Bayesian Classification of Flight Calls with a novel Dynamic Time Warping Kernel

Theodoros Damoulas
Department of Computer Science
Cornell University
Ithaca, 14853 NY, USA
damoulas@cs.cornell.edu

Samuel Henry
Department of Computer Science
Cornell University
Ithaca, 14853 NY, USA
sth56@cornell.edu

Andrew Farnsworth
Cornell Lab of Ornithology
Cornell University
Ithaca, 14853 NY, USA
af27@cornell.edu

Michael Lanzone
Biotechnology and Biomonitoring Lab
Carnegie Museum of Natural History
ARC, Rector, PA 15677, USA
mlanzone@gmail.com

Carla Gomes
Department of Computer Science
Cornell University
Ithaca, 14850 NY, USA
gomes@cs.cornell.edu

***Abstract*—In this paper we propose a probabilistic classification algorithm with a novel Dynamic Time Warping (DTW) kernel to automatically recognize flight calls of different species of birds. The performance of the method on a real world dataset of warbler (Parulidae) flight calls is competitive to human expert recognition levels and outperforms other classifiers trained on a variety of feature extraction approaches. In addition we offer a novel and intuitive DTW kernel formulation which is positive semi-definite in contrast with previous work. Finally we obtain promising results with a larger dataset of multiple species that we can handle efficiently due to the explicit multiclass probit likelihood of the proposed approach[1].**

***Index Terms*—Acoustic Signal Processing, Probabilistic Supervised Learning, Dynamic Time Warping, Kernel Machines**

I. Introduction

Birds are sensitive environmental indicators and among the first animals to respond to changes in local ecosystems and the global climate. Tracking bird populations in space and time is a central challenge for conservation science, particularly in light of the potential changes to the present climate of the planet, and the new field of computational sustainability provides novel insight into pursuing these challenges [1], [2]. The migratory patterns of birds can produce data of interest beyond the ornithological community, and therefore has been the focus of much recent research [3]–[9].

Determining the migration paths of birds at the species level is difficult but one of the most promising methods of tracking avian migrations is by flight call recording. Many species produce flight calls: species-specific vocalizations that vary in duration, frequency, and contour, and are frequently given during nocturnal migration by several hundred species in North America. These signals are the only source of information, at present, for reliably identifying passing nocturnal migrants. Recording stations that can capture such signals are relatively inexpensive and can be deployed remotely [5]. Additionally, these stations can be programmed to record autonomously, facilitating data collection for extended, nocturnal periods.

One of the primary limiting factors to expanding networks of recording stations from single station networks to thousands of microphones is the classification process. Classification of flight calls has traditionally been a labor-intensive and expensive manual process consisting of inspecting spectrograms by trained professionals. In this paper we present a method of automatically classifying such avian flight calls that allows for the deployment of large scale systems of recording stations, see Fig. 1, that have the ability to accurately track the migrations of birds around the world.

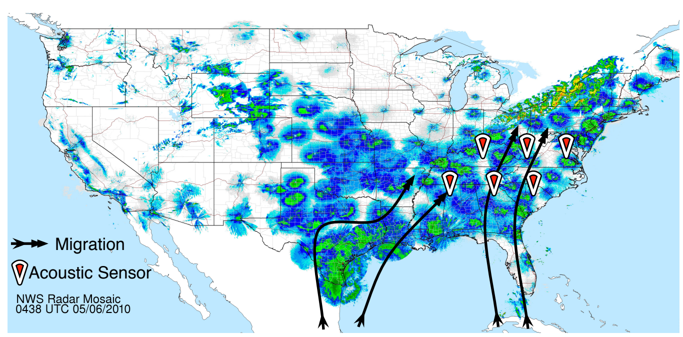

Fig. 1. Developing an efficient classification algorithm allows for large-scale networks of recording stations to be built. This image (United States Weather Surveillance Radar-88Doppler (WSR-88D) network) depicts a frontal boundary with scattered, intense precipitation over Lakes Erie and Ontario, spreading eastward into New York and New England. However, the image is dominated by biological targets, mostly birds, visible across the Great Plains south to Texas and east to the Southeastern Coastal Plain north along the Atlantic coast to Maine. The areas of intense, uniform green representing targets over the radars show bird densities close to 2.5×10^3 birds/km^3.

