final version published in *Remote Sensing of Environment* https://doi.org/10.1016/j.rse.2018.02.071

A continent-wide search for Antarctic petrel breeding sites with satellite remote sensing

Mathew R. Schwaller[1], Heather J. Lynch[1], Arnaud Tarroux[2], Brandon Prehn[3]

[1]Department of Ecology & Evolution, Stony Brook University

[2]Norwegian Polar Institute and Norwegian Institute for Nature Research

[3]Faculty of Forestry, University of British Columbia

[1]Corresponding author: matt.schwaller@gmail.com, +1-301-848-1129

Abstract

The Antarctic petrel (*Thalassoica antarctica*) has been identified as a key species for monitoring the status and health of the Southern Ocean and Antarctic ecosystems. Breeding colonies of the Antarctic Petrel are often found on isolated nunataks far from inhabited stations, some up to hundreds of kilometers from the shoreline. It is difficult therefore to monitor and census known colonies, and it is believed that undiscovered breeding locations remain to be found. We developed an algorithm that can detect Antarctic Petrel colonies and used it to complete a continent-wide survey using Landsat-8 Operational Line Imager (OLI) imagery in Antarctica up to the southernmost extent of Landsat's orbital view at 82.68ºS. Our survey successfully identified 8 known Antarctic Petrel colonies containing 86% of the known population of Antarctic petrels. The survey also identified what appears to be a significant population of breeding birds in areas not known to host breeding Antarctic Petrel colonies. Our survey suggests that the breeding population at Mt. Biscoe (66˚13'S 51˚21'E), currently reported to be in the 1000s, may

CC-BY-NC-ND

actually be on the order of 400,000 breeding pairs, which would make it the largest known Antarctic petrel breeding colony in the world. The algorithm represents a first-ever attempt to apply satellite remote sensing to assess the distribution and abundance of the Antarctic petrel on a continent-wide basis. As such, we note several algorithm shortcomings and identify research topics for algorithm improvement. Even with these caveats, our algorithm for identifying Antarctic petrel colonies with Landsat imagery demonstrates the feasibility of monitoring their populations using satellite remote sensing and identifies breeding locations that should be considered high priorities for validation with directed field surveys.

Keywords: Antarctic petrel, flying seabirds, Landsat, Antarctica, guano, breeding colony

1. Introduction

Effective wildlife management relies on accurate population estimates, but it is a challenge to provide such estimates for seabird species inhabiting remote areas of the Southern Ocean and the Antarctic continent. In particular, continent-wide population estimates for the Antarctic petrel (*Thalassoica antarctica*) are highly uncertain, though the most recent population estimate (van Franeker et al., 1999) of 10-20 million individuals (4-7 million breeding pairs) makes them one of the most abundant birds in Antarctica (Harris et al., 2015). Current population estimates were compiled from ship-based observations in the Weddell Sea, Prydz Bay, and Ross Sea regions and extrapolated to other regions of the Antarctic petrel's circumpolar distribution. Although this population estimate represents the best available science on the abundance of this

CC-BY-NC-ND

species, the authors of the original survey readily acknowledged the "limitations in methodology and interpretation of at-sea censuses." In addition to considerable uncertainty surrounding Antarctic petrel abundance, there is tremendous uncertainty regarding the geographic distribution of their breeding sites. van Franeker et al. (1999) identified 35 breeding locations, but they account for only around 500,000 pairs (i.e., ~1 million breeding individuals), or under a quarter of the estimated total breeding population. This knowledge gap stems from the difficulty of accessing Antarctic petrel breeding locations and the logistical challenges of surveying birds that often nest on remote and inaccessible mountain slopes 200 km (or farther) from shore. Given that vast regions of Antarctica remain largely unexplored, coupled with the striking discrepancy between known breeding locations and estimates of total population, it is likely that a significant number of Antarctic petrel breeding locations remain to be found.

Interest in refining Antarctic petrel population estimates is motivated by their role as a top predator in the Southern Ocean where they act as a central-place forager of Antarctic krill (*Euphausia superba*), fish, and squid during the breeding season (Ainley et al. 1992; Descamps, 2016a; Hodum and Hobson, 2000; Lorentsen et al., 1998; Nicol 1993). Indeed, it has been estimated that 680,000 metric tonnes of krill are consumed by Antarctic petrels each year (Descamps, 2016b). This amount is comparable to the southwest Atlantic's krill "trigger limit" of 620,000 tonnes set by the member nations of the Convention for the Conservation of Antarctic Marine Living Resources (CCAMLR). CCAMLR established such limits to ensure a sustainable krill stock for both fisheries and krill-dependent species (CCAMLR, 2016). To that end, the Antarctic petrel was identified as one of several key "dependent species" included in the CCAMLR

CC-BY-NC-ND

Ecosystem Monitoring Program (CEMP). While CEMP recognizes the need to monitor population changes in this species, it also recognizes that the distribution and abundance of the Antarctic petrel are not well understood (Kock, 2000).

Satellite remote sensing may play an important role in establishing a continent-wide baseline on the distribution and abundance of the Antarctic petrel. Remote sensing is an established tool used to locate and inventory Emperor and Adélie penguin breeding locations (e.g., Fretwell and Trathan, 2009; Schwaller, et al., 2013; Lynch and LaRue 2014). While the technical feasibility of using satellite imagery to identify flying seabird colonies has been recently demonstrated in a study of a single Landsat scene (Fretwell et al., 2015) the use of satellite remote sensing for large-scale Antarctic petrel surveys has not been explored.

Antarctic petrel colonies are visually similar in many ways to penguin colonies but there are some important differences between the two that affect the accuracy and errors of detection. Like those of penguins, Antarctic petrel colonies can be very large and, when closely packed nests are situated on the surface of exposed rock outcrops, they can be identified in satellite images by the associated guano stain. On the other hand, Antarctic petrel colonies are found on steeper slopes than penguins and are therefore subject to more weathering and shadowing, which can obscure the guano stain and decrease the probability of detection. Furthermore, the petrel diet can have a higher proportion of fish-to-krill than the penguin diet (Descamps et al., 2016a; Hodum and Hobson, 2000) and petrel guano could thus be deficient in the chemical constituents (krill carotenoids and chitin) that make penguin guano such a unique and easily detectable target with remote sensing methods.

CC-BY-NC-ND

In this paper, we report on the first attempt to retrieve Antarctic petrel breeding locations from satellite data at the continental scale. We report on the theoretical basis of our Antarctic petrel detection algorithm, the expected colony detection errors of commission and omission, and the overall results of our continent-wide survey using Landsat-8 Operational Line Imager (OLI) data. We provide supplementray materials that allow others to reproduce and verify our results. These materials also identify a large number of potential Antarctic petrel breeding colonies. Thus, this work represents the first steps toward the discovery of previously unknown Antarctic petrel colonies and toward a more comprehensive, routine monitoring of this key Antarctic species.

