Quantitative ornithology with a commercial marine radar: standard-target calibration, target detection and tracking, and measurement of echoes from individuals and flocks

Samuel S. Urmy* and Joseph D. Warren

Stony Brook University, School of Marine and Atmospheric Sciences, 239 Montauk Hwy., Southampton, NY 11901, USA

Summary

1. Marine surveillance radars are commonly used for radar ornithology, but they are rarely calibrated. This prevents them from measuring the radar cross-sections (RCS) of the birds under study. Furthermore, if the birds are aggregated too closely for the radar to resolve them individually, the bulk volume reflectivity cannot be translated into a numerical density.

2. We calibrated a commercial off-the-shelf marine radar, using a standard spherical target of known RCS. Once calibrated, the radar was used to measure the RCS of common and roseate terns (Sterna hirundo L. and Sterna dougallii Montagu) tracked from a land-based installation at their breeding colony on Great Gull Island, NY, USA. We also integrated echoes from flocks of terns, comparing these total flock cross-sections with visual counts from photos taken at the same time as the radar measurements.

3. The radar’s calibration parameters were determined with 1% error. RCS measurements made after calibration were expected to be accurate within ±2 dB. Mean tern RCS was estimated at −28 dB relative to one square meter (dBsm), agreeing in magnitude with a simple theoretical model. RCS was 3–4 dB higher when birds’ aspect angles were broadside to the radar beam compared with head- or tail-on. Integrated flock cross-section was linearly related to the number of birds. The slope of this line, an independent estimate of RCS, was −32 dBsm, within an order of magnitude of the estimate from individual birds, and near the middle of the frequency distribution of RCS values.

4. These results indicate that a calibrated marine radar can count the birds in an aggregation via echo integration. Field calibration of marine radars is practical, enables useful measurements, and should be done more often.

Key-words: echo integration, Great Gull Island, radar cross-section, radar equation, radar ornithology, random forest, seabirds, terns, track while scan

Introduction

Radar has been used to study the movements and distribution of birds since the end of the Second World War (Eastwood 1967), and is a key technique in the growing field of aeroecology (Chilson et al. 2012). At the largest scale, networks of weather radars have revealed continental patterns of bird migration (Bruderer 1997; Chilson et al. 2012; Shamoun-Baranes et al. 2014). Radar has also been applied to ecological problems at more local scales, such as counting cryptic species (Burger 2001), avoiding bird–airplane collisions at airports (Nohara et al. 2005, 2007), and siting wind farms (Desholm et al. 2006). Radar, like other active remote-sensing techniques, returns limited information on the objects or animals under observation, but it has powerful advantages. It can collect near-continuous data over large spatial extents at high spatial and temporal resolution, and is effective when visual observations are not, for instance in darkness, fog or clouds.

To interpret radar echoes from an object, it is necessary to understand how that object scatters electromagnetic (EM) radiation at the frequency of interest. The radar cross-section (RCS) quantifies how strongly a target scatters EM waves, and is a key variable in most radar calculations. RCS, especially for birds, is a complex function of multiple variables. These include the size of the bird relative to the wavelength, the bird’s aspect angle to the incident radiation, the polarization of that radiation, the shape of the bird’s body, and the position of its wings. Accurate RCS values are necessary for remote size identification. They are also necessary to convert volumetric backscatter from distributed targets (e.g. migrating birds on weather radar) to estimates of absolute animal density. RCS can be estimated from a physics-based scattering model, or measured empirically from a known target. Scattering models for birds have mostly been based on extremely simplified geometric shapes (Shaefer 1968; Vaughn 1985), though more sophisticated models have recently appeared (Torvik et al. 2014). Even when theoretical scattering models are available, empirical measurements are needed to validate them.
Marine radars, sometimes modified, have been used frequently by radar ornithologists over the past several decades. They are commercially available, relatively inexpensive, operate at wavelengths appropriate for observing birds, and are designed for reliable operation under harsh conditions at sea or in the field (Larkin & Diehl 2012). In the past, many studies simply counted targets on the radar’s display, or took photographs or video of it. As analog-to-digital converters with the necessary bandwidth have become more widely available and less expensive, more studies have made use of them to digitize and record radar data. However, even when they capture digital data, radar ornithologists rarely calibrate their radars. A number of approaches to radar calibration exist (Atlas 2002), but of these, perhaps the most practical is the use of a standard target with known RCS.