The contributions of this work are:

- Excellent recognition rates (97%) on automatic flight call detection of 5 within-family species (classes), and promising results on a larger dataset of 45 species within and across families. To put the results in perspective

[1]The data and codes are available at www.cis.cornell.edu/ics/projects.php and www.dcs.gla.ac.uk/inference/pMKL

experienced observers classify on average 69% - 91% correctly, while experts (of which there are very few) can classify between 88% - 100% correctly.
- A novel DTW positive semi-definite (p.s.d) kernel ($\text{DTW}_{\text{global}}$) that results in a very discriminative feature space for detection of acoustic signals.
- A successful application of the variational Bayes probabilistic classifier [10] to detect flight calls.
- Publicly available codes and real world datasets.

II. Previous Work

A. Flight Call Classification

There have been attempts to automate flight call classification in the past and the results were promising; but several factors, including robust call measurement and representation of intra- and inter-specific variation in calls, in a computational viable manner, were major unmet challenges. Successful methodologies included template matching schemes [11], [12] and statistical methods for classification [7], [8].

Flight calls must be extracted from the continuous audio streams produced by recording stations. This process is essential even for manual classification, so automated detection and extraction of flight calls from the raw data generated from individual recording stations has been the main focus of prior research [7], [13]. Automatic flight call detection has challenges, in particular the abundance of additional confounding signals in the frequency range of interest for flight calls, but works sufficiently well for the samples discussed in this paper. The detection algorithms are designed to identify high levels of energy with user-defined characteristics in specific frequency bands [5].

The microphones used to capture flight calls vary in type, but an inexpensive and effective design is Pressure Zone Microphones [8]. These microphones can be built with off-the-shelf parts, costing as little as \$50 and can easily be deployed. For more information about the microphones used to capture the raw data, details can be found at [7] and http://www.oldbird.org.

B. Dynamic Time Warping Kernels

Dynamic Time Warping is a well-known dynamic programming method [14] that has been extensively applied to time-series and sequence-based problem domains. It operates by stretching a sequence[2] $\mathbf{x}_i \in \Re^{D_i}$ in order to match another sequence $\mathbf{x}_j \in \Re^{D_j}$ while calculating the cost of alignment $c_{\mathbf{x}_i,\mathbf{x}_j}$ based on an application-specific function or some standard distance measure derived from the warped path(s). A warping path π_{ij} of length $|\pi_{ij}| = p$ is a path in the $\mathbf{x}_i, \mathbf{x}_j$ graph, that can be seen [15] to define a pair of increasing p-tuples (ζ_i, ζ_j) as:

$$\begin{aligned} 1 &= \zeta_i(1) \leq \cdots \leq \zeta_i(p) = D_i \\ 1 &= \zeta_j(1) \leq \cdots \leq \zeta_j(p) = D_j \end{aligned} \qquad (1)$$

[2]As usual x denotes scalar, $\mathbf{x}$ column vector and $\mathbf{X}$ matrix.

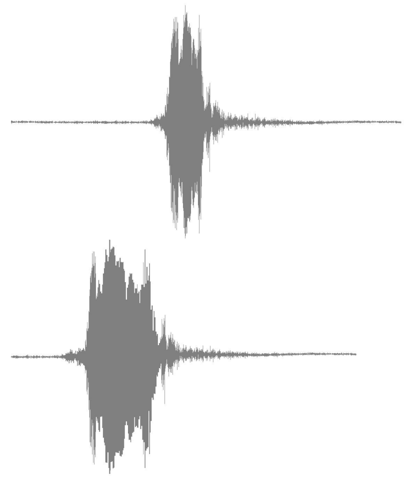

Fig. 2. An example of two flight call signals that are out of phase. The signals have similar structure but will appear very different if point to point comparisons are made.

where ζ_i is the warping transformation of the i^{th} sequence when mapped according to the path π_{ij} and intuitively describes a matching between elements of $\mathbf{x}_i$ and $\mathbf{x}_j$. D_i denotes the dimensionality of vector $\mathbf{x}_i$.

