Ecological Informatics 27 (2015) 55–63

Contents lists available at ScienceDirect

Ecological Informatics

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

Classification of birds and bats using flight tracks

Valerie I. Cullinan [a,*], Shari Matzner [a], Corey A. Duberstein [b]

[a] Pacific Northwest National Laboratory, Marine Sciences Laboratory, 1529 W. Sequim Bay Rd., Sequim, WA 98382, United States
[b] Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, United States

ARTICLE INFO

Article history:
Received 4 February 2015
Received in revised form 19 March 2015
Accepted 25 March 2015
Available online 2 April 2015

Keywords:
Birds
Bats
Automated classification
Offshore wind energy
Thermal video

ABSTRACT

Classification of birds and bats that use areas targeted for offshore wind farm development is essential to evaluating the potential effects of development. The current approach to assessing the number and distribution of birds at sea is transect-surveys conducted by trained individuals in boats or planes, or analysis of imagery collected from aerial surveys. These methods can be costly and pose safety concerns so that observation times are limited to daylight hours and fair weather. We propose an alternative method based on analysis of thermal video that could be recorded autonomously. We present a framework for building models to classify birds and bats and their associated behaviors from their flight tracks. As an example, we developed a discriminant model for theoretical flight paths and applied it to data ($N = 64$ tracks) extracted from 5-minute video clips. The agreement between model- and observer-classified path types was initially only 41%, but it increased to 73% when small-scale jitter was censored and the number of different path types was reduced. Classification of 46 tracks of bats, swallows, gulls, and terns on average was 82% accurate, based on a jackknife cross-validation. Model classification of gulls and swallows ($N \geq 18$) was on average 73% and 85% correct, respectively. Model classification of bats and terns ($N = 4$ and 2, respectively) was 94% and 91% correct, respectively; however, the variance associated with the tracks from these targets is poorly estimated. The models developed here should be considered preliminary because they are based on a small data set both in terms of the numbers of species and the identified flight tracks. Future classification models could be improved if the distance between the camera and the target was known.

1. Introduction

The development of renewable energy sources, including wind, has become part of the future U.S. energy portfolio in an effort to reduce the dependence and the environmental impacts associated with extracting, transporting, and burning fossil fuels. In order to build and operate a wind farm, developers must abide by Federal and State conservation laws including the National Environmental Policy Act, Migratory Bird Treaty Act, and the Endangered Species Act. Each law has its own unique requirements, so the U.S. Fish and Wildlife Service (USFWS) provided guidelines to ease the process for developers of land-based wind energy (USFWS, 2012). These guidelines rely upon pre- and post-construction surveys using field personnel to determine the presence and abundance of birds in order to model and validate the risk to birds. As the annual production of electricity from land-based wind farms has grown, so has the interest in development of production-scale offshore wind energy. However, empirical data describing the distribution, abundance, behavior, and life history is lacking for many bird species. Further, land-based methods to gather these data are either impossible or prohibitively expensive offshore.

European offshore wind energy development precedes the U.S. and already includes 69 wind farms in eleven countries (EWEA, 2014). Methods used to assess risk and effects to bird populations at multiple scales during pre- and post-construction included direct observation or video surveys along boat and aerial transects (Banks et al., 2005; Camphuysen et al., 2004), collision risk models (Band, 2012; Smales et al., 2013), and field experiments to assess turbine avoidance (Guillemette and Larsen, 2002). As research efforts in Europe have gained a greater understanding of effects from wind farms, the need for additional tools and techniques to appropriately assess impacts has become apparent (Bailey et al., 2014). Efforts have been initiated to gather broad-scale wildlife distribution and abundance information off the Atlantic and Pacific coasts of the U.S. to assess potential offshore wind-wildlife conflicts and to aid in the leasing of offshore sites for wind energy development (Adams et al., 2014; Maclean et al., 2009; Normandeau, 2012; USGS, 2014). Research projects are also being conducted to gather empirical data necessary to model risk to birds from collision mortality with wind turbines operating offshore (BOEM, 2014). The limitations of using field personnel for offshore surveys will likely result in an increased use of remote sensing technologies to gather the data necessary to validate these risk models.

* Corresponding author. Tel.: +1 360 681 3662.
E-mail addresses: Valerie.cullinan@pnnl.gov (V.I. Cullinan), Shari.Matzner@pnnl.gov (S. Matzner), corey.duberstein@pnnl.gov (C.A. Duberstein).

http://dx.doi.org/10.1016/j.ecoinf.2015.03.004
1574-9541/

Although remote-sensing tools are already used to characterize offshore wildlife populations, expenses related to data interpretation are still very high and automated detection and classification methods are not yet well developed. For instance, Normandeau (2012) used automated detection algorithms on high spatial resolution imagery gathered for broad-scale assessment to screen out image frames without targets of interest and manually-reviewed the remaining images to glean data. Although Buckland et al. (2012) did not include cost as a factor in comparing digital and visual aerial survey methods, the authors did note that locating and identifying birds in images could be automated from digital survey imagery. Similarly Groom et al. (2013) used large format digital aerial cameras to record aerial surveys for marine birds. Image processing was used to extract regions within images that contained bird-like objects, and these data were then examined by trained ornithologists. Because of the intense effort involved in manual data review, an automated animal-detection algorithm that is both cost-effective and scientifically valid was deemed essential for offshore risk assessment (Normandeau, 2012).

Radar has been used to survey wildlife both day and night for site scale assessments of wind farm collision risk and avoidance behavior. Numerous types of radar have been used to detect, track, identify and study bird behavior including high-resolution marine surveillance radar near offshore wind farms (Gauthreaux and Belser, 2003; Plonczkier and Sims, 2012). Target periodic amplitude modulation of a radar echo has been used as an indicator of wing beat frequency and thus used to identify birds (Bruderer et al., 2010). However, fluctuations due to changes in scattering of the radar energy from the surface of the bird as if flies through a radar beam alter the periodicity of the amplitude modulation and complicate its use as a reliable signature of identity (Torvik et al., 2014). A fundamental piece of information provided by radar is distance to target, which is not available with a single video camera. Radar also provides much greater range than current thermal infrared cameras, but is generally more expensive and requires more power to operate making it difficult to employ at offshore locations that lack infrastructure. Further, target detection with marine radars is challenging due to noise and clutter (Jarrah et al., 2012), and radar echoes off wind turbines could be problematic if operated within the confines of an offshore wind farm.

