

Automatic identification of bird targets with radar via patterns produced by wing flapping

Serge Zaugg^{1,*}, Gilbert Saporta², Emiel van Loon³, Heiko Schmaljohann^{1,4}
and Felix Liechti¹

¹*Swiss Ornithological Institute, 6204 Sempach, Switzerland*

²*Conservatoire National des Arts et Métiers, Paris 75141, France*

³*Universiteit van Amsterdam (UvA), 1012 Amsterdam, The Netherlands*

⁴*Institut für Vogelforschung, 'Vogelwarte Helgoland', 26386 Wilhelmshaven, Germany*

Bird identification with radar is important for bird migration research, environmental impact assessments (e.g. wind farms), aircraft security and radar meteorology. In a study on bird migration, radar signals from birds, insects and ground clutter were recorded. Signals from birds show a typical pattern due to wing flapping. The data were labelled by experts into the four classes BIRD, INSECT, CLUTTER and UFO (unidentifiable signals). We present a classification algorithm aimed at automatic recognition of bird targets. Variables related to signal intensity and wing flapping pattern were extracted (via continuous wavelet transform). We used support vector classifiers to build predictive models. We estimated classification performance via cross validation on four datasets. When data from the same dataset were used for training and testing the classifier, the classification performance was extremely to moderately high. When data from one dataset were used for training and the three remaining datasets were used as test sets, the performance was lower but still extremely to moderately high. This shows that the method generalizes well across different locations or times. Our method provides a substantial gain of time when birds must be identified in large collections of radar signals and it represents the first substantial step in developing a real time bird identification radar system. We provide some guidelines and ideas for future research.

Keywords: radar ornithology; bird identification; pattern recognition; continuous wavelet transform; feature extraction; support vector machine

1. INTRODUCTION

Bird movements have been studied by means of radar for almost 60 years (Eastwood 1967). With radar, birds can be detected over a wide range of distances (from a few metres up to 100 km) during day as well as night and under all weather conditions except heavy precipitation. It allows a precise estimation of target distance and depending on the type of radar and the operational mode also of flight altitude. Thus, the information available from radar is in the best case the position in time and space, and the amount of energy reflected.

Radar has been used: (i) in studies of bird migration (see the reviews by Bruderer (1997*a,b*), Gauthreaux & Belser (2003) and Schmaljohann *et al.* (2007*b*)), (ii) in studies of population dynamics of breeding seabirds (Lilliendahl *et al.* 2003; Hamer *et al.* 2005) or roosting land birds (Russell & Gauthreaux 1998; Bäckman & Alerstam 2002), (iii) monitoring of bird traffic at

prospective locations of power lines, wind farms or bridges (Harmata *et al.* 1999; Desholm & Kahlert 2005) and (iv) in the context of bird strike avoidance near airports and en route (L. Buurma 1996, personal communication).

The potential increase in the construction of wind farms in the coming years is a serious concern for avian conservation (Garthe & Hüppop 2004) and radar methodology has been of great importance for environmental impact assessment (Desholm *et al.* 2006).

An important area of the application of radar ornithology is to quantify the numbers of birds aloft. At a first look the principle seems simple—the targets detected in the radar beam are used to infer the spatial or temporal distribution of birds. But insects, bats, other flying objects and weather phenomena can provide a significant proportion of the signals observed (Eastwood 1967). The major problem in terms of abundance is the distinction between birds and insects. Insect movements are very common (e.g. Riley & Reynolds 1986; Smith *et al.* 1993; Feng *et al.* 2007) and can interfere considerably with the studies on bird movements (Schmaljohann *et al.* in press). Airborne

*Author for correspondence (serge.zaugg@vogelwarte.ch).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2007.1349> or via <http://journals.royalsociety.org>.

insect abundance depends mainly on the geographical area, the season and time of day. Hence, the severity of the problem is likely to be very variable depending on the recording location and time. In the past, either this problem was at least partly overlooked (cf. Schmaljohann *et al.* 2007a) or various means were used to try to separate bird signals from non-bird ones.

- (i) The intensity of radar signals (echo size) has been used jointly with the information of flight speed in several studies to identify birds up to the species level (e.g. Russell & Gauthreaux 1998; Lilliendahl *et al.* 2003). Using signal intensity alone would be questionable since large insects or insect swarms can produce signals of a magnitude similar to that of small birds. In the above-mentioned studies, the information on signal intensity was improved by prior knowledge of the abundance of the species under investigation. Unfortunately, such prior knowledge is rarely available. For the majority of migratory birds (small passerines), the intensity of radar signals will not suffice to distinguish between birds and insects, because there is a large overlap in signal intensity between small birds and large insects (Larkin 1991).
- (ii) Generally, birds fly at higher airspeeds (the speed with respect to the air mass) than insects (Bloch & Bruderer 1982; Larkin 1991; Feng *et al.* 2007). Ground speed (the speed with respect to the ground) together with signal intensity is used in several studies to distinguish between birds and non-birds (Lilliendahl *et al.* 2003; Hamer *et al.* 2005). Ground speed depends on airspeed as well as wind speed and wind direction, and therefore wind information must be taken into account to calculate the airspeed of a bird or insect. Even if very accurate wind profiles were available, a clear separation of birds and insects is difficult due to overlap in airspeed between the two groups (Schmaljohann *et al.* in press). The study of Larkin (1991) shows that this overlap is far from negligible but suggests that airspeed can be a valuable tool for partially discriminating between birds and insects. The severity of the overlap depends on the insect species aloft and therefore information on local insect fauna is needed. The problem is expected to be more severe in regions or during seasons with a high abundance of large flying insects.
- (iii) Recent studies performed with polarimetric weather radar, which transmits and receives horizontally and vertically polarized pulses, indicate that it can be used to discriminate between radar returns produced by birds and insects (Zrnic & Ryzhkov 1998; Bachmann & Zrnic 2007).
- (iv) The temporal pattern of the signal intensity (echo signature) depends mainly on the variation in the shape of the target (e.g. by flapping). In various former studies, we tracked birds and

insects by radar and identified the target through a telescope mounted parallel to the radar beam. This provided radar signals of visually identified targets, which enabled us to learn how the target identity affects the recorded signals. A typical pattern in signals produced by birds is a periodic fluctuation of signal intensity, which is caused by change in body shape due to flapping. For bird targets the frequency of the fluctuation either directly reflects wing flapping frequency or several harmonics (integer multiple of the frequency) are present in the signal. In the latter case wing flapping frequency is easily deduced since it is the fundamental frequency of the harmonics. In our data, we found that insects did occasionally produce periodic signals, but these were close in shape to a pure sine wave (unlike in birds) and their fundamental frequency was generally higher than that in birds. Several studies have pointed out that this pattern can be used to distinguish between the signals of birds and insects, mainly due to the distinctive wing flapping pattern in the bird signals. (Bonham & Blake 1956; Houghton 1964; Bruderer & Joss 1969; Bruderer & Steidinger 1972; figure 1). Bats tracked by radar produced radar signals similar to those of birds (Bruderer & Popa-Lisseanu 2005).

An accurate method to identify bird signals is ‘plot-based labelling’. That is, to look at the plots of the signals. Additionally, plots of the spectral density of the signals (e.g. Fourier transforms or wavelet transforms) can be used. Plot-based labelling has been successfully applied in numerous studies (Bloch *et al.* 1981; Stark & Liechti 1993; Liechti & Bruderer 1995; Bruderer & Liechti 1998; Liechti & Schaller 1999; Nievergelt *et al.* 1999; Hedenström & Liechti 2001; Bäckman & Alerstam 2003; Schmaljohann *et al.* 2007b). However, the method has some limitation—it is very time consuming, it requires a highly experienced person and, although there is a close match between different experts, it is not free of subjective interpretation.

Here, we present a classification algorithm that exploits the temporal pattern in the signals together with absolute signal intensity. Its aim is to automatically identify signals produced by birds. The algorithm is based on a signal processing method (continuous wavelet transform, CWT) and a statistical learning method (support vector classifier, SVC). It is specifically designed to be applicable to large datasets.