Recently there has been a great interest in constructing kernels based on DTW measures [15]–[17]. All the proposed approaches are developed within the Support Vector Machine (SVM) framework and hence cannot explicitly handle uncertainty due to their non-probabilistic nature. The previous kernels are summarized in Table I.

TABLE I
Previous DTW kernels

Work	Kernel function $k(\mathbf{x}_i, \mathbf{x}_j) =$
[16]	$\underset{\pi}{\text{argmax}} \frac{1}{\lvert\pi\rvert} \sum_{p=1}^{\lvert\pi\rvert} \exp(-\frac{1}{\sigma^2} \lvert\lvert\mathbf{x}_{i,\zeta_i(p)} - \mathbf{x}_{j,\zeta_j(p)}\rvert\rvert^2)$
[15]	$\sum_\pi \prod_{p=1}^{\lvert\pi\rvert} \exp(-\beta \lvert\lvert\mathbf{x}_{i,\zeta_i(p)} - \mathbf{x}_{j,\zeta_j(p)}\rvert\rvert^2)$

The kernel function proposed in [16] is not a p.s.d kernel in general [17], and the kernel in [15] is p.s.d "under favorable conditions" but can be diagonally dominant hence requiring additional smoothing. Related past work [18], [19] on constructing kernels for sequence-based problems has employed Hidden Markov Models (HMMs) as the *generative* (in contrast with this work that does not model parametrically the generating distribution) underlying model and have been applied to both supervised and unsupervised learning settings.

III. Method

In order to best present the approach, we first describe the proposed DTW kernel in comparison with previous formulations and then proceed to the adopted implementation for our specific flight call application.

A. A Dynamic Time Warping Kernel from Global Alignment

Both previous approaches construct a DTW kernel by effectively exploiting the path(s) constructed by the dynamic programming steps within a single warping of sequences $\mathbf{x}_i, \mathbf{x}_j$. In [16] this results in a non-p.s.d kernel and in [15] it requires the consideration of all possible paths within the $D_i \times D_j$ matrix.

We instead construct a global alignment DTW kernel from the minimum-cost alignment scores $c^*_{\mathbf{x}_i\mathbf{x}_j}$ of the optimal paths $\{\pi^*_{in}\}_{n=1}^N$ directly (and not the paths themselves), taking into account the optimal N alignments of a signal $\mathbf{x}_i$ *with all* the available sequences in the training set $\{\mathbf{x}_n, t_n\}_{n=1}^N$. Hence we can employ kernel functions that result in a p.s.d kernel, since there is a common metric of alignment to N sequences, such as the Gaussian:

$$\text{DTW}_{\text{global}} \text{ Kernel} \quad k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{||\mathbf{c}^*_{\mathbf{x}_i} - \mathbf{c}^*_{\mathbf{x}_j}||^2}{\theta}\right) \quad (2)$$

where $\mathbf{c}^*_{\mathbf{x}_i} \in \Re^N$ is the vector of minimum-cost alignments through the optimal N warping paths $\{\pi^*_{in}\}_{n=1}^N$ and $\theta \in \Re$ the bandwidth (isotropic case). In the next section we describe explicitly how this cost is calculated in our implementation for flight call detection. It is worth noting that the proposed approach can exploit existing DTW implementations [20] with different cost functions in analogy to the *String* kernels proposed in bioinformatics and the availability of scoring matrices [21].