We are researching the use of thermal infrared video to capture the flight tracks of birds and bats. The use of thermal imaging to detect wildlife may be limited by excessive humidity, distance from the camera, field of view, physical obstruction, and even plumage and pelage characteristics of the animals being studied (Cilulko et al, 2013). However, although optical cameras have a higher resolution than thermal cameras, thermal video can record observations both day and night and in weather conditions that make optical cameras less effective and is relatively inexpensive to collect and easier to interpret compared to radar. We assume that the shape of the flight path and statistics on changes in direction, which are less affected by low resolution than the image of the animal, will contribute to the classification of a target. Normandeau (2014) recorded birds and bats with thermographic cameras. Flight trajectories and animal shape were used to identify objects as foraging bats, bat/bird, or unknown. Erratic flight trajectories were identified as foraging bats, but it is unclear if or how timing and degree of changes in flight direction were quantified or otherwise evaluated. Although the flight path alone does not function as a species or even group level signature, the automated extraction of data from sensor-derived information could provide a method of significant cost savings for related research. The use of flight path as an indicator of identity, even at a coarse scale, will be a step towards eliciting finer scale classification.

Our objectives are to show that it is possible to automate the identification of behavioral flight paths and to classify the associated birds and bats. We present preliminary models for flight track and target classification based on a limited annotated library of tracks and associated characteristics extracted from a single-camera thermal video. Ultimately, our approach will be useful in estimating or validating the magnitude of risk from blade/tower collisions and evaluating avoidance behavior associated with taxa of concern in areas under consideration for wind energy development.

2. Methods

2.1. Data

A FieldPro 5× (Axsys Technologies) thermal video camera with a thermal sensitivity of 0.04 °C, frame rate of 30 Hz, and resolution of 0.03 m/pixel for a target at a distance of 100 m was used to record several hours of video using the sky as background. The camera was mounted on a pan/tilt unit on a tripod, and video was recorded from the shoreline of Sequim Bay and the Straights of Juan de Fuca, WA during the summer months, 2012. The camera was oriented to look just above the horizon for recording nearshore gulls and terns and oriented approximately 45° above the horizontal to record bats and swallows flying among tree tops. The thermal video was recorded directly onto a laptop hard drive which allowed simultaneous viewing of the recording. A field observer equipped with image-stabilized binoculars mounted on a tripod and an audio recorder and laptop conducted a visual survey of the expected field of view. Video clips were made from the larger recording to represent a range of environmental conditions and animal activity.

An annotated library was developed from the observer's field notes and five thermal video clips, each 5 min long cut from the original video. The field notes included the identity of the target, the time it was observed, and the direction of flight in relation to the camera's field of view (e.g., right to left, top to bottom). An additional observer reviewed the thermal video with the field notes and verified the recording of 160 tracks as identified by the field observer ($N = 39$ unknown targets; 2 planes; 1 boat, cormorant, and insect; 4 terns; 8 bats; 31 swallows; and 73 gulls). While viewing the video, the reviewer placed each track into one of five predetermined track shapes (n = 12 angled; 101 linear; 19 quadratic; 14 sine wave; and 14 turnaround) selected to mimic actual flight paths (Fig. 1). For purposes of analysis, it was assumed that target and track type as assigned by both the field observer and the reviewer were 100% correct.

A track-detection algorithm (Matzner et al., in preparation) which set a minimum of six target positions for detection was used on the thermal video. The algorithm extracted the sequential (x, y) positions in two-dimensional space from each of 104 detected tracks (including 8 tracks that were on the edges of the field of view that the field observer and reviewer missed). The algorithm output included a unique track identification number; the time, location, and frame number the target enters and exits the field of view; intensity; cluster dimensions of the pixel blob associated with the target; a sequence of x, y positions in each frame for the centroid (center of mass intensity) of the blob; and a measure of the curviness or sinuosity of the detected track. Sinuosity was defined in the algorithm as the interquartile range of the change in direction between successive centroids. Specifically, the direction of the path at time t (D_t) between the x–y locations from successive centroids at time (t and t − 1) was defined as $D_t = tan^{-1}\frac{y_t - y_{t-1}}{x_t - x_{t-1}}$, and the change in direction (C_t) between successive centroids was then calculated as $C_t = D_{t-1} - D_t$ and used as the basis for discriminating between the different types of flight path.

A theoretical data set of 30 realizations of each track shape ($N = 150$) was created using randomly generated model parameters for each realization in Microsoft Excel (2010). Gaussian (0, 1) noise was then added to each x and y value of the track to maximize the jitter and variability in the flight paths. Tracks were then smoothed using a moving window average with a window size of six x-y values (consecutive time points). Descriptive statistics were then calculated on the smoothed change in direction, for each realization, including the minimum, maximum, mean,

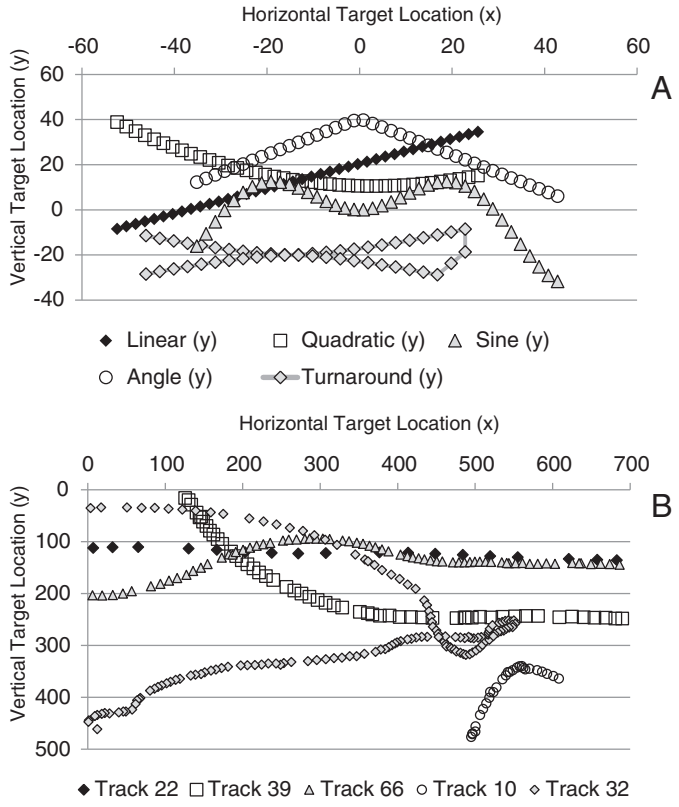

Fig. 1. Theoretical flight track shapes in x–y space (A) and observer-identified flight paths of similar shape (indicated with similar symbols) extracted from thermal video (B).

standard deviation, coefficient of variation (CV), and quartiles (Q1, Q2, and Q3) of the data distribution; the interquartile range (IQR = Q3 − Q1); the difference between the mean and median (central skewness); the range; the sample distribution skewness and kurtosis (a measure of how flat or peaked the distribution of the change in direction is within a track); and the number of times the sign changed between successive x–y elements (ignoring zeros).

