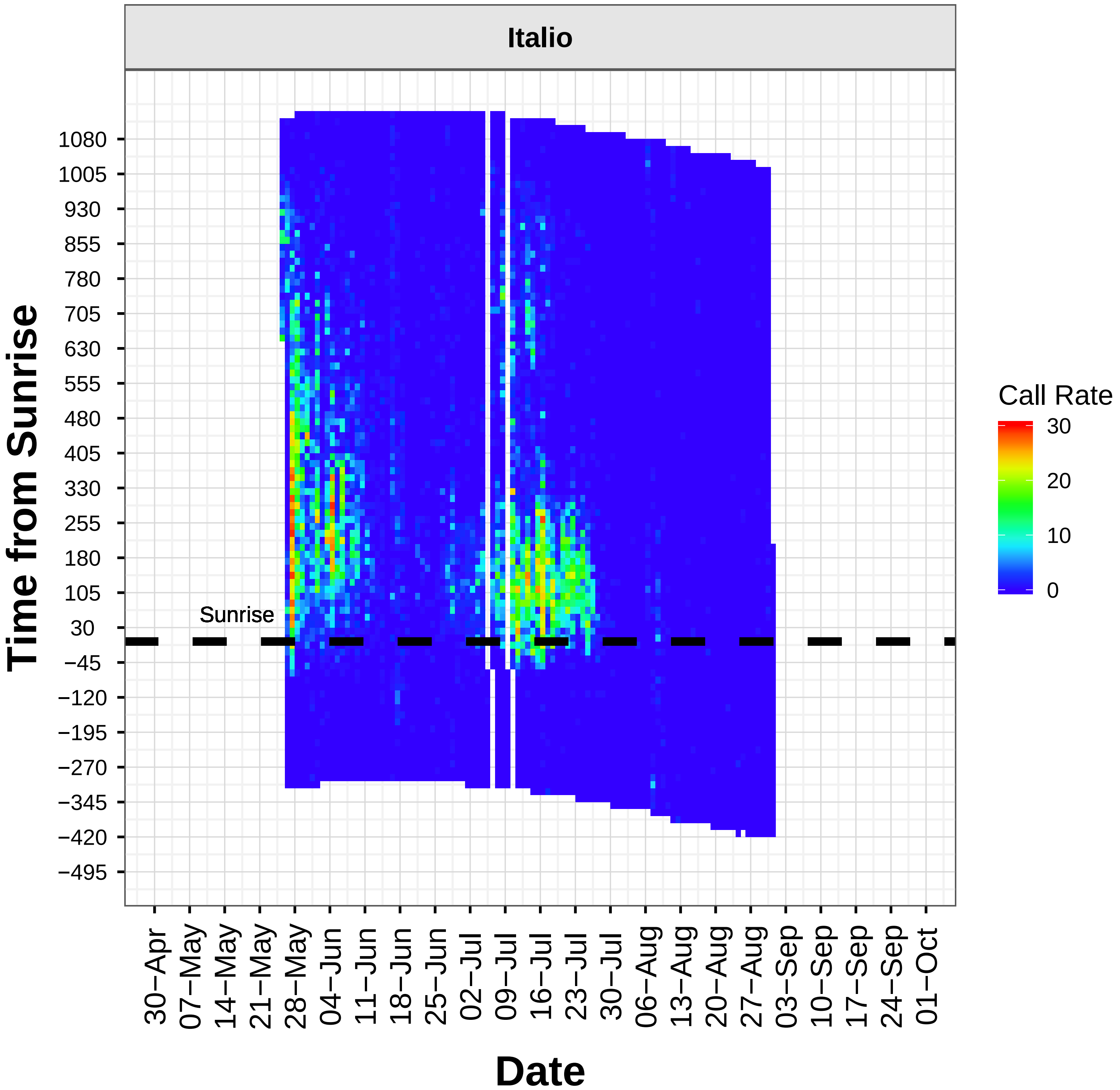


Foul Bay 9
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

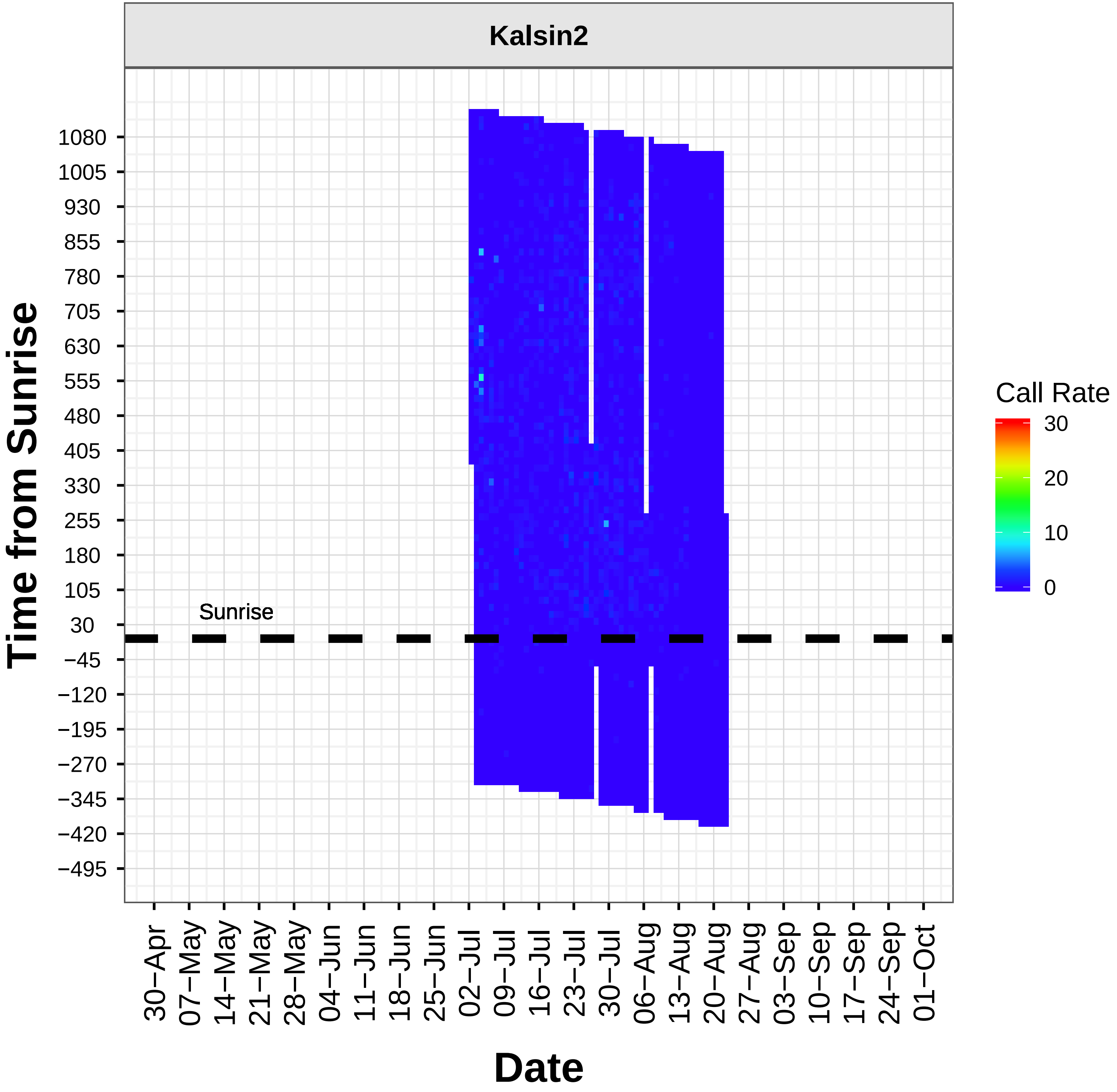
Appendix B


Grassy Island
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

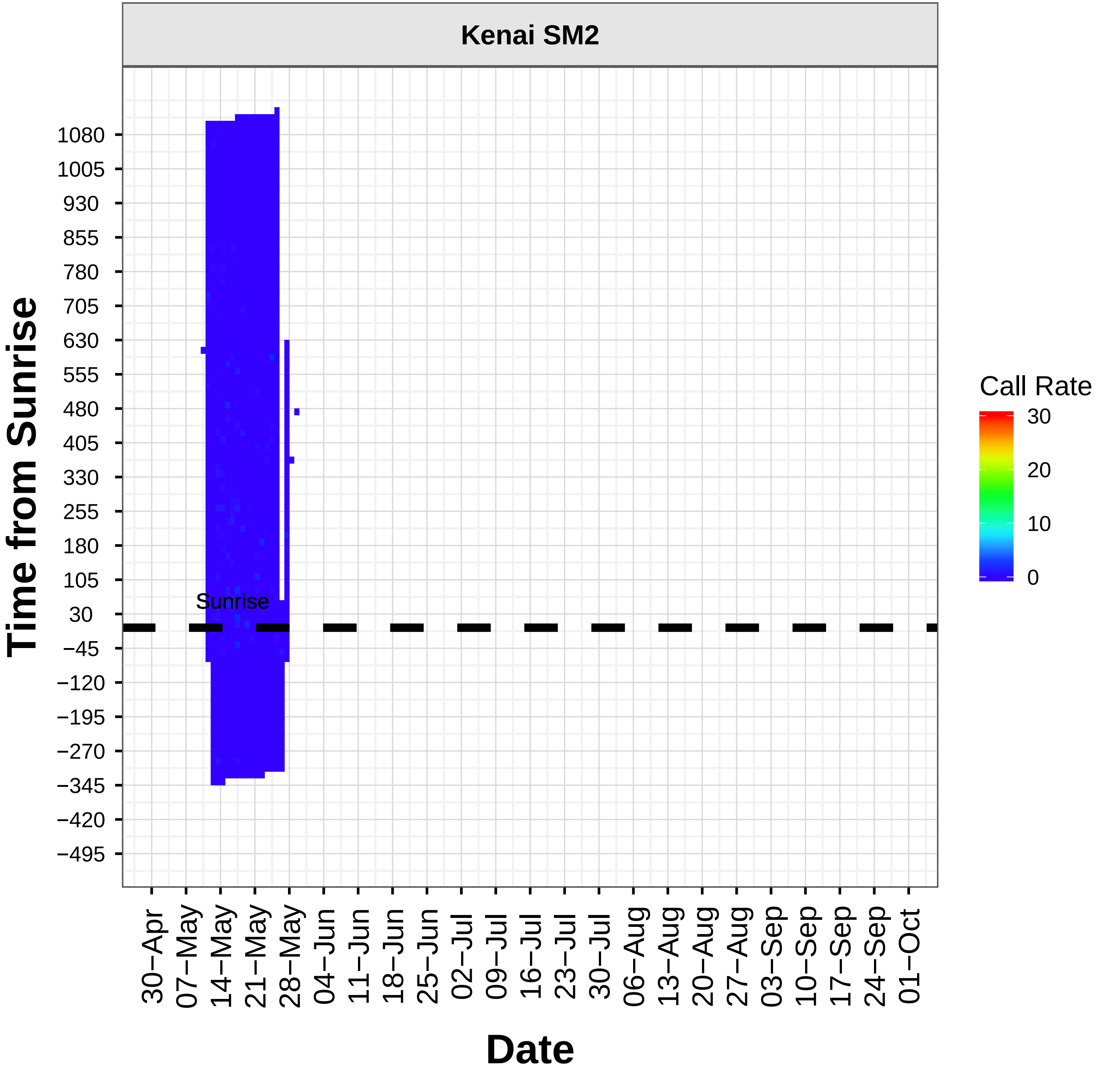
Appendix B


