Remote Sensing of Environment 139 (2013) 353–364

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Continental-scale mapping of Adélie penguin colonies from Landsat imagery

Mathew R. Schwaller [a,*], Colin J. Southwell [b], Louise M. Emmerson [b]

[a] Mail Code 587, NASA/GSFC, Greenbelt, MD 20771, USA
[b] Australian Antarctic Division, 203 Channel Highway, Kingston, Tasmania 7050, Australia

ARTICLE INFO

Article history:
Received 25 January 2013
Received in revised form 1 August 2013
Accepted 2 August 2013
Available online 11 September 2013

Keywords:
Adélie penguin
Antarctica
Landsat
ETM+
Penguin

ABSTRACT

Breeding distribution of the Adélie penguin, *Pygoscelis adeliae*, was surveyed with Landsat-7 Enhanced Thematic Mapper Plus (ETM+) data in an area covering approximately 330° of longitude along the coastline of Antarctica. An algorithm was designed to minimize radiometric noise and to retrieve Adélie penguin colony location and spatial extent from the ETM+ data. In all, 9143 individual pixels were classified as belonging to an Adélie penguin colony class out of the entire dataset of 195 ETM+ scenes, where the dimension of each pixel is 30 m by 30 m, and each scene is approximately 180 km by 180 km. Pixel clustering identified a total of 187 individual Adélie penguin colonies, ranging in size from a single pixel (900 m^2) to a maximum of 875 pixels (0.788 km^2). Colony retrievals have a very low error of commission, on the order of 1% or less, and the error of omission was estimated to be ~3 to 4% by population based on comparisons with direct observations from surveys across east Antarctica. Thus, the Landsat retrievals successfully located Adélie penguin colonies that accounted for ~96 to 97% of the regional population used as ground truth. Geographic coordinates and the spatial extent of each colony retrieved from the Landsat data are available publically. Regional analysis found several areas where the Landsat retrievals suggest populations that are significantly larger than published estimates. Six Adélie penguin colonies were found that are believed to be previously unreported in the literature.

1. Introduction

The Adélie penguin *Pygoscelis adeliae* has a circum-Antarctic distribution (Ainley, 2002) and is widely considered a useful indicator of status and change in the Antarctic and Southern Ocean ecosystems (Ainley et al., 2005; Ainley et al., 2007; Forcada & Trathan, 2009). Various studies have reported the population distribution and abundance of the Adélie penguin, with the most recently published circum-Antarctic compilation by Ainley (2002), which was largely based on a previous synthesis of published and unpublished reports compiled by Woehler (1993). These circum-Antarctic compendia rely on observations acquired over several decades by numerous groups and individuals using ground surveys, surveys from ships, and aerial photogrammetric methods. This set of observations therefore varies greatly in terms of accuracy in the geolocation of Adélie penguin colonies and in the estimates of population size on these areas. While regional and continent-wide estimates of Adélie penguin distribution and abundance have to date been based on direct observations, they are potentially subject to under-estimation bias due to the remoteness and inaccessibility of sections of the Antarctic coastline (Southwell, Smith, & Bender, 2009). As of the 1990s, Woehler (1993) estimated the worldwide Adélie penguin population to be 2,465,800 nesting pairs. More recently, but using mostly the same data, Ainley (2002) estimated the Adélie penguin population to be between 2.4 and 3.2 million breeding pairs worldwide. Ainley (2002) also acknowledged that there are sites recorded in the literature where little is known about the population other than a report that Adélie penguins nest there, that a few sites have not been counted for decades, and there are probably colonies along the Antarctic coastline that still await discovery.

The study described below exploits satellite remote sensing data as an exploration tool for mapping the breeding distribution and spatial extent of the Adélie penguin. The objective of this work is to explore the potential utility of Landsat remote sensing data to contribute to an accurate, continent-wide estimate of the distribution and abundance of the breeding population for this indicator species, and ultimately to the goal of monitoring changes in Adélie penguin population over time.

Several studies have demonstrated that satellite imagery can detect and map the spatial extent of Adélie penguin breeding colonies based on the relatively unique spectral characteristics of penguin guano compared to other targets in the Antarctic environment (Bhikharidas, Whitehead, & Peterson, 1992; Lynch, White, Black, & Naveen, 2012;

* Corresponding author. Tel.: +1 301 614 5382.
E-mail addresses: mathew.r.schwaller@nasa.gov (M.R. Schwaller), colin.southwell@aad.gov.au (C.J. Southwell), louise.emmerson@aad.gov.au (L.M. Emmerson).

0034-4257 Published by Elsevier Inc. Open access under CC BY-NC-SA license.
http://dx.doi.org/10.1016/j.rse.2013.08.009

Mustafa, Pfeifer, Hans-Ulrich, Kopp, & Metzig, 2012; Schwaller, Benninghoff, & Olson, 1984; Schwaller, Olson, Ma, Zhu, & Dahmer, 1989). Satellite imagery has also been used for a recently completed pan-Antarctic survey of emperor penguin (*Aptenodytes forsteri*) colonies using satellite data, which revealed the location of several previously unreported colonies (Fretwell & Trathan, 2009; Fretwell et al., 2012). Because Adélie penguins breed at ice-free sites in contrast to the predominantly fast-ice breeding Emperor penguins, the detection of Adélie penguin guano against the rock background is more challenging. Adélie penguins breed during the austral summer, arriving at the breeding sites in late October until they leave at the end of March. The presence of guano can be detected on the rocky surfaces of their breeding sites, or if winter snows are present, when the accumulated snow melts at the beginning of the breeding season.

The study reported here used Landsat-7 data to generate a synoptic dataset of Adélie penguin colony location and spatial extent along the Antarctic coastline, not including the Antarctic Peninsula (AP) and the Ross Ice Shelf (Fig. 1). The Ross Ice Shelf was omitted because it is not a suitable habitat for Adélie penguin colonization, having no areas of ice-free land. The AP was omitted from the study primarily because the "ground truth" used in this analysis came from east Antarctica and the results obtained from the southern coastline may not be generally applicable to the AP. We also note that the Adélie penguin colonies on the AP appear to be smaller than those along the southern coastline (compare colony sizes in the peninsula in Croxall and Kirkwood (1979) with colony sizes around the same time in east Antarctica in Horne (1983)), and are therefore likely to be more difficult to detect at the 30 m by 30 m pixel resolution of Landsat-7 imagery. Adélie penguin colonies on the peninsula also intermix with colonies of other penguin species, compounding the difficulty of isolating them in remote sensing imagery. Furthermore, the AP is characterized by a different climate regime than the rest of Antarctica (King & Comiso, 2003) and unlike the rest of Antarctica the AP is at least partially vegetated. Given these physical and biological differences, the error characteristics of the algorithm we used to retrieve Adélie penguin colonies from the Landsat data along the southern coastline of Antarctica are likely to be significantly different from those found on the AP. A retrieval of Adélie penguin colonies along the AP is therefore beyond the scope of the investigation presented here.

The remote sensing survey methods described below specifically assess the errors of omission and commission of the algorithm used to retrieve the size and geographic location of Adélie penguin colonies from the Landsat data. Errors of omission may arise from deficiencies in the retrieval algorithm itself, or from a mismatch between the spatial resolution of the satellite imagery and colony size (i.e., by missing small colonies). Errors of commission may arise if the classification algorithm is unable to distinguish Adélie penguin guano from terrain features with similar radiometric properties, for example, from the guano associated with colonies of other land-breeding seabirds such as fulmars and petrels. Error assessment was performed by comparing Landsat colony retrievals with actual Adélie penguin breeding distribution known from recent direct observation over a large section of the east Antarctic coastline. Such errors, if large, could lead to biased estimates of distribution and abundance based on the retrieved colony size and area estimates.

2. Data processing

The sections below on data processing and classification describe the quantitative methods used to separate Adélie penguin colonies from other ground targets in Landsat imagery over Antarctica. Three major data processing steps were taken: the first two reduce radiometric noise, and the third performs the classification. The first processing step transforms the Landsat ETM+ data from raw digital counts to surface reflectance, which is an intrinsic, invariant property of a ground target. In the ideal case, Adélie penguin colony surface reflectance will have a unique and characteristic set of values regardless of the colony location or time of year of the observation. The surface reflectance transformation corrects the data for effects due to instrument calibration and normalizes scene-to-scene and within-scene differences in solar elevation, azimuth and sun–Earth distance. The second processing step performs a spherical coordinate transform on the reflectance data. This transformation is performed because the apparent reflectance in the direction of the ETM+ instrument varies with slope and aspect of the ground surface. The coordinate transformation normalizes the reflectance to eliminate the effects of surface slope and aspect. In short, processing steps 1 and 2 were performed to reduce the radiometric variability from exogenous sources that add noise to the surface reflectance properties of Adélie penguin colonies. The third processing step

