Testing the Efficacy of Acoustic Monitoring for Bonin Petrel at Sand Island on Midway Atoll NWR

Contract #: F15PX02964

For: **Roberta Swift, Region 1 Seabird Coordinator, Office of Migratory Birds and Habitat, US Fish and Wildlife Service**

By: **Abram B. Fleishman, Senior Analyst, Conservation Metrics, Inc.**
Sarah Youngren, Seabird Specialist, Conservation Metrics, Inc.
Dan Rapp, Seabird Specialist, Conservation Metrics, Inc.
Matthew McKown, CEO, Conservation Metrics, Inc.
Paige Roberts, Analyst, Conservation Metrics, Inc.

145 McAllister Way
Santa Cruz, CA 95060
email: matthew.mckown@conservationmetrics.com
tel: (831) 713-5845

Recommended citation:

Fleishman, A.B., S. Youngren, D. Rapp, M. McKown, P. Roberts. 2018. Testing the Efficacy of Acoustic Monitoring for Bonin Petrel at Sand Island on Midway Atoll NWR. Prepared for Office of Migratory Birds and Habitat, U.S. Fish and Wildlife Service, Portland, OR by Conservation metrics, Inc. 145 McAllister Way, Santa Cruz, CA 95060. 21 pages.

Introduction

Many seabirds nest in underground burrows and only attend breeding sites at night making them notoriously difficult to monitor using traditional count-based methods. Current surveys to estimate breeding population of burrow nesters involve selection of random plots across breeding aggregations, counting burrow entrances, and then documenting burrow occupancy (i.e. the proportion of burrows that contain an egg or chick). Measuring occupancy requires physical access to the nest chamber or using technology (fiber-optic burrow-scope) to allow visual confirmation of a breeding event. These counts are labor intensive and time consuming, with potential for the destruction of burrows in friable soils and mortality of petrels themselves.

Nocturnal seabird species are typically quite vocal at breeding aggregations; where acoustic cues are used in breeding displays, mate selection, navigation, and territorial advertising and defense (Bretagnolle 1996). Previous research has shown acoustic activity rates of several seabird species are positively correlated with breeding density at survey points (Borker et al. 2014; Oppel et al. 2014). Thus, acoustic metrics can be used to estimate breeding densities at survey sites. Given the difficulties of estimating burrow nesting seabird density using traditional methods, passive acoustic surveys and automated classification techniques are now being assessed and adopted for seabird monitoring (Buxton and Jones 2012; Borker et al. 2014a; Oppel et al. 2014a; Dufour et al. 2016, Hemson unpublished data). Once the functional relationship between call rates and breeding densities has been established, population size and trend trajectories can be estimated, with proper experimental design, from acoustic surveys alone (Oppel et al. 2014). Resource managers can then implement surveys that are minimally invasive to the focal species, yet still have the statistical power to inform critical management decisions.

The Northwestern Hawaiian Islands (NWHI) host important populations of burrow nesting seabirds. There are currently no surveys to document the spatial extent of breeding aggregations, species diversity, or population trajectories for burrow nesting species in the NWHI. Bonin Petrels are the most abundant of the burrow nesting species in the NWHI and the most recent study estimates that 50-95% (421,120 pairs) of the global population breeds on islands in the NWHI (Pyle and Pyle 2017). However, there is no ongoing population monitoring due to the scale and isolation of the nesting habitat, abundance of burrows, and the difficulties associated with monitoring burrow occupancy in friable sandy soils. Sand Island, which is situated in the Midway Atoll National Wildlife Refuge, is the only human inhabited island within the NWHI and has one of the largest populations of Bonin Petrels: estimated to be ~150,000 pairs (Moore 2009; Pyle and Pyle 2017). Bonin petrel populations dramatically increased after black rats (*Rattus rattus*) were eradicated from Sand Island in the 1990's but surveys to quantify this rebound have been infrequent.

For our study, we evaluated the use of passive acoustics as a cost-effective, non-invasive survey technique for estimating Bonin Petrel breeding densities on Sand Island at Midway Atoll National Wildlife Refuge. Specifically, we tested 1) if Bonin Petrel calls could be detected in the complex soundscape of Midway Atoll NWR and 2) if measures of Bonin Petrel acoustic activity are related to nest density.

Methods

Survey Design

The primary goal of acoustic surveys on Sand Island was to establish the efficacy of acoustic sensors for monitoring Bonin Petrels in the Northwestern Hawaiian Islands. To be a useful method for monitoring Bonin Petrels, 1) acoustic analysis must be able to identify Bonin Petrel calls on recordings accurately and consistently and 2) there needs to be a relationship between the density of occupied burrows near the acoustic sensor and call rates quantified by the acoustic analysis. To test these goals, six survey sites were selected in Bonin Petrel breeding habitat on Sand Island during the fall of 2015. Survey sites were selected to represent a range of burrow densities (2 low, 2 mid, and 2 high density). At each site, burrow occupancy and burrow density were measured in 30m circular plots centered on the acoustic sensor. Sensors were redeployed at the same six sites in January 2017 and one additional site (BOPE7) was added in a new low burrow density area (Figure 1).

Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

Figure 1: Map of acoustic survey sites on Sand Island, Midway Atoll NWR.

Burrow density and occupancy estimates

Burrow surveys were conducted between 8 January and 3 April 2016 (n=4) and 15 February and 31 March 2017 (n=3). We chose to measure burrow density and burrow occupancy in four wedges within a circular plot centered on each sensor because of a) the high density of burrows, b) the difficulty of working within breeding aggregations without collapsing burrows, and c) the labor involved with checking burrows with a fiber-optic camera. Four wedge plots were established at each survey point with their origin at the acoustic sensor, a

9° width, and extending out to 30m in each cardinal direction (Figure 2). The wedge plots were designed to sample ~10% of the total area around each acoustic sensor (36° of 360°). The burrow survey protocol specified that field teams estimate burrow occupancy and census burrows within each wedge plot and within two nested concentric circles with a radius of 5 and 10m from each sensor on each visit. Each wedge plot was subdivided into six 5 m sections to delimit the area within each distance band, and all the burrows within each section were identified and marked. Plot boundaries were sighted-in using a mirror compass ,declination not adjusted (declination for April 10 = 7.49E ± 0.31˚, NOAA), and divided into 5 m increments using premeasured ropes. The borders of each 5 m section were marked using 4" metal staples wrapped in flagging tape. If staples disappeared between visits, we used the premeasured ropes and other corner markers to determine where to replace them. Each survey took 11-13 days to complete, considering the high burrow densities. We used the mid-date of each survey to select when we compared call rates to burrow densities. To validate the plot method, we compared burrow density estimates from that method to those measured by the full burrow census within the 5m and 10m radius circles

The start of a burrow was defined as the point on the ground where a vertical line from the roof of the burrow entrance hit the ground. Burrows that were intersected by the border of a plot were included if ≥50% of the burrow entrance fell within plot (Seto and O'Daniel 1999; AMNWR 2015). A minimum burrow size of ≥15 cm wide and ≥25 cm long was required to be included in density counts (mean entrance size = 16.3 cm wide; Seto and O'Daniel, 1999). All other holes (e.g. old burrows) were not included in counts. All divots were investigated as potential entrances and burrows that required slight excavation (i.e. from recent erosion) were included if they were within the appropriate size range.

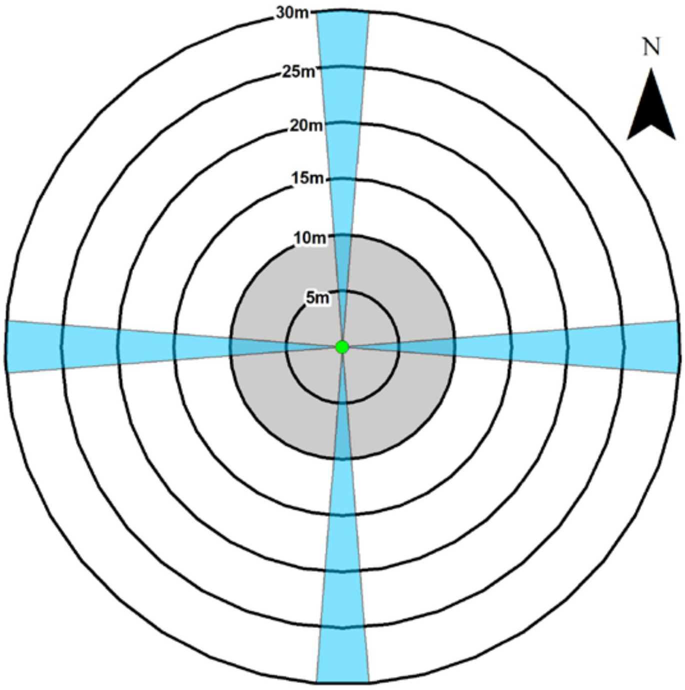

Figure 2: Survey plot design showing the areas surveyed for burrow density in blue (9˚ wedge plots) and grey (5m and 10m circular plots). The acoustic sensor was in the center of the plots.