2. METHODS

2.1. Description of the radar system and data

For this study, we used four datasets consisting of digitized radar signals. The datasets were collected during four radar surveys in the Sahara desert (Mauritania) at two locations in the years 2003 and 2004 (see table 1). The datasets are named A, B, C and D. The radar was operated during all hours of day and night. The height over ground of the detected targets ranges from 20 to 7390 m. This radar system is a

Table 1. Details of the four datasets used in this study. (The absolute number of signals is given for each class and the percentage (equal to class prior) is given in the brackets.)

dataset	recording location	season	BIRD	INSECT	UFO	CLUTTER	total
A	Ouadane	spring 03	3098 (52%)	686 (12%)	442 (8%)	1652 (28%)	5878 (100%)
B	Ouadane	spring 04	4273 (12%)	26 244 (70%)	5465 (15%)	1099 (3%)	37 081 (100%)
C	Ouadane	autumn 03	4946 (14%)	25 964 (73%)	3825 (11%)	612 (2%)	35 347 (100%)
D	Idini	autumn 03	784 (4%)	10 667 (49%)	910 (4%)	9402 (43%)	21 763 (100%)

former military fire-control tracking radar with vertically polarized pulses of 3.3 cm wavelength. The radar was adapted specifically to detect birds (Bruderer *et al.* 1995). Fixed beam measurements (Bruderer 1971) were performed at the oasis of Ouadane (20°56' N, 11°35' W) and Idini (17°57' N, 15°28' W). In fixed beam measurements, the radar is kept stationary pointing to a given direction. All targets moving across the stationary radar beam are recorded, providing information on target distance and variation over time in target reflectivity. To cover all relevant heights, the elevation angle of the beam ranged from 6° to 84° (see electronic supplementary material). The pulse repetition frequency and the sampling frequency of the radar was 2083.33 Hz. In order to improve the signal to noise ratio, the average was taken over adjacent but non-overlapping blocks of 16 consecutive samples. Therefore the final sampling frequency of the signals is 130.21 Hz; 130 Hz will be used for simplicity hereafter. All the signals were digitized and stored automatically. Obvious clutter signals produced by diffuse targets (clouds, rain) were removed manually. Since the sensitivity of a radar increases dramatically at close range, a *sensitivity time control* (Skolnik 1980) was applied to eliminate all targets that would not be detected at a distance of 3 km. Thus, small insects and weak ground clutter were suppressed. For more details on data collection see Schmaljohann *et al.* (2007b).

Each signal is a sequence of numerical values. A signal represents the fluctuation in signal intensity (equal to reflectivity) over time. The time unit is 1/130 s. The primary unit of the radar signal P was measured in milliwatts (mW). Since this quantity is usually very small, we transformed it into a more convenient unit, the dBm. The signal intensity S in dBm is defined as $S = 10 \log(P/1)$, where 1 stands for 1 mW. See the plots of figures 1–3 for examples of radar signals. The duration of the signals ranged from 1.5 to 38 s. This radar recorded the distance of the target with a resolution of 30 m. During seven months of fieldwork, over 5000 single targets were tracked simultaneously with our radar and optically through a telescope. This provided data of visually verified targets. Heiko Schmaljohann (H.S.) did most of this work and used this experience to learn to identify targets based on the plots of the signals and the plots of the spectral density of the signals (fast Fourier transform). This will be referred to as plot-based labelling hereafter. H.S. labelled all the signals used in this study via plot-based labelling. He simultaneously used several criteria. (i) Presence of a wing flapping pattern, which

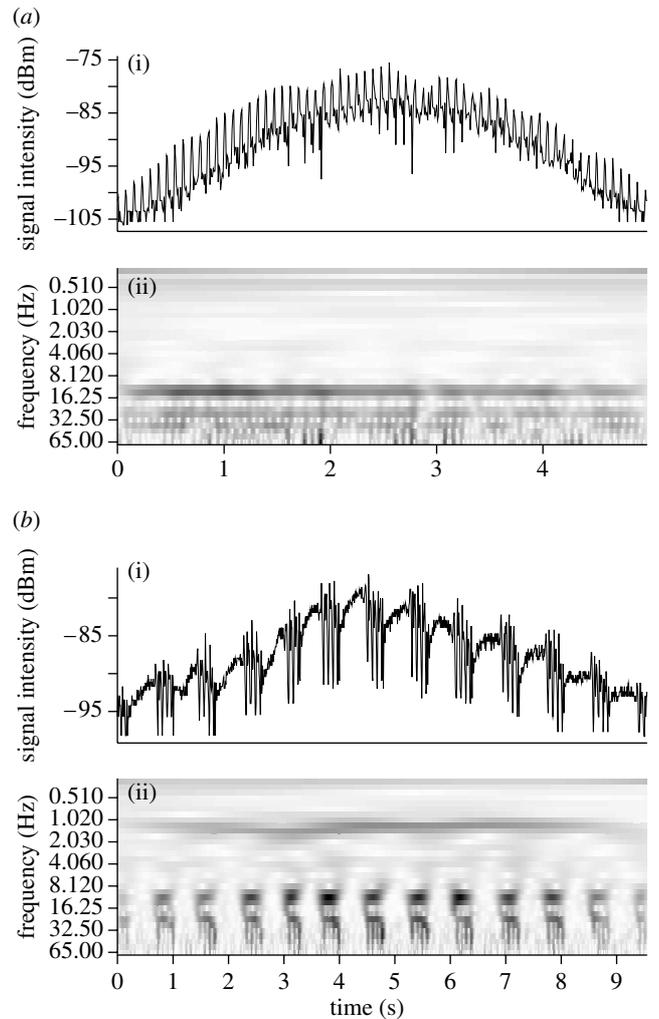


Figure 1. (a, b) Two examples of (i) plots of signals produced by birds flying across the stationary radar beam and (ii) of their CWT (time–frequency representation). Larger values of the CWT are indicated by darker shading. (a) Signal from a continuously flapping bird (wader type) and (b) signal from a bird flapping intermittently (passerine type).

is indicative of a bird (figure 1a, b). The detailed shape of the pattern was also considered. A pattern close in shape to a pure sine function is indicative of an insect. (ii) Presence of repeated clusters of larger fluctuations, which is indicative of passerine-like (figure 1b) or swift-like flapping style. (iii) Value of the fundamental frequency, which reflects the wing flapping frequency. It must be between 2.5 and 30 Hz for birds. (iv) Overall curvature of the signal. A curved signal indicates a target flying through the beam (figures 1 and 2). A straight signal indicates stationary clutter (figure 3).

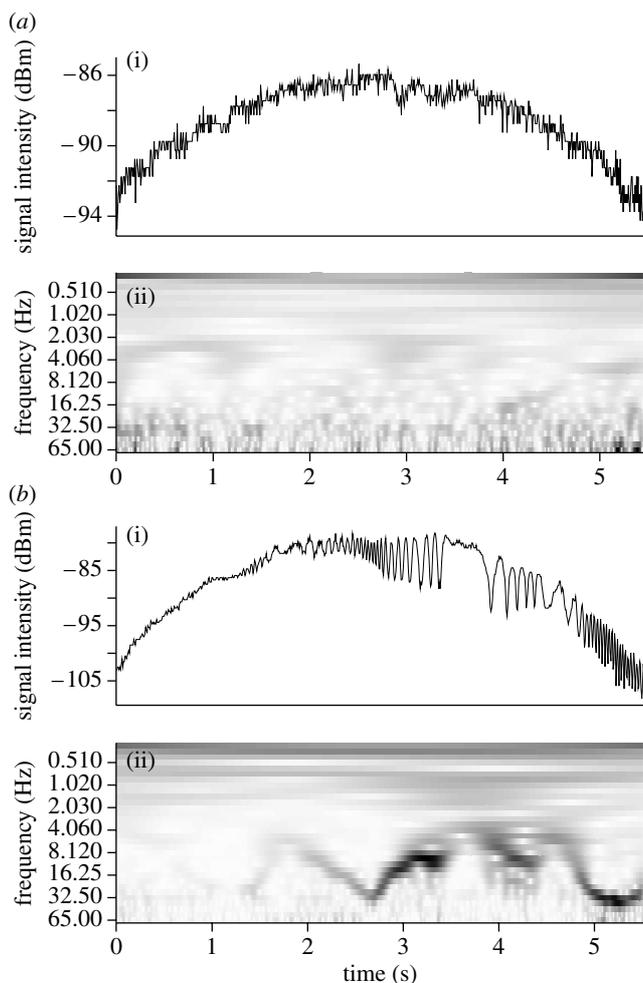


Figure 2. (a,b) Two examples of (i) plots of signals produced by insects flying across the stationary radar beam and (ii) of their CWT (time–frequency representation). Larger values of the CWT are indicated by darker shading. (a) Signal with no dominant frequency and (b) signal with dominant frequency that changes over time.

The four datasets of the present study have been sorted by the plot-based method by H.S. into four classes. Class labels are written in uppercase throughout this article.

BIRD, signals that can unambiguously be assigned to flying birds; INSECT, signals that can unambiguously be assigned to flying insects; CLUTTER, signals that can unambiguously be assigned to objects on the ground (e.g. trees, hills, houses); UFO, signals that are between the typical bird signals and insect signals. They are clearly produced by flying targets but cannot be unambiguously assigned to birds or insects.