This paper has three main objectives. The first is to describe a practical method of capturing digital data from an off-the-shelf marine radar and calibrating these data to an absolute reference with a standard target. The second goal is to empirically measure the RCS of wild terns in the field. The final goal is to assess the practicality of echo integration with such a radar, by comparing the integrated energy from a number of tern flocks with that predicted from the empirical RCS and the number of birds each flock contained.

Materials and methods

STUDY SPECIES AND LOCATION

This study was conducted during the summer of 2014 at Great Gull Island, New York, USA (hereafter GGI). GGI is 7 hectares in area and is located at the mouth of Long Island Sound, at 41°12′7″ N, 72°7′5″ W. A former coastal defense battery, GGI is currently a research station of the American Museum of Natural History. From May through August, it is home to a breeding colony of c. 10 000 pairs of common terns (Sterna hirundo) and 1000 pairs of roseate terns (Sterna dougallii). Both species feed on small fishes and zooplankton by plunge diving from 3–10 m above the water. They make multiple foraging trips per day to feed their chicks, and often form dense circling flocks over concentrations of prey. This project was approved by the Stony Brook University Institutional Animal Care and Use Committee (application number 2014-2101).

RADAR SYSTEM

We used a commercial marine radar (Furuno FR-7252, Furuno Electric Co. Ltd., Nishinomiya, Hyogo Prefecture, Japan) to track the terns. This radar transmitted at a nominal frequency of 9.41 GHz (X-band) with a wavelength of 3.2 cm. Its peak transmit power was 25 kW. The antenna was the stock 1.8-m slotted waveguide bar antenna, transmitting and receiving horizontally polarized radiation. The nominal beam width, between half-power points, was 12° horizontally and 22° in the vertical (i.e. 11° above and below the horizon). The antenna rotated once every 2.4 s, with a nominal pulse repetition frequency (PRF) of 2100 Hz, making a total of 5040 pulses per sweep with an inter-pulse angle of 0.07°. Because the inter-pulse angle was less than the beam angle, even a point-like target would be illuminated by at least 16 pulses as the beam rotated past. The radar was operated in short-range mode with a pulse length of 0.08 μs, giving an effective range resolution of c. 12 m. The radar was mounted on the flat roof of a concrete WWII-era fire control tower, the highest point on GGI at c. 20 meters above sea level (m.a.s.l.).

Radar returns were captured with a digitizing oscilloscope (PicoScope 3405B, Pico Technology, Cambridgeshire, UK). Three signals from the radar’s auxiliary display output were connected to the digitizer via coaxial cables: the ‘heading’ pulse, marking each rotation of the antenna, the ‘trigger’ pulse, marking the transmission of each pulse, and the ‘video’ signal, whose voltage is proportional to the logarithm of the received echo power (Larkin & Diehl 2012). The oscilloscope, operating in ‘rapid block’ mode, captured every pulse from one rotation of the antenna, then wrote this data to disk storage during the subsequent sweep (because of limited data transfer speeds it was not possible to capture every sweep). This lead to an effective temporal resolution of 1 sweep per 4.8 s, or 0.21 Hz. Though the nominal range resolution was 12 m, each pulse was digitized at a sampling rate of 50 MHz, equivalent to a range resolution of 6 m (i.e. oversampled) since the shapes of the transmitted pulse envelopes were closer to Gaussian than square. The oscilloscope was controlled with a custom Python program running on a laptop computer. All electrical power was supplied by a portable gasoline generator through a 24 V AC/DC power converter.

STANDARD-TARGET CALIBRATION

The radar was calibrated on GGI using a hollow, 20.3 cm diameter stainless-steel sphere (Rome Industries #708-S, Peoria, IL, USA). We calculated the sphere’s RCS as 0.0321 m² (321 cm²), using the Mie series solution for a perfectly conducting sphere (Mahafza 2000). The sphere was suspended from the top of a 5.2-m, non-conducting, carbon-fibre pole. An assistant stood in the radar shadow of a hill while holding the pole upright so that the sphere was illuminated above the hill’s crest (Fig. 1). This exercise was repeated at ranges of 166, 309, and 419 m (as measured by the radar) on GGI during August 2014. We conducted additional calibrations in September 2015 in an empty parking lot in Southampton, NY, with the radar mounted on a rolling cart with the antenna 1.25 m above ground. The sphere was supported at the same height by a non-conducting plastic stand and placed 38.62, and 146 m from the radar. Approximately one hundred echoes were recorded from the sphere at each range. Although we did not shield the radar from ground clutter during the parking lot calibration, backscattered energy from the dry asphalt was quite low (at least 46 dB below the peak of the calibration sphere), so it was a relatively minor source of bias.