2.0 Methods

The satellite dataset. Figure 1 llustrates the 1098 locations of the 3944 Landsat-8 Operational Land Imager (OLI) scenes collected for the analysis reported here. The data set consists of scenes acquired between 16 November 2013 and 28 March 2016 from 60.44ºS to the southernmost extent of the Landsat-8 orbital view at 82.68ºS. Only scenes covering known rock outcrops as defined by the Antarctic Digital Database (http://www.add.scar.org) were selected for analysis because the Antarctic petrel does not nest on snow surfaces (Mehlum et al., 1988). Even with this initial data reduction, the resulting dataset consisted of ~4 Tbytes of imagery or approximately $1.5x10^{14}$ pixels.

CC-BY-NC-ND

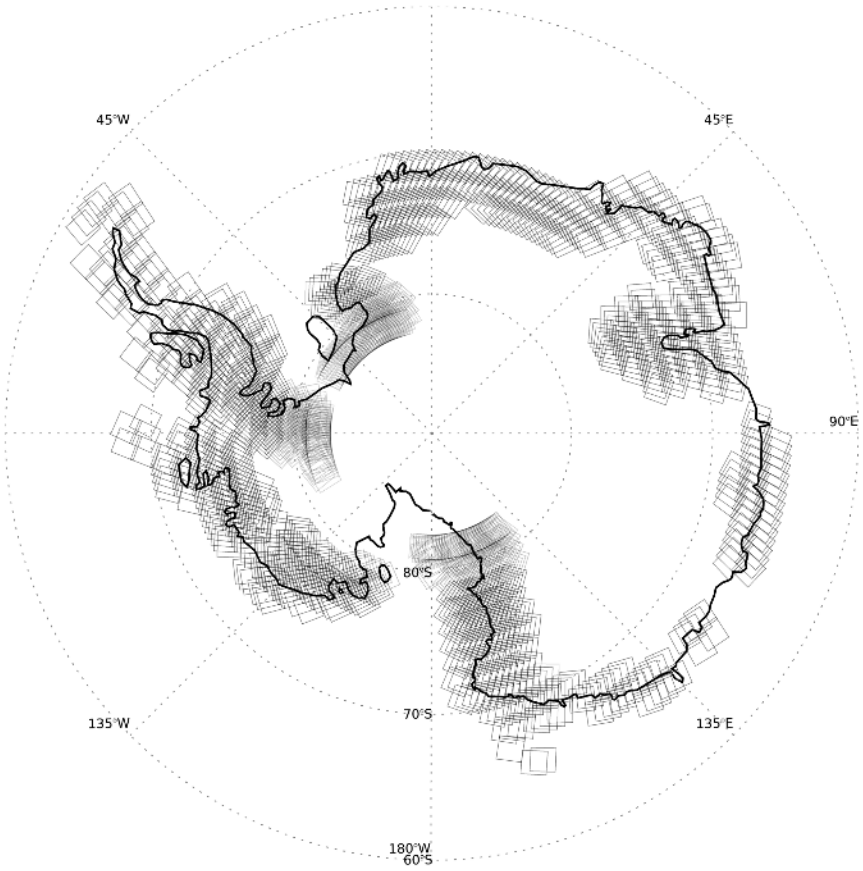

Figure 1. Location of Landsat-8 Operational Line Imager (OLI) imagery used to identify Antarctic petrel nesting areas. The dataset extends to the southern limit of Landsat's orbital view which can be observed in the figure as a broken ring of scene locations at about 82ºS, south of which no imagery is available. Imagery was selected only over known rock outcrops which accounts for the other gaps in data selection.

The search algorithm. The algorithm used to search for Antarctic petrel nesting colonies is based on an Adélie penguin colony search algorithm (Schwaller et al., 2013; Lynch and Schwaller, 2014). The algorithm uses 6 reflectance bands from the Landsat-8 OLI: the blue, green, red, near infrared, shortwave infrared-1, and shortwave infrared-2

CC-BY-NC-ND

bands. The algorithm transforms the 6 reflectance bands of each pixel into 5 spherical coordinate bands since doing so was found to improve the separation of penguin colony pixels from other targets (Schwaller et al., 1989). Using the transformed data, a training set of pixels was selected from known Antarctic petrel breeding areas. The algorithm calculates a ellipsoidal surface in the transformed data space that optimally separates Antarctic petrel colony pixels from pixels selected over other targets, including rock outcrops and soils of various types, snow, ice, and open water (see Schwaller et al. 2013 for details on data transformation and optimization). The equation of the transformed ellipse can be summarized as a 6 by 6 transformation matrix, one dimension greater than the number of spherical coordinate bands. The transformation matrix used in this study is provided in the Appendix. The algorithm retrieval can be reproduced using any Landsat-8 OLI image, the transformation matrix, and the additional steps described in the Appendix.

The training set used to calculate the transition matrix consisted of 206 pixels from three Antarctic petrel colonies at Svarthamaren (135), Mt. Biscoe (61) and Mt. Paterson (10). In addition, 29912 pixels were collected over other ground targets. These samples included areas of soil and rock collected near the research stations San Martin, Mendell, Marimbio, and Davis; pixels collected in the vicinity of Cape Bird, Cape Hallet and Cape Crozier (but not over the Adélie penguin colonies at these sites); and pixels collected in the Price Charles Mountains around 70˚48'S 68˚12'E. Additional pixels were collected for the training set over ice and snow in the vicinity of the Mawson and Juan Carlos research stations. Pixels over snow, ice and rock were collected in the vicinity of Cape Adare (again, not over the penguin colony) and in east Antarctica around 67˚24'S

CC-BY-NC-ND

59˚24'E. The results of the training found that 201 of the Antarctic petrel colony pixels were correctly classified and 5 incorrectly classified as "other ground type;" 29911 of the other ground type pixels were correctly classified and 1 of the other ground type pixels was classified as belonging to the Antarctic petrel colony class. This classification yielded a correlation of 0.985 as calculated by Kendall's tau, which is significant at $p < 0.0001$.

Slope and elevation filtering. The ASTER global digital elevation model (GDEM; Rees, 2012) was used to help filter the pixels identified by the algorithm as belonging to the Antarctic petrel colony class. Filtering was necessary because of the relatively high errors of commission associated with the algorithm. The filter exploits the knowledge that Antarctic petrels nest on elevated cliffs and mountain slopes (Mehlum et al., 1988). The GDEM was used to select Landsat OLI pixels located on surfaces with a slope ≥17˚ and with an average height above terrain (HAT) of 20m. HAT was calculated by selecting a 15 by 15 ASTER GDEM pixel region centered around a given Landsat-8 pixel, then subtracting the center pixel elevation from the average elevation of GDEM pixels in the lowest 20th percentile of the 15x15 pixel region. The GDEM filtering removed flat areas and areas in depressions from consideration as potential Antarctic petrel nesting areas.