2.2. Classification of track shape

A model that discriminates between track shapes was developed using a randomly selected subset ($N = 52$) of the theoretical tracks' descriptive statistics on the smoothed change in direction in Statistica (2013). The remaining tracks were used for verification of the model. Variables going into the forward-stepping discriminant model were standardized and relatively uncorrelated ($r < 0.85$). The track similarity, based on the Euclidean distance and the descriptive statistics of the change in direction for each track shape, was calculated using Minitab (2010), and because the angled, linear, and quadratic tracks were 73% similar (Table 1), a two-step model was constructed. Model 1 was used to discriminate between tracks classified as either sine wave ($N = 12$), turnaround ($N = 12$), or combined angled, linear, and quadratic (ALQ) tracks (ALQ group; $N = 28$). Then Model 2 was used to discriminate between the angled ($N = 8$), linear ($N = 12$), and quadratic tracks ($N = 8$). For Model 1, the maximum probability that a track was likely a member of a given track type had to be greater than 0.7. Otherwise, the track was considered outside of the modeled variability and not classified. For Model 2, if the maximum probability that a track was likely a member of a given track type was less than 0.7, the track was classified as ALQ.

Table 1
Matrix of similarity between track types as assigned by the reviewer; ALQ = the combined angled, linear, and quadratic tracks.

Track Type	Angled	Linear	Quadratic	ALQ	Sine Wave	Turnaround
Angled	80%	70%	74%	–	44%	60%
Linear	70%	69%	70%		41%	54%
Quadratic	74%	70%	83%		45%	58%
ALQ	–			73%	44%	57%
Sine Wave	44%	41%	45%	44%	46%	41%
Turnaround	60%	54%	58%	57%	41%	57%

The final two-step model was then tested on the remaining theoretical tracks and the tracks that were extracted from the thermal video. A minimum of 10 smoothed centroid locations were required to classify the shape of the tracks extracted from the thermal video resulting in $N = 64$ tracks.

The tracks extracted from video were very sinuous compared to the modeled tracks, so the change in direction between successive centroid locations was censored (i.e., set to zero) for absolute values less than 0.3. Classification was conducted with and without censoring.

2.3. Classification of target

A step-wise process to classifying target taxa from the thermal video was developed based on ($N = 46$; 4 bats, 2 terns, 22 gulls, and 18 swallows) flight path discriminant scores from Model 1 and other statistics output from the track-detection algorithm. The remaining 58 tracks detected by the algorithm had either less than 10 smoothed centroids and the track shape could not be classified or the taxon was unknown, one of a kind (boat or cormorant), or missed by the observer.

First, we estimated the probability that the target was near the camera based on a forward-stepping discriminant model and subjective reviewer-assigned estimates of the flight paths being near or far, as seen from the video ($N = 46$). The reviewer's assignment of near or far was relative to other targets and aided by the constant width of a line drawn by the algorithm over detected tracks (Fig. 2). If the wings could be clearly seen outside of the line, then the target was considered nearer to the camera than tracks that were covered by the line. This model was tested on the 58 tracks that could not be used for target classification. Variables used in the model were the standardized track characteristics: third quartile of the blob width (BW); the median number of pixels per blob divided by the number of frames per second the target was in view (B:S); the number of frames per second the target was in view (FPSV); the median BW divided by the median blob height (W:H); and the third quartile of the blob intensity (BI). The pair-wise correlation for each variable in the model was $r < 0.85$.

Second, we developed a model that discriminates between the observer-identified taxa using the second discriminant score from Model 1, the estimated probability that the target was near, and other statistics output from the track-detection algorithm. All variables were standardized. The target classification model was tested on five random samples with replacement ($N > 45$) targets from the modeled data using a jackknife cross-validation process.

3. Results

3.1. Classification of track shape

The classification of the modeled flight paths was much more successful than the classification of the tracks extracted from the thermal video (Fig. 1). Using the modeled flight paths, the first-step discriminant model (Model 1) had a Wilks' lambda (a measure of the proportion of variance that is unaccounted for in the combination of modeled variables; smaller values indicate a better discrimination) of 0.06 ($P < 0.0001$) and retained six of the variables associated with changes in direction (Table 2). The model was able to correctly classify 100% of the ALQ tracks, 75% of the sine wave tracks, and 83% of the turnaround tracks. Two Eigen values, both greater than 1, were able to explain 100%

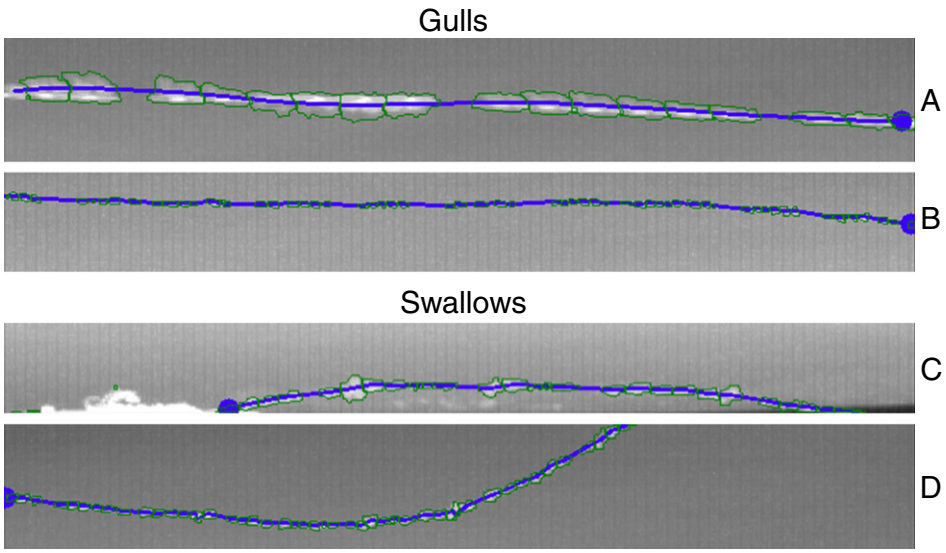

Fig. 2. Tracks detected by the algorithm (blue lines) assigned by the reviewer as near (A and C) or far (B and D) from the camera.

of the variability (Fig. 3). The kurtosis and the minimum change in direction had the greatest discriminating power and coefficients for the discriminant functions with the greatest magnitudes.

The second-step discriminant model (Model 2) had six variables and a Wilks' lambda of 0.295 ($P < 0.01$). The mean change in direction was the variable with the greatest discriminating power. The model was able to correctly classify 100% of the quadratic tracks, 83% of the linear tracks, and 63% of the angled tracks. Only one Eigen value was greater than 1 (explaining 61% of the variability), but both Eigen values explained a significant amount of information (Chi-square; $P < 0.05$) and were used in the classification of the tracks extracted from thermal video.

When the change in direction was not censored, 1 of 64 video-extracted tracks (Track 1) was declared to be outside the modeled variability (Fig. 4). The maximum membership probability for this track was slightly less than 0.7 for the turnaround track type even though the observer classified this flight path as linear and identified the target as an airplane. In this case, the track-extraction algorithm highlights the camera's sensitivity to the changes in atmospheric conditions near and far from the lens. For this track, six observed changes in direction had a magnitude greater than 0.3; thus, even after censoring small changes in direction, this track was still most likely classified as a turnaround track ($P = 0.71$). This type of error can be corrected by setting track types to linear when the magnitude of either the range in x-locations or y-locations of the track is small relative to the size of the camera frame in the corresponding axis.