Italio
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B

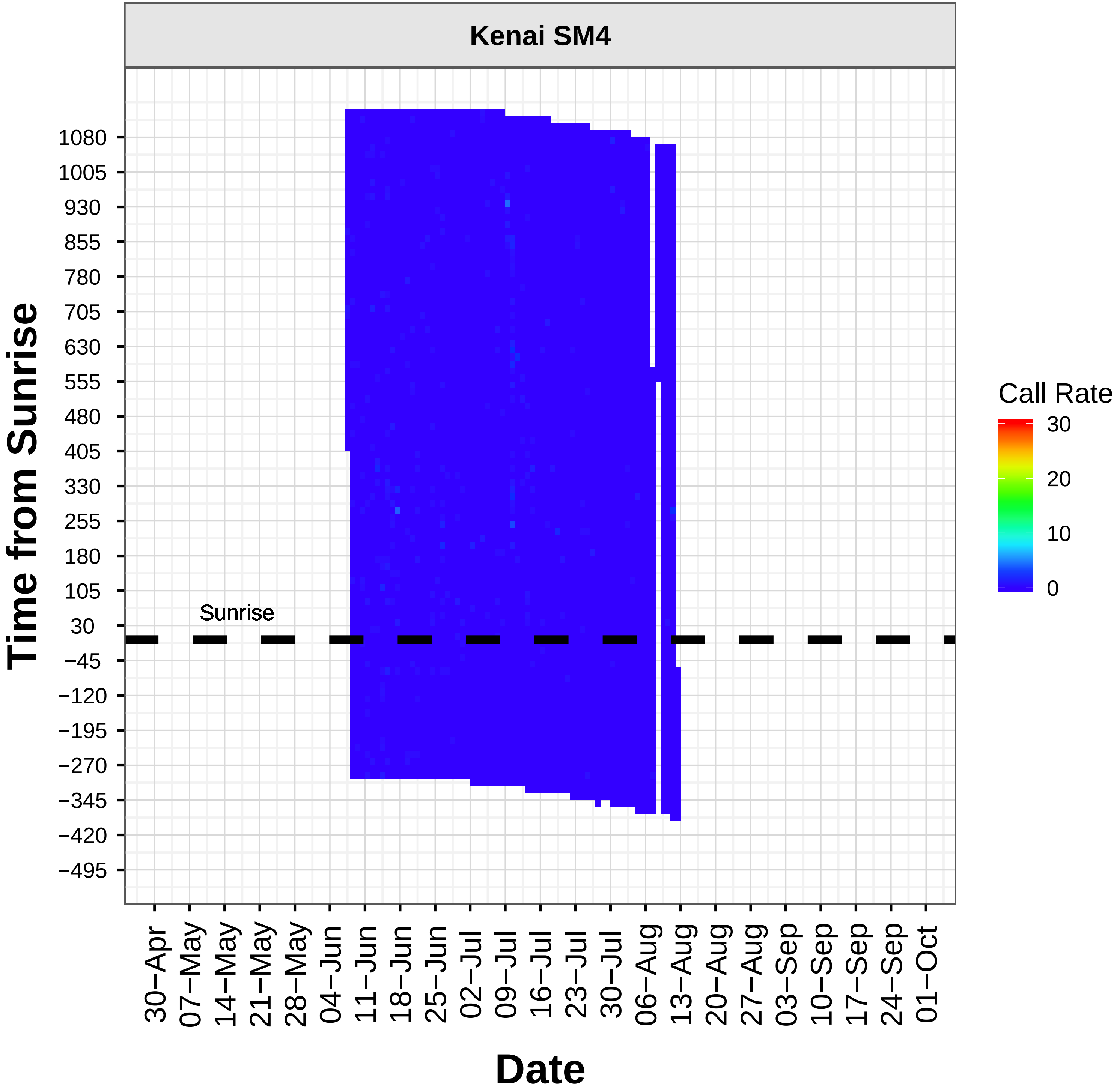

Kalsin2
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B



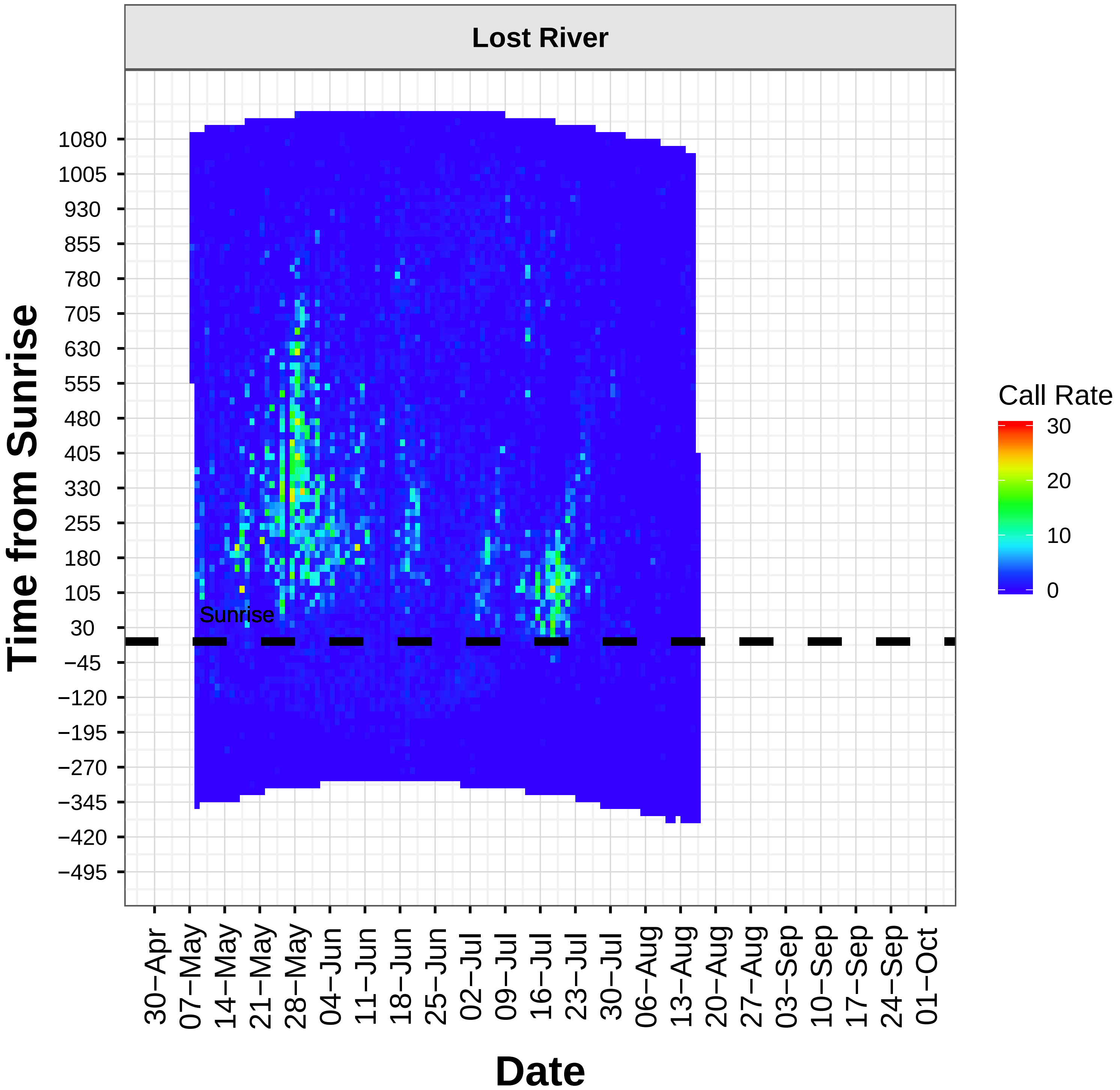
Kenai SM2
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B