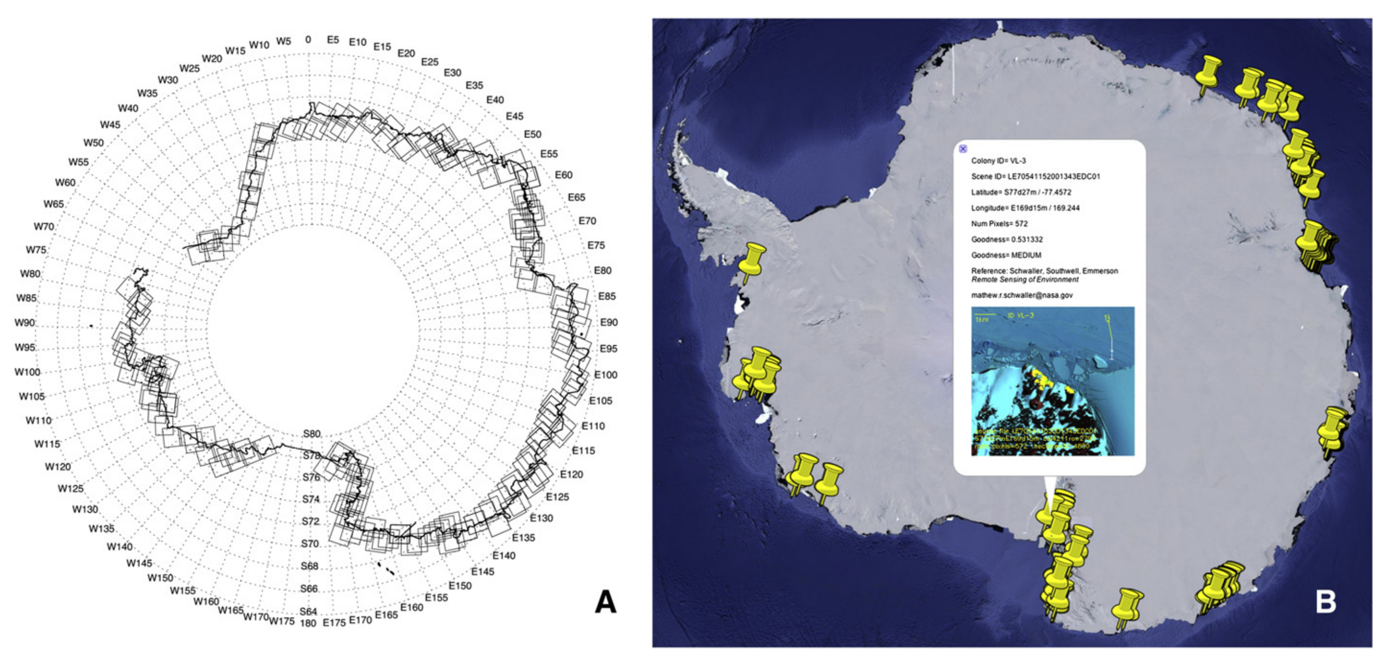

Fig. 1. On the left, location of 195 Landsat-7 ETM+ scenes used in the study. On the right, illustration of the KML placemarks plus a pop-up window that includes a thumbnail image and location information for the Adélie penguin colonies retrieved from the Landsat data (see Appendix A).

performed an affine transformation to normalize the data with respect to the unit sphere. Ground truth data are used to build a single affine transformation matrix that is applied to each pixel in each of the Landsat scenes used in this study. If a transformed pixel falls within the unit sphere (radial distance ≤ 1) the pixel is considered to be part of the "Adélie penguin colony class." This classification method is convenient because—regardless of the satellite instrument, the ground truth data, or the number of bands used—the cut-off criterion for the classification is always equal to 1. This makes the "downstream" processing tasks easier, and it should also make for easier inter-comparison of results from other satellite systems if they are included in subsequent analyses.

This study employed 195 Landsat-7 (L7) Enhanced Thematic Mapper Plus (ETM+) scenes covering the Antarctic continent from the base of the Antarctic Peninsula at W62° eastward to W71° (Fig. 1). The data set was obtained from the Landsat Data Processing System (LDPS) collection of ETM+ Level 1GT data. Scenes were selected if they fell within the era from the beginning of Landsat-7 operations (mid-1999) to the date of the Scan Line Corrector failure (end of May 2003). An effort was made to obtain as many cloud-free scenes as possible covering the study area and to select scenes with dates as close to the austral summer solstice as possible (Table 1). Therefore, some sites were covered by as many as 3 ETM+ scenes, while some were covered by only 1 scene.

2.1. Surface reflectance transformation

At-instrument reflectance data were generated from the raw ETM+ digital counts using the methods described in Chapter 8 of the *Landsat Data Users Handbook* (http://landsathandbook.gsfc.nasa.gov). This conversion normalizes the data with respect to solar zenith angle, which changes within and between scenes, and also compensates for exo-atmospheric solar irradiance, which changes throughout the year as a function of the sun–Earth distance. The conversion thus reduces spectral variability in the data introduced by these exogenous sources. Scaling from reflectance to 16-bit integer data and a non-Lambertian correction were applied as described by Bindschadler et al. (2008) to generate their *Landsat Image Map of Antarctica*. The reflectance normalization used by Bindschadler et al. (2008), which was not physically based, was not applied during ETM+ data processing in this study. Bindschadler et al. (2008) also note that the atmosphere over most of the Antarctic continent is very cold, minimizing the amount of water vapor, and very clean, minimizing the concentration of aerosols. They conclude that atmospheric contributions are negligible and the at-instrument reflectance is a good approximation of the surface reflectance. That conclusion is accepted for this study as well.

2.2. Spherical coordinate transformation

Corrections were applied to the at-instrument reflectance data to minimize the apparent changes in reflectance due to the orientation of ground targets relative to the aperture of the ETM+ instrument. The correction for surface slope and aspect assumes that penguin colony terrain acts like a Lambertian surface. In that case, reflectance in the direction of the satellite is reduced by the cosine of the normal angle of the surface with respect to the aperture of the ETM+. Thus, the normalized reflectance, ρ_n, compared to the reflectance observed, ρ_o, for any given pixel will have the value defined in Eq. (1).

$$\rho_n = \rho_o \cos(\theta) \tag{1}$$

where θ is the angle in the direction of the satellite aperture compared to the normal angle of the ground target.

A spherical coordinate transformation was applied to the ETM+ reflectance data to normalize the terrain effect defined in Eq. (1). This transformation has also been shown to reduce the coefficient of variation in ground-based spectral reflectance measurements of Adélie penguin colonies (Schwaller et al., 1989). The equations for transformation from rectangular coordinates to spherical coordinates are commonly known, but are included below for the sake of completeness. In the equations below the quantities ρ_1, ρ_2, ρ_3, and ρ_4, refer to surface reflectance (unitless) in ETM+ Band 3 (0.63–0.69 μm), Band 4 (0.77–0.90 μm), Band 5 (1.55–1.75 μm), and Band 7 (2.09–2.35 μm), respectively. These bands were selected because they represent distinct portions of the electromagnetic spectrum and as such they are expected to provide relatively independent measurements for the classification of colony and non-colony areas. Band 3 (red) was selected from the visible channels because it measures the longest wavelengths in this portion of the spectrum and therefore is least susceptible to the effects of atmospheric scattering. The red band is also likely to discriminate the often pink-hued color of Adélie penguin colonies. This pigmentation can be attributed to Adélie guano, the contents of which include the partially undigested red-colored carotenoids found in krill (*Euphausia superba*), a major food source for this penguin species. Band 4 was selected because it is the only ETM+ channel in the near-infrared portion of the spectrum. The ETM+ has two channels in the short-wave infrared portion of the spectrum (Bands 5 and 7). Both shortwave infrared bands were included in the classification because they are useful in mapping lithology (Sabine, 1999), and were therefore expected to help separate biological material (guano) from surrounding rock outcrop.

The quantities ϕ_1, ϕ_2, and ϕ_3 are the angular distances (in radians) from the *x*, *y*, and *z* axes to a given point in the transformed, spherical coordinate space. The quantity, *r*, is the radial distance from the origin to a point in the transformed space.

$$\phi_1 = \arctan\left(\frac{\sqrt{\rho_4^2 + \rho_3^2 + \rho_2^2}}{\rho_1}\right) \tag{2}$$

$$\phi_2 = \arctan\left(\frac{\sqrt{\rho_4^2 + \rho_3^2}}{\rho_2}\right) \tag{3}$$

$$\phi_3 = \arctan\left(\frac{\rho_4}{\rho_3}\right) \tag{4}$$

$$r = \frac{\rho_1}{\cos(\phi_1)} \tag{5}$$

The spherical coordinate transform has two benefits. First, assuming Lambertian surfaces, taking the ratio of observed reflectances for a given pixel will numerically cancel out the $\cos(\theta)$ factor associated with pixel orientation to the spacecraft. The transformation thus normalizes the reflectances with respect to the spacecraft and helps to eliminate the scene brightness differences due to surface topography. Secondly, the quantity *r* is the Euclidian distance to a pixel's reflectivity and is a measure of the "brightness" of the pixel due to illumination effects. The coordinates ϕ_1, ϕ_2, and ϕ_3 are measures of the pixel's pure "color" irrespective of illumination. Illumination can vary across a scene due to various effects such as terrain slope, aspect, and atmospheric aerosol content. Therefore, isolation of the illumination effect, and removal of this effect from the scene classification method,

Table 1
Distribution by date of the 195 Landsat-7 ETM+ scenes used in this study.

	January	February	November	December	Total
1999	0	0	0	10	10
2000	21	4	0	0	25
2001	8	3	30	33	74
2002	3	0	23	40	66
2003	15	5	0	0	20
Total	47	12	53	83	195

improves the ability to separate targets of interest from one another based on the available reflectance data.

2.3. Affine transformation classification method

The classification method used in this study is based on the assumption that a unique Adélie penguin colony class can be bounded by an ellipsoid in the 3-dimensional space of the spherically transformed ETM+ reflectance data. A further assumption of the classification is that all other ground targets will fall outside the surface of the ellipsoid in ETM+ spectral space. The equation for the ellipsoid was built using training data from four Landsat-7 ETM+ scenes covering Adélie penguin colonies bordering the Ross Sea. In total, 448 transformed data pixels were selected from the Landsat scenes covering the Adélie penguin colonies on Cape Adare (172 pixels, acquisition date 1/3/2003), Cape Crozier (141, 12/27/2001), Cape Hallet (89, 11/3/2006), and Inexpressible Island (28, 11/3/2006). In addition, pixels were selected from other ground targets within the same scenes: sea ice, glacier and snow, rock outcrop of various types, open water, cloud, and coastline. A total of 144,713 pixels were selected from all other classes.