For the remaining 63 tracks, the model- and observer-classified tracks were in agreement in 26 cases (41%). When the angled, linear, and quadratic track categories were combined (ALQ), the number of tracks in agreement increased to 59%. If only two major categories (ALQ and the combined sine wave and turnaround [ST] tracks) were considered, the number of tracks in agreement increased to 62%. Some of the disagreement in classification was associated with the number and magnitude of changes in direction (jitter) based on the bird movement and the track-extraction algorithm's definition of the centroid. The centroid tracked the movement of the wings and increased the number of changes in direction especially affecting linear flight paths. When changes in direction were not censored, only 12 (44%) of the 27 observer-identified linear flight paths were classified as linear. When the change in direction was censored, 24 (88%) of the 27 linear flight paths were classified as linear (Table 3).

When small changes in direction were censored (absolute values less than 0.3 set to 0), two different tracks out of the 64 were declared to be outside the modeled variability. Both tracks had a maximum membership probability of slightly less than 0.7 for the turnaround track type. Without the censoring, these tracks (both tight "U-shaped") were classified as turnaround tracks with probabilities of membership greater than 80%. In this case, the observer identified both tracks as quadratic flight paths. For the remaining 62 tracks, the model- and observer-classified tracks were in agreement in 35 cases (56%). Using only three categories, the number of tracks in agreement increases to 71%, and with two categories the number of tracks in agreement increases to 73%. With and without censoring small changes in direction, none of the video-extracted tracks were classified as "angled" using the discriminant model. Instead they were classified as linear, quadratic, or

Table 2
Raw mean, standard deviation, and coefficients for the canonical variables for the first discriminant model for sine wave, turnaround, and ALQ tracks.

Standardized descriptive statistic for the change in direction	Mean	Standard deviation	Root 1	Root 2
Kurtosis	3.99	7.49	2.56	0.234
Minimum	−0.86	0.76	2.037	−1.42
Skewness	−0.22	1.48	−0.322	0.600
Median	0.00	0.03	−0.654	−0.121
Number of sign changes	4.56	4.82	0.501	0.162
Central skew	0.00	0.03	−0.096	0.367
Eigen value			4.20	2.12
Cumulative percent of variability explained			66%	100%

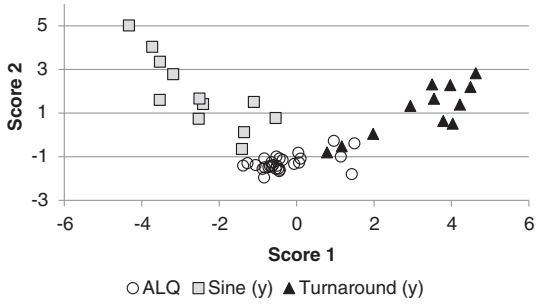

Fig. 3. Plot of the two discriminant scores ($N = 52$ tracks) for the first discriminant model (Table 2) based on modeled flight shapes presented in Table 1.

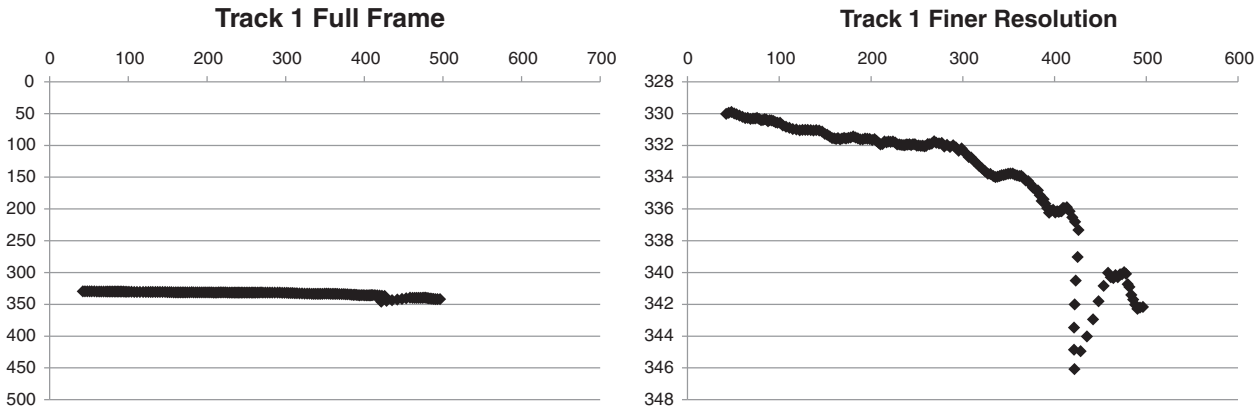

Fig. 4. Plot of the video-extracted x–y location of the centroids for Track 1, identified as an airplane, at the full frame of the camera and at a finer resolution.

turnaround—all associated with the magnitude and the number of changes in direction. Some of the disagreements in classification were associated with the resolution of the track (Fig. 3). Another source of disagreement was the difference between the observer's perceptions of what a track type should look like plus the potential for overlap within flight path types compared to the modeled track types. The extracted tracks (22, 39, 66, 10, and 32) in Fig. 1 (bottom) were classified by the discriminant model without and with censoring small changes in direction as linear, quadratic, linear, quadratic, and turnaround, respectively. For this exercise, the observer was not provided with specific rules for identification of track types and this could be the cause of many of the disagreements.

3.2. Probability of being near

The probability of the target being near the camera was estimated to reduce errors in classification of the target associated with the resolution of the track. We noticed that targets near the camera tended to travel farther per frame as a function of their size-to-speed ratio than targets clearly far from the camera (Fig. 5). However, for some targets the distance was not distinctively clear and there was overlap in the relationship between distance traveled and the size-to-speed ratio. Thus, for purposes of discrimination, a third distance category—"within bounds of potential error"—was assigned to observations with B:S values that fell within the absolute maximum boundary of overlap observed in Fig. 4; defined for modeling as $\log_{10}(\text{B:S}) = 0 \pm 0.25$. The probability of being near the camera, *Prob*(near), was then defined by the following function with multipliers shown in the discriminant score (S_N; Table 4):

$$Prob(\text{near}) = 1 - \sqrt{(S_N+3.24)^2} \Big/ \left\{\sqrt{(S_N+3.24)^2}+\sqrt{(S_N-1.22)^2}+\sqrt{(S_N-1.86)^2}\right\}$$

where $S_N = -2.24\,\text{BW} + 0.66\,\text{B:S} + 0.38\,\text{FPSV} - 0.41\,\text{W:H} - 0.27\,\text{BI}$. This discriminant model had a Wilks' lambda of 0.15 ($P < 0.0001$) and agreed with 86% (12 out of 14) of targets subjectively assigned as being "near" by the observer.