Slope calculation. The ASTER GDEM is stored as 1º by 1º tiles, with 3601 center-referenced pixels per tile. The pixel dimension in the latitude direction is constant (approximately 31 m) but the pixel dimension in the longitude direction varies with latitude. Pixel dimension in the longitude direction is approximately 13 m at 65ºS latitude

CC-BY-NC-ND

but the dimension decreases to approximately 5 m at 80ºS. The slope corresponding to a given Landsat pixel was calculated as the vector sum of the ASTER GDEM slope in the latitude and longitude directions closest to the center coordinate of the Landsat pixel, see Figure 2. As the figure illustrates, two rows of ASTER pixels closest to the center of the Landsat pixel were selected. These two rows of 31 m GDEM pixels represent 62 m in the longitude (*y*-coordinate) direction. Columns of ASTER GDEM pixels were selected adjacent to the Landsat center coordinate as needed to complete an approximately 62 m grid of pixels in the latitude (*x*-coordinate) direction. The slope in the longitude (*y*) direction was calculated as the average elevation difference between pairs of ASTER GDM pixels in each column. The slope in the latitude (*x*) direction was calculated as the average slope computed by a regression of elevation on distance in each row of ASTER GDEM pixels.

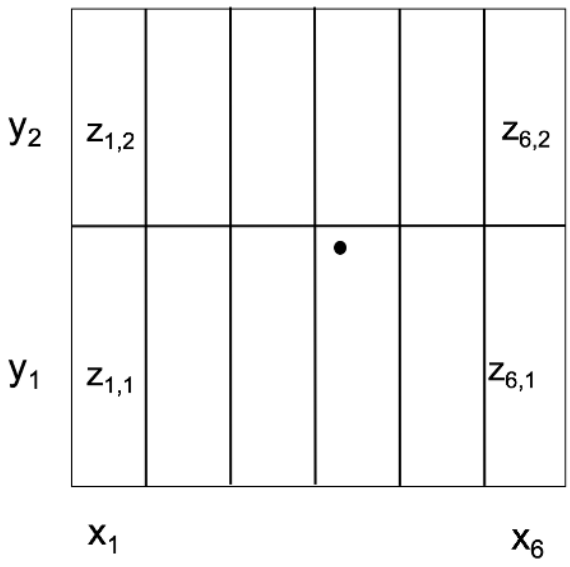

Figure 2. Slope was calculated for the center coordinate of a given Landsat pixel (represented by the dot in the middle of the figure above) using elevation data ($z_{i,j}$) from

CC-BY-NC-ND

the ASTER GDEM. GDEM pixels are represented as rectangles in the figure above because the pixel length in the latitude (y) direction is constant (each ~31 m), but the pixel width in the longitude (x) direction varies with distance from the pole. Two rows of ASTER GDEM pixels were used to calculate slope (~62 m in the y-direction). As many columns of GDEM pixels were used as needed to fill the grid to approximately 62 m in the x-direction, with the grid in this direction centered on the Landsat center coordinate.

CC-BY-NC-ND

3.0 Algorithm Constraints

Given the relatively large dataset of ~150 Gpixels, even small algorithm errors can result in a large number of incorrect classifications when pixels are assigned to the Antarctic petrel "colony" or "not-colony" class. Figure 3 illustrates how certain soils can end up misclassified as Antarctic petrel colony locations. These soils tend to form in depressions and are typically associated with saline deposits (Hambry et al., 2007). They were often misclassified by the algorithm as petrel colony locations. The plots illustrate how the population of pixels collected over these soils intersects with the Antarctic petrel colony pixels. This indicates that there will always be some cases where Landsat observations over these soils will be misclassified as belonging to the Antarctic petrel colony class.

Figure 4 is another geometric example of how the algorithm treats pixel classification success and error. In this case, an ellipsoid was generated to surround training data from known Adélie penguin and Antarctic petrel colonies. The pink- and purple-colored points in the figure represent transformed Landsat-8 OLI reflectance values from several Adélie penguin colonies (Cape Adare, Cape Bird, Cape Crozier, and Cape Hallet) and the green-colored points represent pixel values from Antarctic petrel colonies (Jutulsessen, Mt. Biscoe, Mt. Paterson, and Svarthamaren,). The figure illustrates how pixel values from soil samples (red orbs) are present within the ellipsoid. The soil data points within the ellipsoid can be considered errors of commission since they would be erroneously identified as belonging to the colony class. The figure also indicates a degree of inter-mingling between the Adélie and petrel pixels. Thus, we

CC-BY-NC-ND

expect the classification algorithm to erroneously classify some number of pixels over Adélie colonies as belonging to the Antarctic petrel colony class.

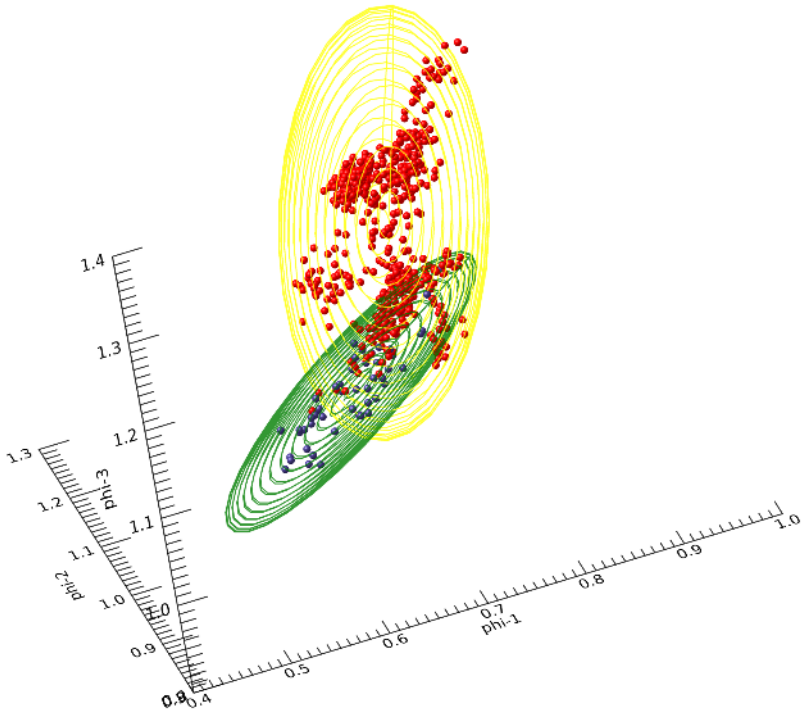

Figure 3. This figure plots the ellipsoid (yellow) that surrounds petrel and Adélie pixels (red points) in 3 of the 5 dimensions of transformed Landsat-8 OLI reflectance data space. Also plotted is an ellipsoid (green) that surrounds data points (blue) collected over soil locations in the Prince Charles mountain rage. Note how the yellow and green ellipsoids intersect, which indicates that some of the pixels imaged over soil sites will end

CC-BY-NC-ND

up classified as belonging to the petrel colony class. An animation that plots all of the dimensions is included with the supplementary materials.