Approximately 12 burrows per survey plot (n = 50 burrows per round per acoustic survey site) were checked using a fiber-optic burrow scope (Pukamanu; Abyssal Hawaii, LLC) to assess occupancy. Occupancy of every selected burrow could not always be determined (e.g., some burrows were too long). In those cases, an additional 5-8 burrows per subsample were selected to compensate for burrows of unknown status. Occupancy data was combined for each radius at each sensor (all four plots) and survey. Burrow occupancy was estimated as:

$$Occupancy = \frac{O_{r,i,s}}{O_{r,i,s} + N_{r,i,s}}$$

where $O_{r,i,s}$ is the number of occupied burrows and $N_{r,i,s}$ is the number of unoccupied burrows detected within a radius r meters from the sensor i during survey s. Burrows whose status could not be determined with a burrow-scope were excluded from the occupancy rate calculation with the assumption that there are equal rates of occupied and unoccupied burrows in the unknown occupancy category. Burrow density was calculated using:

$$Burrow\ Density = \frac{\left(B_{r,i,s} *\ Occ_{r,i,s}\right)}{\left[4 * \left(\frac{9}{360}\right) * \pi * r^2\right]}$$

where$B_{r,i,s}$ is the total number of burrows counted and $Occ_{r,i,s}$ is the estimated occupancy rate in all four wedge plots within a radius r meters from sensor i during survey s.

Recording equipment

Surveys were conducted with six Song Meter 3 (Wildlife Acoustics, Inc.) acoustic sensors in 2015-16 and seven sensors in 2017. Sensors were deployed with two 32 GB SD memory cards and attached to a t-post at a height of approximately 1 m. Sensors recorded in stereo with the internal microphones at a sampling rate of 24,000 Hz and 16 bits, gain was set at +18 dB, and files were stored as uncompressed ".wav" files.

Acoustic Monitoring Schedule

Song Meter clocks were set to GMT-11 and duty-cycled to record for one minute out of every five minutes starting at sunset and ending at sunrise the following morning. Sunset and sunrise were calculated from a central point (Lat: 28.2 N, Lon: 177.38 W) for all sensors. We used *SM3 Configurator.exe* (Version 1.3.1, Wildlife Acoustics) to estimate battery life and memory card space; a single 32 GB memory card was estimated to be filled with recordings after ~35 days and batteries would last 66 days. The recording schedule at the beginning of the 2017 deployment was set incorrectly, with units recorded continuously between sunset and sunrise. We sampled every fifth minute of these data to make the monitoring effort for this period comparable to the rest of the survey effort.

Table 1: Deployment table showing locations of survey sites and dates of deployment. SPID is the survey point name.

Season	SPID	Recording Unit	Northing	Easting	First Recording	Last Recording
2016	BOPE01	BOPE01	462619.114	3121007	12/20/2015 17:58	6/14/2016 02:27
2016	BOPE02	BOPE02	462088.491	3120706	12/20/2015 17:58	6/20/2016 05:49
2016	BOPE03	BOPE03	462075.692	3120564	12/20/2015 17:58	6/20/2016 05:49
2016	BOPE04	BOPE04	463058.763	3120765	12/20/2015 17:58	6/20/2016 05:49
2016	BOPE05	BOPE05	462587.694	3120686	12/20/2015 17:58	6/18/2016 04:59
2016	BOPE06	BOPE06	461996.747	3120058	12/20/2015 17:58	6/20/2016 05:49
2017	BOPE01	BOPE01	462619.114	3121007	1/20/2017 18:20	7/5/2017 01:10
2017	BOPE02	BOPE02	462088.491	3120706	1/20/2017 18:20	6/23/2017 05:50
2017	BOPE03	BOPE03	462075.692	3120564	1/20/2017 18:20	7/7/2017 21:25
2017	BOPE04	BOPE04	463058.763	3120765	1/20/2017 18:20	6/30/2017 03:50
2017	BOPE05	BOPE05	462587.694	3120686	1/20/2017 14:34	6/23/2017 05:35
2017	BOPE06	BOPE06	461996.747	3120058	1/20/2017 18:20	1/24/2017 07:38
2017	BOPE06	PMA2	461996.747	3120058	1/25/2017 14:53	6/23/2017 05:25
2017	BOPE07	BOPE06	463370.42	3121342	2/1/2017 10:00	7/10/2017 20:09

Automated Acoustic Analysis

Conservation Metrics has developed a machine-learning approach for detecting sounds of interest on field recordings. Our method leverages Deep Neural Network (DNN) models, a technique where a classification model is trained to detect a unique combination of spectro-temporal features found in target sounds (i.e. elements of calls from species of interest). Trained models can then be used to search field recordings for sounds with the same combination of features. Deep Neural Networks are widely used for detection and classification challenges in many fields, including speech recognition (Deng et al. 2013) and image recognition (Ciresan et al. 2012).

Bonin Petrels have several known calls (Grant et al. 1983). We built detectors to find a growl (Grant's *churr*) call that is typically made by birds on the ground, and a screech call (Grant's *kuk-u-er cry*) that is typically made by birds in flight. The growl (Figure 3) consists of a band of sound between 300 and 1,000 Hz with an energy peak at 650-700 Hz, often discernable as a staccato purr with a pulse rate of 78-190 pulses per second (Grant et al. 1983). Although this call is most often given on the ground, we found that it was also given in flight, as shown by the presence of a doppler shift. The screech call (Figure 4) is a broad band sound, which is descending in pitch, and between 1,500 and 6,000 Hz. There are few discernable features to this call, aside from a weak energy peak between 2500 and 5500 Hz (Grant et al. 1983).