The equation of the ellipsoid defining the Adélie penguin colony class is found by starting with a unit sphere centered at the origin of the 3-dimensional space (ϕ_1, ϕ_2, and ϕ_3) and transforming the sphere into an ellipsoid that best fits the boundary of the data collected from the colony targets. An affine transformation was used to translate, scale and rotate the unit sphere into the proper location and shape in the transformed data space. The composite transformation matrix, **T**, is defined as the product of three individual matrices, one each for: translation, $\mathbf{T}_T$; scaling, $\mathbf{T}_S$; and rotation, $\mathbf{T}_R$.

$$\mathrm{T} = (\mathrm{T}_R \cdot \mathrm{T}_S)^{\mathrm{T}} \cdot \mathrm{T}_T. \tag{6}$$

The translation matrix, $\mathbf{T}_\mathrm{T}$, is defined as follows, with μ_1, μ_2, and μ_3 equal to the mean of ϕ_1, ϕ_2, and ϕ_3, respectively.

$$\mathrm{T}_T = \begin{bmatrix} 1 & 0 & 0 & \mu_1 \\ 0 & 1 & 0 & \mu_2 \\ 0 & 0 & 1 & \mu_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

The scale, $\mathbf{T}_\mathrm{S}$, matrix is defined below.

$$\mathrm{T}_S = \begin{bmatrix} s_1\sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & s_2\sqrt{\lambda_2} & 0 & 0 \\ 0 & 0 & s_3\sqrt{\lambda_3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

As is the rotation matrix, $\mathbf{T}_\mathrm{R}$.

$$\mathrm{T}_\mathrm{R} = \begin{bmatrix} e_{1,1} & e_{1,2} & e_{1,3} & 0 \\ e_{2,1} & e_{2,2} & e_{2,3} & 0 \\ e_{3,1} & e_{3,2} & e_{3,3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

The matrix $\mathbf{T}_T$ translates the origin of a unit sphere to the center of mass of the penguin colony pixel observations in transformed reflectance space. For the matrix $\mathbf{T}_S$, the values λ_1, λ_2 and λ_3 are the penguin colony transformed pixel sample eigenvalues. Taking the square root of the eigenvalues provides the axis dimensions that convert the unit sphere into an ellipsoid. The scale factors s_1, s_2, and s_3 stretch or shrink the ellipsoid along its axes. The matrix $\mathbf{T}_R$ rotates the ellipsoid using the eigenvectors of the penguin colony sample data $e_{1,1\ldots3}$, $e_{2,1\ldots3}$, $e_{3,1\ldots3}$ to orient the ellipsoid in the direction of the major and minor axes of the sample data.

The scale factors s_1, s_2, and s_3 in Eq. (8) were optimized to find the "best fit" of the ellipsoid to the training data using a Simplex algorithm (Nelder & Mead, 1965). The best fit criterion simultaneously maximizes both the number of penguin colony samples falling within the ellipsoid and the number of non-colony samples falling outside its bounds *and* minimizes both the number of colony samples falling outside the ellipse and the number of non-colony samples within. The specific Simplex algorithm used in this study is based on the one described by Caceci and Cacheris (1984). As the Simplex shrinks the ellipsoid it increases the probability of penguin colony pixels falling outside the bounds of its surface, thus incorrectly classifying them as non-colony (an error of omission). As the ellipsoid expands, it increases the probability of non-colony pixels falling inside the ellipsoid volume and therefore classifying them as penguin colony (an error of commission).

The criterion used to optimize the fit of the ellipsoid to the training data was to maximize Kendall's tau (τ), a measure of rank correlation (Kendall, 1962). In each iteration of the Simplex the penguin colony and non-colony data were classified based on the transformation matrix described above. Points falling within the ellipsoid volume were classified as penguin colony, those falling outside were classified as non-colony. The resulting data were arranged into a 2 × 2 contingency table and τ was calculated as follows:

a = number of penguin colony pixels correctly classified
b = number of non-colony pixels correctly classified
c = number of colony pixels classified as non-colony (error of omission)
d = number of non-colony pixels classified as colony (error of commission).

$$\tau = \frac{(ab - cd)}{\sqrt{(b+c)(d+a)(b+d)(c+a)}} \tag{10}$$

Classification solutions were found using the spherical coordinate data transformed from ETM+ data in bands 3, 4, 5 and 7 as described above. The classifications yielded a "perfect" solution in the sense that it classified all 448 penguin colony pixels and all 144,713 non-colony pixels in the training set without error. The final transformation matrix **T** calculated by this process is provided below.

$$\mathrm{T} = \begin{bmatrix} 0.03116384 & -0.084864540 & -0.045278055 & 0.47614123 \\ 0.38406296 & -0.017599313 & 0.044148981 & 0.72581741 \\ 0.21837277 & 0.043063747 & -0.071185388 & 1.0471284 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \tag{11}$$

To apply this result operationally, the 4 ETM+ reflectance data bands defined above were extracted for each pixel in each Antarctic scene, the spherical transformation was performed, and a vector, $\mathbf{V} = [\phi_1, \phi_2, \phi_3, 1]$ was generated for each pixel in the scene. The transpose of the vector **V** was multiplied by the inverse of the matrix **T** to yield a 4 element array, **A**, whose first 3 elements are the angular distances of the pixel that has been transposed back into the coordinate space of the unit sphere (Eq. (12)). The fourth element of **A** will equal 1.

$$\mathrm{A} = \mathrm{V}^{\mathrm{T}} \cdot \mathrm{T}^{-1} \tag{12}$$

The Euclidian distance, d, of the pixel's ETM+ transformed 4-band reflectance from the origin of the unit sphere is simply the square root of the sum of the squares of the first 3 elements of vector **A**.

The scalar value, d, is used as the Adélie penguin colony classification criterion. If d is equal to 1, the pixel's transformed 4-band reflectance falls on the surface of the unit sphere and is classified as belonging to the Adélie penguin colony class. If $d < 1$ the pixel falls within the volume of the unit sphere and is also classified as "Adélie penguin colony." Note that values of d approaching 0 are considered increasingly

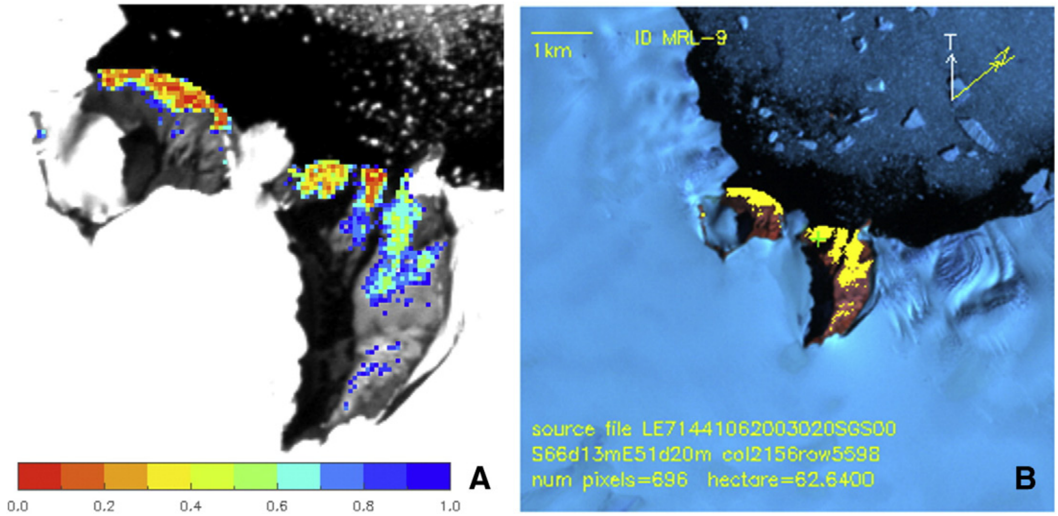

Fig. 2. On the left, Mt. Biscoe image with classified pixels plotted for values of $0 \leq d \leq 1$. Recall that values of d close to 0 are more representative of the Adélie-classification archetype, and that values >1 (not shown here) are considered to be in the "non Adélie penguin colony" class. Although the blue-green pixels are within the range classified as Adélie penguin colonies, the d-values are the least representative of pixels classified as Adélies and may indicate flying seabird guano. On the right, example thumbnail image of the Adélie penguin colony in the Mt. Biscoe area. See the text in Section 3 for an explanation of the image enhancement, notation, and access information.

representative of the Adélie-class archetype. Values of $d > 1$ fall outside of the volume of the unit sphere and are considered to be in the "non-Adélie penguin colony" class. Fig. 2A plots the d-values classified as Adélie penguin colony around Mt. Biscoe (S66°13′ E51°20′).

3. Adélie Penguin colony retrievals

The classification method described above was applied to each of the 195 Landsat-7 ETM+ scenes identified in Fig. 1, a dataset of $>10^{10}$ pixels. That is, the value, d, was calculated by transformation and matrix multiplication of every multispectral pixel in the dataset. The result of the classification was a set of 9143 pixels that met the classification criterion ($d \leq 1$) and that passed the visual review (described below) to indicate that they were part of the "Adélie penguin colony" class, a data reduction of more than 1 in 10^6.

A spatial clustering algorithm was used to group classified pixels into coherent colonies. Although a range of terms and definitions have been used in the literature, the term *colony* used here most closely follows that proposed by Ainley, Nur, and Woehler (1995) which can be summarized as "all penguin breeding areas within 1 km of one another." The haversine formula was used to calculate the distance from a given pixel classified as belonging to the Adélie colony class to every other similarly classified pixel. Classified pixels within 800 m of one another were clustered together to create what is termed an Adélie "colony" throughout the text. In the error analysis below these Landsat-derived colonies are compared to Adélie *breeding sites* that were identified by direct observation from the ground or aircraft as described by Southwell and Emmerson (2013a). In some cases the result is that breeding sites on several adjacent islands in a compact archipelago were merged into a single colony, or that morphologically distinct but proximal breeding sites along the coast were similarly clustered together. Thus, in the text below the term "breeding site" specifically refers to locations identified by Southwell and Emmerson (2013a). The term "colony" is applied to locations derived from the Landsat data and to locations identified by other authors who used this term in their publications. While the colony definition used here was suitable for the analyses described below, it is understood that it may not be appropriate for studies conducted by other investigators. Therefore, the entire set of 9143 classified pixels, including pixel geolocation, pixel d-value, and additional information, is available as a supplemental dataset: http://doi.pangaea.de/10.1594/PANGAEA.804588.

Following spatial clustering of the classified pixels, "thumbnail" images were generated using ETM+ bands 3, 4, and 5 for each colony cluster (see Fig. 2B for an example). Pixels classified as penguin colony (d values ≤ 1) are displayed as bright yellow and are centered in the thumbnail images along with a green fiducial mark at the geometric center of the colony. The ETM+ Level 1GT product used in this study has a demonstrated 1σ geodetic accuracy of 54 m (Lee, Storey, Choate, & Hayes, 2004), and the geolocation of colony centers is expected to be similarly accurate. The thumbnails also include notations defining the latitude and longitude of the fiducial mark, the Landsat scene used, the Landsat scene column and row number of the fiducial mark, the number of colony pixels in the thumbnail, and spatial extent of the colony pixels in hectare (1 ha $= 10^4$ m^2, 0.01 km^2). A colony/thumbnail identification number and km scale are also included, as are arrows pointing to true North (N) and to the top (T) of the Landsat scene. It should be noted that some colonies were covered by multiple L7 scenes. In those cases, the colony thumbnail image with the largest number of retrieved pixels was retained (along with the pixels themselves), while pixels from the other scenes were discarded.