B. Minimum-cost alignment for Flight Calls

Given a training set $\{\mathbf{x}_i, t_i\}_{i=1}^N$, with sequence $\mathbf{x}_i \in \Re^{D_i}$ belonging to class t_i we first construct the spectrogram matrices $\mathbf{S}_{\mathbf{x}_i} \in \Re^{F \times W}$ of the signals where W is the number of windows and F the number of frequency bands from the short time Fourier transformation (STFT).

Having obtained the N spectrograms, we can construct a dissimilarity matrix $\mathcal{D}^{ij} \in \Re^{W \times W}$ for sequences $\mathbf{x}_i$ and $\mathbf{x}_j$ by one minus their (normalized) inner product:

$$\mathcal{D}^{ij}(w,v) = 1 - \frac{\mathbf{S}_{\mathbf{x}_i}(:,w)^{\intercal}\mathbf{S}_{\mathbf{x}_j}(:,v)}{\sqrt{\mathbf{S}_{\mathbf{x}_i}(:,w)^{\intercal}\mathbf{S}_{\mathbf{x}_i}(:,w)\mathbf{S}_{\mathbf{x}_j}(:,v)^{\intercal}\mathbf{S}_{\mathbf{x}_j}(:,v)}}, \quad (3)$$

where $\mathbf{S}_{\mathbf{x}_i}(:,w)$ denotes the w th column of the $\mathcal{D}^{ij}$ matrix and $w, v \in \{1, \ldots, W\}$. The dynamic programming operations of standard DTW procedures can now operate on the dissimilarity matrix $\mathcal{D}^{ij}$ in order to obtain the optimal warping path and the minimum-cost alignment between sequences $\mathbf{x}_i$ and $\mathbf{x}_j$:

$$\begin{aligned} \pi^*_{ij} &= \underset{\pi}{\text{argmax}} \frac{1}{|\pi|} c_{\mathbf{x}_i,\mathbf{x}_j}(\pi) \\ c_{\mathbf{x}_i,\mathbf{x}_j}(\pi) &= \sum_{p=1}^{|\pi|} \mathcal{D}^{ij}_{\pi(p)}. \end{aligned} \quad (4)$$

These are collected to complete the description:

$$\mathbf{c}^*_{\mathbf{x}_i} = \left[c_{\mathbf{x}_i,\mathbf{x}_1}(\pi^*_{i1}), \cdots, c_{\mathbf{x}_i,\mathbf{x}_N}(\pi^*_{iN})\right]^{\intercal}. \quad (5)$$

Eq. 5 can be directly used as an "unkernelized" feature construction ($\text{DTW}_{\text{global}}$) and it is worth noting that the cost is symmetric, i.e. given optimal alignment paths π^*_{ij} and π^*_{ji}, the cost is the same: $c_{\mathbf{x}_i,\mathbf{x}_j}(\pi^*_{ij}) = c_{\mathbf{x}_j,\mathbf{x}_i}(\pi^*_{ji}) \; \forall \, i,j \in \{1, \ldots, N\}$.

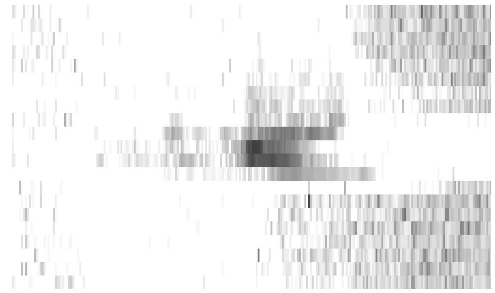

Fig. 3. Example of a spectrogram of the top signal shown in Fig. 2. The x-axis shows the normalized frequency. The y-axis shows time. Each interval on the y-axis represents a window. Shades of gray show the presence of a frequency. Dark gray and black indicate strong presence of the frequency, while light gray to white represents little to no presence of the frequency.