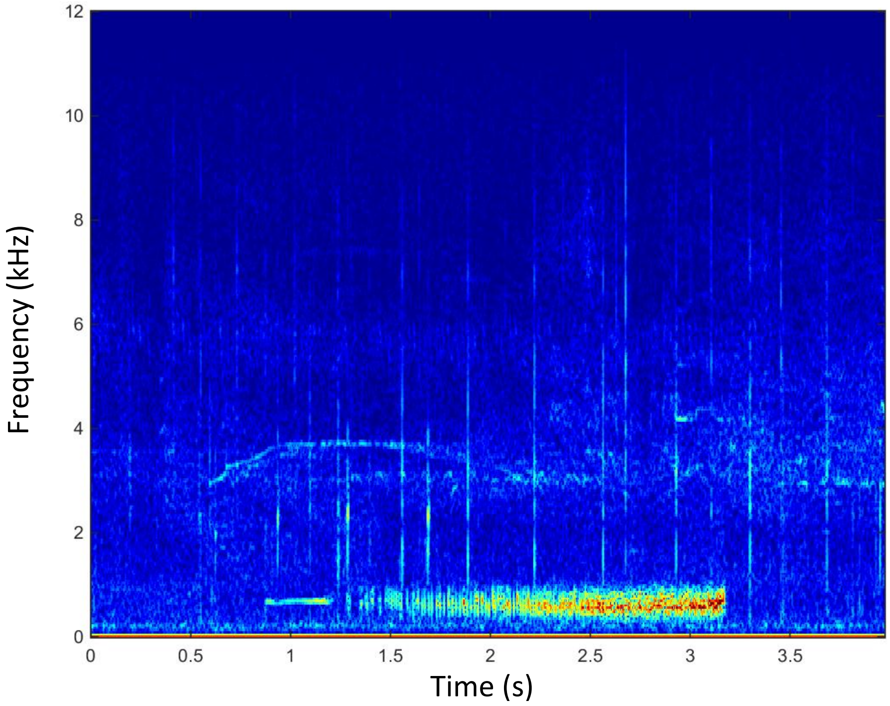

Figure 3: Spectrogram of Bonin Petrel growl vocalization

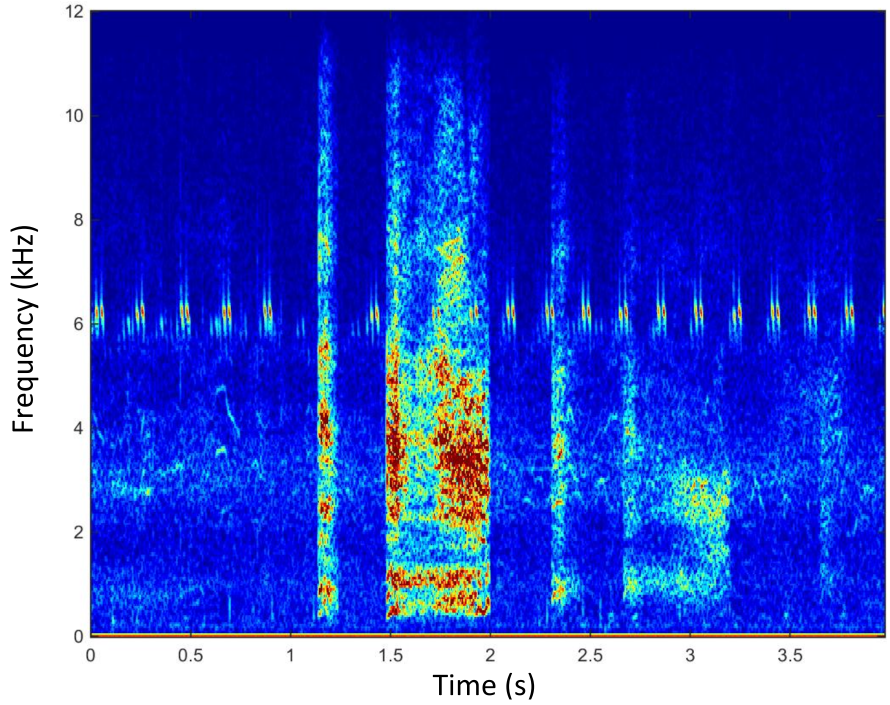

Figure 4: Spectrogram of Bonin Petrel screech vocalization

Model Performance

The accepted method of evaluating real-world performance of a DNN model requires creation of a test dataset that is independent of both the model training and model cross-validation datasets. The model can be run on the independent test dataset, and performance can be measured by calculating ***sensitivity*** (proportion of positives that are correctly identified) and *accuracy* (proportion of detections that are true positives). Ideally, a test dataset should contain a representative sample of data from all monitoring sites, sampled from across the monitoring period, and across the range of acoustic conditions in local soundscapes. It should also contain randomly selected examples of positive events (target species vocalizations), and negative events, in the same proportion that they occur in the natural soundscape. Thus, creation of an ideal test dataset is a challenge that requires manual review and labeling of many thousands of randomly selected clips of acoustic data.

We developed an independent test dataset in two parts. First, we randomly sampled ~250, 2-second clips from acoustic data collected at each of the 7 survey sites, across the range of dates that sensors were deployed. While the resulting dataset contained 2,974, 2-second clips, only 110 growl and 219 aerial calls were identified among the other 2,645 clips of non-target sounds. This portion of the test dataset enables an accurate measure of model accuracy (quantification of false-positives) because it contains a wide variety of sounds that are not target species vocalizations. However, the measure of sensitivity derived from this randomly sampled dataset is not representative of real world performance due to the small number of positive events for each signal.

To improve our measure of sensitivity, we then augmented the randomized test dataset with additional positive events. These events were identified by randomly choosing 25 confirmed positive calls for each signal, from all 13 site/year deployments. We then selected 17-second clips of the acoustic data, centered on those randomly selected calls. Because call presence is temporally auto-correlated, each of those 17-second clips had a high likelihood of containing additional positive events that had not previously been included in our training dataset. The augmented test dataset included an additional 1,081 growl and 1,648 aerial calls to aid the estimation of sensitivity.

For the growl call, we built 17 model versions to iteratively improve model performance. We ultimately selected a model version, BOPE_Growl_V15_29Nov17, with an accuracy of 73.1% and sensitivity of 56.2% on the test data set at a classification threshold of 90% probability. We build eight model versions for the aerial call, and the final detection model (BOPE_Aerial_V6_30Nov17) had an accuracy of 96.8% and sensitivity of 30.5% on the test dataset at a classification threshold of 90% probability.

Bayesian Estimation of Call Rates and Burrow Density

Typically, we summarize acoustic call rates using a simple mean of the number of calls detected for each minute of a given period. This method is subject to many sources of potential bias and error, including variation in recording quality, survey effort, temporal differences in sampling intensity, and environmental factors. To address these issues, we used a Bayesian hierarchical state-space model (initially developed to estimate Hawaiian Petrel and Newell’s Shearwater call rates on Kauaʻi) to estimate Bonin Petrel call rates while accounting for all the above-mentioned sources of bias and variance. The Bayesian modeling approach adds statistical power by partitioning variance into multiple sources of stochasticity that influence seabird behavior and/or acoustic detection probabilities, and especially by disentangling process error (real-world differences in mean call rates and abundance) from observer error (variation in measured call rates at any given point in time and space). The model incorporates fixed and random effects of regional differences, among-site differences, stage of the breeding season, moon illumination, time of day, and microphone quality. By incorporating multiple potential sources of variation, the model can better estimate the true underlying call rates at each site and provide quantitative insights into the various environmental and ecological factors that influence these call rates. In this report, call rates are presented as mean estimated Bonin Petrel growl calls per minute, and 95% credible intervals (where 95% of estimated values fall within this interval in the Bayesian posterior distribution) are presented as an estimate of uncertainty. To compare call rates among sites we used estimated call rates from the 50-110 minutes after sunset peak hour, and selected a date range from 1 February to 31 March, which coincided with the incubation and early chick rearing period (Seto and O'Daniel 1999).

Additionally, when burrow density data corresponding to acoustic survey sites is available, the Bayesian model can be used to estimate a conversion function between acoustic call rates and burrow densities from all survey seasons. This Bayesian estimate of the true underlying burrow density is more statistically powerful than estimates derived from simple linear regression models of acoustics or traditional counts alone. Here, we incorporated acoustic call rates and burrow density data from both survey years to identify whether a statistically significant functional relationship exists between acoustic call rates and count estimates of nest densities. This second Bayesian modelling step uses the posterior distribution of estimated call rates for each site in each year, paired with each burrow survey, to create multiple count "replicates" per mean estimated call rate. Burrow count date (categorical by stage of breeding season) and count radius are then incorporated into Bayesian regression estimation as potential sources of variation in count replicates.

The Bayesian model was implemented using the programs R and JAGS ("Just Another Gibbs Sampler"); JAGS model code is available upon request.

Results

Burrow density and occupancy estimates

Burrow occupancy was measured at each sensor four times in 2016 and three times in 2017. Because burrow occupancy and density were estimated within nested concentric circles, the resulting table is too large for inclusion here and has been attached to this report as a separate excel file. Below we present the burrow occupancy and density for the largest radius circle only (30m).

Burrow occupancy peaked in late February in both years. Occupancy varied among sites from a mean of 31.4% at BOPE01 in 2016 to 78.5% at BOPE02 in 2017. Burrow occupancy differed between years at four sites (Figure 5; increase at BOPE02, BOPE03, BOPE04 and decrease at BOPE05). Occupied burrow density varied among sites and years. Density ranged between 0.056 burrows m^{-2} at BOPE01 in 2016 to 0.248 burrows m^{-2} at BOPE04 in 2017 (Figure 6). Occupied burrow density was higher at all survey sites in 2017 than 2016.