It must be assumed that an unknown proportion of the UFOs were actually produced by birds. It is important to note that the uncertainty in plot-based labelling has been explicitly coded by creating the class UFO. Sand, dust or swarms of insects within the same pulse volume as a bird produce radar returns that partially mask wing flapping pattern. The bird signals from dataset D are the most contaminated with such returns due to frequent sand storms and high abundance of insects. While recording dataset D the

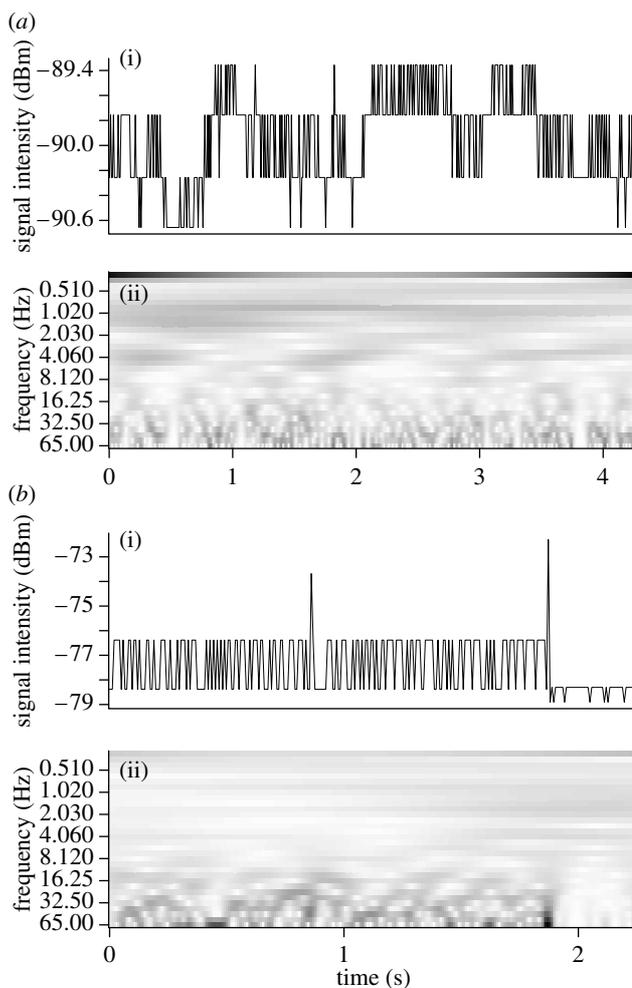


Figure 3. (a,b) Two examples of (i) plots of signals produced by ground clutter and (ii) of their CWT (time–frequency representation). Larger values of the CWT are indicated by darker shading.

airborne insects and sand particles were sometimes so abundant that the signals looked similar to those of clutter and such signals have been labelled as CLUTTER. The signals from datasets C and B are moderately contaminated, mainly due to the abundance of insects. The signals from dataset A are weakly contaminated thanks to the low abundance of insects. The proportion of BIRDS is small in the datasets B, C and D with values between 4 and 14% (table 1).

2.2. Merging classes to obtain a two-class problem

To avoid the additional complication of multiclass models, we recoded the classes. The class BIRD was kept and it is contrasted to the new class OTHER that comprises the INSECT, UFO and CLUTTER classes. Including the UFOs implies that the class OTHER contains an unknown amount of signals that were produced by birds. In the models presented later, the response variable is categorical with two levels (BIRD and OTHER). Therefore the classifier that we construct is aimed at separating signals, which are clearly produced by birds from all other signals.

2.3. Overview of the classification method

The classification method presented in this article is divided into two consecutive steps. (i) A data pre-processing step in which variables are extracted from the signals (§2.4). (ii) A modelling step in which SVC methods are used to train a predictive model that links the variables to the class labels (§2.5).

The ultimate purpose is that the trained model should be used to predict the class of new data (i.e. new signals that have gone through the data pre-processing step) with unknown class label. Therefore an important issue is to quantify the classification performance that can be expected from this classification method. This was done by simulating the situation where we must assign class labels to unlabelled data (§2.8): we first train a SVC with a subset of the data, the training set. Then the trained SVC is used to predict the class labels of another subset of the data, the test set. Hence the test set is treated in the same way as we would treat unlabelled data, but for the test set we know the true class labels. This enables the comparison of the predicted class labels with the true class labels. From this comparison we compute four measures of classification performance (§2.7).

We used the environment for statistical computing R (R Development Core Team 2006) with the packages kernlab (Karatzoglou et al. 2004) for the SVCs and Rwave (Carmona & Whitcher 2004) for wavelet transforms.

2.4. Pre-processing the signals

2.4.1. Distance correction. The intensity of the signals is heavily influenced by the target's distance from the radar. Distance-corrected signals were computed according to the radar equation (R^{-4} law, Skolnik 1980). The distance-corrected signals correspond to what would have been recorded if all the targets were at a distance of 3000 m from the radar. Equations are given in the electronic supplementary material. The distance-corrected signals were used for all further steps.

2.4.2. Variables extraction. Signals are not a suitable input format for our modelling method, the SVC. Therefore variables must be extracted from the signals. The variable extraction results in 67 variables named *maxi*, *sd-stat*, *resolution*, *coefmean-1*, *coefmean-2*, ..., *coefmean-32* and *coefsd-1*, *coefsd-2*, ..., *coefsd-32*. Details are presented below. Variable names are in italic characters throughout this article.

- *maxi*. While a target flies across the fixed beam, the signal intensity will first increase and then decrease. As a consequence the signal's overall shape is curved (figures 1 and 2). The maximum intensity of the distance-corrected signals was used as a variable, which is called *maxi*. This variable depends on the target size, shape, texture, orientation and on how close it flew to the centre of the beam.
- *sd-stat*. We wanted to quantify the variation in the signals that is due to wing flapping. To achieve this, we stationarized the distance-corrected signals

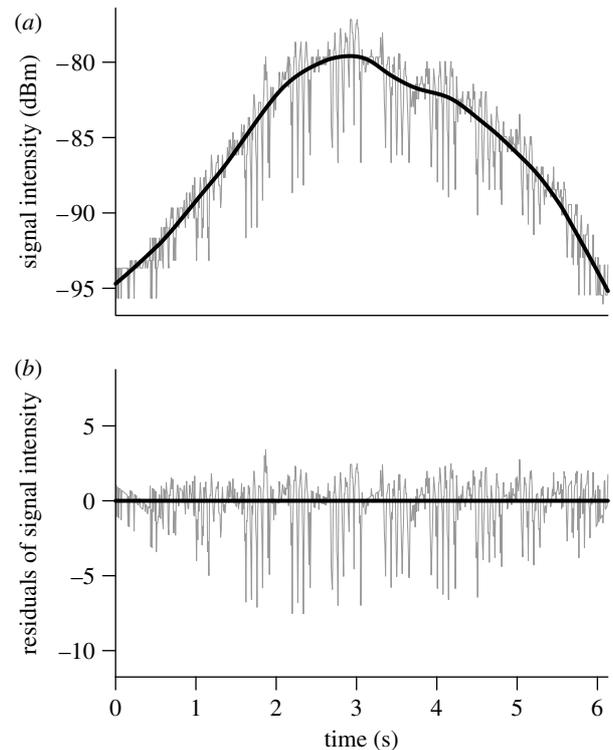


Figure 4. (a) A signal from a bird (grey lines) together with the local regression line (thick black line). (b) The residuals obtained by subtracting the value of the local regression line from the signal. Such residuals were used to extract the variable *sd-stat* from the signals.

(equal to removing the overall curvature) and then computed the standard deviation of the intensity of the stationarized signals. This gives the variable *sd-stat*. Stationarizing was done by taking the residuals from a local regression with a moving window (LOWESS (Cleveland 1981)) with the window size set to 150 samples. As an example, figure 4a shows the distance-corrected signal from a bird and the local regression line and figure 4b shows the stationarized signal. The stationarized signals are not used any further.

- *resolution*. The signal intensity can take only a discrete number of values (a consequence of digitizing). This is obvious in figure 3a where the signal intensity takes only four different values. The number of values represented in a given signal was used as a variable that we call *resolution*. This measures the overall variability in signal intensity.
- *coefmean* and *coefsd*. This group of 2×32 variables was extracted via continuous wavelet transform (CWT), which is a tool for analysing the spectral content of signals. The input of the CWT is a radar signal and the output is a time–frequency representation of the signal (see figures 1–3). A description of this procedure is given next.

Several variants of CWT exist. We computed the CWT with the *Morlet wavelet*. The wavelet methods are described in Carmona et al. (1998), Torrence & Compo (1998) and Addison (2002).