We manually selected the sphere on each radar image, rejecting those where it dipped behind one of the hills or could not
be separated from passing birds. The peak voltage recorded as the beam passed over the sphere was retained. The radar’s video signal $S$ is proportional to the logarithm of the received echo power. The received power is determined by the radar equation, which can be written in simplified form as

$$ P_r = \frac{P_T \sigma G}{r^2} , \quad \text{eqn 1} $$

(Larkin & Diehl 2012), where $P_r$ is the received power density (in W m$^{-2}$), $P_T$ is the transmitted power (in W), $\sigma$ is the target’s RCS (in m$^2$), $r$ is range (in m), and $G$ is a gain factor incorporating all other constants. These include the system gain, antenna beam pattern gain, range weighting gain due to the pulse shape and a calibration offset. If the received level $RL$ (in dB re 1 W m$^{-2}$) is $10 \log_{10}(P_r)$, and $S = a \times RL$, where $a$ is a constant of proportionality relating the logarithm of the received power to the video signal, then

$$ S = a [10 \log_{10}(P_T) + 10 \log_{10}(\sigma) - 40 \log_{10}(r) + G] + \epsilon , \quad \text{eqn 2} $$

where a normally distributed error $\epsilon$ has been introduced. This equation was solved by maximizing the likelihood of the observed echo amplitudes from the calibration sphere, $S$, conditional on $a$ and $G$, using the the ‘optim’ function in R (R Development Core Team 2014).

**EMPIRICAL RCS MEASUREMENTS**

**Target detection and classification**

Individual flying birds were detected and tracked through successive radar sweeps by an automated algorithm. Radar images were first smoothed with a Gaussian kernel with a 2 pixel standard deviation, implemented as part of SciPy (Jones, Oliphant & Peterson 2001), which removed ‘speckles’ from electrical noise and radio-frequency interference. Candidate bird targets were identified on smoothed images as local maxima using the Mahotas computer vision library (Coelho 2013). Targets with peak RCS less than $-60$ dBsm were rejected as too small to be birds.

Many objects besides birds appear as local maxima on the radar image, including boats, reflections from waves (‘sea clutter’), buoys, and electrical or radio interference. Visually, most of these false targets are easily distinguished from birds based on characteristics such as size, shape, peak echo intensity and location (Stepanian, Chilson & Kelly 2014). However, translating what is obvious to the human eye into rules and procedures which a computer can follow is not straightforward. Instead of writing these rules explicitly, we instead opted to use supervised machine learning.

The general goal of a classification algorithm is to separate observations into different classes based on their measured characteristics or features. In this case, the classes of interest were ‘bird’ and ‘other’, and the features were a set of metrics extracted from each local maximum detected on the radar images. A subset of the data, in which a human has labelled each observation, are used to ‘train’ the algorithm, which can then (hopefully) be applied to classify new data.

The classification algorithm we selected was the random forest (Ho 1995; Breiman 2001). Random forests are based on decision trees (Safavian & Landgrebe 1990), which classify observations using a series of if/else statements based on the values of their features, analogous to a troubleshooting flowchart. Random forests are based on an ensemble of decision trees (hence ‘forest’), each one fit to a bootstrapped sample from the training data. When presented with new data, each tree in the forest essentially casts a vote as to how it should be classified. The randomized, ensemble nature of the random forest increases its robustness to overfitting compared to individual decision trees (Ho 1995). In another study using similar radars, Rosa et al. (2016) found random forests were more effective than other approaches to automatic target classification.

To train the random forest, we chose four radar images and detected all local maxima in each image. The area above a threshold of $-60$ dB surrounding each peak was then selected as a ‘target’. We automatically extracted 11 features from each target: its range, azimuth, size (number of pixels), eccentricity (width/depth), and the mean, median, 10th and 90th percentiles, maximum, and logarithmic RCS of the pixels inside the target region. Then, using a custom Python script, we manually clicked all targets which were ‘bird-like’.

Based on our experience collecting data in the field and visualizing it in the lab, birds created distinctive point-like echoes, with extent in range and azimuth similar to the resolution of the radar (i.e. $12$ m $\times$ 2$^\circ$). These echoes were smaller, lower-amplitude, and much less intense than echoes from boats, and larger and rounder than spikes of electrical noise or EM interference. The other major sources of non-biological echoes were land and sea clutter. Land clutter was stationary and obvious,
so we did not select any targets detected over land areas as bird-like. Most sea clutter was likewise easy to identify, since it occurred consistently in well-defined areas where tidal currents flowed over shallow bathymetry (Fig. 3a). When selecting bird-like targets for the training set, we also excluded any targets from these regions. Isolated echoes from waves, known as ‘sea spikes’ in the radar literature (Melfi et al. 2006), were harder to distinguish from birds. The possibility for false detections due to sea spikes was addressed through further filtering procedures, described below.