Thumbnail images were included with files formatted according to the Keyhole Markup Language (KML) convention and bundled together in the KMZ format using the Unix/Linux zip command. The KMZ files can be visualized with common geospatial tools such as Google Earth. Such visualizations were used to help match-up colony locations retrieved from the Landsat data with colony locations identified in other geo-referenced data sets. A KMZ file of Adélie penguin colonies retrieved from the Landsat ETM+ data is available at this url: ftp://trmm-fc.gsfc.nasa.gov/wolff/schwaller_etal_landsat_adelie_colonies-2.kmz and also as an online, supplemental resource. This KMZ file includes ASCII formatted text and png-formated thumbnail images, both of which can be recovered from the KMZ file using the Unix/Linux unzip command. The ASCII text can be parsed to obtain the geographic location, size, and other associated information for each of the colonies retrieved from the Landsat data. For example, the acquisition date of each image used to retrieve colony location and spatial extent can be extracted from the Landsat scene identification number listed in the KML text and on the thumbnail images. The average value of d, the colony classification criterion (Section 3), is also included in the KML text, and is used to make a rough assignment of the "goodness" of the classification. Goodness categories were arbitrarily assigned as follows: average values of d less than or equal to 0.5 are assigned to the "high" category, values between 0.5 and 0.8 are "medium," and values greater than 0.8 are listed as "low." Recall from Section 3 that values of d approaching 0 are considered increasingly representatives of the Adélie-class archetype while values of $d > 1$ are considered to be in the "non-Adélie penguin colony" class. A KMZ file of Adélie penguin colony population estimates collected from the literature is also available at this url: ftp://trmm-fc.gsfc.nasa.gov/wolff/schwaller_etal_Adelie_colonies_from_literature.kmz and also as an online supplemental resource.

Based on a visual analysis of the thumbnail images it was found that a large number of cloud areas were misclassified as Adélie colonies. Preprocessing with a cloud-clearing algorithm would have reduced this error, but identifying and eliminating the cases of misclassification due to clouds were easily accomplished by visual analysis of the multispectral data. Since clouds were easily distinguished from Adélie penguin colonies in the multispectral Landsat imagery, and were cleared prior to subsequent error analysis, the cloud pixels were not counted as errors of commission of the classification method.

4. Error assessment

The ability of the Landsat retrievals to identify Adélie penguin colonies was tested on a large, comprehensive data set of breeding site distribution and order of magnitude abundance from east Antarctica. These data were recently collected by direct observation of occupancy (Southwell & Emmerson, 2013a) and then subsequent counts of adults or occupied nests at actual breeding sites (Southwell & Emmerson, 2013b, 2013c; Southwell et al., 2013). Observations of occupancy and subsequent counts were made by observers from the ground or from aerial photographs. The occupancy survey identified 146 individual Adélie penguin breeding sites from 2303 sites of potential breeding habitat along the Mac. Robertson Land (E59° to E72°), Princess Elizabeth Land (E72° to E87°), Kaiser Wilhelm II Land (E87° to E92°), Queen Mary Land (E92° to E100°) and Wilkes Land (E100° to E140°) coastlines. Counts were made of adults or occupied nests at various times during the breeding season and consequently require standardization to a common population metric (the maximum number of occupied nests, which is equivalent to the maximum number of breeding pairs). Because only some of the standardized counts are currently available (Southwell et al., 2013), we allocated raw count data into half-order of magnitude categories to estimate approximate errors of commission and omission of the Landsat retrieval method in relation to colony size.

As described previously, the Landsat retrieval method spatially aggregated classified pixels found within 800 m of one another into coherent colonies. Therefore, some of the individual breeding sites identified in the direct observation data set had to be merged together to correspond with the spatially aggregated colonies identified in the Landsat retrievals. Furthermore, 6 Adélie penguin breeding sites in the direct observation data set from Kaiser Wilhelm II Land in the area around E93° were not included in the analysis described below for the reasons defined at the end of Section 5. As a result, a total of 119 geographically matched colonies were actually used in the error analysis described below.

Each colony retrieved using the Landsat method was geographically matched to one or more corresponding breeding site(s) in the direct observation data set. Excluding the merged locations, corresponding Landsat colony and direct observation breeding site coordinates matched one another within an average of 179 m, with a standard deviation of 205 m. If Ufs Island is excluded from this calculation (the corresponding geographic coordinates were located at opposite ends of the island) the coordinates match with an average of 158 m and a standard deviation of 133 m.

A summary of the geographic matching results is provided in Fig. 3 and Table 2. As illustrated in the figure, the performance of the Landsat retrieval improves with increasing colony population size. Colonies with population counts in the range of 100–315 had a 0.23 probability of detection, but the probability increased to 0.31 in the range from 316–999, 0.64 in the range from 1000 to 3161, and to 0.97 in the range from 3162 to 9999. All colonies with population counts of ≥10,000 were successfully identified in the retrievals. As summarized in Table 2, the Landsat retrievals successfully identified 63% of the colonies (some, as noted above consisting of merged breeding sites) in the direct observation data set. As illustrated in Fig. 3, the peak of the distribution of breeding site *population* missed by the Landsat retrievals is *an order of magnitude lower* than the population distribution peak for the colonies that were successfully retrieved by the Landsat method. Thus, the breeding sites missed by the Landsat retrieval method had the smallest populations in the region, while those that were found by the Landsat method had by far the largest populations. Although the smallest colonies may be important indicators of new colonization or recent colony abandonment, they contribute little to the population total.

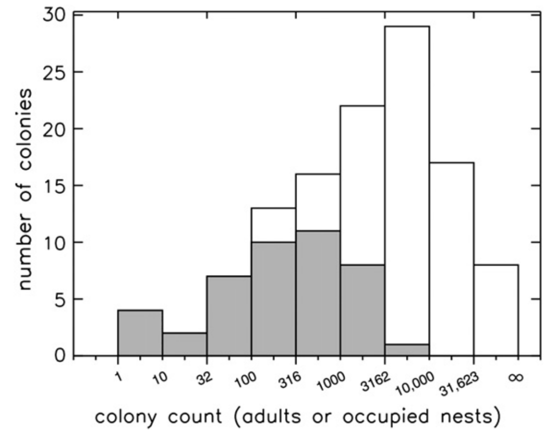

Fig. 3. Population size distribution of the 119 Adélie penguin colonies used for error assessment in this study (see Section 4). Each bin of the histogram represents the number of Adélie penguin colonies located (white) and missed (shaded) by the Landsat retrieval method. The horizontal axis is logarithmic, with boundaries equal to 10^0, 10^1, $10^{1.5}$, 10^2, $10^{2.5}$, ….

The Landsat retrieval method performed well in identifying the Adélie penguin colonies that contribute most to the regional population. The regional population found and missed by the Landsat method was estimated by multiplying the median bin value by the number of colonies per bin in each of the population bins illustrated in the Fig. 3 histogram. Using this approach, the error of omission of the Landsat retrieval method was estimated to be ~3% by population. Thus, the Landsat retrievals successfully located those colonies that account for ~97% of the regional population covered by the direct observation data set. As described above and in Section 5.1, population data from the area around Haswell Island were omitted from this error analysis. If population estimates from this area (some dating from 1980) are included in the error analysis, the error of omission rises to ~4% by population. As for errors of commission, 4 locations consisting of 13 pixels were

Table 2
Summary of Landsat performance in identifying Adélie penguin colonies in east Antarctica. The Landsat retrievals successfully identified 63% of the colonies in the 3 regions identified in the table. The colonies missed by the Landsat method are among the smallest in the region; therefore, the Landsat method successfully located those colonies that account for ~96 to97% of the regional population (occupied nests).

	Wilkes Land	Princess Elizabeth Land	Mac-Robertson Land	Total
Unique colonies from direct observation surveys	19	44	56	119
Landsat-matched colonies	16	31	28	75
Percent matched	84.2	70.5	50.0	63.0

erroneously classified by the Landsat retrieval as penguin colony. This number equates to an error of commission of <1% of the 2933 pixels in the region that were successfully identified as belonging to the Adélie penguin colony class.

Two methods were used to evaluate the correlation between Adélie colony population and the colony area retrieved from Landsat data. The first method employed the raw population count data used to generate Fig. 3, recognizing that the assignment of colonies to the population count bins identified in the figure creates a categorical variable for population count. Two correlation coefficients for categorical variables (Spearman's rho and Kendall's tau) were used to assess the strength of the association between colony population and retrieved area for those colonies found by the Landsat method. 95% confidence limits (in parentheses) were calculated for ρ using Fisher's z-transformation (e.g., Neter & Wasserman, 1974) and calculated for τ using the more exact method of Kendall (1962). Spearman's ρ found a correlation of 0.833 (0.741, 0.889), Kendall's τ found a correlation of 0.709 (0.621, 0.785), with a sample size $n = 74$ in both cases. A second correlation assessment was performed using linear regression on population data from the Ross Sea area published in Woehler (1993), again compared to colony spatial extent from the Landsat retrievals. These population data were used because, as noted above, only some standardized counts for of the east Antarctica dataset are currently available, and the categorical data are not suitable for linear regression. Therefore, the population data in Woehler's (1993) compendium are, at present, the best available published source for this comparison, which is illustrated in Fig. 4. As illustrated in the figure, a strong linear correlation was observed when performing regression through the origin: $r^2 = 0.9943$ with upper and lower 95% confidence intervals of 0.9988 and 0.9723, for a sample size of $n = 9$.

5. Regional results

It is not practical for this publication to list all of the 187 Adélie penguin colonies retrieved from the Landsat data along the Antarctic coastline from W62° eastward to W71°. As previously mentioned, however, the geographic coordinates and the spatial extent of each colony are included in KMZ files (see Appendix A). These files can be used with common geospatial viewers to explore individual colonies and to compare the distribution of retrieved locations to those identified in the literature. Based on this dataset, some findings of interest on a regional scale are summarized below.