The time resolution of the signals is 130 Hz and due to the Nyquist limit they cannot represent frequencies above 65 Hz. Frequencies below 0.31 Hz were not considered because some of the signals were not long enough to detect lower frequencies. Therefore we considered 32 frequencies between 0.31 and 65 Hz. The frequencies were selected on an exponential scale, which is usual with the CWT. For a signal composed of n samples the output of the CWT is a matrix with 32 rows and n columns. The numerical values in this matrix are called wavelet coefficients and they represent the signal amplitude at a given frequency and position in time.

Before computing the CWT, the signals were prepared in two steps. (i) The distance-corrected signals were demeaned and then normalized with *sd-stat* (i.e. divided by *sd-stat*). The wavelet coefficients are proportional to the amplitude of the signal. As explained above, *sd-stat* quantifies the amplitude due to wing flapping and therefore normalization with *sd-stat* ensures that the wavelet coefficients related to wing flapping (i.e. the upper frequencies of our range) have comparable magnitudes across the signals. There is no need to remove the overall curvature before computing the CWT. The curvature produces non-zero coefficients for low frequencies (typically below 0.31 Hz), but it does not affect the coefficients for higher frequencies, which are related to wing flapping. (ii) In order to reduce edge effects, the signals were padded on both ends with 1000 samples equal to the first or last value represented in the signal. The wavelet coefficients corresponding to these padded values were removed afterwards.

Next, 2×32 variables were extracted from the matrices obtained by CWT as follows. For every row of these matrices, the mean and standard deviation of the wavelet coefficients were computed. In other words, for each of the 32 frequencies, we computed the mean and standard deviation over time of the wavelet coefficients. These variables are called *coefmean 1–32* and *coefsd 1–32*. Smaller numbers are used to name higher frequencies. An overview of the variable names and associated frequencies is given in the electronic supplementary material. These 64 variables contain information on the periodic patterns in the signals.

2.4.3. Transformation and standardization of the variables. The variables had skewed distributions and several standard transformations were applied to obtain more symmetric distributions (see electronic supplementary material). All variables were centred and scaled to mean = 0 and standard deviation = 1.

2.5. Support vector classifiers

As a statistical modelling method, we used the SVC (Cortes & Vapnik 1995; Vapnik 2000; Hastie *et al.* 2001). The SVC accepts several predictor variables and one binary categorical response variable. In our case, the response has the two classes BIRD or OTHER. During a training phase, a function that links the response to the variables is estimated via a dataset with known class labels. The resulting function is called a

trained SVC and can be used to predict the class of new data instances with unknown class labels. A trained SVC does not directly predict the class. Its basic prediction is a numerical value, which we call the ‘score’. The score provides a ranking of the data instances. Data instances with a larger score (more positive) are more likely to be BIRDS and data instances with a smaller score (more negative) are more likely to be OTHERs (see figures 5 and 6). In order to predict the class (class-level prediction), a decision threshold must be chosen on the score. For example, setting the threshold to zero means that data instances that have a predicted score greater than zero are assigned to the BIRD class and data instances that have a predicted score smaller or equal to zero are assigned to the OTHER class. In SVCs a decision threshold equal to zero leads to the lowest overall error rate (i.e. the proportion of data instances whose class is wrongly predicted). A threshold equal to zero was used for all the analyses presented in this article.

There exist several subtypes of SVCs, which are specified by a kernel function (short: kernels). In preliminary work we assessed the performance of SVCs with four different kernels: ‘linear’, ‘polynomial’, ‘radial basis function’ and ‘Laplace’. The Laplace kernel gave the best performance and was used for this study.

The SVC with Laplace kernel has two tuning parameters named C and σ which control the capacity of the classifier (i.e. its ‘flexibility’). They are called tuning parameters because values yielding good classification performance must be chosen by the operator (see below).

2.6. Model selection

A model is defined as a subset of the 67 variables and by specified values for the tuning parameters C and σ . A model selection was performed with dataset B to choose a subset of variables and values of the tuning parameters that yield the best classification performance. Details are given in the electronic supplementary material. Specifications of the selected model are given in §3.

2.7. Four ways to measure classification performance

An unbiased estimation of classification performance is obtained by predicting the class of data instances that were not used for training and compare these predictions with the actual class labels. For this comparison we used four performance measures: the area under the ROC curve (AUC), the accuracy (ACC), positive predictive value (PPV) and false negative rate (FNR).

The AUC (Fawcett 2006) quantifies how well two classes are separated along the score axis. The AUC is estimated with the predicted scores and the true class labels of test data instances. AUC takes values between 0 and 1. A classifier that classifies at random has an AUC of 0.5, and a classifier that achieves perfect separation has an AUC of 1. An AUC below 0.5 indicates that the classifier does worse than random

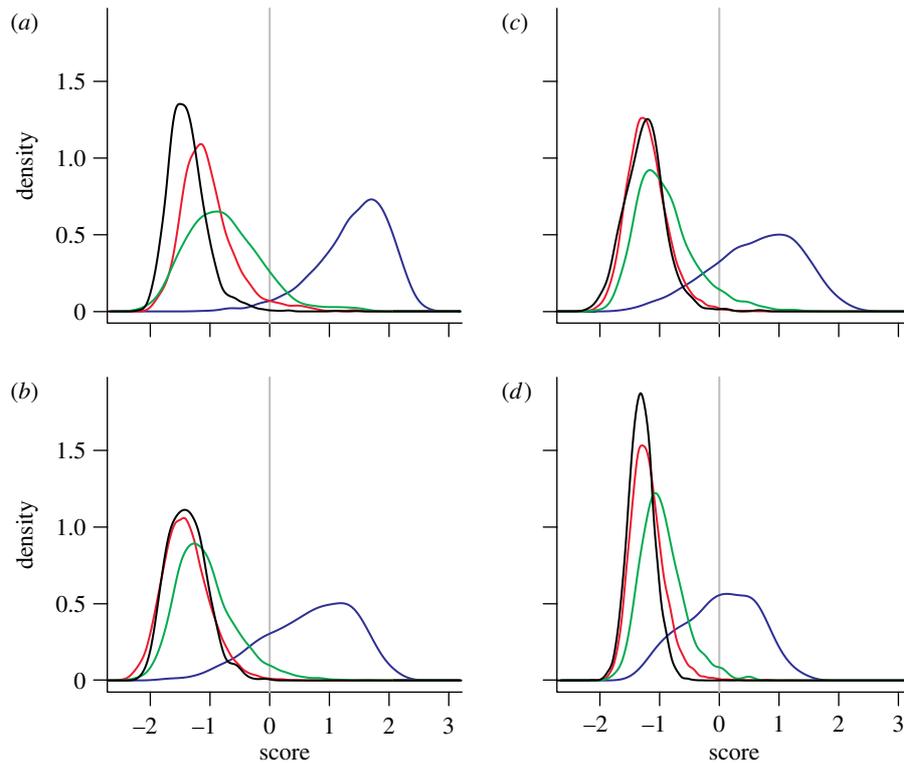


Figure 5. Within dataset method. (a)–(d) Datasets A–D: plots of the estimated density over the score for the four classes (BIRD, blue; UFO, green; INSECT, Red; CLUTTER, black). Densities were estimated with kernel density estimation. The class-conditional densities are shown and hence the class priors are not deducible from the plots. The class priors are given in table 1. The grey vertical line represents the decision threshold equal to zero used to obtain the class-level predictions.

guessing. AUC offers two important advantages. (i) It does not depend on the *a priori* frequency of the classes in the dataset (called ‘class priors’ hereafter). Thus it is the only performance measure that allows comparison across different datasets, since class priors differ between the datasets (table 1). (ii) It informs on the separating ability of the score independently of a particular decision threshold on the score. A technical description is given in the electronic supplementary material.

When class-level prediction has been made, a test dataset can be divided into four groups. (a) BIRDS that were predicted to be BIRDS. (b) OTHERS that were predicted to be OTHERS. (c) BIRDS that were predicted to be OTHERS. (d) OTHERS that were predicted to be BIRDS. Let a , b , c and d represent the number of data instances found in the four groups, then

$$\text{Accuracy : } \text{ACC} = (a + b)/(a + b + c + d),$$

$$\text{Positive predictive value : } \text{PPV} = a/(a + d),$$

$$\text{False negative rate : } \text{FNR} = c/(a + c).$$

A more detailed description is found in (Fawcett 2006).