The final training data set included 107,602 individual targets, of which 3071 were manually identified as bird-like. Twenty percent of these targets were randomly removed from the training set and held for out-of-sample validation. A random forest of 100 trees, including a maximum of four predictor features at each split and with a maximum tree depth of 30, was fit to the training set, using the open-source TrackPy library (Pedregosa et al. 2012). The fitted model was tested against the validation set to estimate its sensitivity and specificity (i.e. the true positive and true negative rates) and its overall accuracy. We also calculated the ‘mean decrease impurity’ metric for each variable, which measures how much each the accuracy of the classifier decreases when the variable is permuted after training but before classification (Louppe et al. 2013).

**Target tracking and RCS measurement**

These target detection and tracking procedures were applied to the area around GGI on the morning of 22 July 2014. During this time, winds recorded at the National Oceanic and Atmospheric Administrations’s nearby data buoy (station number 44060, at 41°16’’ N, 72°4’’ W) were between 0 and 1·5 m s⁻¹. The sea state was near calm (corresponding to number 0–1 on the Beaufort wind force scale). Examining echoes from areas of water outside the tidal rips, containing no birds or other sources of backscatter, we found that energy returned from the sea surface was even less than that due to electrical noise in the radar system. In fact, sea clutter in these areas was difficult to detect in the noise without averaging over hundreds or thousands of pixels. As a result, we consider sea clutter as negligible outside of the tidal rips (Fig. 3a).

Two hundred sweeps were analysed, representing 16 min of data collection. In each sweep, targets (i.e. local maxima) were detected, and then classified as ‘bird’ or ‘other’ by the random forest. Retained bird targets were linked together into tracks using the open-source TrackPy library (Allan et al. 2014). Only tracks with five or more detections were retained for analysis. Tracks were also eliminated if their speed was not between 7 and 15 m s⁻¹, or if any frame-to-frame turning angles were <90°. These filtering steps were intended to eliminate any false tracks generated by the linking algorithm joining sea spikes or other random sources of point clutter. After all linking and filtering was done, 3665 bird targets, linked into 707 tracks, were available for analysis.

The peak voltage of each echo in the retained bird tracks was converted to an RCS using (2). After track identification, further restrictions were applied. We plotted the frequency distribution of their RCS, and calculated their summary statistics. These are reported in decibels relative to 1 m², but all calculations were performed in the linear domain. We also examined the dependence of RCS on aspect angle – i.e. the angle between a bird’s direction of flight and the radar beam – by fitting a simple loess smoother to RCS as a function of aspect angle. Measured RCS values were compared to the geometric cross-section of a spheroid (Mahafza 2000) with major and minor radii of 13 and 3 cm, approximately the size of a tern’s body (Cabot & Nisbet 2013). While this model is an extreme simplification of actual microwave backscatter from a bird, it provides a convenient order-of-magnitude reference.

**Ground truth**

Because species identification is not possible from radar echoes alone, stationary visual transects (Dokter et al. 2013) were used to estimate the species composition of the radar targets. In each set of observations, a telescope (Nikon Fieldscope III, Nikon Corporation, Tokyo, Japan) with 75× magnification and 2° field of view was aimed from the top floor of the radar’s observation tower (elevation ≈18 m) at several conspicuous landmarks. These were the Montauk Lighthouse (bearing 123° true), a ruined fortification at the northern end of Gardiner’s Island (198°), a water tower on Plum Island (251°), and a smokestack at the Millstone Nuclear Power Plant in Connecticut (342°). For 4 min at each landmark an observer looked through the telescope, calling out the species of each bird crossing the field of view, while a second observer tallied these counts. Birds were identified to the genus level, since species identification was not always reliable during the short time each bird was in view. The main groups were Sterna terns (common and roseate, S. hirundo and S. dougallii), Larus gulls (American herring, great black-backed and ring-billed gulls, L. smithsonianus, L. marinus, and L. delawarensis), double-crested cormorants (Phalacrocorax auritus) and others (mostly small passerines). Thirty-eight sets of four transects were conducted on 16 days in 2014, at times ranging from 05.00 to 20.00h.