The wedge plots consistently underestimated burrow density within 5- and 10-meter radii from the acoustic sensor by -61.7 +/- 73.1% and -21.4 +/- 32.6% respectively (Figure 7). This is likely driven by the small total area sampled by the wedge plots at small radii (i.e. 7.85m^2 and 31.4m^2 for wedge plots within the 5 and 10m radii respectively). Our assumption is that wedge plots would be comparable to full plots at the larger radii, as the overall survey area increased; however, we were unable to carry out counts of full circles at the 20m and 30m radius considering the labor required and the potential impact to nesting birds.

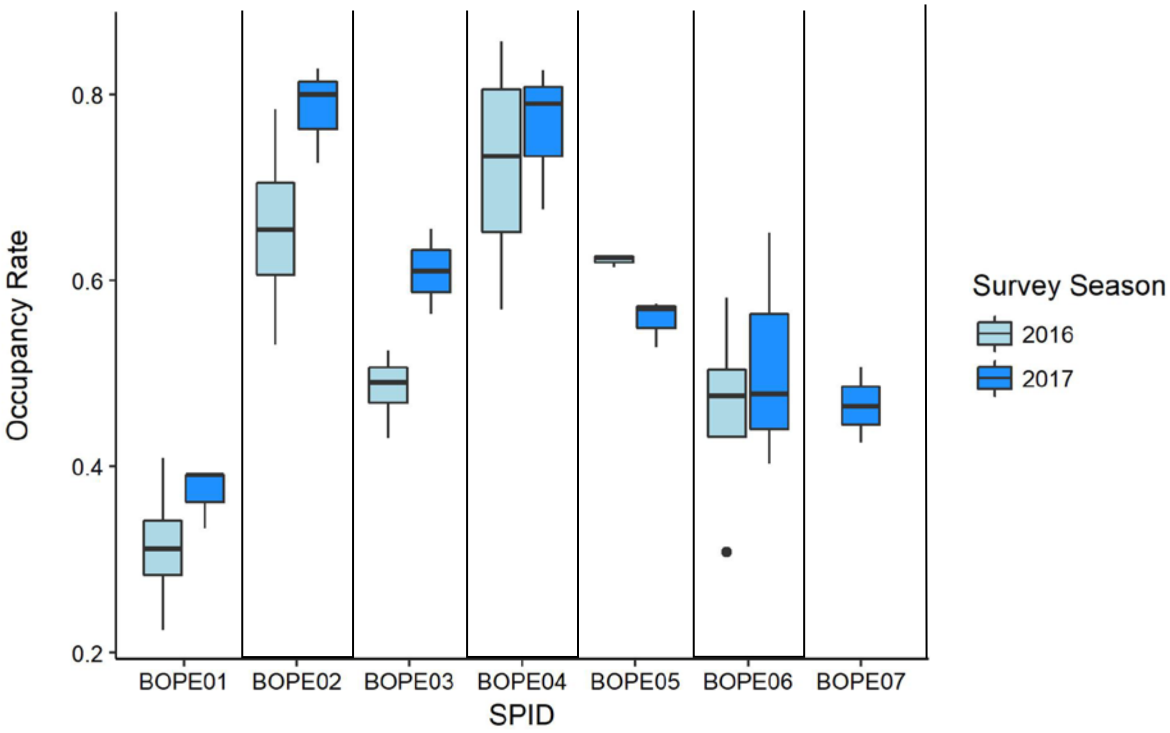

Figure 5: Bonin Petrel burrow occupancy rate in the wedge plots within 30m of the acoustic sensor between 8 January and 3 April in 2016 (n=4) and 15 February and 31 March in 2017 (n=3).

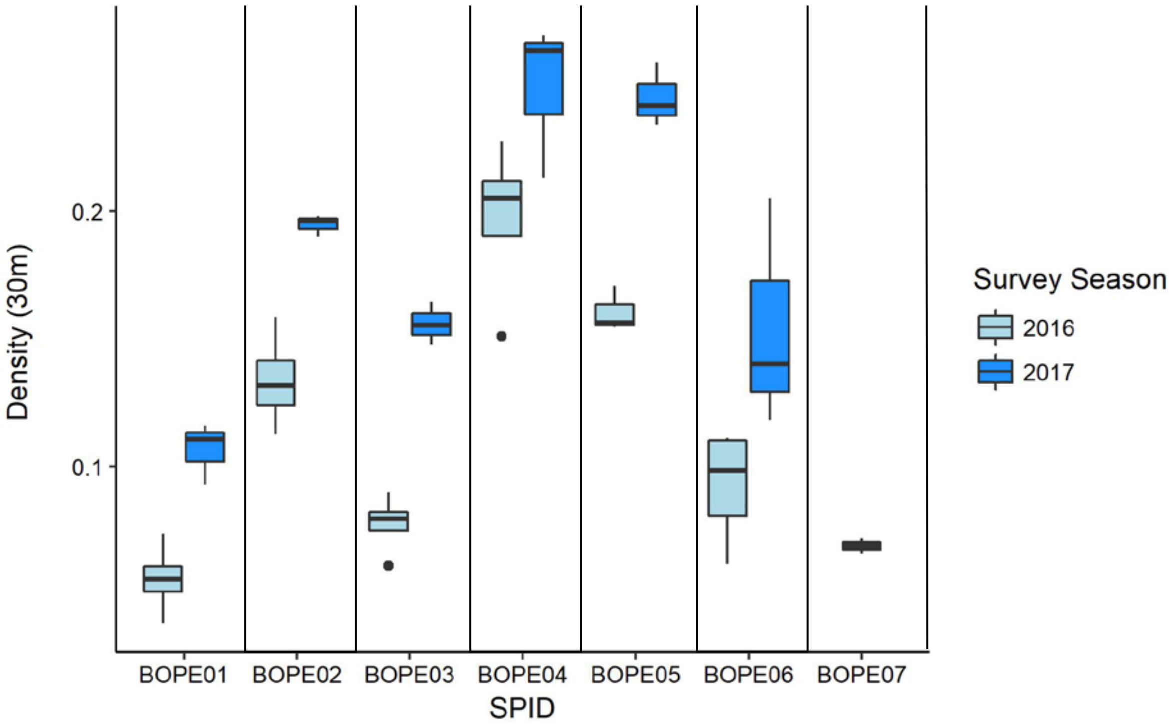

Figure 6: Bonin Petrel occupied burrow density within 30m of the acoustic sensor in the wedge plots between 8 January and 3 April in 2016 (n=4) and 15 February and 31 March in 2017 (n=3).

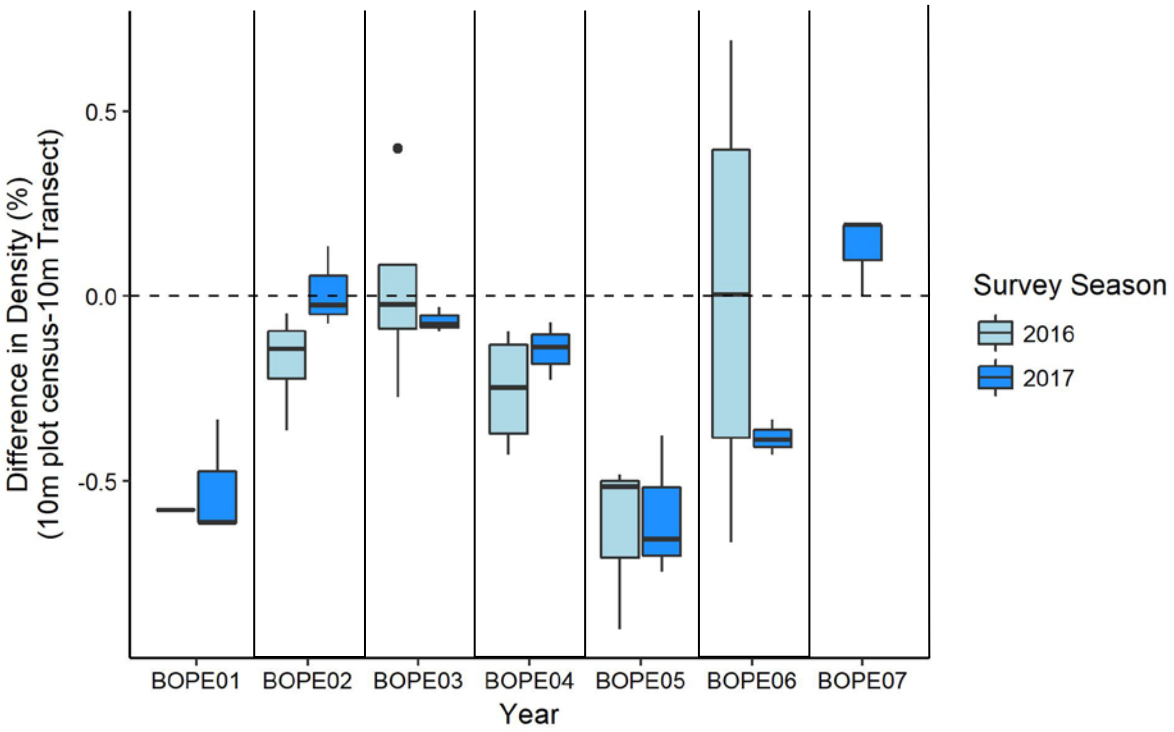

Figure 7: Difference in estimated density between the full circle burrow census and the wedge plot method at 10 m between 8 January and 3 April in 2016 (n=4) and 15 February and 31 March in 2017 (n=3).