The accuracy (ACC) is the overall proportion of correctly classified data instances. It is the ‘usual’ performance measure. However in the case of unbalanced class priors the accuracy must be interpreted with caution. High values of the accuracy are often a direct consequence of the fact that one class is over-represented and do not reflect the separating ability of the classifier. As an example consider dataset D, where only 4% of the signals are BIRDS and 96% are OTHERS (table 1). Imagine a very unspecific classifier that would

assign all the signals to the OTHERS. Such a classifier would still achieve an accuracy of 96%, but it would be useless to identify BIRDS in the data.

The PPV is the proportion of data instances correctly predicted to be BIRDS among all the data instances that have been predicted to be BIRDS. Since our purpose is to isolate a subset which contains mostly BIRDS we wish PPV to be as high as possible.

The FNR is the proportion of BIRDS that are wrongly predicted to be OTHERS among the total number of BIRDS in the test set. That is, it is the proportion of BIRDS that are ‘lost’ when the classifier is used for prediction, thus we wish FNR to be as small as possible.

The PPV and the FNR are useful when the interest is on one of the two classes as is the case here. ACC, PPV and FNR are specific for a given decision threshold, unlike the AUC that is independent of the decision threshold. It is important to realize that PPV and FNR are interrelated via the decision threshold. This means that PPV can be improved by moving the decision threshold to more positive values, but this will inevitably give a worse value for the FNR.

2.8. Assessing the classification performance of the selected model on different datasets

2.8.1. *The principle of cross validation.* The aim of cross validation is to estimate the expected classification performance in the prospective situation where the classifier will be used for prediction on new data. The principle is to predict the score or class of a subset of data which was not used for training (a test set). Then,

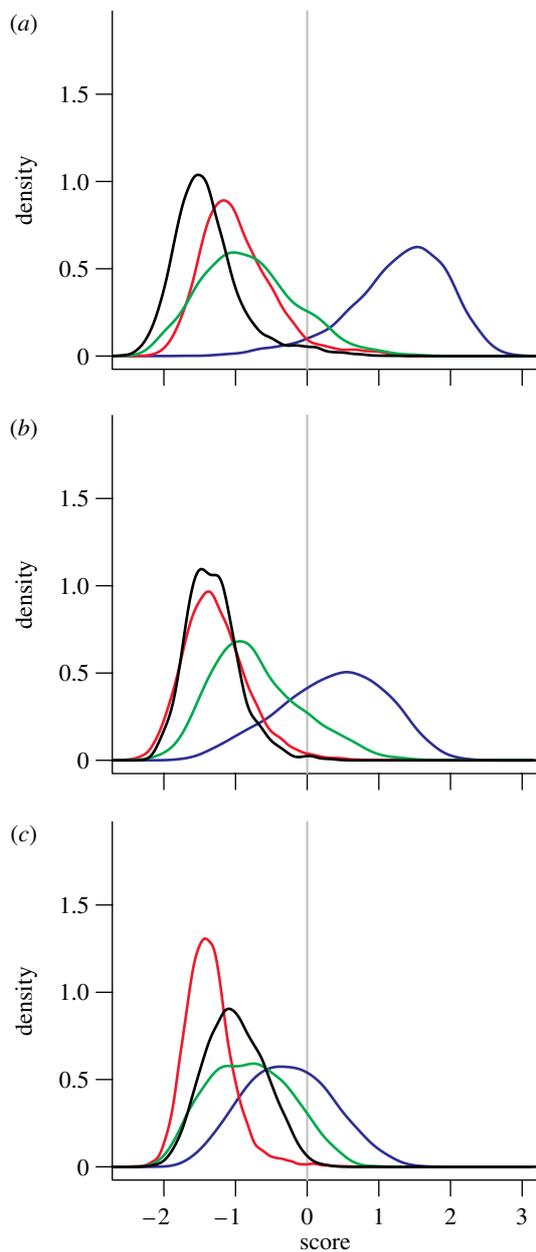


Figure 6. Cross dataset method. (a)–(c) Datasets A, C and D: plots of the estimated density over the score for the four classes (BIRD, blue; UFO, green; INSECT, Red; CLUTTER, black). Densities were estimated with kernel density estimation. The class-conditional densities are shown and hence the class priors are not deducible from the plots. The class priors are given in table 1. The grey vertical line represents the decision threshold equal to zero used to obtain class-level predictions.

the predictions made by the classifier on the test set are compared with the true class labels in order to estimate the classification performance (i.e. AUC, ACC, PPV or FNR). Cross validation leads to an unbiased estimation of classification performance. We computed confidence intervals around the performance estimators (tables 2 and 3) using the method described by Nadeau & Bengio (2003). We have chosen to use the 95% confidence bounds, i.e. intervals that include the true (but unknown) classification performance in at least 95 out of 100 cases. See the electronic supplementary material for a more formal development.

2.8.2. Cross validation to assess the ‘within dataset method’. This procedure involves data from one dataset at a time. It was performed separately for the four datasets. It can be divided into three steps. (i) A training set of sample size N_{train} was randomly sampled without replacement from a dataset. The SVC was trained with this training set. (ii) A test set of sample size N_{test} was randomly sampled without replacement from the remaining signals of the dataset. The score and class of the elements of the test set were predicted with the trained classifier. (iii) The four performance estimators were computed from these predictions and the class labels. Steps (i)–(iii) were repeated 30 times and the average over the repeats was used as the final estimator (table 2). N_{train} was 4000 and N_{test} was 5000 except for the dataset A, where N_{test} was 1878. (Additional results with $N_{\text{train}}=1000$, 2000 and 8000 are given in the electronic supplementary material.)

2.8.3. Cross validation to assess the ‘cross dataset method’. This procedure involves data from two datasets at a time. The training data were always taken from dataset B and the test data were taken from one of the three other datasets A, C or D. The ‘cross dataset method’ was performed separately for the datasets A, C or D. The procedure is slightly more complicated than the ‘within dataset method’ due to the fact that class priors are different in the four datasets. It can be divided into three steps. (i) A training set of size N_{train} was randomly sampled with replacement from dataset B with the constraint that the class priors are equal to those of the test dataset. The SVC was trained with this training set. (ii) A test set of sample size N_{test} was randomly sampled without replacement from dataset A or C or D. The score or class of the elements of the test set was predicted with the trained classifier. (iii) The four performance estimators were computed from these predictions and the class labels. Steps (i)–(iii) were repeated 30 times and the average over the repeats was used as the final estimator (table 3). N_{train} was 4000 and N_{test} was 5000. (Additional results with $N_{\text{train}}=1000$, 2000 and 8000 are given in the electronic supplementary material.)

2.9. Visualization of classification performance

In order to visualize the classification performance, we plotted the distribution of the test data points over the score separately for each class (see figures 5 and 6). The distributions were estimated by kernel density estimation with a Gaussian kernel. Kernel bandwidth was selected according to the method of Scott (1992). Note that the distributions are normalized to have a surface of 1. Thus the plots do not account for the differences in class priors.

3. RESULTS

3.1. Model selection

The selected model (i.e. the best performing model) includes 43 out of the 67 variables: *resolution*, *maxi*, *sd-stat*, *coefmean* 32–25 and 16–1 and *coefsd* 32–29 and

Table 2. Performance estimators and conservative 95% confidence limits when the within dataset method was used. ($N_{\text{train}}=4000$, $N_{\text{test}}=5000$ (except for dataset A, $N_{\text{test}}=1878$).)

dataset	AUC	ACC	PPV	FNR
A	0.9953 ± 0.0021	0.9711 ± 0.0053	0.9733 ± 0.0056	0.0283 ± 0.0066
B	0.9850 ± 0.0020	0.9691 ± 0.0024	0.9287 ± 0.0089	0.2010 ± 0.0128
C	0.9748 ± 0.0034	0.9594 ± 0.0026	0.9174 ± 0.0075	0.2184 ± 0.0203
D	0.9656 ± 0.0105	0.9796 ± 0.0014	0.8691 ± 0.0263	0.4851 ± 0.0505

Table 3. Performance estimators and conservative 95% confidence limits when the cross dataset method was used. ($N_{\text{train}}=4000$, $N_{\text{test}}=5000$. Here the training data are always from dataset B.)

dataset	AUC	ACC	PPV	FNR
A	0.9898 ± 0.0033	0.9497 ± 0.0084	0.9601 ± 0.0068	0.0563 ± 0.0160
C	0.9496 ± 0.0068	0.9310 ± 0.0065	0.7932 ± 0.0266	0.3207 ± 0.0304
D	0.8873 ± 0.0200	0.9627 ± 0.0046	0.5060 ± 0.0618	0.6618 ± 0.0498

Table 4. The classification performance (AUC) of some of the models considered in the model selection procedure (including the finally selected model). (The results presented in this table are based on the subset of 6000 signals from dataset B that was set aside for model selection. The AUC is given for the best combination of tuning parameters (C and σ). When several values of the tuning parameters give the same maximum AUC, we show the smaller value of C and σ .)