**Flock integration**

Visual scans with 8 × 40 lensatic binoculars were made every 30 min while the radar was collecting data. Any tern feeding flocks present were recorded, along with their compass bearing, approximate range, and estimated number of birds. Flocks were also photographed with a Nikon 7100 DSLR with a 70–300 mm zoom lens and polarizing filter. These photos were later examined in detail; if the flock was well-lit and in clear focus, the birds in the flock were counted manually. The photo’s time of capture was also noted from its metadata. These visual counts were compared with integrated echo energy from their corresponding flocks.

Before making this comparison, however, it is necessary to modify the usual radar equation for the unusual geometry in this study. We will first outline the physical intuition behind the standard version of the radar equation for distributed scatterers. We then describe the unique aspects of echo integration
for tern flocks, and show how they require a modified radar equation to account for them.

**Standard radar equation for distributed targets**

When there are multiple scatterers in a single pulse volume, the total measured reflectivity at that range is equal to the total number of scatterers \( N \) multiplied by their mean RCS, \( \langle \sigma \rangle \).

\[
P_r = \frac{P_T N \langle \sigma \rangle G}{\rho^4}.
\]

The number of scatterers is in turn the product of their numerical density \( n \) (number per m\(^3\)) and the pulse volume, \( V \) (m\(^3\)).

If the radar has horizontal and vertical beam widths of \( h \) and \( \phi \), the pulse volume is given by

\[
V = \frac{\pi}{4} h^2 \left( \frac{\phi}{2} \right) \sin \left( \frac{\phi}{2} \right).
\]

where \( c \) and \( t \) are the speed of light (in m s\(^{-1}\)) and the pulse length (in s). Using the fact that \( N = n V \), we can rewrite Equation 3 in terms of the numerical density of birds as

\[
P_r = \frac{P_T n \pi r^2 \sin (\theta/2) \sin (\phi/2) \langle \sigma \rangle G}{\rho^4}
\]

\[
= \frac{P_T n \langle \sigma \rangle G'}{\rho^4},
\]

where the constant factors from (4) have been incorporated into a new constant \( G' \). Intuitively, the energy losses due to two-way geometrical spreading still scale as \( r^{-2} \), but the number of birds (and hence reflectivity) increases proportional to \( r^2 \). The partial cancellation of these two scaling relations leaves \( r^2 \) in the denominator.

**Modified radar equation**

For the present purpose, however, a different version of the radar equation is necessary. Feeding flocks, depending on the number of terns, are on the order of 10s to 100s of meters across. However, feeding terns rarely fly higher than about 6 m above the surface (Cabot & Nisbet 2013). Because the radar beam is vertically fan-shaped, feeding flocks will fill it horizontally, but not vertically. As a result, while the pulse volume still increases \( \propto r^2 \) due to vertical and horizontal spreading, the illuminated volume of the flock, \( V_{\text{flock}} \), only increases in one direction, horizontally. As a result, the number of birds illuminated is

\[
N = n V_{\text{flock}} = n c h r \sin \left( \frac{\theta}{2} \right),
\]

where \( h \) is the average height of the flock. Substituting (5) into (3), we arrive at the correct (if unusual) radar equation for this geometry, with a factor of \( r^2 \) in the denominator,

\[
P_r = \frac{P_T n \langle \sigma \rangle G}{\rho^3},
\]

where, once again, all constants have been gathered together into \( G \) for simplicity.

**Integration and regression**

The radar reflections from each photographed flock were outlined by hand using a lasso tool in a custom Python script. The mean volume reflectivity inside each outline was multiplied by the total outlined area to give an integral over this area. The five radar sweeps recorded closest in time to each photo were integrated and averaged together. Only flocks whose echoes could be clearly separated from sea clutter and other targets such as boats were used. Integrated echo energy was regressed on the number of birds counted in the corresponding flock. The slope of this line is expected to equal the average RCS of a single bird, \( \langle \sigma \rangle \) (Diehl, Larkin & Black 2003). Based on the diagnostics run on a preliminary ordinary least-squares regression, several of the data points had outsized influence (Cook’s distance \( > n/4 \)). To ensure the reliability of our results, we fit the model, using robust regression (iterated re-weighted least squares), as implemented in the R function ‘rim’ in the ‘mass’ package (Venables & Ripley 2002). The scripts used for this analysis, and all other analyses described here, are available in the online Supporting Information for this article (Data S1).