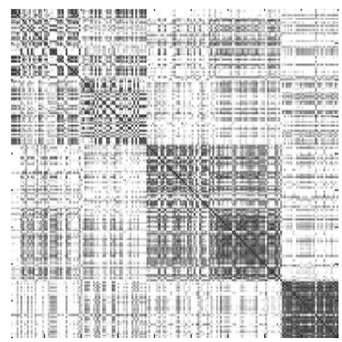

Fig. 4. The $\text{DTW}_{\text{global}}$ kernel from the primary flight call data (N=200, class sorted). The block structure captures the similarity of flight calls within and across species.

C. Probabilistic Multiple Kernel Learning

Having described the underlying DTW kernel representation we now turn to the probabilistic classifier that is adopted for the flight call detection. We employ the variational Bayes (VBpMKL) classifier recently proposed in [10] in order to take advantage of the uncertainty quantification measures (posterior variance) that are explicitly provided and can assist in decision making (costs and risks of misclassification). Furthermore, the methodology allows for inclusion of prior knowledge, which is vital for the spatio-temporal migration patterns of species of birds and their prior probability of being present and hence detected by such flight call systems.

Finally, the multiple kernel learning capabilities of the method provide a promising tool for integrating other sources of information for flight call detection such as spatio-temporal (time stamp and station that flight call is detected), radar traces, and even weather based information.

The approach is a kernel formulation of a generalized linear model (GLM) with an explicit multiclass probit likelihood that can handle multiple classes with a single underlying model. The likelihood is given by:

$$P\Big(t_i = m|\mathbf{W}, \mathbf{k}_i^{\boldsymbol{\beta\theta}}\Big) = \mathcal{E}_{p(u)} \left\{ \prod_{l \neq m} \Phi\Big(u + (\mathbf{w}_m - \mathbf{w}_l)\, \mathbf{k}_i^{\boldsymbol{\beta\theta}}\Big) \right\}, \tag{6}$$

where $\mathbf{w}$ are the regression coefficients, $\boldsymbol{\beta}$ specifies the kernel combination parameters for MKL that are not employed in this work, $\boldsymbol{\Phi}$ is the normal cdf function, m, l denote classes, and $u \sim \mathcal{N}(0, 1)$. Approximate Bayesian inference is performed via a variational treatment for the posterior distribution which is described in detail in [10].

IV. Pseudo-Code

The pseudo-code implementation of the proposed classification method is listed in Algorithm 1. Standard cross-validation procedures need to be adopted as usual.

Algorithm 1 Probabilistic DTW kernel classifier

```
Initialization and pre-processing
# STFT Spectrograms
for all X ∈ ℜ^{N×D_{x_i}} (Flight call sequences) do
    S_{x_i} ← spectrogram(x_i)
end for
# Dissimilarity (Cost) Matrix and DP
for i = 1 to N do
    for j = i to N do
        D^{ij} ←
        c*_{x_i} ←
    end for
end for
# Create the DTW Kernel
for i = 1 to N do
    for j = i to N do
        k(x_i, x_j) ←
    end for
end for
# Comment
P(t_i = m|W, k_i) ←
```

V. Experimental Setup and Feature Construction

The performance of the proposed approach is compared to several other classification methods and feature constructions on real world flight call datasets. All the reported results are for 10 replications of 10-fold Cross-Validation (10×10CV) unless otherwise stated and *baseline* denotes performance by assignment to largest populated class.

A. Data

Two flight call datasets are presented in this work. All the samples were collected via extraction from recording stations, recording captive birds, or recording and labeling via direct observation. The primary dataset consists of 5 classes (species of bird) with 40 samples (flight calls) from each class for a total of 200 samples. The five bird species recorded for the primary dataset are listed in Table II.