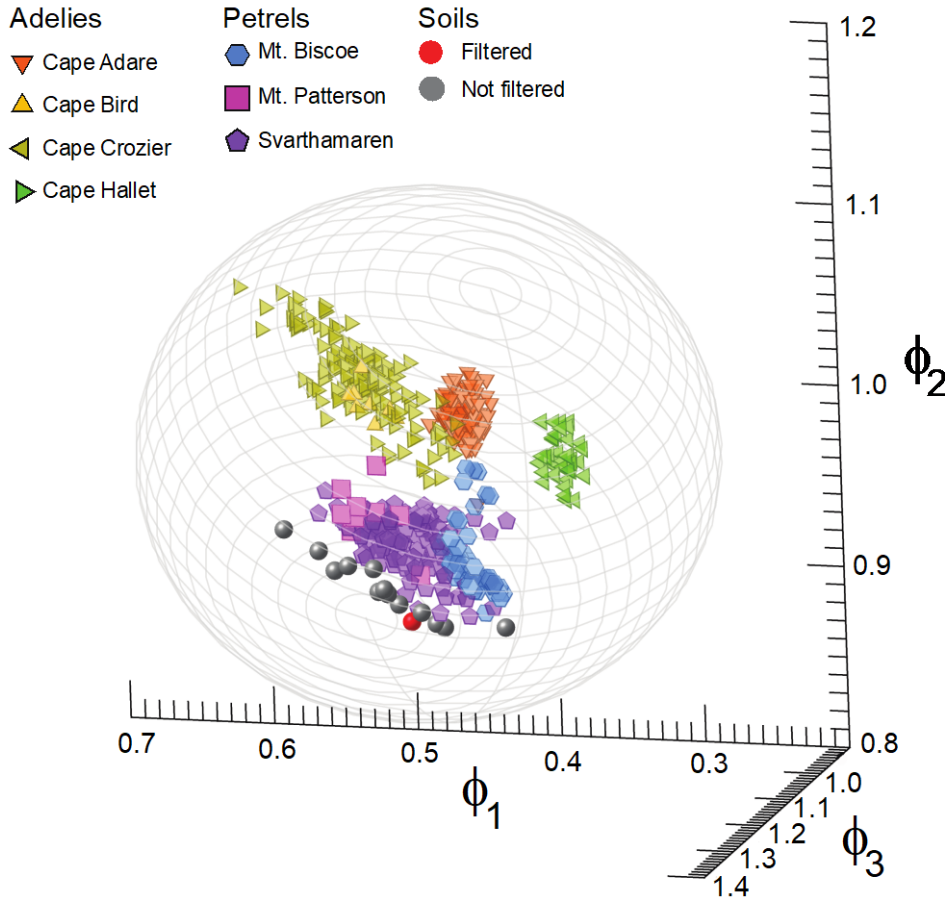

Figure 4. Geometric ellipses surrounding Adélie and Antarctic petrel colonies in the transformed data space. Soil samples from the Prince Charles Mountains are illustrated as grey orbs if they could be eliminated by elevation and HAT filtering or by a single red orb in the case where the sample was not identified by filtering. Note that the Adélie data points fall closest to the centroid of the ellipse (and overlap the petrel data points

CC-BY-NC-ND

somewhat) while the petrel data points fall closer to the margin and appear to mix with the soil samples.

Figure 4 also illustrates the effectiveness of slope and elevation filtering to reduce errors of commission. The 35 red orbs in Figure 4a represent Landsat pixels collected over non-colony soils that fall within the colony classification space. Figure 4b identifies the 3 red orbs remaining after filtering using the slope and HAT criteria. The filtering criteria thus eliminated ~90% of the errors of commission associated with pixels collected over soils in the Prince Charles mountains that were incorrectly assigned to the colony class.

Following the slope and HAT filtering, pixels were grouped into coherent colonies using the clustering algorithm described by Schwaller et al. (2013). Only colonies with 2 or more pixels were retained for further analysis because many of the 1-pixel colonies were likely to be errors of commission. The resulting colonies were analyzed visually and any that were identified as clouds or cloud-covered were also removed from further analysis.

4. Results

Landsat classification and subsequent filtering identified 3227 Landsat-8 OLI pixels clustered into 328 colonies as belonging to the Antarctic petrel colony class. Records for with each of these pixels are available as a supplementary dataset. Each record includes the following data fields: a colony identification number, a Landsat source file identifier, the radial distance of the pixel in the transformed data space (see the Methods section), the scene row and column indices, the latitude and longitude of the

CC-BY-NC-ND

Landsat pixel's center coordinates (in both northing/easting and geographic latitude/longitude), the pixel's associated elevation and slope, the pixel's HAT, an ASTER GDEM scene identifier, the ASTER GDEM stacking number, and the 6 Landsat OLI reflectance values associated with the pixel. Details on each data field are included in the file header.

Figure 5 shows the continent-wide distribution of pixel clusters identified by the algorithm as belonging to the Antarctic petrel class, together with locations where Antarctic petrels have been reported from ship-based observations. The locations of the colonies retrieved from the Landsat OLI data are available as a Google Earth kmz file which is included in the supplementary material. The kmz file includes images of the colonies derived from the Landsat data as well as summary information including colony latitude, longitude, and colony area.