Acoustic Survey Effort

We collected and analyzed 5,052.18 hours of recordings during 2,198 combined survey nights at seven survey sites. Microphones failed during six deployments and there were moderate data quality reductions (>15% of the data) at two sites: BOPE05-2017 and BOPE 04-2016 (Figure 8, Table 2). All low-quality data were identified and incorporated into the Bayesian call estimates, which accounts for lower detection probabilities during periods with poor-quality recordings.

Table 2: Acoustic survey effort table with all raw data and reduced dataset after removal of data collected with malfunctioning microphones.

Season	SPID	Total Nights	Total Hours	Survey Nights	Survey Hours	Nights Low Quality	% Nights Low Quality	Hours Low Quality	% Hours Low Quality
2016	BOPE01	173	403.07	172	402.39	1	0.01	0.68	0.00
2016	BOPE02	184	435.16	184	431.62	0	0	3.54	0.01
2016	BOPE03	184	435.21	184	431.15	0	0	4.06	0.01
2016	BOPE04	168	383.95	168	325.25	0	0	58.7	0.15
2016	BOPE05	182	430.92	182	424.96	0	0	5.96	0.01
2016	BOPE06	184	435.06	184	416.67	0	0	18.39	0.04
2017	BOPE01	167	377.95	167	375.27	0	0	2.68	0.01
2017	BOPE02	155	354.67	155	350.88	0	0	3.79	0.01
2017	BOPE03	169	383.24	169	376.57	0	0	6.67	0.02
2017	BOPE04	162	366.42	162	346	0	0	20.42	0.06
2017	BOPE05	155	354.74	154	235.42	1	0.01	119.32	0.34
2017	BOPE06	155	350.51	154	336.76	1	0.01	13.75	0.04
2017	BOPE07	160	341.28	159	327.61	1	0.01	13.67	0.04
Total		2198	5,052.18	2,194	4,780.55	4	0.002	271.63	0.05

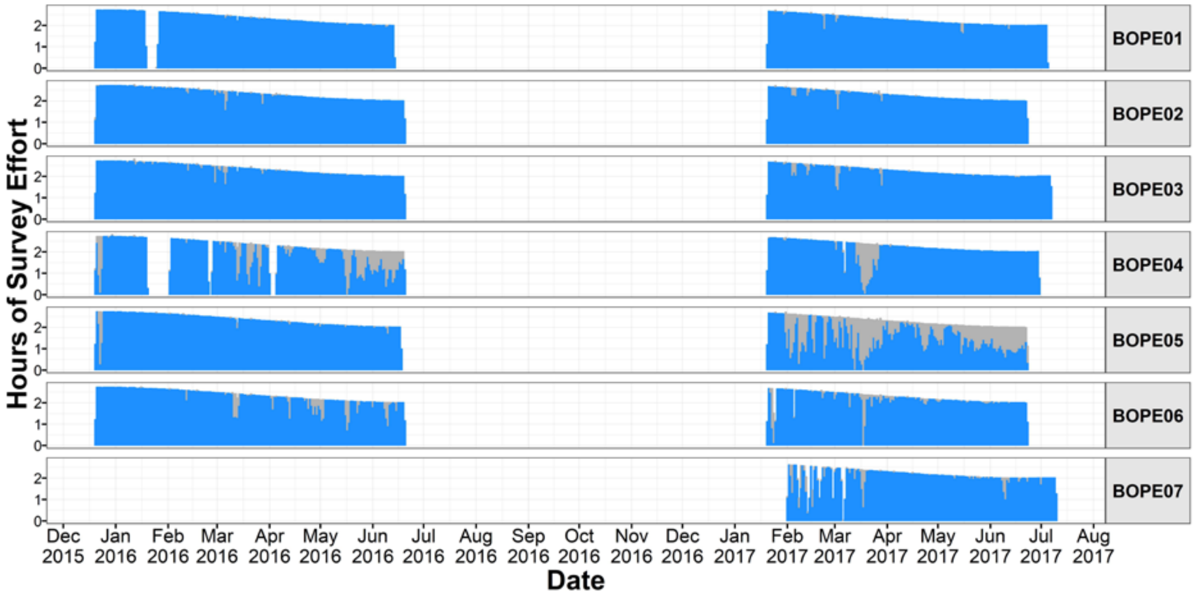

Figure 8: Total hours of acoustic recording effort at each survey site, by date. Bars represent total hours of survey effort. Grey portions of bars represent poor quality recordings based on spectral measurements of microphone quality.

Bonin Petrel acoustic activity

Although model performance on the test dataset was higher for the aerial call, we predicted that the growl call would correlate better with burrow density. The growl and aerial calls showed similar temporal and relative abundance patterns, suggesting both signals provide similar information. Below we present figures and analysis for the growl call only.

Bonin Petrel calls were detected at all survey sites. On a daily basis, estimated call rates peaked 50 to 110 minutes after sunset and slowly tapered throughout the remainder of the night (Figure 9, Figure 10). Activity varied seasonally, with the highest call rates in February and March 2016 and late January and February 2017 (Figure 9, Figure 10). The lull in activity in January 2016 may be a result of reduced bird numbers during the pre-lay exodus. Sensors were deployed in the middle of the pre-lay exodus in 2017, as indicated by the sharp increase in estimated call rate from mid to late January (Figure 9, Figure 10). Mean egg laying date is normally toward the end of January (Seto and O'Daniel 1999) which may have been detected as the late January - early February peak in estimated call rates (Figure 9, Figure 10). Vocal activity waned in early May in both years. Moon illumination had a strong effect on vocal activity and rates were 2.10 times higher during a new moon compared with a full moon (Figure 11).

The highest activity sites were BOPE01, BOPE04, BOPE05 and BOPE06 in 2016 and BOPE05 in 2017 (Figure 12, Table 3). Vocal activity rates declined from 2016 to 2017 at four of the six sites (BOPE01, BOPE03, BOPE04, and BOPE06). There was no significant relationship between estimated Bonin Petrel call rates and burrow densities by site (Bayesian r^2 = 0.0076, Figure 13). In general, the density estimates are highly uncertain; whereas, call rate estimates had less variance. That said, BOPE01 2016 showed high variability in call rate estimates.