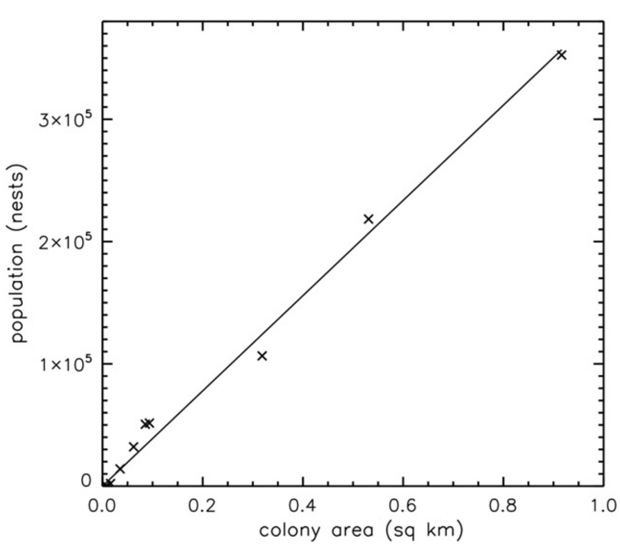

Fig. 4. Linear regression through the origin of Adélie penguin colony area (km^2) compared to population in the Ross Sea area as reported by Woehler (1993). The equation of the regression line is $\hat{y} = 0.389x$ where x is expressed in m^2.

5.1. East of the Antarctic Peninsula, Coats Land, and western Dronning Maud Land (W62° to E35°)

No Adélie penguin colonies were identified in the Landsat retrievals for the vast section of coastline from the base of the Antarctic Peninsula at W62° eastward to E35°. Only one colony of just 3 breeding pairs has previously been reported in the literature (in 1983) in this region (Woehler, 1993).

5.2. Eastern Dronning Maud Land and Enderby Land (E35° to E45°)

The Landsat retrievals located only two of 14 colonies previously identified in the region (Hoshiai, Sweda, & Tanimura, 1984). All of the 12 missed colonies had small populations (<1000 individual penguins) and were all below the typical limit of resolution for the Landsat retrieval method.

5.3. Mac. Robertson Land, Princess Elizabeth Land, and Wilkes Land (E59° to E72°, E72° to E87°, E100° to E142°)

Fig. 5 illustrates the distribution of retrieved colony locations in the area around Cape Batterbee, where Adélie penguin colonies are relatively unexplored. A bold + marks the location of Adélie penguin colonies identified in published compilations (Ainley, 2002; Woehler, 1993). Each Landsat pixel identified as belonging to the Adélie penguin colony class is marked with a small red square. The published compilations report 5000 pairs of Adélie penguins on Proclamation Island, identify their presence (but no counts) at Cape Batterbee, and provide approximate geographic coordinates (Cooper, 1985). The Landsat retrievals accurately identify the geographic location and spatial extent of the colonies at these two sites, as well as on several unnamed surrounding islands and on nearby coastal rocks. The colonies illustrated in Fig. 5 cover a spatial extent of 0.222 km^2 (247 pixels). By comparison, the east and west colonies at Cape Crozier occupy 0.515 km^2 (572 pixels), and had a population reported by Woehler and Croxall (1997) of 118,220 pairs in 1991. The strong correlation between Adélie penguin population and retrieved colony area reported above (Section 4) suggests that the published population estimate for the Proclamation Island/Cape Batterbee area was a significant underestimation, perhaps by an order of magnitude, of the true Adélie breeding population in that location at the time of Cooper's (1985) observation in 1985, or that the population has increased dramatically in the last 30 years. Note that the Landsat retrievals identified 3 additional colony locations (5 breeding sites) not specifically identified in the Ainley (2002) and Woehler (1993) compendia. We refrain from calling these “new colonies” because they there are all in the vicinity of Cape Batterbee, which is included in these publications.

The Landsat retrievals also found that the Adélie population at Mt. Biscoe (S66°13′ E51°20′) may be significantly larger than previously reported. Bassett, Woehler, Ensor, Kerry, and Johnstone (1990) reported the presence of an Adélie penguin colony on Mt. Biscoe and a nearby massif to the west, and from a brief visit very early in the breeding season suggested the combined colony population to be at least 5000 breeding pairs. The Landsat retrievals classified 696 pixels (0.626 km^2) as Adélie penguin colony at this location (see Fig. 2). This spatial extent may include considerable contributions from nesting areas of flying seabird species. As illustrated in Fig. 2A, a large fraction of the pixels retrieved around Mt. Biscoe had low classification d-values. These low d-values (blue colors) were generally associated with areas of high relief on the slopes of Mt. Biscoe as illustrated in a supplemental animation (ftp://trmm-fc.gsfc.nasa.gov/wolff/schwaller_etal_biscoe_movie.mpg). In the animation, Landsat retrievals are superimposed on a digital elevation model obtained from the global DEM dataset derived from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data (Hirano, Welch, & Lang, 2003). Note that a 2× elevation exaggeration is used in the Mt. Biscoe animation. It seems likely that the areas of

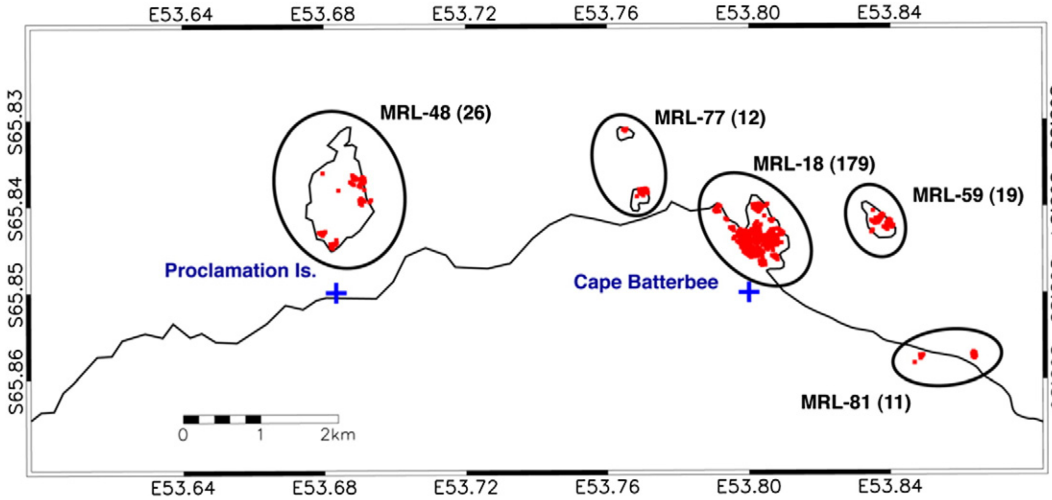

Fig. 5. Map of the distribution of Adélie penguin breeding areas retrieved from the Landsat data in the area around Proclamation Island and Cape Batterbee. A bold + marks the location of Adélie colonies identified in published compilations. Each Landsat pixel identified as belonging to the Adélie penguin colony class is marked with a red square. Individual pixels clustered into coherent colonies are indicated by ellipses, along with the colony's placemark number from the KMZ file (with number of pixels in parentheses).

high relief and low *d*-values are nesting areas of bird species other than the Adélie penguin. Indeed, an early ship-based observation of Mt. Biscoe reported the presence of "countless flocks" and "clouds of birds" (Falla, 1937; Mawson, 1930). The verification of the spatial extent of the Adélie colony and of other seabird species on Mt. Biscoe clearly deserves further study.

5.4. Kaiser Wilhelm II Land and Queen Mary Land (E87° to E100°)

No Adélie penguin colonies were identified in the Landsat retrievals along the coastline of this region. The direct surveys found 6 occupied breeding sites in this region between E92°30′ and E93°33′. All but two of these sites had relatively small populations (<1000 adults). A count of the largest population on Haswell Island (S66°32′ E93°00′) was not obtained in the recent direct observational survey, but Starck (1980) counted 24,700 adults in 1978–79. The Landsat ETM+ imagery collected in December 2001 and 2002, during what should be the active breeding season, shows that there was a region of fast ice extending >20 km from Haswell Island to open water. The imagery also shows no obvious leads of open water between the islands and the sea. It seems possible that Adélie penguin occupancy on Haswell and its surrounding islands was limited by the presence of extensive fast ice during the breeding season for the era in which the Landsat images used in this study were collected. Adélie penguins are known to regularly travel across extensive fast ice to reach breeding sites in some regions, e.g. at Mawson and Syowa (Emmerson & Southwell, 2008; Kato, Watanuki, & Naito, 2003), but it is possible that prolonged, unusually extensive fast ice over several years could result in breeding sites being temporarily abandoned, as breeding success is known to decrease in the Mawson region during years of extensive fast ice (Emmerson & Southwell, 2008). The occupation and breeding success of Adélie penguins on Haswell and nearby islands is clearly a topic for further investigation. Given the discrepancies noted above, however, these 6 breeding sites were omitted from the error analysis described in Section 4.