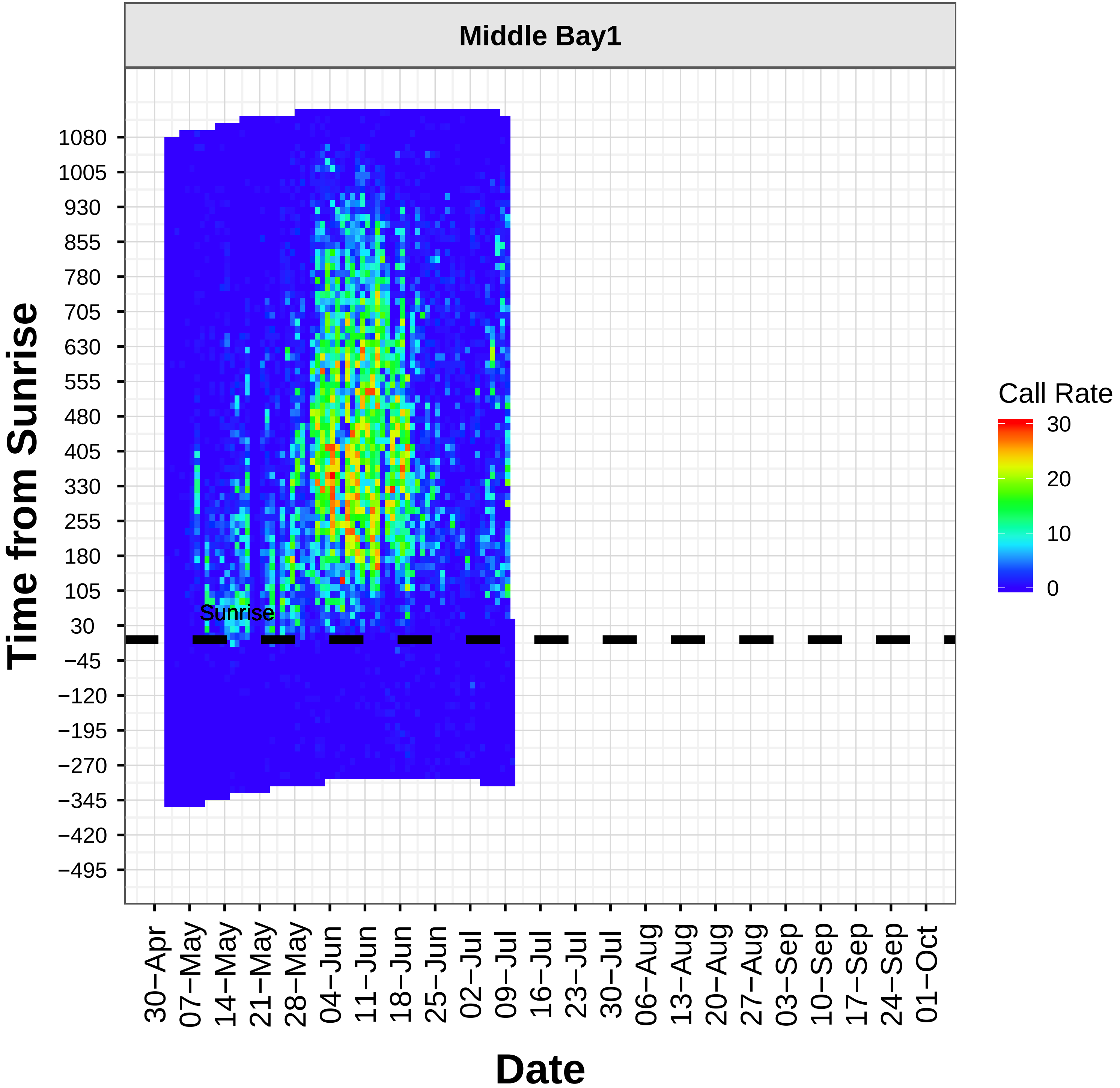


Kenai SM4
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B

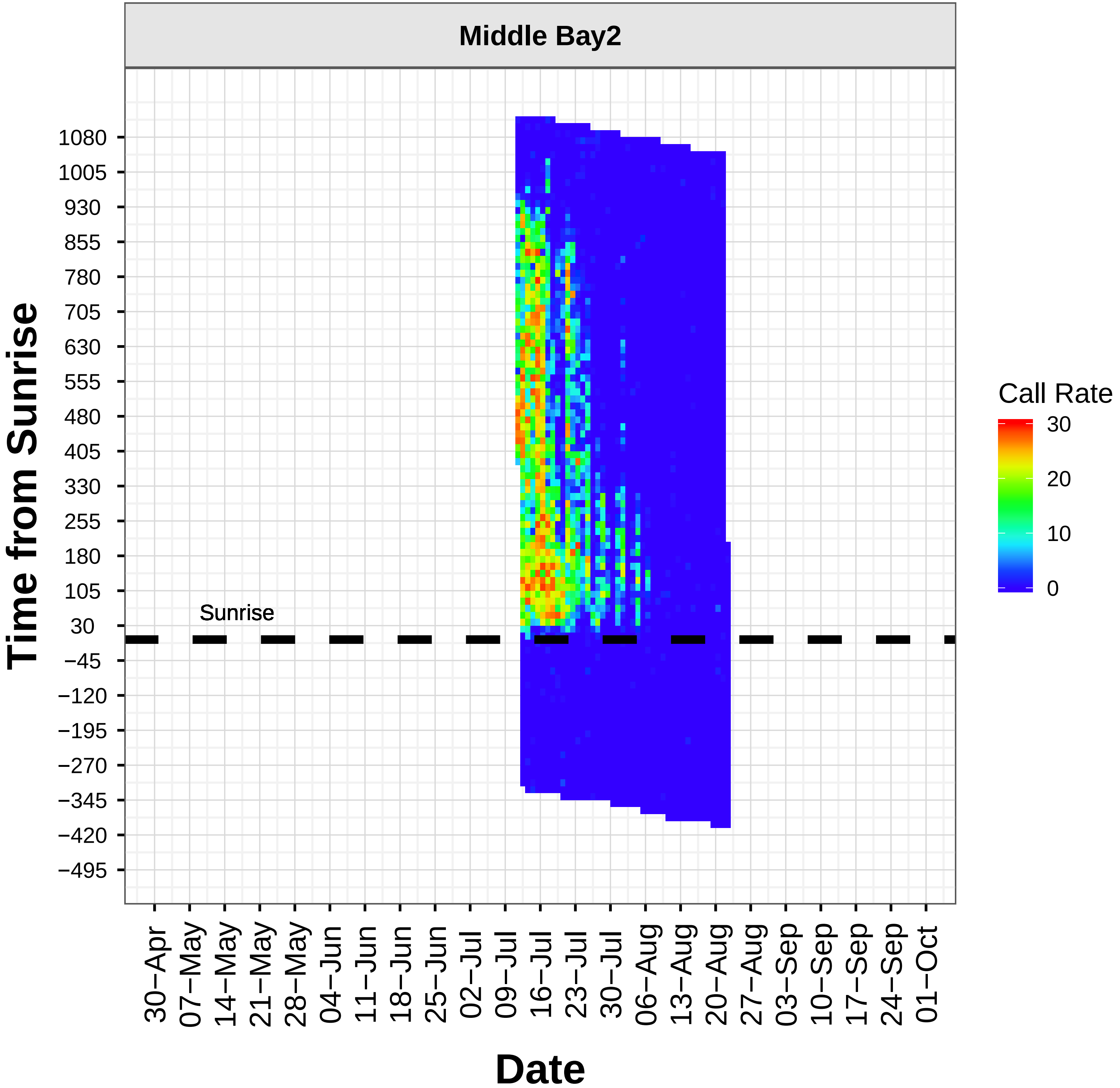

Lost River
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B