The near/far discriminant model was then used on 58 detected targets with short tracks or unknown targets not used for model building. The calculated probability of being near produced a bimodal distribution well separated at 0.6. Thus, the proximity to the camera was assigned as being "near" if the probability of being near was at least 0.6; otherwise it was assigned as being "far." The model-predicted assignments of distance were in agreement with 85% of the observer-assigned distances. However, because the assignment of near and far is subjective (just like the earlier classification of track types), a plot of the distance traveled per frame relative to the size-to-speed ratio observed in the modeled data (Fig. 5) was constructed with the test data (Fig. 6). In general, the relationship between variables was maintained in the test data. A notable exception is the unknown target that has a value of 15 for the distance traveled per frame and was estimated to be far from but observed to be near the camera and outside of the bounds of potential error. Only a subset of this flight path was detected ($N = 11$ elements), and the body blob characteristics may be poorly estimated.

3.3. Classification of target

For the classification of targets a discriminant model was developed for the three track types (ALQ, sine wave, and turnaround tracks) using the observer-identified taxa. The variables in the model output from the track-extraction algorithm were the third quartile of the BI and distance traveled. The remaining variables in the model were calculated from information output from the algorithm, the second discriminant score from Model 1 of the track classification, or the probability of being

Table 3
Confusion matrix for the classification of flight path types for tracks with at least 10 smoothed centroid x–y locations with and without censoring the change in direction.

Classification method and classified path type		Observed path type					Classified type total
		Angled	Linear	Quadratic	Sine	Turnaround	
Classification without censoring	Linear	1	12	2	4	0	19
	Quadratic	2	6	4	1	0	13
	Sine	0	2	0	0	1	3
	Turnaround	3	7	7	1	10	28
	NA	0	1	0	0	0	1
Classification with censored directional changes <0.3	Linear	3	24	6	6	0	39
	Quadratic	0	0	1	0	0	1
	Sine	0	2	1	0	1	4
	Turnaround	3	2	3	0	10	18
	NA	0	0	2	0	0	2
Observed type total		6	28	13	6	11	64

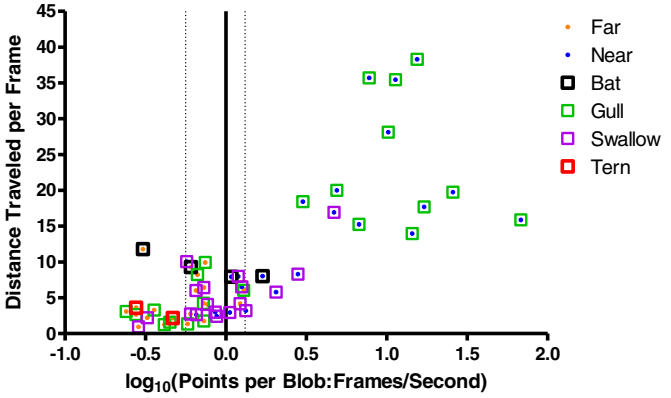

Fig. 5. Tracks identified by taxa and proximity to camera used to model the probability that the target was near the camera; dotted lines at (−0.25 and 0.12) depict the minimal bounds of overlap between near and far targets based on the minimum and maximum x-values in each class, respectively.

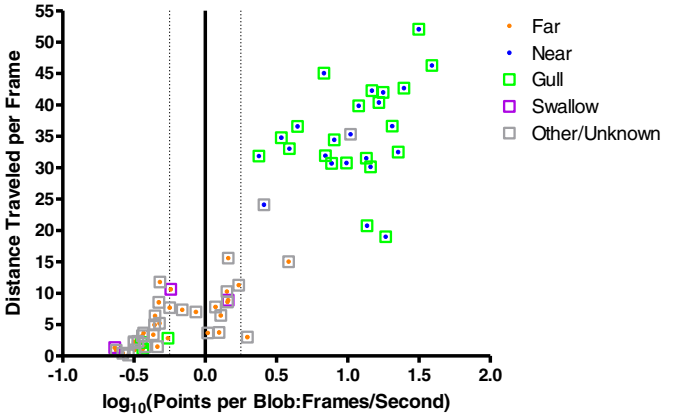

Fig. 6. Test tracks identified by taxa or unknown and the assignment of proximity to the camera based on the estimated probability that the target is near the camera; dotted lines depict the bounds of overlap between far and near targets based on the modeled "within error" category.

near the camera. The blob height CV was defined as the median blob height divided by the blob height standard deviation. The speed CV was defined as the mean speed (pixels/s) divided by the speed standard deviation. The proportional change in blob height was defined as the range in blob height divided by the maximum blob height. The resulting model had a Wilks' lambda of 0.15 ($N = 46$; $P < 0.0001$) and 87% correct classification (Table 5 and Fig. 7). Both Eigen values explained a significant amount of information (Chi-square; $P < 0.01$).

3.4. Error analysis of target classification

Errors in the classification of tracks unrelated to the detection errors from the extraction algorithm include misidentification of tracks, observed flight paths outside of the variability characterized by the five modeled track types, and associating types of tracks to either birds or bats based on behavior. The latter error is associated with the amount of data (tracks both detected and classified) from known birds and bats (correctly identified by humans, recorded in an annotated library of observed flight paths) and whether observations are representative of each target species' flight behavior.

Human errors include misidentification in observation and recording in an annotated library, observing flight paths outside the field of view or too far away from the camera, missing flight paths recorded by the camera, and the misidentification of track type and nearness to the camera. The latter two errors are based on subjective decisions and are difficult to assign as an error. It is difficult to bound the camera's field of view in the open sky and explicitly define the video resolution into human terms. Because of this difficulty, the field observer may detect targets just outside the field of view of the camera. When more targets are observed than captured on the video, it is difficult to assign the correct field observation to that recorded on video.

Finally, errors are made when classifying the target from classified tracks. Five bootstrap samples of the 46 tracks used to build the target discriminant model were created to assess four types of uncertainty in the model-building process: correlation structure and values used for standardization of data, coefficients for the discriminant model, jackknife cross-validation of target classification errors, and characteristics of tracks that were misclassified. The five samples were created by merging a random sample of 60% of the classified tracks with replacement and a stratified random sample of 50% of bats, swallows, and gulls, and 100% of terns. This sampling strategy guaranteed that there were at least two tern and two bat flight tracks in each cross-validation sample.

The mean and standard deviation of most of the parameters for the five samples were not significantly different from the values used to standardize the model data (t-test; minimum $P = 0.13$ and $P = 0.09$, respectively). The mean standard deviation for distance traveled was nearly significantly different between test samples (t-test; $P = 0.06$), and the mean standard deviation for distance traveled per frame was significantly different ($P = 0.02$). The largest correlation was between the probability of being near and the ratio of the BW third quartile and the proportional change in blob height extremes ($r = 0.91$, $N = 5$ samples). The second largest correlation was between the standard deviation of BW and the B:S ratio ($r = 0.86$).