5.5. Adélie Land and King George V Land (E140° to E153°)

The Landsat retrievals identified 33 Adélie colonies in this region, while 10 locations were previously documented in the literature (Barbraud, Delord, Micol, & Jouventin, 1999). Fig. 6 maps the distribution of colonies retrieved from the Landsat data along part of the Way Archipelago where there are relatively few reports of Adélie penguins in the literature. Barbraud et al. (1999) identified Adélie penguin presence on 3 islands in the southeast of the Way Archipelago as illustrated in the figure, although no counts were provided. They also counted 1600 chicks at Garnet Point, but incorrectly list the geographic coordinates of this feature, locating it 18.6 km too far to the southeast. Earlier, Ensor and Bassett (1987) directly counted 1868 chicks at Stillwell Island. They also made ship-based counts of other islands

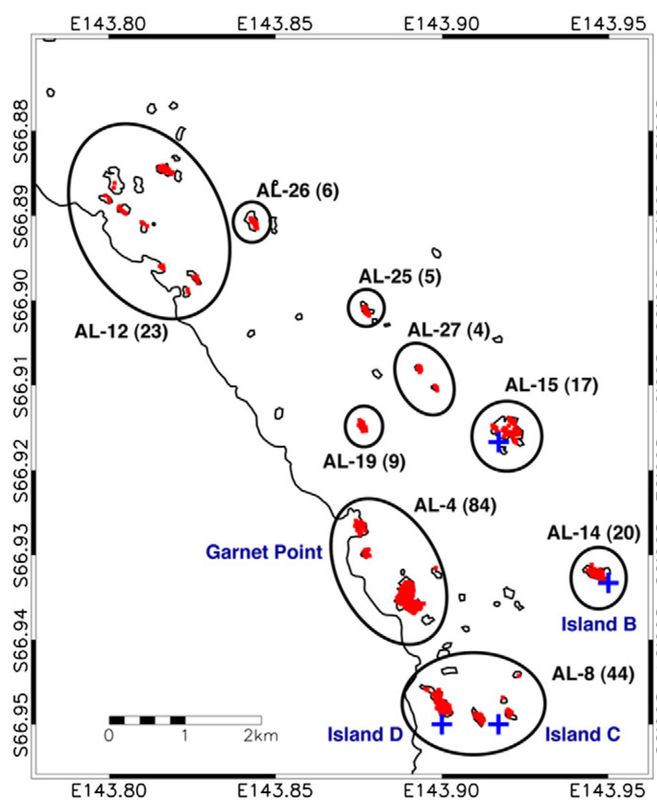

Fig. 6. Distribution of Adélie penguin breeding areas retrieved from the Landsat data in the Way Archipelago. Adélie colonies identified by Ensor and Bassett (1987) and Barbraud et al. (1999) are marked by a bold + along with their respective names. Each Landsat pixel identified as belonging to the Adélie penguin colony class is marked with a small a red square. Individual pixels clustered into coherent colonies are indicated by ellipses, along with the colony's placemark number from the KMZ file (with number of pixels in parentheses).

within the Way Archipelago, but the geographic locations of these islands were not provided in their manuscript. Using the definition described in Section 3, the Landsat retrievals identified 5 colonies and 19 breeding locations in the Way Archipelago that are not specifically identified in the Ainley (2002) or Woehler (1993) compilations. Again, we refrain from calling these "new colonies" because they are part of the Way Archipelago, which is referenced in both of these publications as an Adélie penguin breeding area. The total area classified as Adélie penguin colony in the Way Archipelago (the area east of Cape Denison and west of Watt Bay) equaled 0.262 km^2 (289 pixels). The spatial extent of the Adélie penguin colonies retrieved from the Landsat data suggests a recent population that is more than double the 23,084 pairs in 1982 that Woehler (1993) inferred from chick counts performed by Ensor and Bassett (1987).

Even in the relatively well explored parts of this region, the Landsat method was found to provide better geolocation accuracy for known Adélie colonies, particularly around Cape Hunter, the Mackellar Islands, Cape Denison, Cape Gray, Cape Pigeon Rocks, and Garnet Point. An average geolocation difference of 2.2 km was found between the published locations and the locations obtained from the Landsat retrievals for these sites. Based on the results reported above for Mac. Robertson Land, the Landsat data improves the geolocation accuracy of known colonies in these areas by about an order of magnitude.

The Landsat retrievals classified 255 pixels on the Mackellar Islands (S69°58′ E142°38′), as belonging to the Adélie colony class, the fifth largest colony complex by area of all those found in this study. Ensor and Bassett (1987) counted 27,260 chicks on the 3 Mackellar islands, but noted in their report that earlier expeditions estimated 100,000 pairs in the 1913–14 season and 200,000 penguins in 1930–31 (Falla, 1937). The Landsat area estimate is more consistent with the colony sizes that would support the earlier estimates. However, it is also possible that Landsat classification overestimated the actual colony size due to redistribution of guano from sub-colonies located on numerous small ridges to surrounding areas on the islands (Paul Ensor, personal communication). This is another case where further study is warranted.

5.6. Oates Land (E153° to E160°)

The Landsat retrievals classified 3 colonies in this region that are believed to be previously unreported in the literature. All 3 were found in the proximity of the previously reported Aviation Island colony, with the nearest 15 km and farthest 44 km away. These 3 assumed Adélie penguin colonies are listed below (with KMZ file identification numbers in parentheses).

Mount Archer/Harald Bay, S69°12′ E157°41′, 88 pixels, 0.079 km^2 (VL-8)
Kartografov Island/Harald Bay, S69°12′ E157°42′, 1 pixel, 900 m^2 (VL-243)
Arthurson Ridge, S69°22′ E158°31′,18 pixels, 0.016 km^2 (VL-23).

5.7. Victoria Land (E160° to E180°)

The performance of the Landsat retrieval method was hampered somewhat in the Ross Sea region of Victoria Land where most of the colonies are along shoreline oriented north–south, in contrast to the rest of the southern coastline which is oriented predominantly east–west. Colonies with a slope and aspect oriented to the south and west will tend to be in shadow, since the solar orientation is to the north and east during a Landsat daytime overpass. In particular, colonies on the south coast of Coulman Island [S73°30′ E169°50′; 16,302 pairs in 1989 reported by Woehler and Croxall (1997); 21,887 pairs reported by Ainley (2002)] and the west coast of Franklin Island [S76°07′ E168°15′; 54,753 pairs in 1989 reported by Woehler and Croxall (1997); 55,600 pairs reported by Ainley (2002)] were not retrieved from the Landsat data due to shadowing. Shadows cast over the Adélie penguin colonies altered the multispectral reflectance observed by the satellite to the point where these targets were incorrectly classified as belonging to the non-Adélie penguin colony class. The colony at Downshire Cliffs was also not retrieved from the Landsat data. This may be due to the proximity of this colony to the Ross Sea polynya, and the persistent cloud cover that is often associated with such features (Minnett & Key, 2007).

5.8. Marie Byrd Land (W103° to W180°)

The Landsat retrievals identified 19 Adélie colonies in this region, while 8 locations were previously documented in the literature (Woehler, 1993; Woehler & Croxall, 1997). A scientific cruise in 1992 reported the presence Adélie penguins in Pine Island Bay (S74°W104°), specifically on the Lindsey, Edwards, and Brownson Islands, but no counts were made (Anonymous, 1992). Based on this report, Woehler and Croxall (1997) estimated the Adélie population of the region at "several hundred pairs," while Ainley (2002) lists the population size of the Lindsey Islands at 5000. The Landsat retrievals map out a large number of colonies in the Pine Island Bay area, with a total spatial extent of 0.749 km^2 (833 pixels), see location MBL-1 and nearby placemarks in the KMZ files. By way of comparison, the Landsat retrieval method found the colony at Cape Adare (the largest Adélie colony in Antarctica) covers an area of 0.788 km^2 (875 pixels). Given the strong correlation between colony extent and population, it seems likely that the Adélie breeding population in the Pine Island Bay area is extensive, and could exceed 100,000 occupied nests.

Three assumed Adélie penguin colonies were found in the region of Marie Byrd Land and are believed to be previously unreported in the literature. These are listed below with their KMZ file identification numbers in parentheses.

Waite Island, S72°43′ W103°27′, 62 pixels, 0.056 km^2 (MBL-16)
Molar Island, S73°40′ W101°40′, 14 pixels, 0.013 km^2 (MBL-14)
Sims Island, S73°16′ W78°32′, 62 pixels, 0.056 km^2 (MBL-7).

6. Discussion

The most significant contribution of the work described in this paper is a set of physically based methods to retrieve the geographic location and spatial extent of Adélie penguin colonies from Landsat ETM+ data. These methods were applied in a synoptic survey over approximately 330° of longitude along the coastline of Antarctica where Adélie penguins are known or believed to breed.

Based on the Landsat retrievals, 9143 individual pixels were classified as belonging to an Adélie penguin colony out of the entire dataset of 195 ETM+ scenes. These individual pixel locations and associated information are publically available via Pangaea (http://doi.pangaea.de/10.1594/PANGAEA.804588). Pixel clustering yielded a total of 187 individual colonies, ranging in size from a single pixel (900 m^2) to a maximum of 875 pixels (0.7875 km^2) for the colony located at Cape Adare. Geographic coordinates, spatial extent, and a thumbnail image of each colony are embedded in KMZ files that are publically available as part of this publication. The KMZ files can be used with common geospatial viewers to explore individual colonies and to compare the distribution of retrieved Adélie penguin colony locations to those identified in the literature.

The results of this research demonstrate that the Landsat retrievals can be used as an exploration tool to identify the location and spatial extent of Adélie penguin colonies previously unreported in the literature. Six locations were identified in the Landsat retrievals that are assumed to be previously unidentified Adélie penguin colonies, three in Oates Land and three in Marie Byrd Land. These colonies are "new" in the strictest sense possible: there is no record in the literature of any Adélie penguin breeding population in their vicinity. The 6 colonies occupy a combined spatial area of 245 pixels (0.22 km^2), which is approximately equal to the area occupied by the Adélie penguin colony on Cape Hallet

(235 pixels, 0.21 km^2). Results reported here demonstrate a strong correlation between retrieved colony area and Adélie penguin breeding population. Given this correlation, we can assume that the area occupied by these 6 colonies will support an Adélie penguin breeding population comparable to that of Cape Hallet, where 56,153 pairs were reported in 1988 (Woehler & Croxall, 1997) and 43,942 pairs were reported by Ainley (2002).

The Landsat retrievals were also found to be a valuable tool for discovering new breeding areas in locations where Adélie penguins were previously known to exist, but where specific locations and counts have not been previously reported. The best example of this is illustrated in the Pine Island Bay area of Marie Byrd Land. In this case a single previous report identified the presence of Adélie penguin colonies in the area, but no counts were made (Anonymous, 1992). The most recently published estimates of the breeding population in this area range from several hundred pairs (Woehler & Croxall, 1997) to 5000 pairs on the Lindsey Islands (Ainley, 2002). As illustrated in the KMZ files, the Landsat retrievals found an extensive array of what are assumed to be Adélie breeding locations on 11 separate colonies. The retrievals in the Pine Island Bay area account for a colony spatial extent of 0.751 km^2. This spatial extent is comparable to that retrieved over Cape Adare (0.788 km^2), which had an Adélie population of 272,388 pairs in 1988 (Woehler & Croxall, 1997) and 169,200 pairs reported by Ainley (2002). Again, we assume a comparable population in the Pine Island Bay area based on the strong correlation observed between population and the colony spatial extent. In this case, the population in the Pine Island Bay area is likely to be on the order of 150,000 pairs, or around 30 times what is currently reported in the literature.