CC-BY-NC-ND

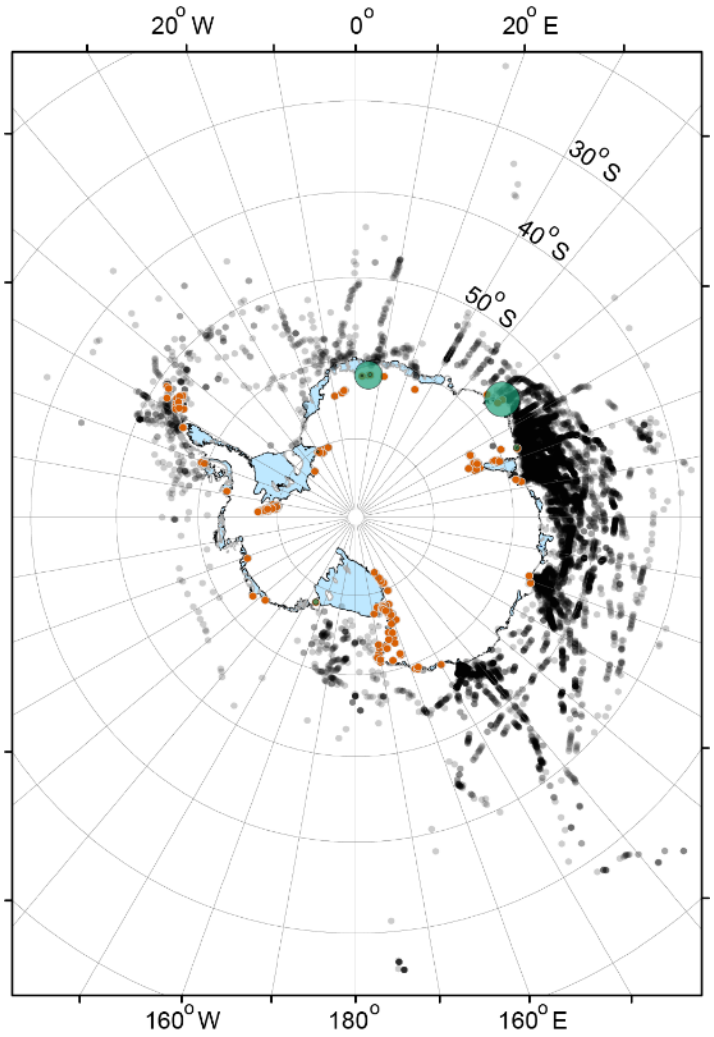

Figure 5. Continent-wide distribution of the possible Antarctic petrel colonies identified by the Landsat algorithm (orange dots), with the locations of Svarthamaren and Mt. Biscoe colonies highlighted in green. Locations of Antarctic petrels reported from ship-based observations (GBIF Secretariat, 2017) are also identified (locations indicated by black dots with partial transparency to facilitate interpretation in areas where sightings overlap). Note that there is a significant observation bias in the at-sea observations with many more ship-based transects conducted in East Antarctica than around the rest of the continent.

CC-BY-NC-ND

final version published in *Remote Sensing of Environment* https://doi.org/10.1016/j.rse.2018.02.071

Table 1 identifies the Antarctic petrel breeding colonies listed in van Franeker's 1999 compendium that were identified by the Landsat classification algorithm and the population in breeding pairs associated with each colony. A total population figure of 495,162 pairs for all Antarctic petrel breeding colonies was calculated by adding breeding populations of each individual colony listed by van Franeker, substituting 3000 pairs for the two colonies listed as "1000s," and substituting 50 pairs for the two colonies listed as "10s." The colonies identified by the Landsat algorithm and listed in Table 1 account for 425,500 breeding pairs out of this total, thus the algorithm identified Antarctic petrel colonies that make up 86 percent of the population reported by van Franeker et al. (1999).

CC-BY-NC-ND

final version published in *Remote Sensing of Environment* https://doi.org/10.1016/j.rse.2018.02.071

Table 1. This table identifies the Antarctic petrel breeding colonies that were identified by the Landsat classification algorithm. The population in breeding pairs associated with each colony is taken from van Franeker (1999); the total population of all Antarctic petrel breeding colonies cited by this reference is 495,162 pairs. The algorithm identified colonies associated with 425,500 pairs, about 86% of the total breeding population. The identification numbers (ID column) can be used to find these colonies in the Google Earth kmz file available as a supplementary dataset.

Colony Name	Population	ID	#Pixels
Svarthamaren	250,000	S1-4	188
Scullin Monolith	157,000	S1-59	12
Mt. Paterson	10,000	S1-72	9
Murray Monolith	3,500	S1-287	2
Mt. Biscoe	1,000s	S1-2	306
Jutulsessen West	1500	S1-288	2
Kvitholten	300	S1-130	4
Mt. Provender	200	S1-228	2

CC-BY-NC-ND

5. Discussion

Satellite remote sensing was employed here in a first-ever attempt to assess the distribution and abundance of the Antarctic petrel with this technology. The methods described in this paper enabled us to perform a continent-wide retrieval of colony locations from Landsat-8 Operational Line Imager (OLI) satellite data. Previous methods for assessing Antarctic petrel populations in Antarctica relied on discovering individual colonies and amassing counts, or on extrapolating counts made from ship-based transects. Both of the traditional methods have their limitations. Due to Antarctica's vast and inhospitable terrain it would be impossible for expeditions to systematically explore the isolated mountain tops and nunataks where these birds often nest. Indeed, many of the known Antarctic petrel colonies were found during geologic field campaigns or other expeditions that were not primarily focused on biology, and it is believed that many have yet to be discovered. Similarly, population estimates based on at-sea observations have significant limitations. Ship cruises are expensive to undertake, by necessity they are a "snapshot" in time that may not be optimal for population estimation, and they also require an extrapolation from a narrow transect to an entire sea or even the entire continent. Of course, remote sensing methods have their own set of limitations, in particular, a high commission error that results in a large number of sites erroneously identified as Antarctic petrel colonies. However, there are significant benefits to adding satellite data analysis to the set of tools used in monitoring Antarctic seabirds. Landsat-8 collects images over the entire Antarctic continent every year, typically with many cloud-free scenes over any given site. Using these data for continent-wide exploration is relatively easy, and once a baseline of colonies is established it can be easily monitored.

CC-BY-NC-ND

Even if the commission rate is high, it should be noted that we began the investigation with 3944 Landsat-8 OLI scenes consisting of some $1.5x10^{14}$ pixels and ended up with a set of just 3227 pixels clustered into 328 possible colonies. This represents a 10 orders-of-magnitude reduction in search volume. Thus, satellite remote sensing methods can turn the search for new colonies from an impossible problem to one that is systematic and manageable.

The Antarctic petrel colony search algorithm generally performed well in terms of finding the largest of the known Antarctic petrel colonies, and therefore the vast majority of the breeding population. In general, seabird populations follow a log-normal distribution where the majority of the population is concentrated in a small number of colonies (Forbes et al., 2000; Grecian et al., 2012; Jovani et al., 2008). This is certainly true in Antarctica where 90% of the Adélie penguin population, for example, is concentrated in 28% of the colonies (Appendix A of Lynch and LaRue, 2014). Based on previous records of colony sizes (van Franeker, 1999) we can assume Antarctic petrels also display a long-tailed colony size distribution. We suggest that remote sensing can be combined with ground-based or other methods to identify and track the continent-wide distribution and abundance of the largest Antarctic petrel colonies. In this case it should be possible to assess population trends over time with far greater precision than has been possible to date.