	$maxi$	$sd-stat$	$resolution$	$coefmean$	$coefs_d$	C	σ	AUC
full model	y	y	y	1–32	1–32	10	0.1	0.9848
model A	—	—	—	1–32	1–32	10	0.1	0.9783
model B	y	y	y	—	1–32	100	0.01	0.9741
model C	y	y	y	1–32	—	100	0.01	0.9831
model D	y	y	y	—	—	10 000	0.001	0.8724
model E	—	—	—	1–32	—	100	0.1	0.9731
model F	—	—	—	—	1–32	10 000	0.0001	0.9386
selected model	y	y	y	1–16 and 25–32	1–12 and 29–32	100	0.1	0.9891

12–1. The selected tuning parameters for this model are $C=100$ and $\sigma=0.1$. This model underlies all the results presented hereafter. The performance of a selected set of models is summarized in table 4.

3.2. Performance of the within dataset method

The smallest accuracy is obtained for dataset C (ACC=0.9594) and the highest for dataset D (ACC=0.9796). The classification performance measured with AUC is highest on the dataset A followed by B, then C and finally D (table 2). Interestingly, dataset A with the highest proportion of birds resulted in the highest AUC and D with the lowest proportion of birds resulted in the lowest AUC. Overall AUC ranges between 0.9656 and 0.9953. The best performance for PPV and FNR resulted in the dataset A. The subset of data which is predicted by the classifier to be of the class BIRD contains 97.3% (PPV=0.9733) of signals, which truly are of the class BIRD, whereas only 2.8% (FNR=0.0283) of the total amount of BIRDS in the dataset have been wrongly predicted to be of the class OTHER. The lowest performance for PPV and FNR resulted in the dataset D. The subset of data which is predicted by the classifier to be of the class BIRD contains 87% (PPV=0.8691) of signals, which truly are of the class BIRD. But 49% (FNR=0.4851) of the total amount of

BIRDS in the dataset have been wrongly predicted to be of the class OTHER (table 2).

3.3. Performance of the cross dataset method

Names of datasets in this section always refer to the data used for testing. The training set was always dataset B. Qualitatively the results of the cross dataset method are similar to those of the within dataset method, but the performance is lower. The AUC is highest on the dataset A, followed by C and finally D. Note that this is the same ranking as for the within dataset method (table 3).

3.4. Plots of class-specific distribution of the score

In all the four datasets, the relative position of the estimated distributions was qualitatively similar (figures 5 and 6). In all the datasets and for both the within and cross dataset methods, the class UFO presents the largest overlap with the BIRDS.

The plots illustrate the results obtained with the AUC, which can be interpreted as a measure of the overlap between two classes along the score. The overlap between BIRDS and OTHERs is lowest for dataset A and highest for dataset D. This holds for the within (figure 5) as well as the cross dataset methods (figure 6).

Hence the overlap is related to the proportion of BIRDS in the datasets (cf. table 1). In terms of overlap, the cross dataset method is generally worse. For dataset D the cross dataset method suffers from a large overlap between BIRDS and the other three classes, which results in a poor performance (cf. table 3).

4. DISCUSSION

4.1. General aspects

The results clearly demonstrate that an automatic classifier for radar signals can be successfully established. A useful property of the method is that the output is a score, which provides a ranking of the data instances according to the reliability of the predictions. The decision threshold on the score can be set to other values than 0. For example, setting it to higher values improves the PPV but deteriorates the FNR (figures 5 and 6). This means that the selected subset will contain a higher proportion of birds with the disadvantage that more of the overall amount of the birds are lost to the rejected subset. It is also possible to define two decision thresholds on the score. An upper threshold can be chosen to identify a subset of data with a high proportion of birds (high PPV). A lower threshold can identify a subset of the data which includes only very few birds (low FNR). Finally, the data subset that is defined in between the two thresholds can either be treated separately in a specific analysis or be labelled additionally by the plot-based method.

In all cases a certain proportion of the BIRDS are not included in the selected subset. A desired property of the classifier would be that the excluded BIRDS represent a random subsample of all the BIRDS. Presently, we cannot say whether this is the case. A large part of the classifier's recognition ability is based on the pattern produced by wing flapping. Therefore it is possible that birds that do not produce a clear wing flapping pattern (e.g. soaring birds, dense bird flocks) will get, on average, a lower predicted value for the score than single birds, which are flapping most or all of the time. If this is the case, our method would selectively exclude soaring bird species and species that fly in dense flocks. This important point will hopefully be clarified in future studies.

Our approach uses a signal pre-processing procedure and then a nonlinear modelling method (SVC), which is necessary to achieve a high classification performance. A disadvantage of such an approach is that it makes it hardly possible to understand intuitively the process by which the classifier assigns a score to the signals.

The classifier resulting from the training process is dependent on the training dataset and thus on the correctness of labelling by the expert. There was a constraint in our expert learning, because we had to extrapolate from signals identified visually during daytime to nocturnal signals; though in a few field projects we could also verify nocturnal signals with infrared and light beam observations (Liechti *et al.* 1995). Since the flight mode of birds is basically the same during day and night (except soaring birds), we think that this extrapolation is unproblematic.

4.2. Performance

In the datasets B, C and D, the target class (BIRD) is underrepresented (table 1). This is an unfavourable situation because smaller class priors of the class BIRD *per se* decrease the PPV. In this context the PPV should be interpreted by considering the improvement it brings when compared with the proportion of BIRDS in the whole dataset (i.e. the class prior of the BIRDS).

Our principal purpose is to identify and select a subset of the data, which should contain a high proportion of BIRDS while rejecting another subset, which should contain as few BIRDS as possible. Let us take dataset B for illustration. Drawing biological conclusions on birds based on the complete dataset would be misleading since only 12% of the signals are from BIRDS. When the classifier is applied, the data are split into a selected subset and a rejected subset. The proportion of BIRDS in the selected subset is over 92% (PPV=0.9286), while 20% of the total number of BIRDS are lost to the rejected subset (FNR=0.2010). Drawing conclusions based on the selected subset seems much more reasonable. The results obtained for dataset C are very similar. For dataset A the proportion of BIRDS in the selected subset is over 97% and less than 3% of the total amount of BIRDS is lost to the rejected subset. This performance is above our expectation. It shows that when class priors are balanced, the practical usefulness of the classifier is excellent. For the least performing dataset D, the proportion of BIRDS in the selected subset is 87%, while 49% of the total amount of BIRDS is lost to the rejected subset. If we consider that the initial proportion of BIRDS in the complete dataset was only 4%, this represents a large improvement of the data quality, but the price paid in terms of the amount of birds lost to the rejected subset is large. Dataset D was recorded under frequent sandstorms and an extremely high abundance of insects, which reduces the classification performance (see below). In our experience, such situations are rare, and the performance obtained on datasets A, B and C is more representative of the practical usefulness of our method.

Using a classifier trained with data from another dataset (cross dataset method) provided qualitatively similar results as above, but with a consistently lower classification performance (table 3). It performed well when the proportion of BIRDS was approximately 50% (dataset A), but insufficiently when the proportion of BIRDS was low (datasets C and D). In the cross dataset method, it may be that the separability differs between the test set and the training set. Hence, estimations of PPV and FNR obtained from a cross validation on the training set will not necessarily represent the PPV and FNR obtained on the test set. More tests with other datasets are needed to conclude whether the cross dataset method is a reliable solution, at least when the proportion of BIRDS is high (e.g. dataset A). The big advantage of the cross dataset method is that there is no need to create a further training set via the time-consuming plot-based labelling.

It is known that an approximately correct specification of the class priors of the test set does improve the reliability of the predictions (Saerens *et al.* 2002). In our cross dataset approach, we assumed knowledge of the

class priors in the test set. In practice this may not always be available (for example for real time classification). But other sources of information may be used to estimate priors. For example, temperature may be used to estimate insect activity or preliminary knowledge about the seasonal or local abundance of insects may be available.

Regardless of the training method used, the classifier achieved the best results (highest AUC) with dataset A, followed by B and C and finally D. This ranking is consistent with the decreasing proportion of BIRDS in the datasets (table 1), which was caused by a tremendous increase in the number of insects and amount of airborne sand. Therefore, overlap of different signals occurred frequently in the datasets B, C and D. The presence of insects and/or sand in the same pulse volume as a bird is expected to partially mask the wing flapping pattern. But the recognition ability of the classifier is mostly based on the wing flapping pattern in the signals. Therefore it is no surprise that as the patterns become less clear the classification performance decreases.