The under-estimation of Adélie breeding population described above for the Pine Island Bay area is evident in other regions around Antarctica. The Landsat retrievals found 5 colonies (19 breeding locations) in the Way Archipelago and 3 more colonies (5 breeding locations) in the Proclamation Island/Cape Batterbee region that are not explicitly reported in the literature. The breeding populations on the Mackellar Islands and Mt. Biscoe colonies are also likely to be underestimated in the literature, again, based on the spatial extent of these colonies as observed in the Landsat retrievals. These cases may be due to earlier under-estimation of population size, significant population increase between the time of earlier observations and the time of the Landsat survey, or both. In addition to this evidence of under-estimation of Adélie penguin populations, it should also be noted that the geodetic accuracy of colony location reported in the literature is often rather poor. In contrast, the Landsat data have a consistent 1σ geodetic accuracy of 54 m. This level of accuracy is, to cite one example, more than an order of magnitude better than the geodetic accuracy of published locations for Adélie penguin colonies located in Adélie Land.

One of the stated objectives of the research performed here is to exploit the Landsat retrievals for developing a timely, accurate, and continent-wide estimate of the Adélie penguin breeding population. As a first priority, an estimator on this scale must establish a sampling frame: an enumeration of the sampling units of the population. As described above, the Landsat retrieval method is an effective tool for locating all but the smallest Adélie penguin colonies and breeding areas, and for estimating their spatial extent. These methods can therefore contribute to a synoptic sampling frame for population estimation, particularly in remote areas where ground or ship-based observations are infrequent or non-existent.

Once a sampling frame is established, the Landsat retrievals can contribute to a statistical estimator of Adélie penguin breeding population within the limits of the frame. Previous investigations at the colony and sub-colony level found a strong linear correlation between the number of nests per sub-colony and sub-colony area [$r^2 = 0.964$, Woehler and Riddle (1998); $r^2 = 0.80$, Naveen, Lynch, Forrest, Mueller, and Polito (2012)]. Results demonstrated here found that there is a similarly strong correlation between population and colony area at the colony or regional scale ($\rho = 0.833$, $\tau = 0.709$, $r^2 = 0.994$). The retrieval of colony area from satellite data is based on the relatively unique spectral characteristics of penguin guano compared to other targets in the Antarctic environment. It is likely that the strongest signal observed in satellite imagery will come from the freshest guano and that the chemical content of the material will change over time with weathering. At present, however, there is only a single report of laboratory-based spectrophotometric measurements of such targets (Schwaller et al., 1984). Additional laboratory and in situ measurements are needed to better understand the spectrophotometric qualities of seabird guano and how these change with species, diet and time.

The strong correlation observed between population and colony spatial extent translates into a relatively constant packing density of Adélie penguin nests within sub-colonies and colonies. This characteristic should allow colony spatial extent derived from satellite imagery to be used as an auxiliary variable to estimate population. Simply stated, Adélie penguin colony spatial area retrieved from satellite data multiplied by an estimate of packing density equals population. More formally, a population estimator based on regression or ratio techniques can be used to estimate the population total and to place an error bound on the estimate. Indeed, the ratio technique is a best linear unbiased estimator in the case where the relationship between the dependent and auxiliary variable is linear through the origin and where variance increases with increasing values of the auxiliary variable (Cochran, 1977), a model that most likely fits the area-population relationship described here.

In practice, it is likely that a synoptic estimator of Adélie penguin population will rely on a synthesis of direct observations in accessible areas and satellite data elsewhere. The satellite data itself will likely include moderate resolution Landsat data for large regional surveys of breeding locations and spatial extent, and sub-sampling of population with either direct observations or high spatial resolution satellite data. The relationship between breeding area spatial extent and population can then be used to estimate the synoptic population. One would envision relatively frequent repeat surveys at a subset of sites to capture inter-annual variations in Adélie penguin populations, while less frequent remote sensing surveys would be required to capture the slowly varying secular trend in colony spatial extent.

While only a limited set of Adélie population data was used to establish the correlation between colony area and population as described here, the colony location and spatial extent data generated from the Landsat retrievals are embedded in KMZ and Pangaea files published along with this manuscript. These data are easily accessible for verifying the area-population relationship as new and better population data become available. To further promote the reproducibility of the results reported here, the computer code for performing the Landsat retrievals, written in ENVI+IDL, is available by request to the authors.

Although the Landsat retrievals may be used as an auxiliary variable to estimate Adélie penguin populations, there are several physical and biological factors that may bias the estimate. These factors are considered below along with mitigation strategies to help minimize their influence.

Considerable effort was spent on estimating the retrieval errors of commission and omission which, if large, could lead to a positive or negative bias in the estimates of Adélie penguin distribution and abundance. To estimate errors, the Landsat retrievals were compared to direct surveys of Adélie penguin populations conducted across east Antarctica during the Austral summers from 2005/06 to 2011/12. These "ground truth" data were compared to the Landsat retrievals of Adélie penguin colony presence, which was based on imagery collected during the era from 1999 to 2003. Although these two eras do not overlap, the variation in colony spatial extent is expected to be relatively low over this time period, particularly in comparison to the inter-annual variation in colony population. This slight mismatch in the timing of satellite and direct searches is unlikely to affect the assessment of distribution because the processes of colony extinction and foundation occur over relatively long time periods (Southwell & Emmerson, 2013a).

After cloud clearing, the Landsat retrieval of Adélie penguin colonies was found to have a very low error of commission, on the order of 1% or less when compared to an east Antarctica direct observation dataset. As for errors of omission, the Landsat retrieval method performed best in identifying the Adélie colonies that contributed most to the regional population. The Landsat method identified colonies with population counts in the range of 100–315 adults or occupied nests with a 0.23 probability of detection, but the probability of detection increased steadily with increasing colony size, and all colonies with counts of ≥10,000 were successfully identified in the retrievals. Overall, the Landsat method successfully retrieved 63% of the colonies in the east Antarctica dataset. Because the Landsat retrievals consistently identified the largest of these colonies, the overall error of omission of the Landsat retrieval method was estimated to be ~3% by population. The Landsat retrievals were therefore able to successfully locate those colonies that accounted for ~97% of the population in the east Antarctica dataset. It should be noted that population data from the area around Haswell Island were omitted from this error analysis for the reasons detailed in Section 5.1. If population estimates from this area (some dating from 1980) are included in the error analysis, the error of omission rises to ~4% by population.

The errors of omission noted above for the smaller colonies are almost certainly due to the relatively coarse 30 by 30 m pixel resolution of the Landsat ETM+ data. The smallest colonies resolved by the Landsat method (in the range of 100–315 adults or occupied nests) typically corresponded to only 2–3 Landsat pixels. The relatively low success rate for resolving colonies of this size is most likely due to the inclusion and mixing of non-colony terrain in the Landsat pixel field of view. These smaller colonies may well be identified by high spatial resolution imaging satellites that have ground resolutions on the order of 200 pixels for each corresponding Landsat pixel. While the methods developed for this investigation were only applied to Landsat data, they can be adapted to other multispectral satellite data types with higher spatial resolution as well as to future, planned high resolution multispectral imaging satellites.

Although the Landsat retrievals had a relatively low error of omission in the regional comparisons for east Antarctica, there were individual cases in Victoria Land where the method failed to retrieve comparatively large Adélie penguin colonies. This failure can be predominantly attributed to the north–south orientation of the colonies in the region, and to the fact that the colonies were obscured by shadow. Persistent cloud cover in some parts of Victoria Land may also have contributed to omission errors. The omission errors due to shadowing and cloud cover may be mitigated by repeated sampling and by off-nadir sampling with satellites that have this capability. Such sampling can collect observations at a variety of sun angles and elevations. This strategy may allow for full solar illumination of a colony in a single scene, or for integration of a complete view from a series of observations.

A positive bias was observed in the Landsat retrievals that can be attributed to the presence of seabirds other than Adélie penguins nesting on ice-free coastal rock outcrops. This bias was most clearly observed in the Mt. Biscoe colony where a considerable extent of the retrieved colony area is almost certainly occupied by flying seabirds. In general, however, the number and spatial extent of seabird breeding colonies are considerably smaller than for Adélie penguins (Creuwels, Poncet, Hodum, & van Franeker, 2007; van Franeker, Gavrilo, Mehlum, Veit, & Woehler, 1999) although this could in part be an artifact of limited search effort for flying seabirds. In addition, the often steep slopes that flying seabirds nest on would minimize the two-dimensional area of flying seabird breeding colonies. Thus, the retrieval error by area attributed to flying seabirds is likely to be relatively small when estimating Adélie colony spatial area on a large regional or continent-wide basis. Furthermore, it should be possible to mitigate the flying seabird bias by introducing digital elevation and spatial data into the Adélie colony retrieval algorithm. Flying seabirds tend to nest on higher elevations, steeper slopes, and at locations farther from the ocean than Adélie penguins. These criteria could be added to the Adélie penguin classification algorithm to improve classification accuracy. It may also be possible to develop an exploration tool to locate flying seabird colonies by applying a different retrieval algorithm to the same satellite remote sensing and DEM data.

Even with the errors noted above, and while other potential sources of bias could be cited, the Landsat retrievals are largely free of the "incomplete search effort" bias (Southwell et al., 2009) that has characterized previous estimates of Adélie penguin breeding distribution and abundance. The retrievals of Adélie penguin colony location and spatial extent can easily scale to any number of scenes covering an arbitrarily large area. Once the data are selected, the retrievals are completely automated (although some post-retrieval cloud screening must be conducted). The Landsat retrievals extend to large regional or continental scales much more easily than observations based on ground, ship, or even aerial surveys, which are typically limited to the vicinity of inhabited stations.

Certainly a promising area for future research is to extend the Landsat retrievals to the Antarctic Peninsula. The peninsula was omitted from this study primarily because the "ground truth" for the analysis came from Victoria Land and east Antarctica. The biology of the Antarctic Peninsula is different from the rest of Antarctica: Adélie penguin colonies on the peninsula intermix with those of other species, and the peninsula is at least partially vegetated. The Antarctic Peninsula is characterized by a different climate regime than the rest of Antarctica, with wetter surface conditions and more persistent cloud cover. Furthermore, the generally north–south orientation of the coastline may result in greater shadowing of Adélie penguin colonies. This increases the difficulty of penguin colony detection as described in the results presented here for retrieval of the Adélie penguin colonies in Victoria Land, which are also characterized by a similar coastline orientation. Given these physical and biological differences, the error characteristics of the algorithm used to retrieve Adélie penguin colonies from the Landsat data along the southern coastline of Antarctica are likely to be significantly different from those found on the Antarctic Peninsula. A retrieval of Adélie penguin colonies along Antarctic Peninsula deserves further study, but it is beyond the scope of the investigation presented here.