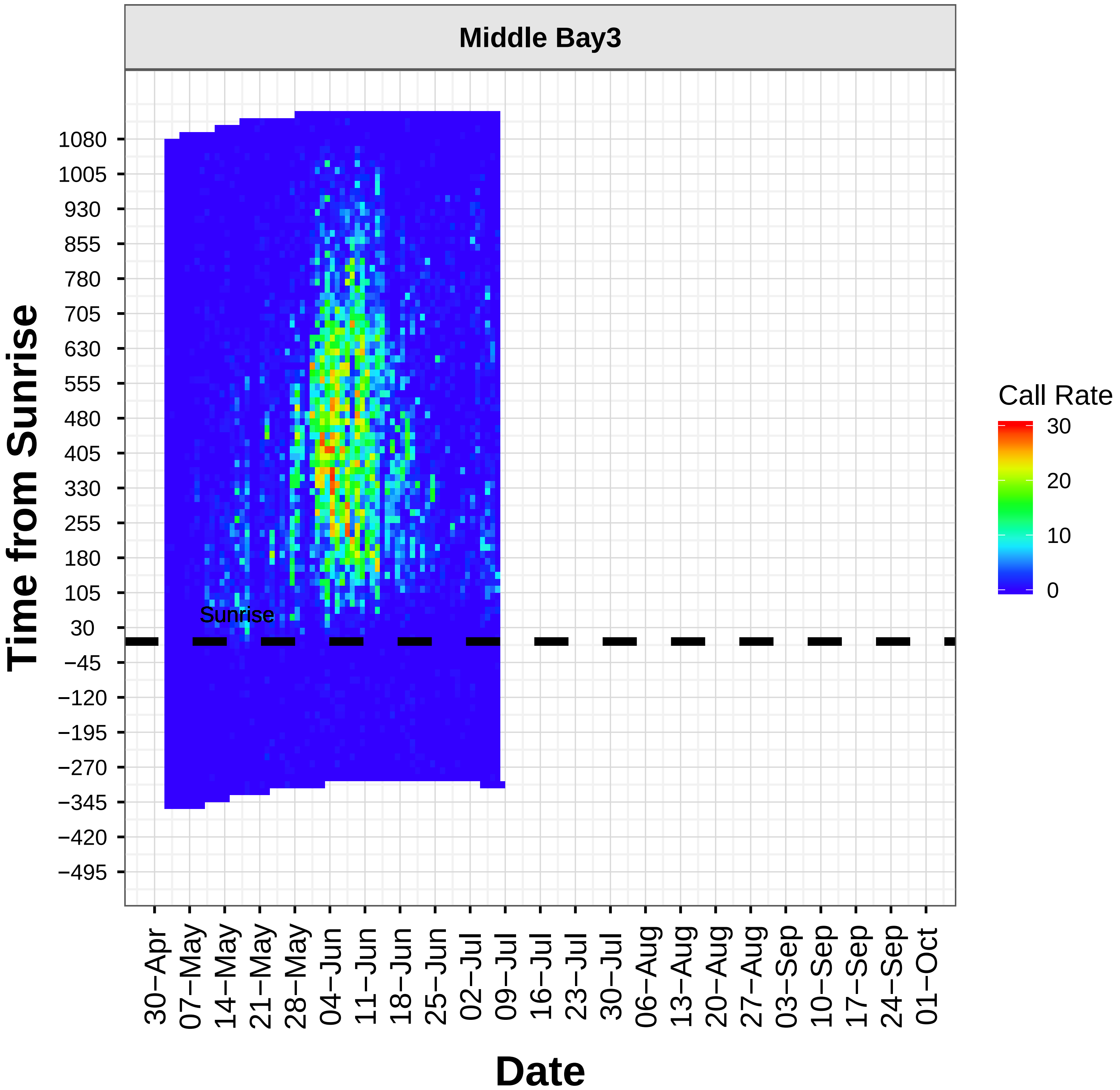
Middle Bay1
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B



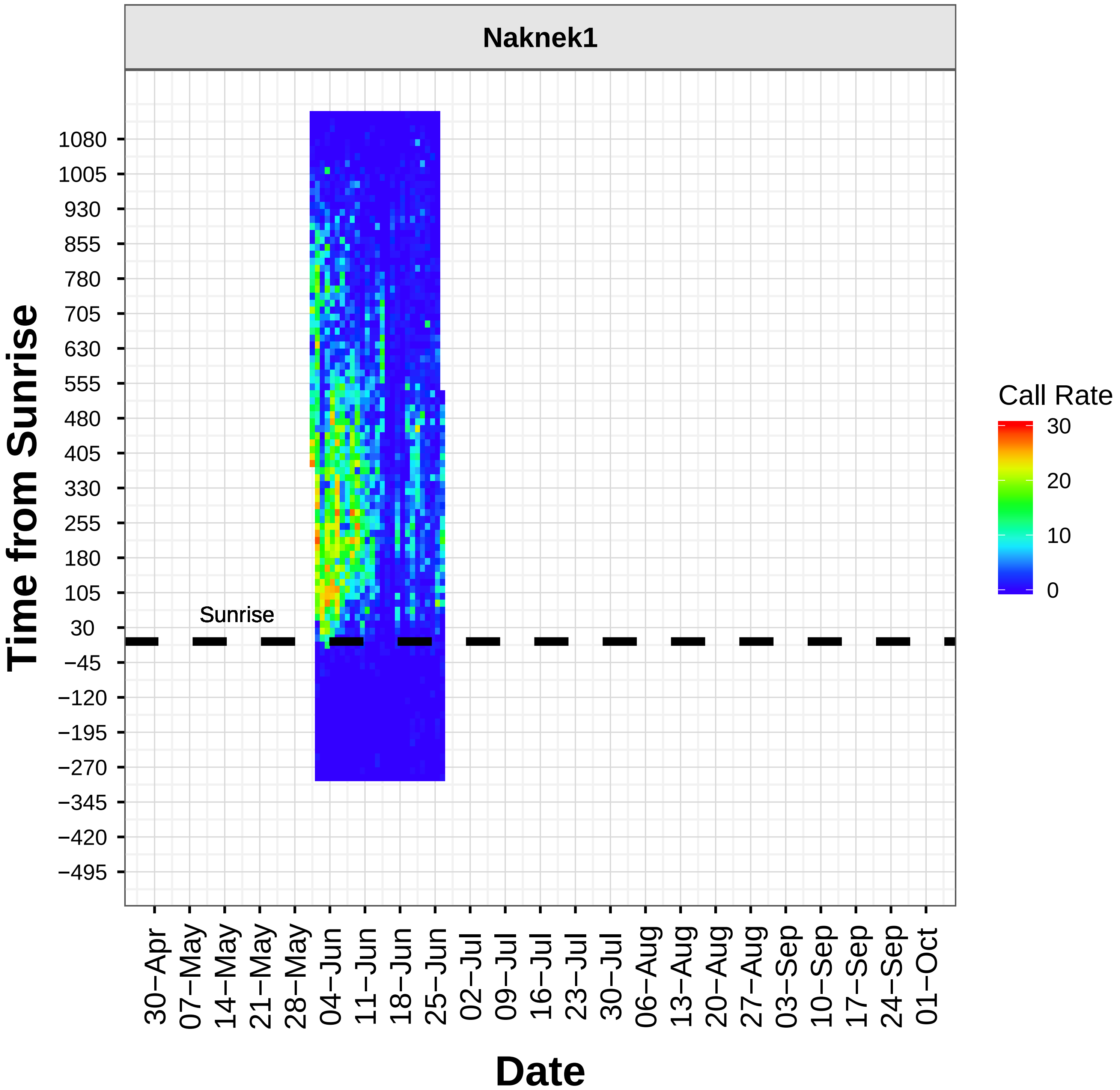
Middle Bay2
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B