The magnitude of the standardized coefficients for the discriminant model (Table 5) was calculated for the model data and each sample, then ranked and compared using Kendall's test of concordance. The standardized coefficients with the greatest rank have the greatest weight in the calculation of the discriminant score. The test of concordance tests the null hypothesis that the ranks of the coefficients between the six data sets were not associated. The rankings of the coefficients for the first score were not detected to be significantly different ($P = 0.15$), but the rankings for the second score were significantly

Table 4
Raw mean, standard deviation, and coefficients for the canonical variables for the discriminant model for targets near, within error, or far.

Standardized descriptive statistics for the track elements	Mean	Standard deviation	Root 1
Blob width Q3	14	12.7	−2.25
Median number of pixels:frames/s	5	11.1	0.66
Frames/s	27	6.5	0.38
Median blob width:median blob height	2	0.9	−0.41
Blob intensity Q3	1	0.1	−0.27
Eigen value			4.98
Cumulative percent of variability explained			98%

Table 5
Raw mean, standard deviation, and coefficients for the canonical variables for the discriminant model for tracks classified as angled, linear, or quadratic tracks and observed taxa.

Standardized descriptive statistics for the track elements	Mean	Standard deviation	Root 1	Root 2
Blob intensity Q3	1	0.1	−1.69	−0.49
Distance traveled	338	181	−0.56	0.31
Model 1 track discriminant score 2	−0.30	2.34	0.42	−0.19
Blob height CV	0.31	0.14	0.09	0.41
Speed CV	0.56	0.26	0.03	0.90
Probability (near)	0.45	0.26	0.23	1.57
Proportional change in blob height	0.75	0.13	−0.36	−1.05
Median blob width:median blob height	2	0.9	−0.26	−0.79
Eigen value			1.93	0.88
Cumulative percent of variability explained			64%	94%

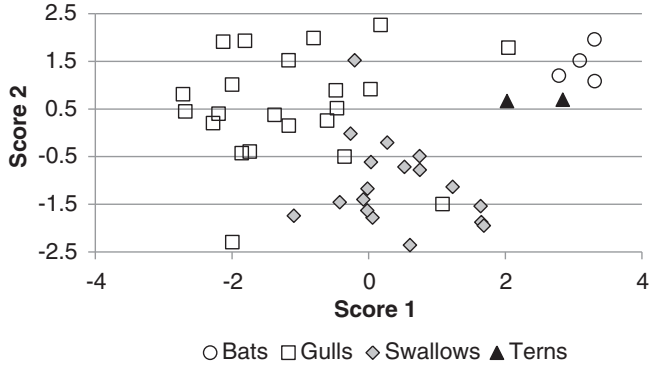

Fig. 7. Plot of the two discriminant scores for all classified tracks ($N = 46$) using the model coefficients and parameters presented in Table 5.

different ($P = 0.01$). The proportion of the variance explained by the first root ranged in the samples from 50% to 74% and had a mean of 60% compared to 64% for the modeled data. The proportion of the variance explained by the second root ranged in the samples from 17% to 40% and had a mean of 31% compared to 29% for the modeled data. The coefficients for the variables in the model were affected by the sample.

The jackknife cross-validation of target classification entailed recalculating the discriminant function repeatedly by omitting one observation at a time and then classifying the omitted observation. In general, groups were classified less successfully in the cross-validation (the minimum correct classifications for gulls and swallows were 58% and 75%, respectively) than when all data were used (Table 6). Bats and terns on average had 94% and 91% correct classifications using cross-validation, but there were only four bat and two tern tracks in the model data. Thus, the variance associated with the tracks from these targets is poorly estimated. Gulls and swallows, on the other hand, may be fairly well represented with a sample size of at least 18 even though on average the model-derived classifications for these species were 73% and 85% correct, respectively. Independently collected validation data would provide a better estimate of the accuracy of the models.

Ten tracks were consistently misclassified during the cross-validation. Four of them were classified with likelihoods less than 60% of belonging to the wrong group. These tracks would have been considered too far away from any class centroid to have been classified. Three of the tracks were classified with likelihoods greater than 90% of belonging to the wrong group. These tracks had characteristics that were indistinguishable from the misclassified group, suggesting that more track and blob characteristics would be beneficial for target discrimination. The remaining three tracks had likelihoods between 70% and 76% of belonging to the wrong group. These tracks had small differences in the Euclidean distance to the correct and the misclassified group centroids.

Table 6
Cross-validation of target classification using the model data and five bootstrap samples and the model defined by Table 5.

Data	Analysis	Correct (%)	Bat	Gull	Swallow	Tern
Modeled	*N*	46	4	22	18	2
	Full	87%	100%	82%	89%	100%
	Cross-validated	78%	100%	68%	83%	100%
Sample 1	*N*	48	3	26	17	2
	Full	83%	100%	73%	94%	100%
	Cross-validated	75%	100%	69%	77%	100%
Sample 2	*N*	47	3	21	18	5
	Full	100%	100%	100%	100%	100%
	Cross-validated	92%	67%	91%	100%	80%
Sample 3	*N*	54	5	25	20	4
	Full	85%	100%	80%	85%	100%
	Cross-validated	78%	100%	72%	75%	100%
Sample 4	*N*	51	5	22	21	3
	Full	90%	100%	82%	95%	100%
	Cross-validated	86%	100%	77%	95%	67%
Sample 5	*N*	46	5	12	23	6
	Full	85%	100%	75%	83%	100%
	Cross-validated	80%	100%	58%	83%	100%
Average	Full	88%	100%	82%	91%	100%
	Cross-validated	82%	94%	73%	85%	91%

4. Discussion

Automated identification of birds and bats in recorded video has the potential to provide useful information for assessing the risks posed by wind energy development. Classification of flying fauna to a species or small group of similar species from sensor-derived data likely requires the examination of many characteristics exhibited during the observation. Body shape, flock formation and behavior, wing beat frequency and rhythm, speed, and flight path often provide clues to the identity of a bird or bat and to its behavior at the time of detection (Bruderer et al., 2010; Desholm et al., 2006; Liechti and Bruderer, 2002; Pennycuick, 1996, 2001; Zaugg et al., 2008). Wing beat frequency was used by Li and Song (2014) to identify 32 bird species in video clips. These authors found that the ability to classify birds was dependent on the uniqueness of their wing beat frequencies compared to other candidate species in the set. Their lack of precision was more a result of the magnitude of overlapping wingbeat frequencies among species than their ability to estimate the frequency from the video. The addition of other attributes such as flight path shape may increase success of species identification. For instance, Gillson (2001) described behavioral differences between the closely related sooty shearwater (*Puffinus griseus*) and short-tailed shearwater (*Puffinus tenuirostris*). Sooty shearwaters flight is relatively smooth, repetitious, and undulating while short-tailed shearwater flight is laborious, erratic and includes more directional changes. Descriptive information describing flight behavior such as this is often used by birders to differentiate among species and could be used for automated classification if it was quantified.