**Results**

Echoes from the calibration sphere at different ranges were well described (\( R^2 = 0.85 \)) by the theoretical curve (Equation 2). The optimized values (\( \pm 95\% \) confidence interval) of \( a \) and \( G \) were 0.0126 (\( \pm 0.0003 \)) V dB\(^{-1}\) and 144.7 (\( \pm 2.1 \)) dB (Fig. 2). Though the parameters were fairly well constrained, the likelihood surface showed an inverse relationship between them.
indicating that the two parameters could trade off against each other. The echoes were also variable in amplitude, with a residual standard deviation of 0.06 V. The variability in the echoes appeared greatest at the longest ranges.

When tested against the validation data set ($n = 21,521$), the random forest correctly rejected 20,861 of 20,929 non-bird targets, for a specificity of 99.7%, and detected 423 of 529 bird-like targets, for a sensitivity of 71% (Fig. 3). The overall accuracy (correct/total classifications) was 99.99%. The most important features for classification in the random forest were mean echo level, azimuth angle, 90th percentile of echo level, and eccentricity (Table 1). When features were correlated (e.g. mean echo level, echo peak, and RCS), the random forest tended to rely on them roughly evenly, as seen in their similar importance weights.

The RCS of birds tracked on the radar ranged from −60 to −6 dBsm, with an overall mean and median of −25 and −34 dBsm (Fig. 4). In linear terms, these correspond to cross-sections of 32 and 4 cm$^2$. RCS was 3–4 dB higher at broadside incidence than head- or tail-on (Fig. 4a), though this was minor

Fig. 3. Detection, classification, and tracking of bird targets from radar images. (a) Radar image showing Great Gull Island (GGI, bright patch at centre), Little Gull Island (LGI, smaller patch to the northeast of GGI), sea clutter due to rough water in tidal rips (areas of diffuse scattering along diagonal line from SW to NE), and terns departing GGI. The radius of the displayed image is 1.6 km. (b) Same image, with all automatically detected local maxima marked by white points. Many of these detections are spurious (i.e. not birds), especially over land, over tidal rips, and at longer ranges, where the signal-to-noise ratio begins to fall. (c) Same image, but with estimated non-bird targets removed by a random-forest classifier. Note the rejection of almost all targets associated with areas of land and sea clutter. (d) Final set of bird tracks from this area during the sampled time period (c. 16 min). Most are flying towards or away from the colony, consistent with the behaviour of foraging terns.
effect: the loess model explained only 1% of the overall variance. Most birds were flying towards or away from GGI, so aspect angles near 0° and 180° (head-on and tail-on) were most common. The geometric cross-section of the tern-sized ellipsoid was greatest at broadside incidence and lowest end-on (−12 and −38 dBsm). Birds identified in visual transects were overwhelmingly terns: out of 5563 total birds counted, 5394, or 97%, were terns. Of the remainder, 86 were gulls, 72 were cormorants, and 11 were other species.

Sixteen flocks had sufficiently clear photographic and radar images to be counted and integrated. Integrated flock reflectivity was positively related to the number of birds (Fig. 5). The slope of the regression line was $6 \times 10^4$ m$^2$ per bird with a standard error of $1 \times 10^2$ m$^2$ per bird. This corresponds to a logarithmic RCS of −32 dBsm, or 6 cm$^2$. The difference of this slope from zero was highly significant ($P < 0.001$). The fitted intercept, $6 \times 10^{-3}$, was not significantly different from zero ($P = 0.77$).

Discussion

The theoretical calibration curve fit the raw echoes from the calibration sphere well. As an aside, the Mie-series solution for the sphere’s RCS was within 1% of its simple geometric cross-section (i.e. $\pi r^2$), so this approximation would have been adequate in this situation, with the sphere’s circumference more than 20 times the wavelength. The estimates of the two calibration parameters were fairly well constrained, with coefficients of variation (i.e. standard error/fitted value) both near 1%. However, the variable voltage of the raw echoes, especially at long ranges, limited the final accuracy of the calibration. The maximum-likelihood values of the two parameters were negatively related, indicating a minor degree of indeterminacy: an increase in the value of one parameter could be partly offset by a decrease in the other. Based on the precision of the estimate of the gain $G$, RCS measurements made with this radar system had a 95% confidence interval of ±2 dB. It is worth emphasizing that these calibration values are only valid for this instrument, and would have to be determined independently for other radars. They might also be expected to shift over time as the radar’s magnetron ages.