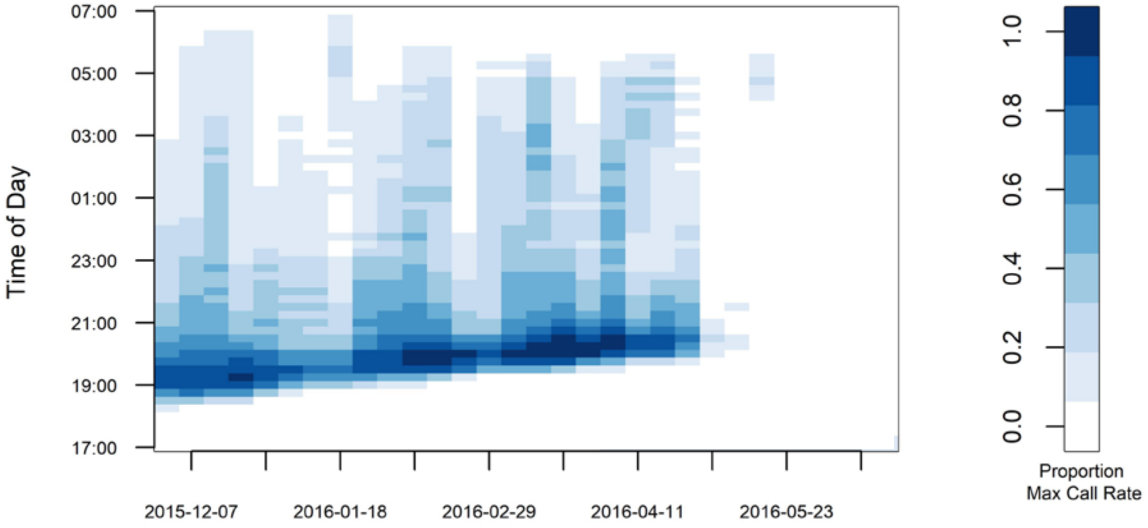

Figure 9: Relative Bonin Petrel vocal activity (growl call) by day and time during 2016. Color represents the proportion of the maximum call rate exhibited by the data during the specific period (15-minute, weekly time scale).

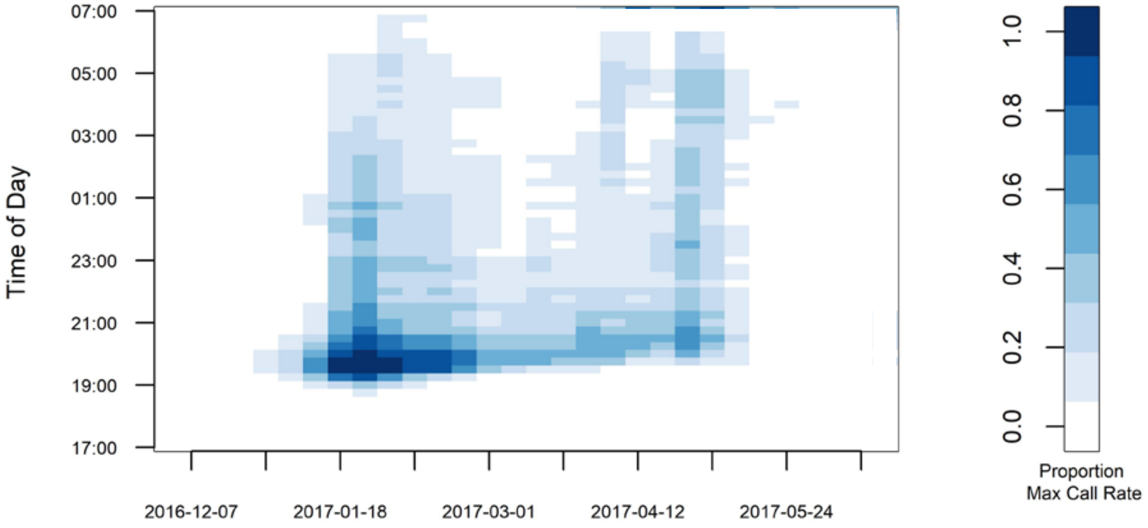

Figure 10: Relative Bonin Petrel vocal activity (growl call) by day and time during 2017. Color represents the proportion of the maximum call rate exhibited by the data during the specific period (15-minute, weekly time scale).

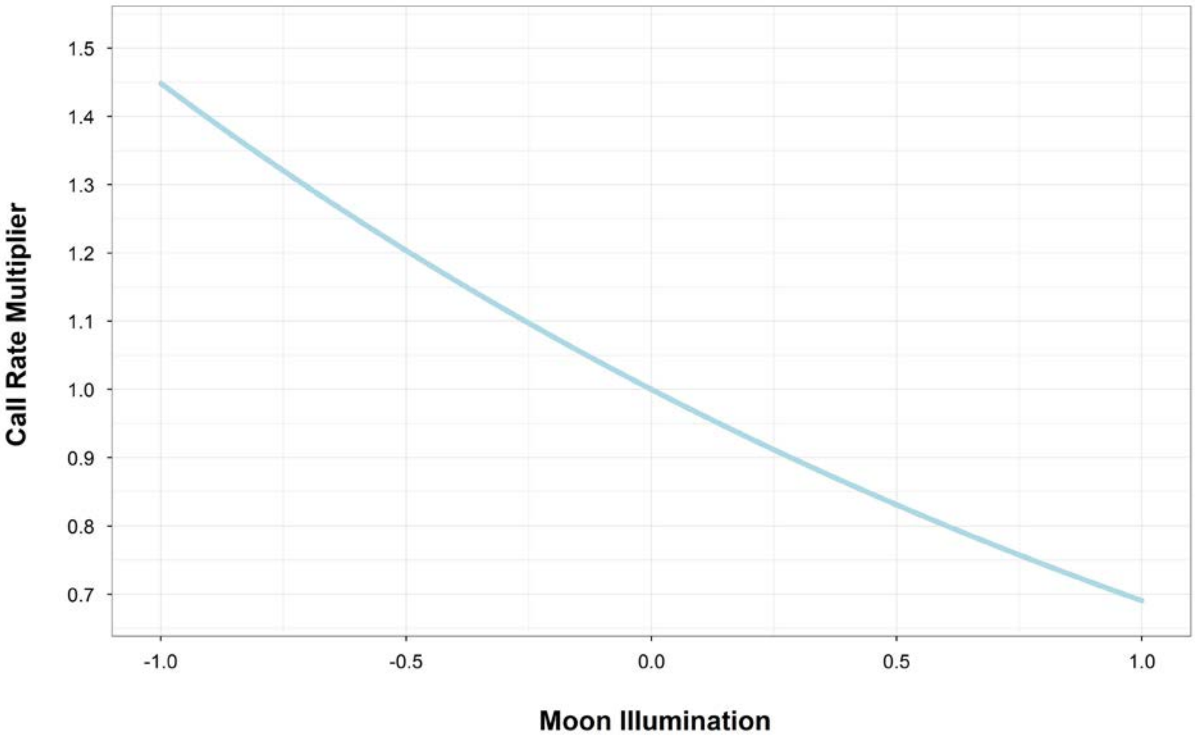

Figure 11: Effect of moon illumination on Bonin Petrel acoustic activity in 2016 measured as the average change in acoustic activity levels (Call Rate Multiplier). A new and full moon have a moon illumination value of -1 and 1, respectively. Effect of moon illumination in 2017 was not statistically different from 2016.