Although the algorithm missed Antarctic petrel colonies associated with 14 percent of the currently estimated population it is likely that the algorithm found a significant number of unreported breeding pairs. In particular, Mt. Biscoe in Enderby Land (66˚13'S 51˚21'E) appears to support a much larger breeding population than the

CC-BY-NC-ND

1000s currently reported in van Franeker's (1999) compendium (Table 1). The Landsat algorithm identified 306 pixels on Mt. Biscoe and nearby highlands as belonging to the Antarctic petrel breeding colony class. This is nearly 60 percent more than the number of pixels associated with the colony at Svarthamaren, currently the largest known Antarctic petrel colony, with a population estimated at 250,000. The population figure of 1000s in van Franeker's (1999) compendium is attributed to Bassett et al. (1990). These authors reported on their visit to Mt. Biscoe on 27 and 29 October 1985, which is significantly earlier than the Antarctic petrel breeding season that, in other colonies, begins at the end of November (Lorensen and Røv, 1995; van Franeker et al., 2001). Even with this limitation, Bassett et al. (1990) observed "vast numbers" of Antarctic petrels although from their vantage point "many sites were probably hidden by the aspect and rocky terrain," and only a limited time was available to assess the area occupied by these birds. Indeed, they conclude that "our observations suggest that a substantial number [of Antarctic petrels] may breed in the Mount Biscoe area." These observations are supported by Falla's (1937) report (p. 157) of the Mawson expedition: "When off Cape Ann, Enderby Land, on 14 January, 1930, we were visited by large flocks of *Thalassoica* coming from the direction of the land … many of them came from the direction of Mt. Biscoe, from discoloured patches on the rocky sides of which clouds of birds were seen rising." This is further supported by Mawson's (1930) own account: "Most of the northerly face [of Mt. Biscoe] is encrusted with guano, for countless flocks of Antarctic petrels and other sea birds resort there during the nesting season." Given these observations and our algorithm's identification of extensive guano-like signatures in the available Landsat imagery, we conclude that the Antarctic petrel population on and

CC-BY-NC-ND

around Mt. Biscoe is substantially greater than current estimates. Assuming a nest density equal to that at Svarthamaren (250,000 breeding pairs over 188 pixels) we suggest that the Antarctic petrel breeding population at Mt. Biscoe (306 pixels) is on the order of 400,000 pairs. Additional efforts are needed to validate of this assertion.

The algorithm also located an Antarctic petrel colony on Mt. Provender (80˚23'S 29˚57'W, colony S_228, 2 pixels) that had previously been reported by sightings of birds in flight (Wright and Wyeth, 1971), but the colony has never been actually observed or visited. The precise location provided by the algorithm enables future expeditions to the area to validate the existence of the colony.

Although the algorithm successfully identified a number of Antarctic petrel colonies as noted above, a large fraction of the 328 identified "colony hits" are almost certainly errors of commission. For example, there are a large number such hits in the Prince Charles Mountains in Mac. Robertson Land, starting at 70˚S and extending to 74˚S more or less along 68˚E longitude. Indeed, the most southerly of these hits (colony S1_111 at 74˚21'S 66˚39'E, see the supplementary kmz file) is 675 km from open water and is very unlikely to be a bird colony. As noted in the Methods section, there are soil types in this region that the algorithm often mistakes for Antarctic petrel colonies. On the other hand, bird colonies (e.g., snow petrels *Pagodroma nivea*) are known to exist in the Pagodroma Gorge of the Prince Charles Mountains, 250 km from shore (Brown, 1966; Heatwole et al., 1991) and a snow petrel colony was found as far south as the Greenall Glacier (73˚15'S), which is ~440 km from the coast (Goldsworth and Thomson, 2000). There are even reports of bird sightings as far south as 73˚39'S 68.25'E – 500 km from open water in Prydz Bay (Kuehn, 1998). Thus, although most of the hits identified in the

CC-BY-NC-ND

Prince Charles Mountains are almost certainly errors of commission, it is possible that some of them are legitimate colony locations. New methods are required to validate, verify, and otherwise separate commission errors from actual colony locations.

Sites identified by the algorithm in a number of other areas are also likely errors of commission. A total of 84 locations in the Ellsworth Mountains were identified as colony hits, but even the northernmost of these is >400 km from open water. Furthermore, we found no reports in the literature of Antarctic petrel sightings in this area, and no Antarctic petrel colonies or birds in flight were observed by private expeditions conducted between Patriot Hills (80˚18'S 81˚21'W) and Vinson Massif (78˚52'S 85˚37'W, David Rootes personal communication). Similarly, there is a set of 8 hits, one with 60 pixels, located in the vicinity of 73˚50'S 5˚12'W in Dronning Maud Land that are ~350 km from open water. A cursory examination suggests that these may be alluvial soils or melting snow drainage features rather than seabird breeding colonies. Additional commission errors are associated with Adélie penguin colonies that were incorrectly assigned to the petrel colony class. These can be identified in the kmz visualization: the ground overlay images include green boxes that identify Adélie colony locations. Adjacent to the boxes are 4-letter colony codes that uniquely identify the colony name (for colony codes see Lynch and LaRue, 2014; Appendix 2).

Finally, the algorithm identified many hits in the Transantarctic Mountains of Victoria Land. In the nearby Ross Sea, large numbers of Antarctic petrels were found in ship-based observations with a regional population estimated at ~5 million birds (Ainley et al., 1984). These authors note that densities of this species tend to be highest in the central and eastern parts of the Ross Sea (from 76° S, 170° W to 73° S 175° E), and the

CC-BY-NC-ND

observed densities fall off closer to the Victoria Land coastline. While many, if not most, of the areas identified as guano in the Transantarctic Mountains are likely errors of commission, the large populations observed in the Ross Sea suggest that some of these hits may be in fact be previously undiscovered Antarctic petrel colonies. Some of these hits may also identify colonies of other flying seabird species. As mentioned above, further work is needed to separate commission errors from true Antarctic petrel colony locations.

Our algorithm's high rate of commission error is partly by design. In developing the algorithm we tried to minimize the classification errors of *omission* at the cost of increased errors of *commission*. The goal was to minimize the number of colonies missed by the algorithm even if a relatively large number of colony “hits” turn out to be false positives that need to be evaluated and eliminated by further examination. An additional weakness is that the algorithm appears to under-perform in moderate-sized colonies. For the colonies at Murray Monolith, Scullin Monolith, Mt. Paterson and Jutulsessen, for example, the number of pixel “hits” appears to be significantly smaller than what would be expected given the reported breeding populations at these sites. This under-performance could also be a result of the DEM filtering that we apply to the pixels retrieved by the algorithm.

The algorithm errors observed in the retrieval of Antarctic petrel colonies are considerably greater than the 1% commission and 3-4% omission errors found in similar algorithm retrievals of Adélie penguin colonies (Schwaller et al., 2013). This disparity may be partially a consequence of differences in the diet of these two species. While difference in diet composition between these two krill predators is difficult to quantify

CC-BY-NC-ND

and varies regionally, Antarctic petrels may, on average, consume a greater percentage of fish than Adélie penguins. The Adélie's high-krill diet results in guano that is rich in undigested krill carotenoid and chitin that, when deposited in sufficiently high concentrations, is relatively easy to identify by spectrophotometric methods (Rees et al., 2017). Lower concentrations of distinctive chitin and carotenoid, and higher concentrations of hydroxyapatite, the primary mineral constituent of bone and fish scales, may explain why petrel guano looks more "soil-like" than penguin guano in a spectrophotometric sense (Figure 4), and is therefore more subject to classification errors.