4.3. Possible improvements

Our datasets were dominated by nocturnal migrants, consisting mostly of small to medium sized birds flying singly. Signals of large birds (e.g. soaring birds) and flocks of birds were relatively rare (less than 1% of the BIRDS). Therefore, the present method was developed mainly to identify signals from small- to medium-sized single birds. In general, large birds and bird flocks produce stronger signals than small single birds, but unambiguous wing flapping patterns can hardly be detected. In flocks the signal represents the superimposed wing flapping of all individuals in the radar pulse volume, while large birds are often gliding or soaring without flapping. In addition interferences between the pulses (i.e. the electromagnetic wave) reflected by several targets within the same pulse volume will also render the signals more difficult to analyse.

Our classifier uses 43 variables. The signal intensity (echo size) is only represented in one variable ‘*maxi*’. The remaining variables contain information on the variability or on the shape of the signals, which both reflect the wing flapping. Therefore, our method heavily relies on information related to the wing flapping and probably does not fully exploit the potential of using the signal intensity. An improved version of the classifier should use the information on signal intensity more efficiently than in the present study. This is expected to result in a classifier with a good performance on bird flocks and large birds.

In general, the frequency represented by the overall curvature of the signal intensity is outside the range of frequencies analysed by our algorithm. However, this curvature together with target distance and width of the radar beam could be used to extract a variable representing the flight speed of the targets. Since birds have a higher airspeed than insects, this could further improve the classification performance.

A serious problem is the co-occurrence of intense insect and bird migration. In such situations, many signals of birds are contaminated by radar returns from

insects, making them very difficult to be labelled unambiguously by the human expert. Accordingly, such signals were labelled as UFO. Thus an unknown proportion of the UFOs were actually produced by flying birds. These signals may contain characteristic patterns of birds that were not recognized by the human expert, but that are recognized by the classifier. Under this assumption we expect that the classifier has more difficulties at separating the BIRDS from the UFOs than the BIRDS from the INSECTS or CLUTTERS, which is the case in our data (see figures 5 and 6). At the moment, we can only speculate that UFO signals with high scores were actually produced by birds. But to confirm this we need additional independent information (e.g. parallel infrared images).

4.4. Applications

4.4.1. *When is the exclusion on non-bird signals important?* We are convinced that for any radar study dealing with the movements of small- to medium-sized birds, a discrimination between bird and non-bird (in particular insect) signals is a prerequisite. Of course the exclusion of non-bird signals becomes less important when the overall proportion of birds is close to 100%. But in most situations such high proportions are not seen. In our study, which we consider representative for arid sub-tropical climates in general, birds represent 4–52% of the signals. The study by Larkin (1991) showed that in temperate regions insect contamination can be 24–28% (up to nearly 100% on specific nights), and even in the Arctic insect signals were recorded regularly by radar (Gudmundsson et al. 2002).

4.4.2. *Application across different radar systems.* The method is feasible for all kind of radar systems, which allow a target to be recorded continuously for at least a few seconds. We might wish to train a classifier with data from a given radar system and use this classifier to identify birds in a dataset recorded with another radar system. But the following aspects must be considered. (i) The process by which wing flapping produces the periodic fluctuations in the signal is not fully understood. It is known that wavelength and polarization of the pulses do affect this process considerably (Bruderer 1997b; Bachmann & Zrnic 2007). (ii) The overall power of the received signals depends on technical settings, which can differ significantly between radar systems, and even the relative fluctuations of the signal are dependent on the dynamic range of the amplifier. (iii) The time resolution of the signals can differ depending on the signal processing units of different radar systems. Based on these three points, we think the training data should be recorded with the same radar system as the data on which the classifier is applied.

4.4.3. *Real time classification.* Radar measurements are used for bird strike warnings, mainly in military aviation. Up to now these warnings are based on the expert knowledge of the radar operators. In order to introduce an automatic classifier, radar signal identification should be in real time or close to it (within

minutes). Such real time warnings could also be used to mitigate detrimental effects of wind turbines by stopping them for the time of mass migration. For real time classification we have to rely on a classifier that was trained on data recorded in the past. Thus, there is a strong demand to improve the performance of the cross dataset method.

The differences in flapping styles between bird species are reflected in the shape of the recorded signals, and hence in the values of the extracted variables. It is reasonable to assume that, in general, the species composition of the birds aloft differs between recording locations and periods. We conclude that differences in species composition between the datasets are the probable cause of the lower performance of the cross dataset method as compared with the within dataset method. Based on this we suggest to improve the cross dataset method by creating a ‘standard’ training dataset where all different kinds of birds (small to large sized, continuous and intermittently flapping) are more or less equally represented. Such a standard training dataset could be built up over time by combining datasets from different sites, until repeated comparisons with the within dataset method would indicate a satisfactory performance.

The cross dataset method could be implemented in real time to provide a relative quantification of migration intensity (i.e. allowing relative comparisons between time periods). It could be used, for example, to detect a fast increase in the abundance of birds aloft and help us react by giving an alarm.

4.5. Perspectives

A promising approach for the future would be to combine several existing radar methods that provide information on target identity. (i) The method presented here, which uses wing flapping pattern. (ii) Signal intensity (echo size). As mentioned above signal intensity is underexploited in our method. (iii) Airspeed. (iv) Signal dissimilarity due to varying polarization (dual polarization radar, [Bachmann & Zrnic 2007](#)). As at least some of the information gathered by these methods is complementary, combining the methods might improve classification performance. Whether it is technically feasible to record simultaneously all this information with one radar system is questionable.

The time resolution of our recording system is restricted to 130 Hz, implying an upper bound at 65 Hz on the frequencies that can be analysed. Using a higher time resolution would allow the detection of higher frequencies and this might provide additional information on the wing flapping of some insects and improve the classification performance.

The pattern produced by the wing flapping of birds contains information related to the wing flapping frequency and the mode of flapping (continuous, intermittent). This information can be used to divide the birds’ signals into several subclasses according to their wing flapping style, for example wader-like, passerine-like or swift-like ([Bloch *et al.* 1981](#)). This is of great interest for the studies on bird migration. It would be interesting to see whether the division into subclasses can be automated as well.

4.6. Conclusion

It has been known for many years that bird signals can be recognized accurately by human experts thanks to the excellent pattern recognition ability of humans. The main finding of our study is that this recognition task can be automated by using machine learning methods such as the SVC. This considerably reduces the dependency on human experts and will allow analysis of large datasets of automatically recorded radar signals in almost real time. We are convinced that this kind of classification will be an indispensable tool for future radar applications dealing with biological targets.

We would like to thank Erich Bächler, Bruno Bruderer, Alexandros Karatzoglou, Beat Naef-Daenzer, Dieter Peter, Thomas Steuri and Barbara Trösch for their assistance and support. We would like to thank Mark Desholm and two anonymous referees for their constructive comments on the manuscript.

The algorithm was implemented in the free and open source statistical programming language R ([R Development Core Team 2006](#)). Researchers interested in the algorithm (R-code) are encouraged to contact us.

REFERENCES

- Addison, P. S. 2002 *The illustrated wavelet transform handbook*. Bristol, PA: Institute of Physics Publishing.
- Bachmann, S. & Zrnic, D. 2007 Spectral density of polarimetric variables separating biological scatterers in the VAD display. *J. Atmos. Oceanic Technol.* **24**, 1186–1198. (doi:10.1175/JTECH2043.1)
- Bäckman, J. & Alerstam, T. 2002 Harmonic oscillatory orientation relative to the wind in nocturnal roosting flights of the swift *Apus apus*. *J. Exp. Biol.* **205**, 905–910.
- Bäckman, J. & Alerstam, T. 2003 Orientation scatter of free-flying nocturnal passerine migrants: components and causes. *Anim. Behav.* **65**, 987–996. (doi:10.1006/anbe.2003.2119)
- Bloch, R. & Bruderer, B. 1982 The air speed of migrating birds and its relationship to the wind. *Behav. Ecol. Sociobiol.* **11**, 19–24. (doi:10.1007/BF00297661)
- Bloch, R., Bruderer, B. & Steiner, P. 1981 Flugverhalten nächtlich ziehender Vögel—Radardaten über den Zug verschiedener Vogeltypen auf einem Alpenpass. *Vogelwarte* **31**, 119–149.
- Bonham, L. L. & Blake, L. V. 1956 Radar echoes from birds and insects. *Sci. Mon.* **82**, 204–209.
- Bruderer, B. 1971 Radarbeobachtungen über den Frühlingzug im Schweizerischen Mittelland. (Ein Beitrag zum Problem der Witterungsabhängigkeit des Vogelzugs.) *Ornithol. Beob.* **68**, 89–158.
- Bruderer, B. 1997a The study of bird migration by radar. Part 2: major achievements. *Naturwissenschaften* **84**, 45–54. (doi:10.1007/s001140050348)
- Bruderer, B. 1997b The study of bird migration by radar. Part 1: the technical basis. *Naturwissenschaften* **84**, 1–8. (doi:10.1007/s001140050338)
- Bruderer, B. & Joss, J. 1969 Methoden und Probleme der Bestimmung von Radarquerschnitten freifliegender Vögel. *Rev. Suisse Zool.* **76**, 1106–1118.
- Bruderer, B. & Liechti, F. 1998 Flight behaviour of nocturnally migrating birds in coastal areas—crossing or coasting. *J. Avian Biol.* **29**, 499–507. (doi:10.2307/3677169)