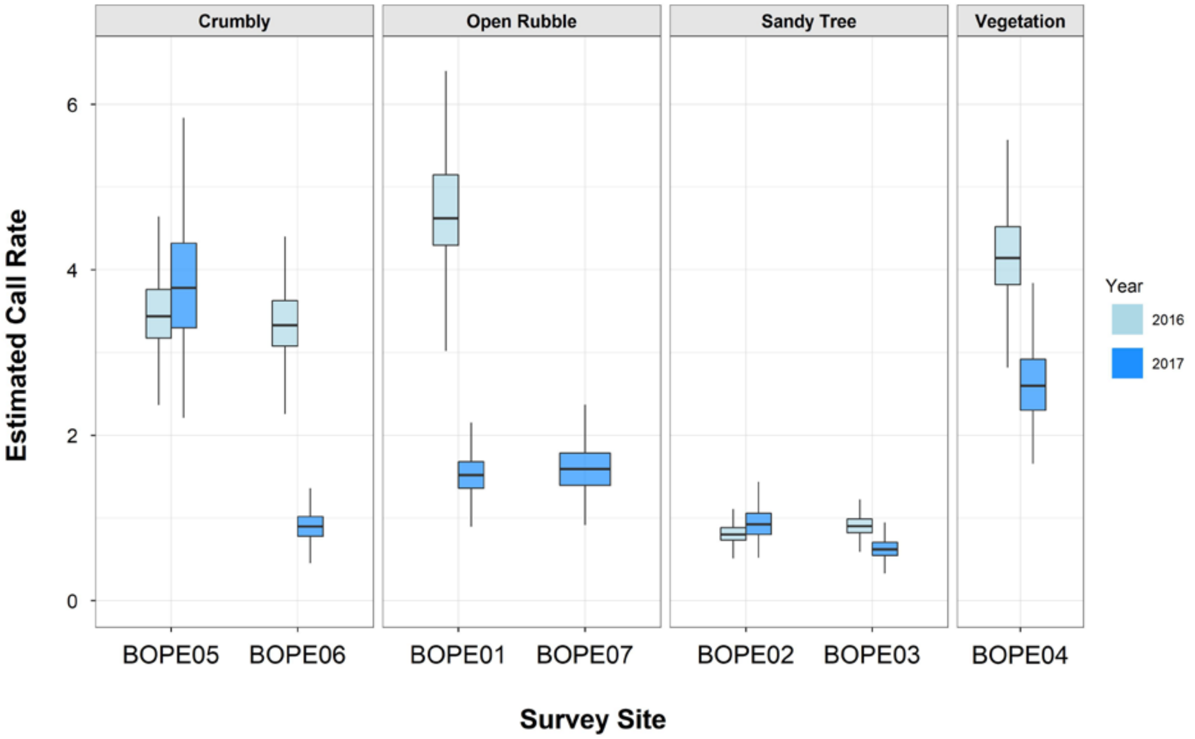

Figure 12: Estimated Bonin Petrel vocal activity (growl call) by year during peak hour (50 to 110 minutes after sunset) and peak seasonal activity period (1 February to 31 March) grouped by habitat type.

Table 3: Estimated call rate per minute for Bonin Petrel growl calls during peak hour (50 to 110 minutes after sunset) and peak seasonal activity period (1 February to 31 March).

Season	SPID	Rate Per Min	Lower 95% CI	Upper 95% CI
2016	BOPE01	4.84	3.68	7.35
2017	BOPE01	1.53	1.10	2.08
2016	BOPE02	0.81	0.60	1.08
2017	BOPE02	0.93	0.63	1.35
2016	BOPE03	0.92	0.69	1.24
2017	BOPE03	0.63	0.43	0.96
2016	BOPE04	4.25	3.28	5.98
2017	BOPE04	2.63	1.84	3.86
2016	BOPE05	3.51	2.69	4.83
2017	BOPE05	3.85	2.53	5.78
2016	BOPE06	3.40	2.55	4.69
2017	BOPE06	0.90	0.61	1.26
2017	BOPE07	1.60	1.10	2.25

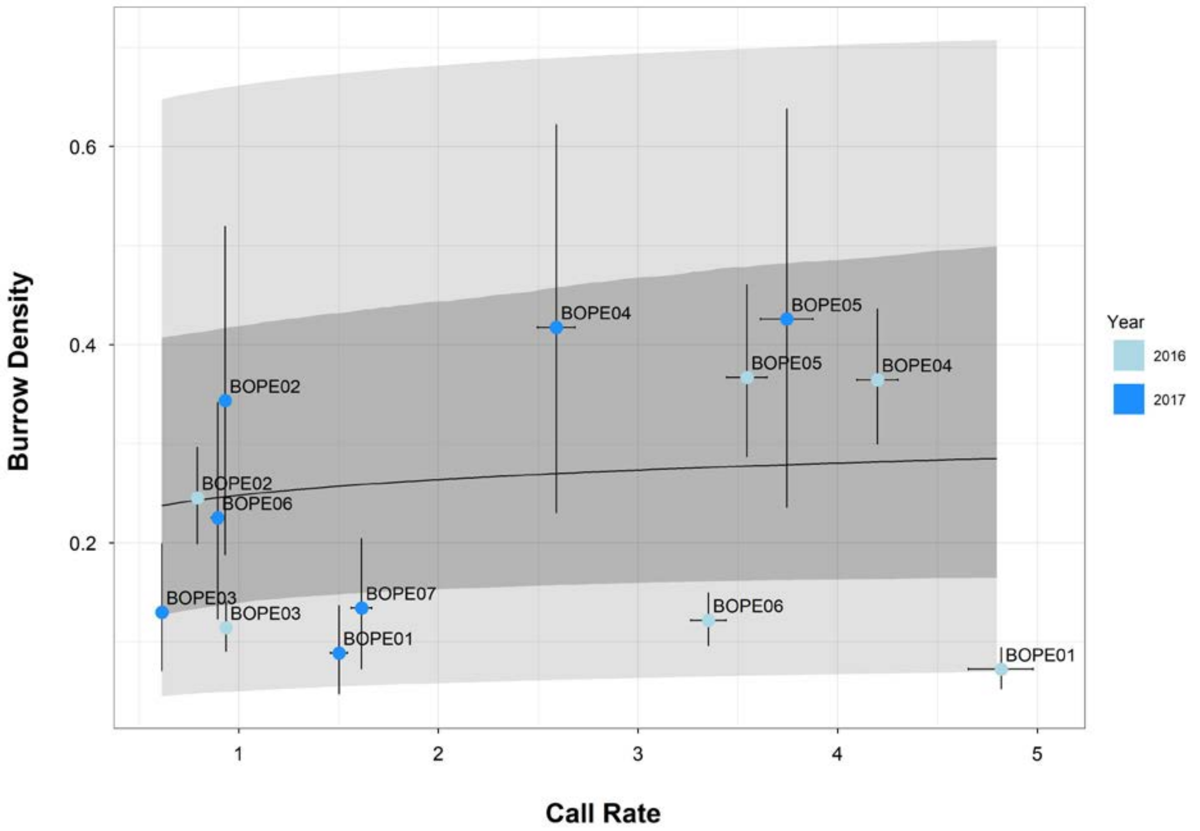

Figure 13: Call rate to burrow density conversion function. The function uses model estimates for Bonin Petrel growl call rates and compares them to multiple burrow density estimates at each site, over both survey years. Blue points are estimated mean call rates and nest densities for each site and year. The light grey uncertainty band represents the 95% credible interval for a function to estimate overall density estimation from call rates, and the dark grey band represents the 90% credible interval. Black error bars on each point represent 95% credible interval for mean call rate estimation (horizontal lines) and mean density estimation (vertical lines) at each site.

Discussion

The soundscape of Midway Atoll is dominated by vocalizations from individuals of multiple species that breed in high densities on the island, including the largest nesting population of Laysan Albatross and one of the largest breeding populations of Bonin Petrels in the world. Despite the complex soundscape, we were able to detect two Bonin Petrel calls with our DNN classifiers. Both calls produced similar diel and seasonal patterns as well as similar measures of relative abundance.

We found no relationship between burrow density and either the growl or aerial call rates within 5 to 30m of the sensor. This result differs from similar studies of several other colonial seabirds where a significant functional relationship has been found between call rates and breeding densities, including Forster's Tern (*Sterna forsteri*, Borker et al. 2014), Cory's Shearwater (*Calonectris borealis*, Oppel et al. 2014), Black Noddy (*Anous minutus*, Hemson unpublished data), Brown Booby (*Sula leucogaster*, Hemson unpublished data), and Wedge-tailed Shearwater (*Ardenna pacifica*, Hemson unpublished data, McKown unpublished data). One difference may be that *Pterodroma* petrels often call while on the wing, as well as on the ground and inside burrows. Individuals vocalizing while flying often transit large loops over the colony (>200m). Thus, if the signal being detected by passive acoustic surveys is produced by birds in the air instead of on the ground, this may preclude any relationship between call rates and breeding densities at smaller scales (i.e. ≤30m). Although we attempted to track a signal that was produced by birds on the ground (the growl call), observations by field personnel make it clear that this signal is produced both by birds in the air and on the ground. This behavioral difference may therefore mean that passive acoustic surveys are not effective for estimating the breeding densities of some species with aerial displays, or that any density relationship would likely occur over a larger spatial scale (>200m). One interesting result supports the possibility that aerial displays are contributing to call rates at our fixed survey sites - call rates were generally lower at survey points in forested areas (BOPE02 and BOPE03) with medium to high breeding densities, than in other more open habitat types (Figure 12). Conversely, low-density sites in open habitat (BOPE01, BOPE07) had much higher call rates than expected from burrow density alone. This supports anecdotal observations by field personnel who noticed that the sites in open areas often had elevated flight activity not paired with high levels of ground activity.