For site-scale risk assessment, algorithms to automate animal detection and classification have not been widely published and are considered intellectual property. Groom et al. (2013) detected 93% of the birds (that were also detected manually by observers) by using an automated extraction algorithm of selected data features (potential birds) from aerial digital images of two existing wind farms off the coast of England. However, much of the data was removed from analysis because of glare and wave action. It is not clear from Groom et al. (2013) what the percentage of correct detection would be in the area removed from analysis but the authors state that the rate of false positives increased in areas affected by glare. A baseline monitoring study of raptor migration routes in relation to a planned offshore wind farm (Skov et al., 2012) used radar and algorithms to automate track detections in real time with one observer entering data into a database and another observer attempting to find and identify the object in the field with binoculars or a telescope. Only those tracks that the second observer was able to identify were used in their analysis. There were no estimates of the false positive or negative rates of the track-detection algorithm. Current programs funded by the Federal Aviation Administration and the National Wildlife Research Center are assessing the accuracy and detection error rates of radar with changing atmospheric conditions, bird range, altitude, and size, for which there is little published information (FAA, 2010; Herricks et al., 2012; King, 2013; USDA, 2014).

The path a flying animal travels as it passes a thermal sensor was used in our study to help classify the target but could also provide information about its activity. For instance, foraging for insects during flight requires a wing that provides maximum maneuverability. Most North American bats are insect feeders that hunt flying insects during flight. The lack of feathers and proportionately less flight muscle per unit weight than comparably sized birds results in bats generating less power per wing flap than birds (Norberg and Norberg, 2012). Smaller

bats also have proportionally lower wing loading because of their light weight and large wings, which enables the slow flight and high maneuverability necessary for effective foraging via short-range echolocation of prey. The path of a bat foraging for airborne insects would be expected to be highly tortuous and exhibit a complex zig-zag pattern. Conversely, the path of birds soaring on thermal air currents would enable one to exclude any number of species that would not be expected to exhibit such behavior.

Animals that pass very near a sensor would likely appear as a linear track regardless of the species or the behavior the animal may be engaged in when observed. All of our targets exhibited a linear flight path at least 10% of the time. The amount of time an animal exhibits straight flight daily or even seasonally is unknown because activity budgets containing detailed flight information simply are not available and would be prohibitively expensive to derive. In our video, gulls and terns were observed more often flying in a linear path than bats and swallows.

A single camera does not provide range information, and without an estimate of distance, a linear track has little meaning. Additional sensors such as radar have been coupled with camera technologies to provide target range (Gauthreaux and Livingston, 2006), and have also enriched the meaning of track shape. Much of the modeling presented in Section 3 would be altered if we had a measure of distance since this would allow an estimate of body size. A model including body size would likely not have misclassified gulls as swallows and visa-versa (Fig. 7).

Opportunities to further develop the application of thermal video include the extraction of additional information to increase classification accuracy, such as wing beat frequency. Track shape can be assessed as an indicator of a blade collision or avoidance behavior. Establishing a list of candidate species expected to occur at a particular location and time (of day or year) could allow for prior probabilities of identity to be incorporated into the classification process. Clearly more information would increase the success rate of classification to a group or species level or at a minimum determine that the detected target is not a species of concern.

5. Conclusion

Using an annotated library of flight paths by bats, swallows, gulls, and terns, we have shown that it is possible (with 82% accuracy overall) to classify the targets based on tracks detected with video from a single thermal camera. The models we have developed for track shape and species classification could be added to the track-detection algorithm for automation purposes. However, these models should be considered preliminary even though they do provide evidence of species classification from thermal video. Future research to increase the robustness of these models includes the extraction of additional measurements from the imagery and the expansion and diversification of the annotated library. A larger annotated library would allow for the splitting of data for model development and independent model validation instead of the jackknife cross-validation estimates of the classification errors.

Variables including ambient weather conditions and time of day relative to sunrise and sunset should be recorded with sensor data to complement information contained within the flight track shape. The direction and strength of ambient winds present when animal flight is observed would also affect flight speed and behavior and likely affect track shape; thus wind characteristics should be evaluated for their effect on correct classification.

Limitations imposed by the use of a single camera must be considered by others who may want to attempt a similar classification of birds and bats using thermal video. Without a measure of range, estimates of wing beat frequency and prior probabilities of target occurrence will likely have a greater influence on determining target identity than flight track shape. However, once the target is identified track shape can be used to estimate behavior and collision risk.

Acknowledgments

This research was funded by the Wind and Water Power Technologies Office within the U.S. Department of Energy-Office of Energy Efficiency and Renewable Energy. This funding source is not responsible for the content or design of this study.

References

Adams, J., Felis, J.J., Mason, J.W., Takekawa, J.Y., 2014. Pacific Continental Shelf Environmental Assessment (PaCSCEA): Aerial Seabird and Marine Mammal Surveys Off Northern California, Oregon, and Washington, 2011–2012. Bureau of Ocean Energy Management, Pacific OCS Region, Camarillo, CA (BOEM 2014-003. 266 pp. Available at: http://www.boem.gov/2014-003/).

Bailey, H., Brookes, K.L., Thompson, P.M., 2014. Assessing environmental impacts of offshore wind farms: lessons learned and recommendations for the future. Aquat. Biosyst. 10, 8.

Band, W., 2012. Using a collision risk model to assess bird collisions for offshore windfarms. Report Commissioned by Strategic Ornithological Support Services (SOSS). http://www.bto.org/sites/default/files/u28/downloads/Projects/Final_Report_SOSS02_Band1ModelGuidance.pdf. Last accessed March 31, 2015.

Banks, N.A., Burton, N.H.K., Austin, G.E., Carter, N., Chamberlain, D.E., Holt, C., Rehfisch, M.M., 2005. The potential effects on birds of the Greater Gabbard offshore wind farm report for February 2004 to March 2005. BTO Research Report No. 419. British Ornithological Trust (Available at: http://www.bto.org/sites/default/files/u196/downloads/rr419.pdf).

Bruderer, B., Peter, D., Boldt, A., Liechti, F., 2010. Wing-beat characteristics of birds recorded with tracking radar and cine camera. IBIS 152, 272–291.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E., Woodward, R., 2012. Aerial surveys of seabirds: the advent of digital methods. J. Appl. Ecol. 49, 960–967.

Bureau of Ocean Energy Management (BOEM), 2014. BOEM Environmental Studies Program Ongoing Study: Developing and Applying a Vulnerability Index for Scaling the Possible Adverse Effects of Offshore Renewable Energy Projects on Seabirds on the Pacific OCS (PC-12-01). (Available at: http://www.boem.gov/pc-12-01/).

Camphuysen, C.J., Fox, A.D., Leopold, M.F., Petersen, K., 2004. Towards Standardized Seabirds at Sea Census Techniques in Connection With Environmental Impact Assessments for Offshore Wind Farms in the U.K. Report to COWRIE in Conjunction With NIOZ. (Available at: http://tethys.pnnl.gov/sites/default/files/publications/COWRIE_Seabird_Census.pdf).

Cilulko, J., Janiszewski, P., Bogdaszewski, M., Szczygielska, E., 2013. Infrared thermal imaging in studies of wild animals. Eur. J. Wildl. Res. 59 (1), 17–23.