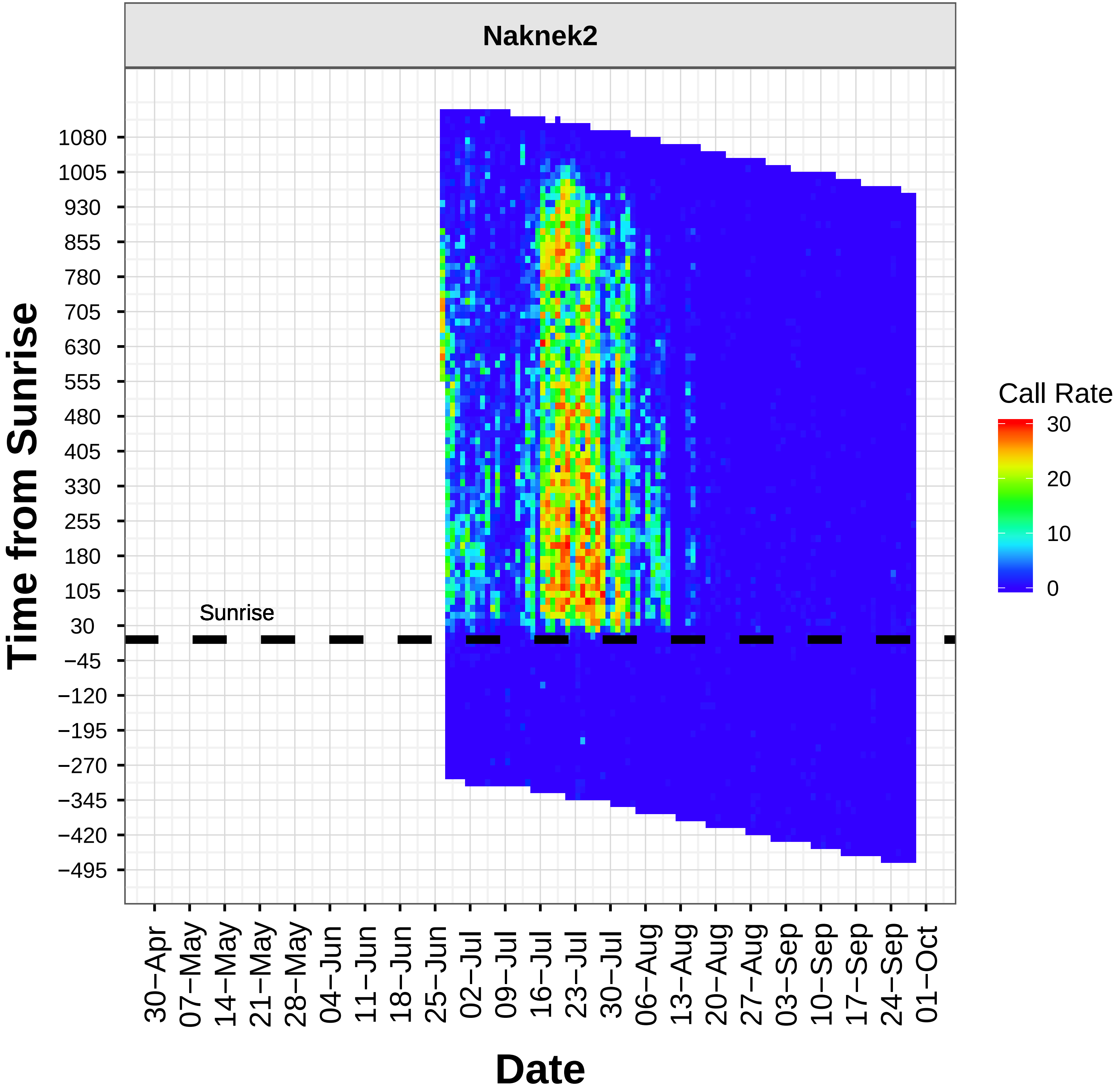
Middle Bay3
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B



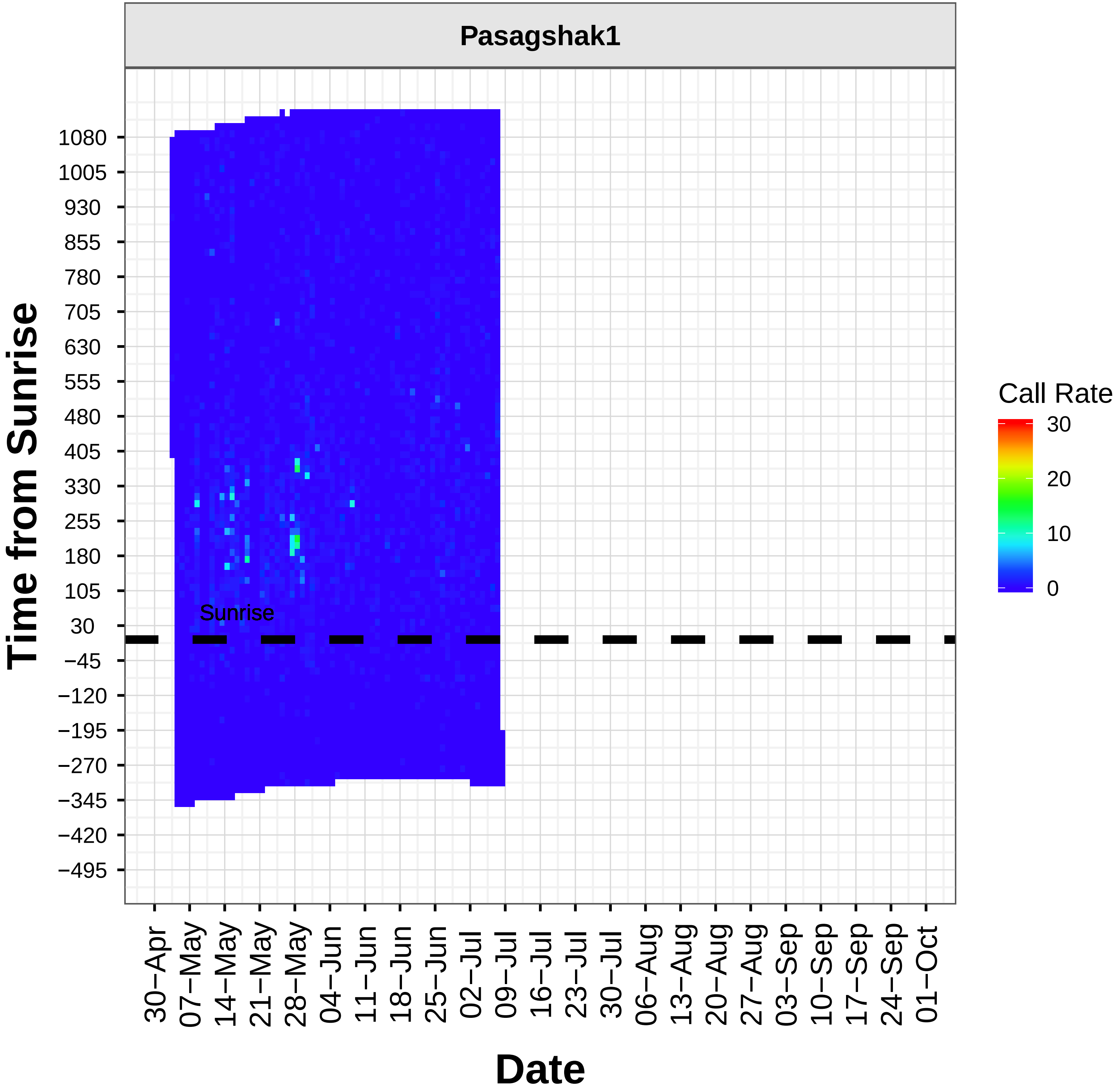
Naknek1
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B