The uncertainty in the calibration parameters is due the variable peak voltages of the raw echoes from the calibration sphere. A certain amount of variability in the radar’s video signal was due to electrical noise. This radar was designed for reliable performance in a shipboard environment at reasonable cost, not for precise reflectivity measurements, so some of the electrical noise in the system is probably a result of the radar’s original design goals and compromises. Also, the gasoline

Table 1. Relative importance of features included in random forest classifier for bird targets

<table>
<thead>
<tr>
<th>Feature</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.18</td>
</tr>
<tr>
<td>Azimuth</td>
<td>0.13</td>
</tr>
<tr>
<td>90th percentile</td>
<td>0.11</td>
</tr>
<tr>
<td>Eccentricity</td>
<td>0.11</td>
</tr>
<tr>
<td>Echo peak</td>
<td>0.10</td>
</tr>
<tr>
<td>RCS</td>
<td>0.08</td>
</tr>
<tr>
<td>Range</td>
<td>0.07</td>
</tr>
<tr>
<td>Size</td>
<td>0.06</td>
</tr>
<tr>
<td>SD</td>
<td>0.05</td>
</tr>
<tr>
<td>Median</td>
<td>0.05</td>
</tr>
<tr>
<td>10th percentile</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Features were extracted automatically from local maxima on radar images. Importance weights are rounded for display and do not sum exactly to 1. RCS, radar cross-sections.
generator, while practical on an island with no outside power supply, was not as electrically quiet as a direct current, or a better-filtered alternating current, system could have been.

The higher variance in echo voltage at longer ranges is harder to explain, since electrical noise should be constant at all ranges. One possibility is that the sphere was always within a few degrees of the beam’s centre (the geometry of the radar’s beam is discussed more in detail below). Another possibility is a change in temperature. On the afternoon of 17 August 2014, when measurements were made at the two longest ranges, ambient temperatures were near 25°C, at least 5°C higher than on any of the other calibration dates, and the temperature inside the radar scanner, in exposed sunlight, may have been higher. Thermal noise in a conductor is proportional to temperature (Nyquist 1928), and this temperature difference could possibly account for higher noise in the longest-range measurements. Another possibility is that the radar beam’s side lobes may have transmitted or collected returns from objects off-axis, possibly contributing to the variability.

The random forest classifier had a much higher specificity (99.9%) than sensitivity (71%). Put differently, these numbers imply ‘type I’ and ‘type II’ error rates of <0.001 and 0.29. In other words, it was highly conservative, and passed very few false targets into the analytic sample. When combined with the additional filtering performed by the target tracking algorithm, in particular the rejection of short tracks and those with unrealistic speeds for terns, we can be even more confident that the large majority of tracked targets were birds. Finally, nearly all the birds observed visually near GGI were terns, supporting the assumption that all bird-like targets were due to terns. This conclusion is corroborated by the fact that most of the tracks were flying along paths to and from the island (Fig. 3d), consistent with behaviour by foraging terns. While there are probably a few returns from other species, and false detections due to sea clutter or external noise, they are not expected to affect the overall picture of tern RCS.

We intentionally selected a period with low winds and flat water to measure tern RCS. The performance of our algorithms, both in terms of sensitivity and specificity, would be expected to drop during rougher conditions, when energy scattered by the sea surface would overwhelm echoes from birds. Our goal in this study was to obtain the best possible measurements of tern RCS. For this reason, we only used data from a calm period. Assessing the performance of the target detection, classification, and tracking methods under varying conditions of wind and sea state is an important objective on its own, and will be the topic of a future study using this data set.

Empirical RCS values for tracked terns fell in the expected range for birds of their size (Vaughn 1985). RCS was highly variable, ranging across nearly five orders of magnitude, though this is expected for radar targets with complex shapes which change as they flap their wings (Shaefer 1968; Vaughn 1985; Chilson et al. 2012). Terns have a wingbeat frequency on the order of 1 Hz, which is much lower than the radar’s PRF (2100 Hz) and higher than its sweep frequency (0.2 Hz), meaning that our RCS measurements should be approximately independent snapshots of the birds’ wingbeat cycles. Measured RCS values were in reasonable agreement with the range of values predicted by the simple geometric model, though there was only a weak dependence of RCS on aspect angle. Birds are generally expected to scatter more strongly near broadside (Shaefer 1968), but this trend was evidently masked by other sources of error. These could include variability in the birds’ tilt angle (which the radar could not measure), system noise in the radar, and imprecision in estimates of the birds’ flight direction. Wind drift could also lead to track headings that differed from the birds’ true orientation. Finally, it is important to note that all RCS values here are for horizontally polarized radiation, and would be expected to change under alternate polarizations.