Given the classification errors noted above there is clearly room for algorithm improvement. The current algorithm is based on physical principles but it fundamentally relies on training data and assumptions about breeding biology. The algorithm would benefit from a better characterization of the physical and biological properties of Antarctic petrel breeding areas. Spectrophotometric measurement of guano-covered and guano-free soils in and around the colonies would help to establish a better physical basis for the algorithm. Such measurement may also suggest techniques to help reduce commission errors without adding to omissions. Better biological characterization of Antarctic petrel breeding areas is also needed. Nest density measurements would help provide evidence for extrapolation from pixel area to population. Assessment of dietary preferences (especially fish to krill ratios) may help explain algorithm under-performance. New methods are also needed to verify and validate the 328 possible colonies identified by the Landsat algorithm. High-resolution commercial satellite data may well play a role in this process, although the spatial resolution of this imagery is problematic because the (guano) signal to noise ratio in each individual pixel is low and

CC-BY-NC-ND

thus pixels are difficult to classify; as such, automated methods for interpretation have not yet been developed. These limitations may be overcome with imagery from long-range drones that can have an arbitrarily fine spatial resolution, although there may be operational issues associated with flying in mountainous areas and imaging the cliff sides where these birds often nest.

In presenting this version of the retrieval algorithm we make the results available as supplementary materials, and in the Appendix we describe how these results may be reproduced. By doing so we encourage others to verify and validate our results, to contribute new results that will lead to algorithm improvement, and to employ such improvements in a successively more precise understanding of the distribution and abundance of the Antarctic petrel.

Acknowledgements

We wish to thank Woody Turner and Cindy Schmidt of NASA's Ecological Forecasting Program for funding and for enlightened management of this work under grant NNX14AC32G. We also thank Andy Michaelis from NASA Ames Research Center for help in navigating the Pleiades supercomputer, and Joey Griebel from the Harris Corporation for help with licensing ENVI+IDL on Pleiades.

References

CC-BY-NC-ND

Ainley, D.G., O'Connor, E.F. and Boekelheide, R.J. (1984). The marine ecology of birds in the Ross Sea, Antarctica. Ornithological monographs, 32, iii-97.

Ainley D.G., Ribic, C.A. and Fraser W.R. (1992). Does prey preference affect habitat choice in Antarctic seabirds? Marine Ecology-Progress Series, 90, 207-207.

Bassett, J.A., Woehler, E.J., Ensor, P.H., Kerry, K.R. and Johnstone, G.W. (1990). Adélie penguins and Antarctic petrels at Mount Biscoe, western Enderby Land, Antarctica. Emu-Austral Ornithology, 90, 58-60.

Brown, D.A. (1966). Breeding biology of the snow petrel *Pagodroma nivea* (Forster). ANARE Science Report Series B, 89,1-63.

CCAMLR 2016. Krill Fishery Report. (2016). Commission for the Conservation of Antarctic Marine Living Resources. https://www.ccamlr.org/en/document/publications/krill-fishery-report-2016

Descamps, S., Tarroux, A., Lorentsen, S.H., Love, O.P., Varpe, Ø., and Yoccoz, N.G. (2016b). Large-scale oceanographic fluctuations drive Antarctic petrel survival and reproduction. Ecography, 39, 496-505.

CC-BY-NC-ND

Descamps, S., Tarroux, A., Cherel, Y., Delord, K., Godø, O.R., Kato, A., Krafft, B.A., Lorentsen, S.H., Ropert-Coudert, Y., Skaret, G. and Varpe, Ø. (2016a). At-sea distribution and prey selection of Antarctic petrels and commercial krill fisheries. PloS one, 11, p.e0156968.

Falla, R.A. (1937). BANZ Antarctic Research Expedition 1929-1931. Reports-Series B, 11, 1-288.

Forbes, L.S., Jajam, M., and Kaiser, G.W. (2000). Habitat constraints and spatial bias in seabird colony distributions. Ecography, 23, 575-578.

Fretwell, P.T., Phillips, R.A., Brooke, M.D.L., Fleming, A.H., and McArthur, A. (2015). Using the unique spectral signature of guano to identify unknown seabird colonies. Remote Sensing of Environment, 156, 448-456.

Fretwell, P.T., and Trathan, P.N. (2009). Penguins from space: Faecal stains reveal the location of emperor penguin colonies. Global Ecology and Biogeography, 18, 543–552.

GBIF Secretariat records for *Thalassoica antarctica* (Gmelin, 1789). (2017). GBIF Backbone Taxonomy. Checklist Dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2017-10-20.

CC-BY-NC-ND

Goldsworthy, P.M. and Thomson, P.G. (2000). An extreme inland breeding locality of snow petrels (*Pagodroma nivea*) in the southern Prince Charles Mountains, Antarctica. Polar Biology, 23, 717-720.

Grecian, W.J., Witt, M.J., Attrill, M.J., Bearhop, S., Godley, B.J., Grémillet, D., Hamer, K.C. and Votier, S.C. (2012). A novel projection technique to identify important at-sea areas for seabird conservation: An example using Northern gannets breeding in the North East Atlantic. Biological conservation, 156, 43-52.

Hambrey, M.J., Glasser, N.F., McKelvey, B.C., Sugden, D.E. and Fink, D. (2007). Cenozoic landscape evolution of an East Antarctic oasis (Radok Lake area, northern Prince Charles Mountains), and its implications for the glacial and climatic history of Antarctica. Quaternary Science Reviews, 26, 598-626.

Harris, C.M., Lorenz, K., Fishpool, L.D.C., Lascelles, B., Cooper, J., Coria, N.R., Croxall, J.P., Emmerson, L.M., Fijn, R.C., Fraser, W.L., Jouventin, P., LaRue, M.A., Le Maho, Y., Lynch, H.J., Naveen, R., Paterson-Fraser, D.L., Peter, H.-U., Poncet, S., Phillips, R.A., Southwell, C.J., van Franeker, J.A., Weimerskirch, H., Wienecke, B., and Woehler, E.J. (2015). *Important Bird Areas in Antarctica 2015*. BirdLife International and Environmental Research & Assessment Ltd., Cambridge.

Heatwole, H. Betts M., Webb J., Crosthwaite P. (1991). Birds of the northern Prince Charles Mountains, Antarctica. Corella, 14, 120-122.

CC-BY-NC-ND

Hodum, J.P. and Hobson, K.A. (2000). Trophic relationships among Antarctic fulmarine petrels: insights into dietary overlap and chick provisioning strategies inferred from stable-isotope ($\delta^{13}C$ and $\delta^{15}N$) analyses. Marine Ecology-Progress Series, 198, 273-281.

Jovani, R., Mavor, R. and Oro, D. (2008), Hidden patterns of colony size variation in seabirds: a logarithmic point of view. Oikos, 117, 1774-1781.

Kock, K.H., ed. (2000). *Understanding CCAMLR's approach to management*. CCAMLR, Hobart.