The results of the investigation presented here demonstrate the utility of multispectral satellite remote sensing data for mapping the distribution of major Adélie penguin colonies on large regional and even continental scales. Given the strong correlation between retrieved colony area and population also demonstrated here, it seems entirely possible that satellite remote sensing can contribute to a continent-wide Adélie population estimate. Such an estimate would likely be based on a combination of regional-scale surveys using ground or aircraft observations in accessible areas, while relying on satellite data to estimate Adélie breeding populations in the many remote regions of Antarctica.

Acknowledgments

The direct survey in east Antarctica was conducted under AAD ASAC project 2722 in accordance with permits issued under the Antarctic Treaty (Environmental Protection) Act 1980 and was approved by the Australian Antarctic Ethics Committee. We thank Mr. Woody Turner of NASA Headquarters for funding the publication of this paper. We also thank three anonymous reviewers for their insightful comments and suggestions, as well as Paul Ensor for sharing his observations of Mt. Biscoe and the Mackellar Islands in the 1980s, and Gary Geller, Heather Lynch, and Phil Trathan for helpful comments on a preliminary version of the manuscript.

Appendix B. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.rse.2013.08.009. These data include an animation illustrating the Adélie penguin colony area around Mt. Biscoe. Google Map visualizations of the Adélie penguin colonies retrieved from Landsat ETM+ data and the location of penguin colonies reported in the literature are also available as online supplementary datasets.

References

Ainley, D.G. (2002). *The Adélie penguin: Bellwether of climate change.* New York: Columbia University Press.

Ainley, D., Ballard, G., Ackley, S., Blight, L. K., Eastman, J. T., Emslie, S. D., et al. (2007). Paradigm lost, or is top-down forcing no longer significant in the Antarctic marine ecosystem? *Antarctic Science*, *19*, 283–290.

Ainley, D.G., Clarke, E. D., Arrigo, K., Fraser, W. R., Kato, A., Barton, K. J., et al. (2005). Decadal-scale changes in the climate and biota of the Pacific sector of the Southern Ocean, 1950s to the 1990s. *Antarctic Science*, *17*, 171–182.

Ainley, D.G., Nur, N., & Woehler, E. J. (1995). Factors affecting the distribution and size of pygoscelid penguin colonies in the Antarctic. *Auk*, *112*, 171–182.

Anonymous (1992). The southern rim of the Pacific Ocean: Preliminary geologic report of the Amundsen Sea—Bellinghausen Sea cruise of the Polar Sea, 12 February–12 March 1992. *Antarctic Journal of the United States*, *27*, 11–14.

Barbraud, C., Delord, K. C., Micol, T., & Jouventin, P. (1999). First census of breeding seabirds between Cap Bienvenue (Terre Adélie) and Moyes Islands (King George V Land), Antarctica: New records for Antarctic seabird populations. *Polar Biology*, *21*, 146–150.

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

Bhikharidas, A. K., Whitehead, M.D., & Peterson, J. A. (1992). Mapping Adélie penguin rookeries in the Vestfold Hills and Rauer Islands, east Antarctica, using Spot HRV data. *International Journal of Remote Sensing*, *13*, 1577–1583.

Bindschadler, R., Vornberger, P., Fleming, A., Fox, A., Mullins, J., Binnie, D., et al. (2008). The Landsat image mosaic of Antarctica. *Remote Sensing of Environment*, *112*, 4214–4226.

Caceci, M. S., & Cacheris, W. P. (1984). Fitting curves to data. *Byte*, *9*, 340.

Cochran, W. G. (1977). *Sampling techniques.* : John Wiley & Sons.

Cooper, J. (1985). Adélie penguins breeding in eastern Enderby Land, Antarctica. *Emu*, *85*, 205–206.

Creuwels, J. C. S., Poncet, S., Hodum, P. J., & van Franeker, J. A. (2007). Distribution and abundance of the Southern Fulmar Fulmarus glacialoides. *Polar Biology*, *30*, 1083–1097.

Croxall, J. P., & Kirkwood, E. D. (1979). *The distribution of penguins on the Antarctic Peninsula and islands of the Scotia Sea.* Cambridge: British Antarctic Survey.

Emmerson, L., & Southwell, C. (2008). Sea ice cover and its influence on Adélie penguin reproductive performance. *Ecology*, *89*, 2096–2102.

Ensor, P. H., & Bassett, J. A. (1987). *The breeding status of Adélie penguins and other birds on the cost of George V Land, Antarctica.* Kingston, Tasmania: Australian National Antarctic Research Expeditions (ANARE) (Research Notes 50).

Falla, R. A. (1937). *Birds.* Adelaide: The B.A.N.Z.A.R. Expedition Committee.

Forcada, J., & Trathan, P. N. (2009). Penguin responses to climate change in the Southern Ocean. *Global Change Biology*, *15*, 1618–1630.

Fretwell, P. T., LaRue, M.A., Morin, P., Kooyman, G. L., Wienecke, B., Ratcliffe, N., et al. (2012). An emperor penguin population estimate: The first global, synoptic survey of a species from space. *PloS One*, *7*.

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

Hirano, A., Welch, R., & Lang, H. (2003). Mapping from ASTER stereo image data: DEM validation and accuracy assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*, 356–370.

Horne, R. S.C. (1983). *The distribution of penguin breeding colonies on the Australian Antarctic Territory, Heard island, the McDonald Islands, and Macquarie Island.* : Australian National Antarctic Research Expeditions (ANARE) (Research Notes 50).

Hoshiai, T., Sweda, T., & Tanimura, A. (1984). Adélie penguin census in the 1981–82 and 1982–83 breeding seasons near Syowa station, Antarctica. *Memoirs of the National Institute of Polar Research*, *32*, 117–121.

Kato, A., Watanuki, Y., & Naito, Y. (2003). Annual and seasonal changes in foraging site and diving behavior in Adélie penguins. *Polar Biology*, *26*, 389–395.

Kendall, M. G. (1962). *Rank correlation methods.* New York: Hafner Publishing Company.

King, J. C., & Comiso, J. C. (2003). The spatial coherence of interannual temperature variations in the Antarctic Peninsula. *Geophysical Research Letters*, *30*.

Lee, D. S., Storey, J. C., Choate, M. J., & Hayes, R. W. (2004). Four years of Landsat-7 on-orbit geometric calibration and performance. *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 2786–2795.

Lynch, H. J., White, R., Black, A.D., & Naveen, R. (2012). Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar Biology*, *35*, 963–968.

Mawson, D. (1930). The Antarctic cruise of the "Discovery," 1929–1930. *Geographical Review*, *20*, 535–554.

Minnett, P. J., & Key, E. L. (2007). Meteorology and atmosphere—Surface coupling in and around polynyas. In W. O. SmithJr., & D.G. Barber (Eds.), *Polynyas windows to the world* (pp. 127–161). Amsterdam: Elsevier.

Mustafa, O., Pfeifer, C., Hans-Ulrich, P., Kopp, M., & Metzig, R. (2012). Pilot study on monitoring climate induced changes in penguin colonies in the Antarctic using satellite imagesIn N.C.a.N.S. German Federal Ministry of the Environment (Ed.), .

Naveen, R., Lynch, H. J., Forrest, S., Mueller, T., & Polito, M. (2012). First direct, site wide penguin survey at Deception Island, Antarctica, suggests significant declines in breeding chinstrap penguins. *Polar Biology*, *35*, 1879–1888.

Nelder, J. A., & Mead, R. (1965). A simplex-method for function minimization. *The Computer Journal*, *7*, 308–313.

Neter, J., & Wasserman, W. (1974). *Applied linear statistical models.* Homewood, Illinois: Richard D. Irwin, Inc.

Sabine, C. (1999). Remote sensing strategies for mineral exploration. In A. N. Rencz (Ed.), *Manual of remote sensing* (pp. 375–447). New York: John Wiley & Sons, Inc.

Schwaller, M. R., Benninghoff, W. S., & Olson, C. E. (1984). Prospects for satellite remote-sensing of Adélie penguin rookeries. *International Journal of Remote Sensing*, *5*, 849–853.

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

Southwell, C., & Emmerson, L. (2013a). Large scale occupancy surveys in East Antarctica discover new Adélie penguin breeding sites and reveal an expanding breeding distribution. *Antarctic Science*, *25*, 531–535.

Southwell, C. J., & Emmerson, L. M. (2013b). New counts of Adélie penguin populations at Scullin and Murray monoliths, Mac. Robertson Land, East Antarctica. *Antarctic Science*, *25*, 381–384.

Southwell, C., & Emmerson, L. (2013c). First population counts at newly discovered Adélie Penguin *Pygoscelis adeliae* breeding sites along the Wilhelm II, Queen Mary and Wilkes Land coastlines, east Antarctica. *Marine Ornithology*, *41*, 87–89.

Southwell, C., McKinlay, J., Low, M., Wilson, D., Newbery, K., Lieser, J. L., et al. (2013). New methods and technologies for regional-scale abundance estimation of land-breeding marine animals: Application to Adélie penguin populations in East Antarctica. *Polar Biology*, *36*, 843–856.

Southwell, C., Smith, D., & Bender, A. (2009). Incomplete search effort: A potential source of bias in estimates of Adélie penguin breeding populations in the Australian Antarctic Territory. *Polar Record*, *45*, 375–380.

Starck, W. (1980). The avifauna of Haswell Island (East Antarctica) in summer 1978/1980. *Polish Polar Research*, *1*, 183–196.

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

Woehler, E. J. (1993). *The distribution and abundance of Antarctic and Subantarctic penguins.* Cambridge: Scientific Committee on Antarctic Research, Scott Polar Research Institute.

Woehler, E. J., & Croxall, J. P. (1997). The status and trends of Antarctic and sub-Antarctic seabirds. *Marine Ornithology*, *25*, 43–66.

Woehler, E. J., & Riddle, M. J. (1998). Spatial relationships of Adélie penguin colonies: Implications for assessing population changes from remote imagery. *Antarctic Science*, *10*, 449–454.