Desholm, M., Fox, A.D., Beasley, P.D.L., Kahlert, J., 2006. Remote techniques for counting and estimating the number of bird-wind turbine collisions at sea: a review. Ibis 148, 76–89.

European Wind Energy Association (EWEA), 2014. The European Offshore Wind Industry—Key Trends and Statistics 2013. (Available at: http://www.ewea.org/fileadmin/files/library/publications/statistics/European_offshore_statistics_2013.pdf).

Federal Aviation Administration (FAA), 2010. Airport avian radar systems. AC 150/5220-25. Released 23 November 2010. http://www.faa.gov/airports/resources/advisory_circulars/index.cfm/go/document.current/documentNumber/150:5220-5225.

Fish, U.S., Wildlife Service (USFWS), 2012. U.S. Fish and Wildlife Service Land-based Wind Energy Guidelines. OMB Control Number 1018-0148 (Available at: http://www.fws.gov/windenergy/docs/WEG_final.pdf).

Gauthreaux Jr., S.A., Belser, C.G., 2003. Radar ornithology and biological conservation. Auk 120 (2), 266–277.

Gauthreaux Jr., S.A., Livingston, J.W., 2006. Monitoring bird migration with a fixed-beam radar and a thermal-imaging camera. J. Field Ornithol. 77, 319–328.

Gillson, G., 2001. Separation of sooty and short-tailed shearwaters. Available at:. http://thebirdguide.com/pelagics/book/id_sosh_stsh.htm.

Groom, G., Stjernholm, M., Nielsen, R.D., Fleetwood, A., Petersen, I.K., 2013. Remote sensing image data and automated analysis to describe marine bird distributions and abundances. Ecol. Inform. 14, 2–8.

Guillemette, M., Larsen, J.K., 2002. Post development experiments to detect anthropogenic disturbances: the case of sea ducks and wind parks. Ecol. Appl. 12 (3), 868–877.

Herricks, E.E., Bunch, W., Osmek, S., Svoboda, F., 2012. Using an avian radar to supplement an airport wildlife hazard assessment. U.S. Department of Transportation, Federal Aviation Administration. Technical Report No. DOT/FAA/TC-TN12/27 (Available at: http://www.airporttech.tc.faa.gov/safety/downloads/).

Jarrah, A., Jamali, M., Ross, J., Gorsevski, P., Frizado, J., Bingman, V.P., 2012. Sensitivity analysis for optimal parameters for marine radar data processing. Am. J. Signal Process. 3 (3), 78–83.

King, R.E., 2013. Research on bird-detecting radar. U.S. Department of Transportation, Federal Aviation Administration. Interim Technical Report No. DOT/FAA/TC-13/3 (Available at: http://www.airporttech.tc.faa.gov/safety/downloads/).

Li, W., Song, D., 2014. Automatic bird species detection from crowd sourced videos. IEEE Trans. Autom. Sci. Eng. 11 (2), 348–358.

Liechti, F., Bruderer, L., 2002. Wingbeat frequency of barn swallows and house martins: a comparison between free flight and wind tunnel experiments. J. Exp. Biol. 205, 2461–2467.

Maclean, I.M.D., Wright, L.J., Showler, D.A., Rehfisch, M.M., 2009. A review of assessment methodologies for offshore windfarms. British Trust for Ornithology Report Commissioned by COWRIE Ltd. (Available intermittently at: http://www.offshorewind.co.uk)

Matzner, S., Cullinan, V.I., and Duberstein, C.A. A new thermal video processing method to support risk assessment for birds and bats. (in preparation).
Norberg, U.M.L., Norberg, R.A., 2012. Scaling of wingbeat frequency with body mass in bats and limits to maximum bat size. J. Exp. Biol. 215, 711–722.
Normandeau Associates, Inc. (Normandeau), 2012. High-resolution Aerial Imaging Surveys of Marine Birds, Mammals, and Turtles on the US Atlantic Outer Continental Shelf—Utility Assessment, Methodology Recommendations, and Implementation Tools. Prepared for the U.S. Department of the Interior, Bureau of Ocean Energy Management Under Contract # M10PC00099.
Normandeau Associates, Inc. (Normandeau), 2014. Acoustic Monitoring of Temporal and Spatial Abundance of Birds Near Outer Continental Shelf Structures: Synthesis Report. U.S. Dept. of the Interior, Bureau of Ocean Energy Management, Herndon, VA (BOEM 2014-004. 172 pp. Available at: http://www.data.boem.gov/PI/PDFImages/ESPIS/5/5349.pdf).
Pennycuick, C.J., 1996. Wingbeat frequency of birds in steady cruising flight: new data and improved predictions. J. Exp. Biol. 199, 1613–1618.
Pennycuick, C.J., 2001. Speeds and wingbeat frequencies of migrating birds compared with calculated benchmarks. J. Exp. Biol. 204, 3283–3294.
Plonczkier, P., Simms, I.C., 2012. Radar monitoring of migrating pink-footed geese: behavioural responses to offshore wind farm development. J. Appl. Ecol. 49, 1187–1194.
Skov, H., Leonhard, S.B., Heinänen, S., Zydelis, R., Jensen, N.E., Durinck, J., Johansen, T.W., Jensen, B.P., Hansen, B.L., Piper, W., Grøn, P.N., 2012. Horns rev 2 monitoring 2010–2012. Migrating birds. Orbicon, DHI, marine observers and biola. Report commissioned by DONG Energy. Available at:. http://www.ens.dk/sites/ens.dk/files/undergrund-forsyning/vedvarende-energi/Migrating_birds_monitoring_2012_Final_ver_3%20(2).pdf.
Smales, I., Muir, S., Meredith, C., Baird, R., 2013. A description of the Biosis model to assess risk of bird collisions with wind turbines. Wildl. Soc. Bull. 37 (1), 59–65.
Torvik, B., Knapskog, A., Lie-Svendsen, O., Olsen, K.E., Griffiths, H.D., 2014. Amplitude modulation on echoes from large birds. Proceedings of the 11th European Radar Conference (Rome, Italy. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6991236).
U.S. Department of Agriculture (USDA), 2014. "FAA Funds Avian Radar Evaluation" in Jan. 30, 2014 USDA APHIS, Wildlife Damage Spotlight. (Available at: http://www.aphis.usda.gov/).
U.S. Geological Survey (USGS), 2014. Pacific Continental Shelf Environmental Assessment (PaCSEA): Aerial Seabird and Marine Mammal Surveys off Northern California, Oregon, and Washington, 2011–2012. Prepared for the Bureau of Ocean Energy Management under Interagency Agreement M10PG00081 (Available at: http://www.boem.gov/2014-003/).
Zaugg, S., Saporta, G., Van Loon, E., Schmaljohann, H., Liechti, F., 2008. Automatic identification of bird targets with radar via patterns produced by wing flapping. J. R. Soc. Interface 5, 1041–1053.