Lorentsen, S.H., Klages, N., and Røv, N. (1998). Diet and prey consumption of Antarctic petrels Thalassoica antarctica at Svarthamaren, Dronning Maud Land, and at sea outside the colony. Polar Biology, 19, 414-420.

Kuehn G. 1998. Southern Prince Charles Mountains, Mawson Escarpment 1997-1998. Unpublished Australian National Antarctic Research Expedition (ANARE) Report.

Lorentsen, S.H. and Røv, N. (1995). Incubation and brooding performance of the Antarctic petrel *Thalassoica antarctica* at Svarthamaren, Dronning Maud Land. Ibis, 137, 345-351.

CC-BY-NC-ND

Lynch, H.J. and LaRue, M.A. (2014). First global census of the Adélie Penguin. The Auk, 131, 457-466.

Lynch, H.J. and Schwaller, M.R. (2014). Mapping the abundance and distribution of Adélie Penguins using Landsat-7: first steps towards an integrated multi-sensor pipeline for tracking populations at the continental scale. PloS one, 9, p.e113301.

Mawson, D. (1930). The Antarctic Cruise of the "Discovery," 1929-1930. Geographical Review, 20, 535-554.

Mehlum F., Gjessing Y., Haftorn S. and Bech C. (1988). Census of breeding Antarctic petrels *Thalassoica antarctica* and physical features of the breeding colony at Svarthamaren, Dronning Maud Land, with notes on breeding Snow petrels *Pagodroma nivea* and South Polar skuas *Catharacta maccormicki*. Polar Research, 6, 1-9.

Nicol, S. (1993). A Comparison of Antarctic petrel (*Thalassoica antarctica*) diets with net samples of Antarctic krill (*Euphausia superba*) taken from the Prydz Bay region. Polar Biology 13, 399-403.

Rees, W. G. (2012). Assessment of ASTER global digital elevation model data for Arctic research. Polar Record, 48, 31-39.

CC-BY-NC-ND

Rees, W.G., Brown, J.A., Fretwell, P.T. and Trathan, P.N. (2017). What colour is penguin guano? Antarctic Science, 29, 1-9.

Schwaller, M.R., Olson, C.E., Ma, Z., Zhu, Z. and Dahmer, P. (1989). A remote sensing analysis of Adélie penguin rookeries. Remote Sensing of Environment, 28, 199-206.

Schwaller, M. R., Southwell, C.J. and. Emmerson, L.M. (2013) Continental-scale mapping of Adélie penguin colonies from Landsat imagery. Remote Sensing of Environment, 139, 353-364.

van Franeker, J.A., Creuwels, J.C., Van Der Veer, W., Cleland, S. and Robertson, G. (2001). Unexpected effects of climate change on the predation of Antarctic petrels. Antarctic Science, 13, 430-439.

van Franeker, J.A., Gavrilo, M., Mehlum, F., Veit, R.R., and Woehler, E.J. (1999). Distribution and abundance of the Antarctic petrel. Waterbirds, 22, 14-28.

Wright, G.K. and Wyeth, R.B.. 1971. Shackleton Range 197071. Dog sledge journey—geological. British Antarctic Survey internal report AS6/2Z/1970/K11. British Antarctic Survey, Cambridge.

Wood, J. (1996). The geomorphological characterisation of digital elevation models. University of Leicester Ph.D. dissertation. http://hdl.handle.net/2381/34503.

CC-BY-NC-ND

final version published in *Remote Sensing of Environment* https://doi.org/10.1016/j.rse.2018.02.071

CC-BY-NC-ND

Appendix

This appendix describes the steps used for processing top of the atmosphere (TOA) reflectance data from the Landsat-8 Operational Line Imager (OLI) to determine if any given pixel falls into the "Antarctic petrel colony" class. Only reflectance data from the OLI blue, green, red, near-infrared, short-wave-infrared-1, and short-wave-infrared-2 channels were used; these bands are designated respectively as ρ_1, ρ_2, ρ_3, ρ_4, ρ_5, ρ_6. The Landsat-8 coastal and cirrus bands are not included in the analysis.

The spectral bands were arranged in the order below before transforming the data into spherical coordinates. This arrangement was found to provide an optimum separation between Antarctic petrel colony area and surrounding targets.

$$[\rho_5 \quad \rho_2 \quad \rho_1 \quad \rho_3 \quad \rho_4 \quad \rho_6]$$

The vector $\mathbf{V} = [\phi_1 \quad \phi_2 \quad \phi_3 \quad \phi_4 \quad \phi_5 \quad 1]$ was computed by a spherical coordinate transformation of the TOA reflectance bands as follows.

$$\phi_1 = arctan\left(\frac{\rho_5}{\rho_2}\right),$$

$$\phi_2 = arctan\left(\frac{\rho_5 + \rho_2}{\rho_1}\right),$$

$$\phi_3 = arctan\left(\frac{\rho_5 + \rho_2 + \rho_1}{\rho_3}\right),$$

and so on through ϕ_5.

The transition matrix, **T**, defines a unit spheroid in the transformed spectral space that separates Antarctic petrel colony pixels from all other targets in the Landsat data (see Schwaller et al. 2013 for details on how **T** is calculated). **T** =

CC-BY-NC-ND

$$\begin{bmatrix} -0.098455263 & -0.085666526 & 0.0070475469 & -0.0060604455 & 0.00050076071 & 0.45407839 \\ -0.061222895 & 0.034763469 & -0.027884789 & -0.011445202 & -0.0069187189 & 0.92826559 \\ -0.044248551 & 0.046176384 & 0.044881258 & -0.0075065201 & -6.8737621e-05 & 1.0786967 \\ -0.033644091 & 0.024852252 & -0.016621225 & -0.0064340191 & 0.011656478 & 1.1291415 \\ 0.12814319 & -0.026740639 & 0.0032261196 & -0.014405842 & 0.00011587505 & 1.1299354 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix, **A**, is the set of angular distances to a given pixel in the transformed space, and is calculated as follows.

$$\mathbf{A} = \mathbf{V}^{\mathbf{T}} \cdot \mathbf{T}^{\mathbf{-1}}$$

If we define $\mathbf{A} = [a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad 1]$, the Euclidian distance, *d*, to a pixel in the spherical coordinate data space is calculated as the square root of the sum of squares of the first 5 elements of the matrix **A**.

$$d = \sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2 + {a_4}^2 + {a_5}^2}$$

The value of *d* is used as the decision rule to determine whether a pixel belongs to the Antarctic petrel colony class. If $d \leq 1$ then the pixel is categorized as belonging to the Antarctic petrel colony class since the pixel falls within the volume or on the surface of the spheroid in the transformed spectral data space that defines this class. If $d > 1$ the pixel is assigned to the "not a colony" class. This procedure can be applied to any Landsat-8 OLI scene: pixels within the scene that yield a value $d \leq 1$ are assumed to be part of the Antarctic petrel colony class.

CC-BY-NC-ND