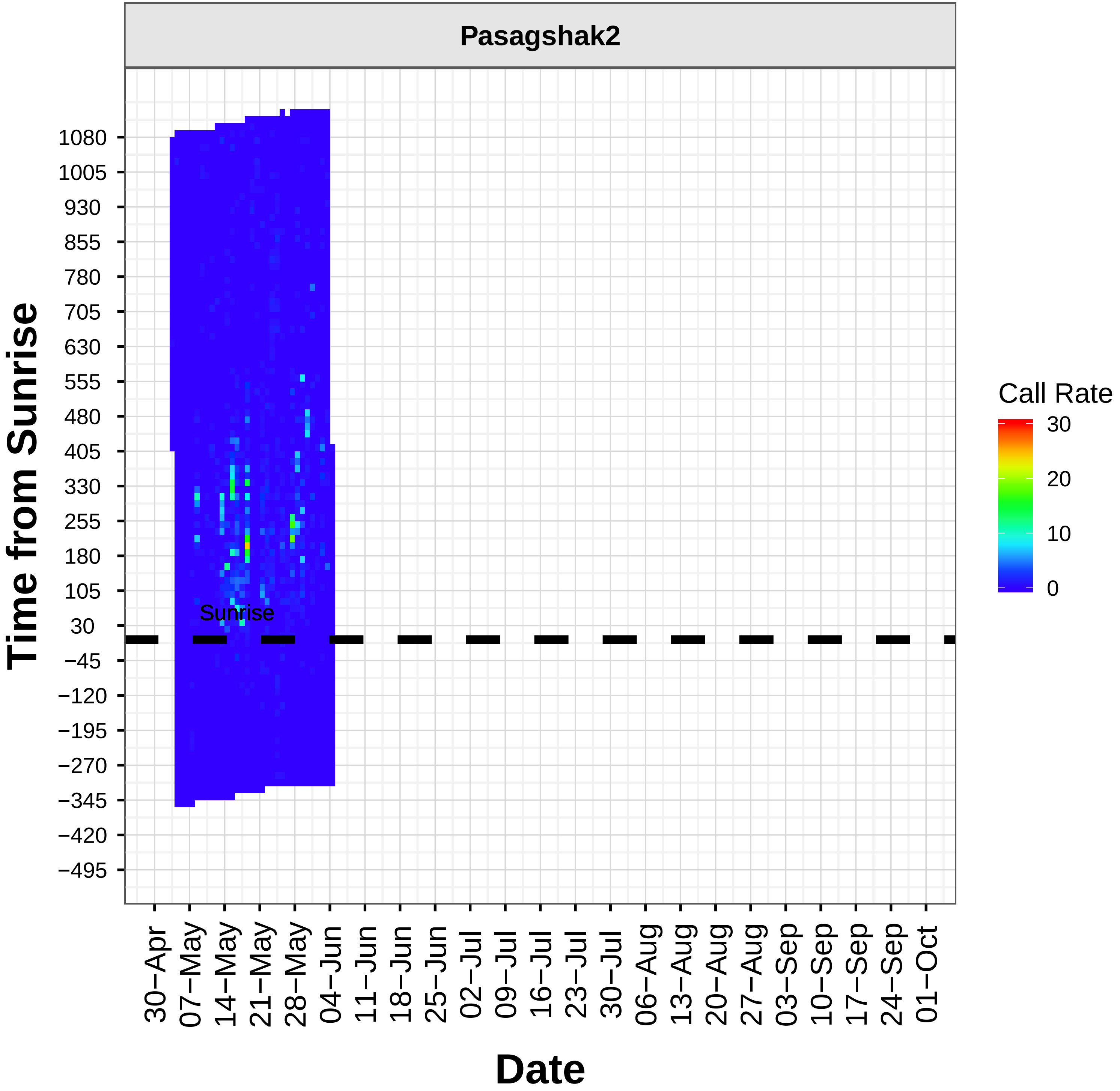


Naknek2
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B

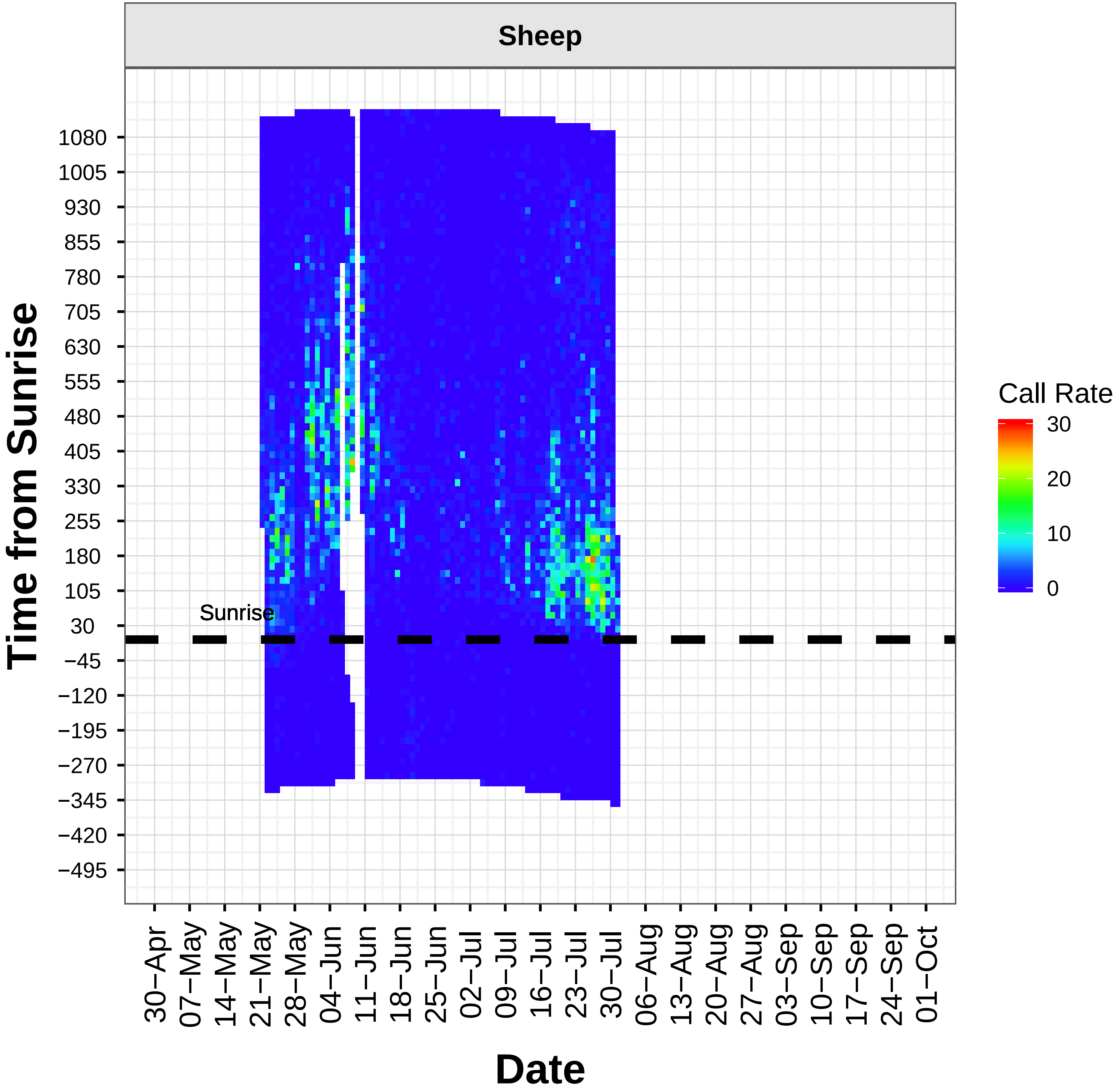

Pasagshak1
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B



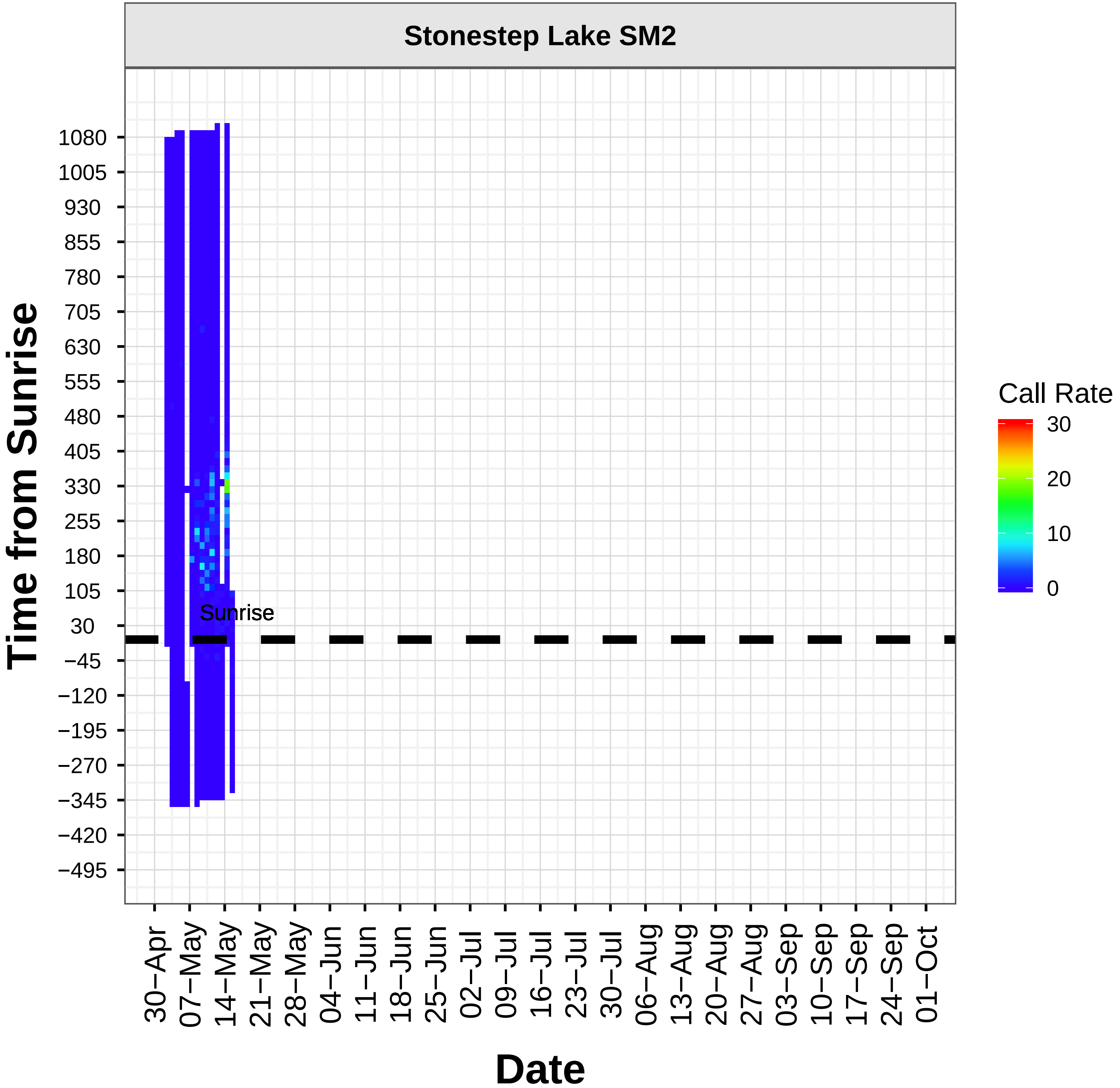
Pasagshak2
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B