In other *Pterodroma* species, call rates have been successfully related to burrow density. For Black-capped Petrels in the Dominican Republic and Hawaiian Petrel on Kaua'i, burrow density within ~200 m of the acoustic sensor has been shown to be related to call rate; whereas, relationships were either not present or much weaker at smaller radii (Raine, personal communication 2018; McKown et al. 2015). However, there are fundamental differences in the colony structure and overall density of Bonin Petrels nesting at Midway compared to these two species. For example, both of the other petrel species currently breed in high mountainous areas with dense vegetation and mean burrow densities are several orders of magnitude lower than that found on Midway Atoll.

It would be interesting to try to test for a density relationship on Midway at larger spatial scales. However, the extremely fragile soil structure and relatively high burrow density make it problematic for field crews to count all burrows and estimate occupancy at greater than 30-meter radius from the acoustic sensor. In fact, burrow counts and occupancy estimates were difficult even at the 30m radius. We developed a sample-based method for censusing burrows in wedge shaped plot that allowed for several density estimates throughout the season. It is clear that at small scales (5 m), the wedge samples under-estimated occupied burrow densities compared to full circle counts. However, this sampling error decreased as the survey radius increased, which was likely a result of the increase in total area surveyed in the larger wedges. We predict that the estimated burrow density from the wedge plot method at larger radii (>=20 m) may accurately represent burrow density within the circle. That said, surveying wedges to a 200m radius is not feasible.

Finally, the activity data indicate several interesting patterns in the phenology of the species, diel attendance patterns, and the effect of moonlight on colony activity. In addition, we collected one other dataset, imagery from an infrared camera (FLIR Systems, Inc.) that was pointed skyward at each survey point. Bonin Petrels flying over the survey sites appear as clear thermal targets. It would be interesting to analyze these data and compare them to call rates at each site to see if both metrics are correlated.

Future Directions

Other call types:
Acoustic surveys produce a huge quantity of raw data. A benefit of this monitoring approach is that the raw data can be archived, and reanalyzed with the latest technology or to answer new questions. We analyzed recordings for two call types, however, Bonin Petrels make several other calls (Grant et al. 1983) and there is a possibility that one of these other calls may show a relationship with burrow density. However, as these call types are used less frequently, creating a functional detection model could require a significant time investment to find enough examples for a training dataset.

Chick calls might also be a useful signal to track. For example, in some seabird species such as the Wedge-tailed Shearwater (*Ardenna pacifica*), chicks are vocal enough to be detected and call rates have been shown to be correlated with burrow densities (unpublished data Queensland Parks and Wildlife Service). One challenge on Midway might be the difficulty of differentiating between calls of chicks from different species in this multi-species colony, as chick calls can vary as chicks mature.

The extent to which Bonin Petrel chicks call is unknown. It may be worth exploring whether chick calls can be detected on the existing recordings, because if they can be detected, we have an existing burrow density dataset to which we can compare chick call rates.

Additional questions:
We can think of three primary questions that arise from the current survey results:

1. **Is acoustic activity correlated with breeding density at larger spatial scales?** The pilot project found no relationship between call rates and Bonin Petrel burrow densities measured within 30m of the acoustics sensors. However, there was evidence that acoustic metrics were influenced by individuals vocalizing outside the 30m burrow count circles, including birds in flight and birds attending burrows outside the count

circle. Are Bonin Petrel call rates correlated with burrow densities over a larger area than the 30m radius circular plot used in the pilot study? And if so, at what distance?

2. **Are acoustic activity rates correlated with total individuals?** Both breeding and non-breeding individuals call during the breeding season. Are Bonin Petrel call rates correlated with the total number of Bonin Petrel individuals in a given area (i.e. as opposed to breeding pairs at burrow sites)? And, if so, could total individuals be a useful metric to monitor Bonin Petrel populations?

3. **Can seasonal activity patterns be used to monitor breeding phenology?** Season-long acoustic surveys can provide information about the initiation of breeding activity, end of breeding activity, and duration of peak activity. These metrics can be useful for tracking long-term phenology.

The first question, focused on exploring larger spatial scales, would require a new survey design to quantify burrow density and occupancy over larger spatial scales (Sand Island or a large section of Sand Island) coupled with more acoustic survey sites. Although logistically challenging, burrow densities could be estimated based on a series of random plots sampling burrow density across the different habitat strata/regions of the island, similar to the methods employed by Moore (2009) or Rayner et al. (2007). With acoustic sensors in the field at the same time as a properly designed large-scale burrow survey, we could test for a relationship between acoustic metrics and burrow densities at larger spatial scales (200m) than was investigated in the pilot study.

To address Question 2, acoustics would need to be coupled with a survey that accurately provided estimates of the number of individuals around a sensor. One way this could be undertaken is to couple a new survey with infrared cameras (FLIR) to quantify the number of individuals flying and on the ground in a given area.

It is more difficult to address question 3 because we would need a longer time series (number of seasons) to understand if we can detect meaningful changes in the timing and magnitude of calling activity, and if those differences are related to reproductive parameters of interest that can be used to address management goals. For example, is the timing of the acoustic peak in call rate during the incubation period earlier by 2 weeks in one year vs another, and does that predict reproductive effort or success?

Acknowledgements

Funding for this work was provided by USFWS Region 1 Inventory & Monitoring Program. Special thanks to Roberta Swift, Meg Duhr, Bob Peyton, and Kelly Goodale who supported this project. FWS staff & Volunteers at Midway Atoll National Wildlife Refuge made data collection and island life enjoyable during both field seasons.

Literature Cited

AMNWR (2015) Wildlife Inventory Plan, Alaska Maritime Wildlife Refuge, Protocol #13, Version 1.2. https://absilcc.org/science/amnwr/Protocols/13 Burrow Nester Populations Protocol v1.4.pdf. Accessed 15 Sep 2015

Borker AL, McKown MW, Ackerman JT, et al (2014) Vocal activity as a low cost and scalable index of seabird colony size. Conserv Biol 00:1–9. doi: 10.1111/cobi.12264

Bretagnolle V (1996) Acoustic communication in a group of nonpasserine birds, the petrels. In: Kroodsma DE, Miller EH (eds) Ecology and Evolution of Acoustic Communication in Birds. Comstock Publishing/Cornell University Press, pp 160–177

Buxton RT, Jones IL (2012) Measuring nocturnal seabird activity and status using acoustic recording devices: applications for island restoration. J F Ornithol 83:47–60.

Ciresan D, Meier U, Schmidhuber J, et al (2012) Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Providence, RI, pp 3642–3649

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. IEEE, Vancouver, BC, pp 8599–8603

Dufour O, Gineste B, Bas Y, et al (2016) First automatic passive acoustic tool for monitoring two species of procellarides (Pterodroma baraui and Puffinus bailloni) on Reunion Island, Indian Ocean. Ecol Inform 35:55–60. doi: 10.1016/j.ecoinf.2016.08.004

Grant GS, Warham J, Pettit TN, Whittow GC (1983) Reproductive Behavior and Vocalizations of the Bonin Petrel. Wilson Bull 95:522–539. doi: 10.2307/4161825

McKown M, Fleishman AB, Earl AD (2015) Acoustic Surveys for Black-capped Petrel on Hispaniola - November, 2015. 13.

Moore J (2009) A comparative analysis of population estimation methods for a burrow-nesting seabird: A novel ground-count method and closed population capture-recapture modelling. 1–135.

Oppel S, Hervías S, Oliveira N, et al (2014) Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. Nat Conserv 7:1–13. doi: 10.3897/natureconservation.7.6890

Pyle RL, Pyle P (2017) The birds of the Hawaiian Islands: Occurrence, history, distribution, and status, Verison 2. B.P. Bishop Museum, Honolulu, HI, U.S.A.

Rayner MJ, Clout MN, Stamp RK, et al (2007) Predictive habitat modelling for the population census of a burrowing seabird: A study of the endangered Cook's petrel. Biol Conserv 138:235–247. doi: 10.1016/j.biocon.2007.04.021

Seto NWH, O'Daniel DL (1999) Bonin Petrel (Pterodroma hypoleuca). Birds North Am Online. doi: 10.2173/bna.385