- Bruderer, B. & Popa-Lisseanu, A. G. 2005 Radar data on wing-beat frequencies and flight speeds of two bat species. *Acta Chiropt.* **7**, 73–82. (doi:10.3161/1733-5329(2005)7[73:RDOWFA]2.0.CO;2)
- Bruderer, B. & Steidinger, P. 1972 Methods of quantitative and qualitative analysis of bird migration with a tracking radar. In *Animal orientation and navigation* (eds S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs & R. E. Belleville), pp. 151–167. Washington, DC: NASA.
- Bruderer, B., Steuri, T. & Baumgartner, M. 1995 Short-range high-precision surveillance of nocturnal migration and tracking of single targets. *Isr. J. Zool.* **41**, 207–220.
- Carmona, R. (S original) & Whitcher, B. (R port) 2004 *Rwawe: time-frequency analysis of 1-D signals*, R package version 1.22. See <http://www.orfe.princeton.edu/~rcarmona/>.
- Carmona, R., Hwang, W. L. & Torr sani, B. 1998 *Practical time-frequency analysis*. New York, NY: Academic Press.
- Cleveland, W. S. 1981 LOWESS: a program for smoothing scatterplots by robust locally weighted regression. *Am. Stat.* **35**, 54. (doi:10.2307/2683591)
- Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Mach. Learn.* **20**, 273–297.
- Desholm, M. & Kahlert, J. 2005 Avian collision risk at an offshore wind farm. *Biol. Lett.* **1**, 296–298. (doi:10.1098/rsbl.2005.0336)
- 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**(Suppl. 1), 76–89. (doi:10.1111/j.1474-919X.2006.00509.x)
- Eastwood, E. 1967 *Radar ornithology*. London, UK: Methuen.
- Fawcett, T. 2006 An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874. (doi:10.1016/j.patrec.2005.10.010)
- Feng, H. Q., Zhang, Y. H., Wu, K. M., Cheng, D. F. & Guo, Y. Y. 2007 Nocturnal windborne migration of ground beetles, particularly *Pseudoophonus griseus* (Coleoptera: Carabidae), in China. *Agric. For. Entomol.* **9**, 103–113. (doi:10.1111/j.1461-9563.2007.00326.x)
- Garthe, S. & H ppop, O. 2004 Scaling possible adverse effects of marine wind farms on seabirds: developing and applying a vulnerability index. *J. Appl. Ecol.* **41**, 724–734. (doi:10.1111/j.0021-8901.2004.00918.x)
- Gauthreaux Jr, S. A. & Belser, C. G. 2003 Radar ornithology and biological conservation. *Auk* **120**, 266–277. (doi:10.1642/0004-8038(2003)120[0266:ROABC]2.0.CO;2)
- Gudmundsson, G. A., Alerstam, T., Green, M. & Hedenstr m, A. 2002 Radar observations of arctic bird migration at the northwest passage, Canada. *Arctic* **55**, 21–43.
- Hamer, T. E., Schuster, S. M. & Meekins, D. 2005 Radar as a tool for monitoring Xantus’s murrelet populations. *Mar. Ornithol.* **33**, 139–146.
- Harmata, A. R., Podruzny, K. M., Zelenak, J. R. & Morrison, M. L. 1999 Using marine surveillance radar to study bird movements and impact assessment. *Wildl. Soc. Bull.* **27**, 44–52.
- Hastie, T., Tibshirani, R. & Friedman, J. 2001 *The elements of statistical learning, data mining, inference, and prediction*. Springer series in statistics. Berlin, Germany: Springer.
- Hedenstr m, A. & Liechti, F. 2001 Field estimate of body drag coefficient on the basis of dives in passerine birds. *J. Exp. Biol.* **204**, 1167–1175.
- Houghton, E. W. 1964 Detection, recognition and identification of birds on radar. In *Proc. 11th Weather Radar Conference*, pp. 14–21. Toronto, Canada: University of Toronto Press.
- Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. 2004 Kernlab—An S4 package for kernel methods in R. *J. Stat. Softw.* **11**, 1–20. (<http://www.jstatsoft.org/v11/i09/>)
- Larkin, R. P. 1991 Flight speeds observed with radar, a correction: slow “birds” are insects. *Behav. Ecol. Sociobiol.* **29**, 221–224. (doi:10.1007/BF00166405)
- Liechti, F. & Bruderer, B. 1995 Direction, speed and composition of nocturnal bird migration in the south of Israel. *Isr. J. Zool.* **41**, 501–515.
- Liechti, F. & Schaller, E. 1999 The use of low-level jets by migrating birds. *Naturwissenschaften* **86**, 549–551. (doi:10.1007/BF00166405)
- Liechti, F., Bruderer, B. & Paproth, H. 1995 Quantification of nocturnal bird migration by moonwatching: comparison with radar and infrared observations. *J. Field Ornithol.* **66**, 457–468.
- Lilliendahl, K., Solmundsson, J., Gudmundur, A. G. & Taylor, L. 2003 Can surveillance radar be used to monitor the foraging distribution of colonially breeding alcid? *Condor* **105**, 145–150. (doi:10.1650/0010-5422(2003)105[145:CSRBT]2.0.CO;2)
- Nadeau, C. & Bengio, Y. 2003 Inference for the generalization error. *Mach. Learn.* **52**, 239–281. (doi:10.1023/A:1024068626366)
- Nievergelt, F., Liechti, F. & Bruderer, B. 1999 Migratory directions of free-flying birds versus orientation in registration cages. *J. Exp. Biol.* **202**, 2225–2231.
- R Development Core Team 2006 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. See <http://www.R-project.org>.
- Riley, J. R. & Reynolds, D. R. 1986 Orientation at night by high-flying insects. In *Insect flight, dispersal and migration* (ed. W. Danthanarayana), pp. 71–87. Berlin, Germany: Springer.
- Russell, R. W. & Gauthreaux, S. A. 1998 Use of weather radar to characterize movements of roosting Purple Martins. *Wildl. Soc. Bull.* **26**, 5–16.
- Saerens, M., Latinne, P. & Decaestecker, C. 2002 Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure. *Neural Comput.* **14**, 21–41. (doi:10.1162/089976602753284446)
- Schmaljohann, H., Liechti, F. & Bruderer, B. 2007a An addendum to “Songbird migration across the Sahara: the non-stop hypothesis rejected!”. *Proc. R. Soc. B* **274**, 1919–1920. (doi:10.1098/rspb.2007.0188)
- Schmaljohann, H., Liechti, F. & Bruderer, B. 2007b Songbird migration across the Sahara—the non-stop hypothesis rejected! *Proc. R. Soc. B* **274**, 735–739. (doi:10.1098/rspb.2006.0011)
- Schmaljohann, H., Liechti, F., B chler, E., Steuri, T. & Bruderer, B. In press. Quantification of bird migration by radar—a detection probability problem. *Ibis*. (doi:10.1111/j.1474-919x.2007.00797.x)
- Scott, D. W. 1992 *Multivariate density estimation. theory, practice and visualisation*. New York, NY: Wiley.
- Skolnik, M. I. 1980 *Introduction to radar systems*, 2nd edn. New York, NY: McGraw-Hill Book Company.
- Smith, A. D., Riley, J. R. & Gregory, R. D. 1993 A method for routine monitoring of the aerial migration of insects by using a vertical-looking radar. *Phil. Trans. R. Soc. B* **340**, 393–404. (doi:10.1098/rstb.1993.0081)
- Stark, H. & Liechti, F. 1993 Do Levant Sparrowhawks *Accipiter brevipes* also migrate at night? *Ibis* **135**, 233–236.
- Torrence, C. & Compo, G. P. 1998 A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78. (doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
- Vapnik, V. N. 2000 *The nature of statistical learning theory*, 2nd edn. Berlin, Germany: Springer.
- Zrnec, D. S. & Ryzhkov, A. V. 1998 Observations of insects and birds with a polarimetric radar. *IEEE Trans. Geosci. Rem. Sens.* **36**, 661–668. (doi:10.1109/36.662746)