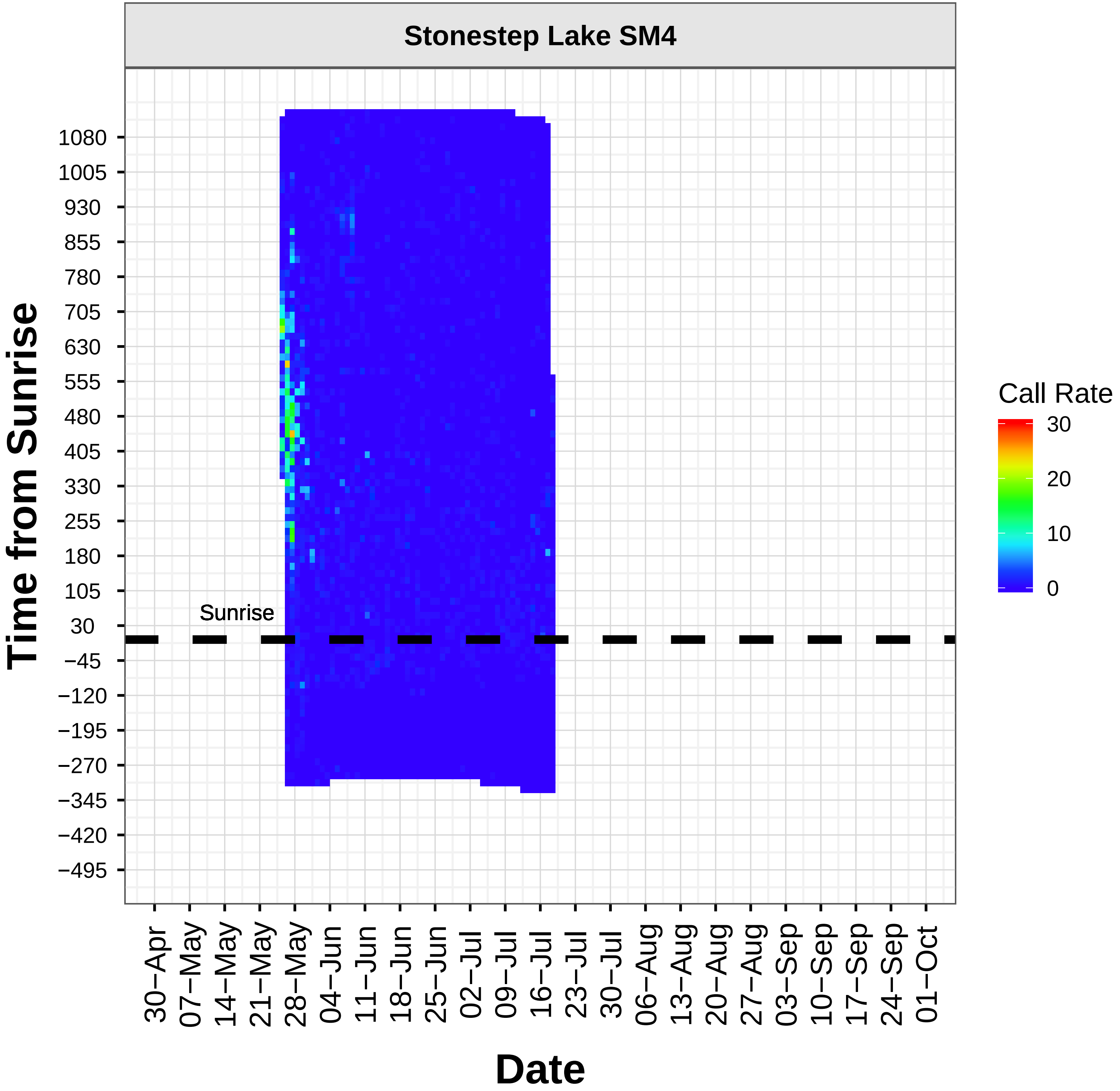


Sheep
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B

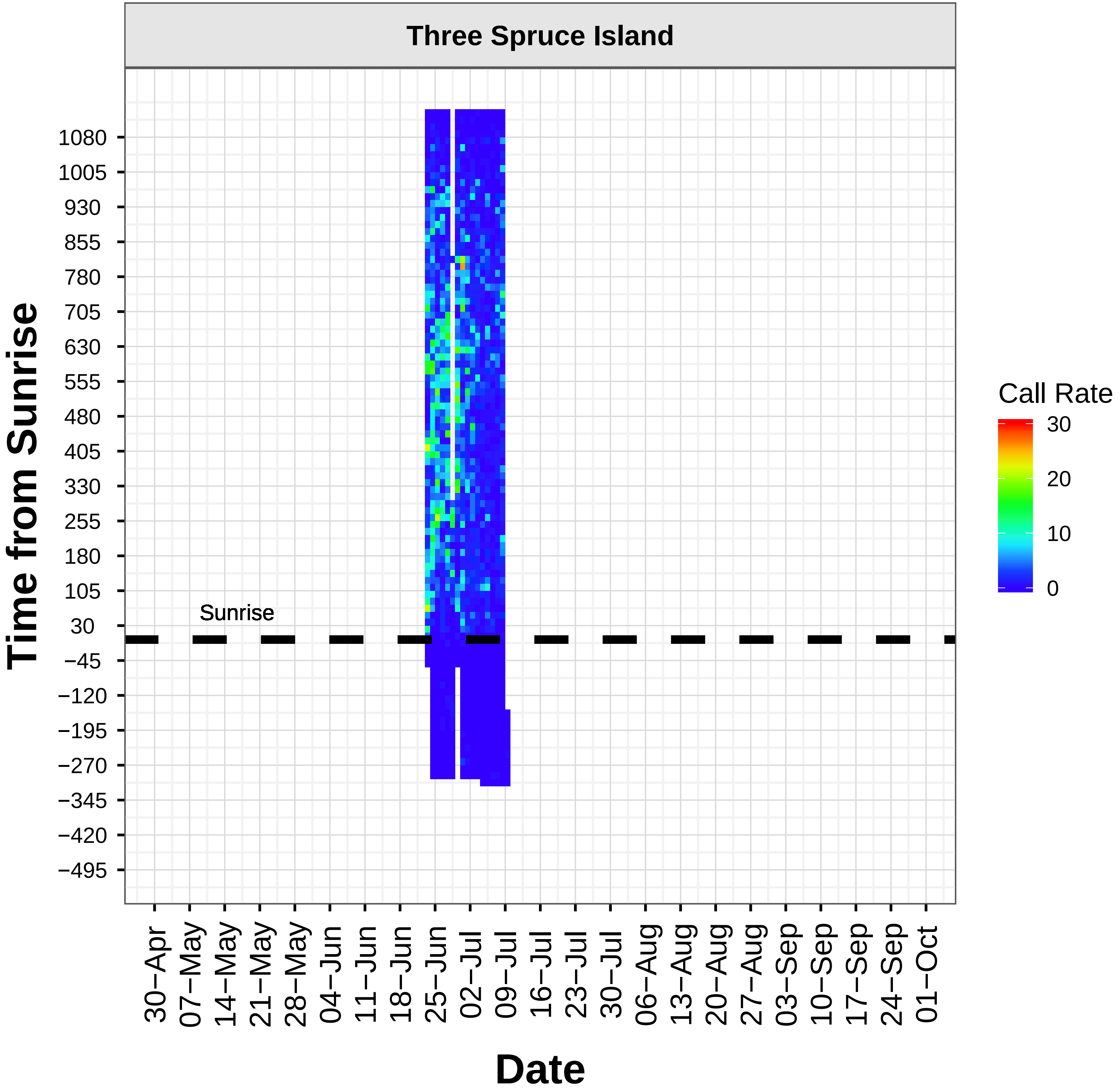

Stonestep Lake SM2
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B



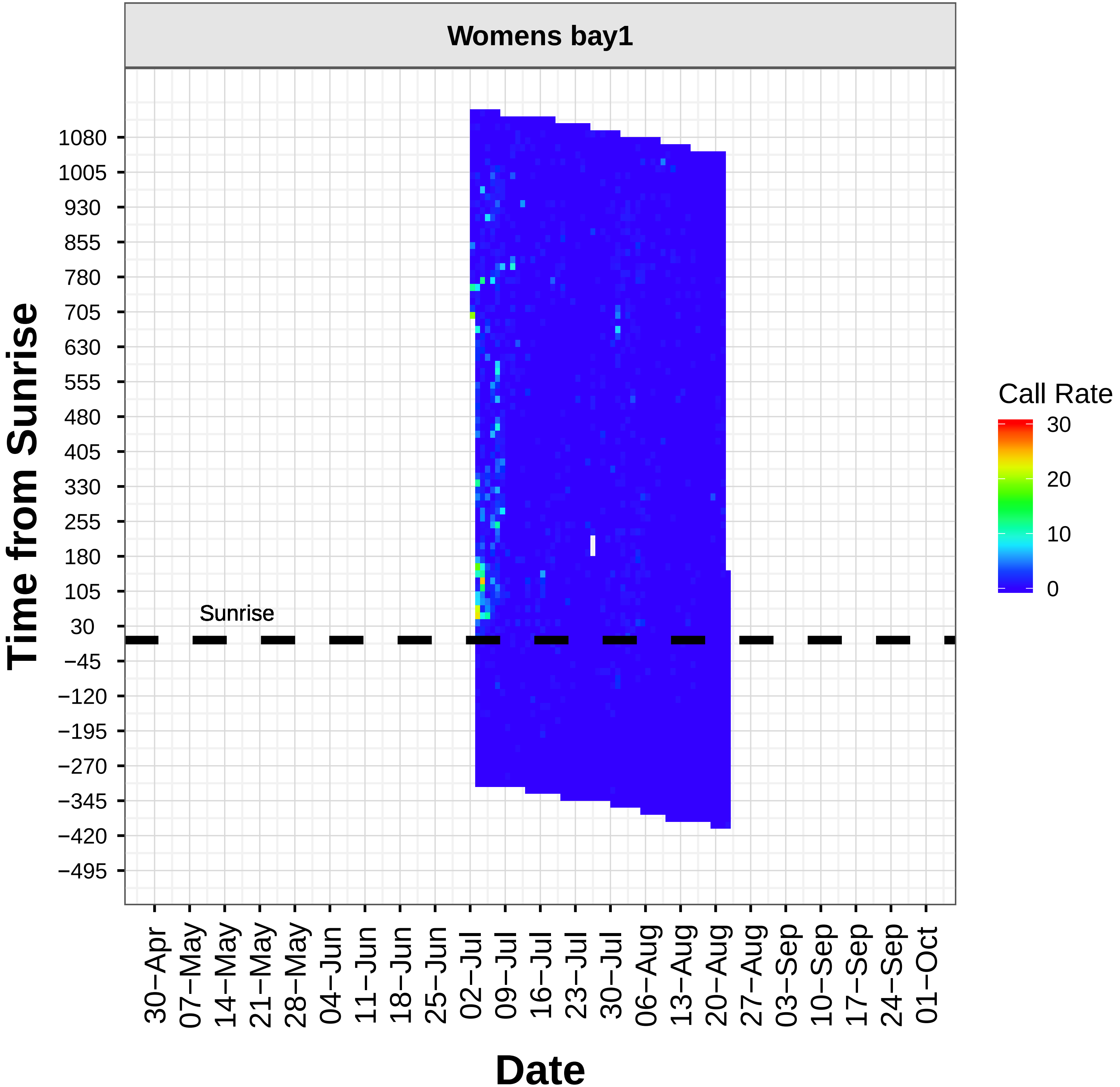
Stonestep Lake SM4
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0

Appendix B



Three Spruce Island
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Date
Call Rate
30
20
10
0

Appendix B



Womens bay1
Time from Sunrise
1080
1005
930
855
780
705
630
555
480
405
330
255
180
105
30
−45
−120
−195
−270
−345
−420
−495
Sunrise
Date
30−Apr
07−May
14−May
21−May
28−May
04−Jun
11−Jun
18−Jun
25−Jun
02−Jul
09−Jul
16−Jul
23−Jul
30−Jul
06−Aug
13−Aug
20−Aug
27−Aug
03−Sep
10−Sep
17−Sep
24−Sep
01−Oct
Call Rate
30
20
10
0