Total flock RCS was related linearly to the number of birds in each flock. The slope of this line, an independent estimate of mean tern RCS, was near the centre of the distribution measured from individual terns. Flocks displayed substantial variability in their integrated RCS from one radar image to the next. This variability could be due to several factors. The echo from multiple scatterers at the same range is the sum of their individual echoes, which may interfere constructively or destructively depending on their positions in space. This is why it is necessary to average several successive echoes to estimate the flock’s total RCS. Birds could also leave or join the flock in between radar sweeps, changing the total number. Finally, the average orientation of birds could change from one sweep to the next, altering the mean RCS without changing the number of birds. Regardless of the cause or causes of within-flock variability, the slope of the fitted regression line was in agreement with the individual RCS measurements, indicating that echo integration with a marine radar is a viable method to estimate the number of seabirds in a near-surface feeding flock.

One aspect of the radar system we ignored in this work was its beam pattern. In general, the energy returned from a target depends on its angular distance from the centre of the beam. The horizontal beam pattern could be safely ignored, since the high PRI meant each target was ‘painted’ with at least 16 pulses (PRF × rotation period × horizontal beam width/360°), ensuring one pulse would land on-axis, or very near it. The effects of vertical position in the beam pattern were also likely minor, though for different reasons. Foraging terns habitually fly close to the water, rarely rising more than 5–10 m above the surface. If the radar is at height $h_1$ and a tern is flying at range $R$ and height $h_2$, then its off-axis angle is $\tan^{-1}((h_1 - h_2)/R)$. At ranges longer than a few hundred meters, this means that nearly all terns will be within 1–2° of the beam’s vertical axis. Since the beam is so wide vertically, the beam pattern is quite flat on axis, so this will be a minor source of error in most of the surveyed area. For the RCS measurements reported here, 90% of which come from ranges of 300 m or more, the effect is negligible. Closer to the radar, however, the beam pattern would have to be taken into account.
The shape of the radar’s pulse was also largely ignored in this study. As mentioned previously, the pulse generated by the magnetron is not square, ramping up and down from its peak transmit power over tens of nanoseconds. We also over-sampled the video signal with respect to the nominal range resolution. The combination of these factors meant that each bird target typically returned some energy in each of 4–6 consecutive range bins, with its highest return (used to calculate RCS) in one of the central bins. Temporal oversampling had a similar effect to painting the target with multiple pulses: at least one digitization bin was likely close to the true peak echo. Oversampling was not as great in range as it was in azimuth, however, and so the uncertainty as to where exactly the target fell in the range gate is another source of uncertainty both in the calibration and in measuring RCS. While oversampling in range and azimuth reduces uncertainty in RCS measurements, it also increases the data storage and processing requirements dramatically. These tradeoffs need to be considered when planning studies that depend on measuring RCS precisely.

If birds are flying overhead, measuring their RCS becomes much more challenging, since their position in the beam is hard to determine. Phased array radars are capable of measuring the target’s position electronically, but have historically been too expensive for all but military applications. However, phased array systems based on less-expensive solid-state amplifiers are beginning to be used in meteorological research. There is a possibility that they will become available to radar biologists as well at some point in the not-too-distant future.

This paper demonstrates a practical method to make marine radar measurements truly quantitative by calibrating the radar with a standard target. The procedure is straightforward and does not require a great deal of time. Field calibration can be performed in a few hours or less, and the analysis required is no more complex than fitting a nonlinear regression model. The benefits are considerable – even with an off-the-shelf marine radar, empirical RCS measurements can be taken, and the size of flocks estimated accurately from their integrated echoes.

The last major review of RCS in radar biology (Vaughn 1985) is over 30 years old. The physics of radar scattering from biological targets is poorly understood, and most models to date have been based on simple shapes like sphere, cylinders, and spheroids. To our knowledge only one group has calculated scattering from a bird model with realistic shape and material properties (Törvik et al. 2014). More theoretical and modelling work is clearly needed, as are field measurements of as many species as possible. Accurate RCS values are needed to estimate the density of animals in distributed bioscatter, for instance in large-scale studies of bird migration, using weather radars. And knowledge of the scattering properties of individual organisms is the first step on the difficult path to remotely classifying them. Before one can obtain these necessary data, calibration is the necessary procedure.


Received 6 July 2016; accepted 20 October 2016
Handling Editor: Francesca Parrini

**Supporting Information**

Additional Supporting Information may be found online in the supporting information tab for this article:

**Data S1.** Analysis scripts.