TABLE II
Bird Species

Common Name	Scientific Name	No. Samples
Magnolia Warbler	Dendroica magnoliae	40
Nashville Warbler	Vermivora ruficapilla	40
Chestnut-sided Warbler	Dendroica pennsylvanica	40
American Redstart	Setophaga ruticlla	40
Yellow-rumped Warbler	Dendroica coronata	40

The data is digitized and stored as .wav files. The calls range from between 1916 and 6037 features (sampling every 4.5×10^{-5} seconds); each call was padded with zeros to create a uniform length dataset. As an extension to the primary data set, an auxiliary data set consisting of 42 classes (species information available from our data repository), 1180 samples, and 15075 features. Samples per class range from 10 to 40, and call lengths range from 1175 to 15075. *Both datasets are publicly available at www.cis.cornell.edu/ics/projects.php*. One of the challenges of the primary data set is that most of the samples are warblers, which tend to have structurally similar flight calls. The auxiliary data set contains a larger variety of species, and therefore greatly varying samples of flight calls.

B. Global Features

When classifying flight calls, humans tend to look for a few global features of the calls. These attributes were extracted as an alternative feature construction to the proposed $\text{DTW}_{\text{global}}$ features and kernel: average energy value of the call, length of the flight call, maximum amplitude of the call and number of peaks.

C. Down Sampling

Down sampling, by averaging the signal over fixed-length intervals, can capture the general structure of the signal while reducing the total number of features. Frequency based feature extractions are based on standard Fast Fourier Transformations (FFT) with best performing settings.

D. Competing Classifiers

In order to examine the performance of VBpMKL and the various feature extraction methods, classification was performed with several other Weka classifiers [22] and an SVM Multiclass implementation [23]. The classifiers employed are given below with their implementation details and cover a range of popular classification techniques.

VI. Results

A. Primary Data Set

The technique described in this paper produces excellent results with a 97.6% average percent correct classification for the primary dataset under consideration and promising results with the 45 class flight call data. Furthermore, the underlying $\text{DTW}_{\text{global}}$ feature extraction method (kernelized or not) proves to be very discriminative for most classifiers employed. For Tables IV to VI we report only the mean of the 10×10CV as

TABLE III
CLASSIFIERS USED FOR COMPARISON

Classifier Type	*Implementation*	*Reference*
C4.5 Decision Tree	J48	[24]
Nearest Neighbors	Kstar	[25]
Bayesian Network	BayesNet	[26]
Regression	Simple Logistic	[27], [28]
Decision Table	Decision Table	[29]
Ensemble Decision Tree	Random Forest	[30]
Boosting Decision Stumps	Logit Boost	[31]
Ensemble Decision Tree	Rotation Forest	[32]
Support Vector Machine	$SVM^{multiclass}$	[23]

TABLE IV
AMPLITUDES AND DOWN SAMPLING, CORRECT CLASSIFICATION

Classifier	10	25	50	100	250	Raw
J48	57.5	**66.0**	61.0	60.0	53.0	55.0
Kstar	60.0	**67.5**	66.5	65.0	66.5	20.0
BayesNet	46.5	52.5	55.0	49.0	50.5	**67.5**
Simple Logistic	54.5	59.5	63.0	58.5	**64.5**	46.0
Decision Table	46.0	54.5	**52.5**	52.0	47.0	**52.5**
Random Forest	60.0	64.5	67.0	**73.0**	66.5	58.0
Logit Boost	58.0	59.0	62.5	60.0	55.5	**66.5**
Rotation Forest	61.5	67.0	68.5	66.5	66	**69.5**

TABLE V
FREQUENCY (FFT) AND DOWN SAMPLING, CORRECT CLASSIFICATION

Classifier	10	25	50	100	All
J48	73.35	**74.65**	73.65	73.65	75.2
Kstar	78.85	81.9	**83.05**	84.2	19.25
BayesNet	67.15	70.6	70.6	**72.2**	74.1
Simple Logistic	70.3	74.35	77.45	75.0	**84.4**
Decision Table	62.5	66.1	64.2	64.5	**67.85**
Random Forest	82.05	83.05	83.2	**83.55**	80.6
Logit Boost	80.25	80.9	**82.1**	80.75	82.0
Rotation Forest	83.35	84.9	85.6	84.8	**86.7**

TABLE VI
GLOBAL FEATURES, CORRECT CLASSIFICATION

Classifier	10	25	50	100	250
J48	62.5	**71.0**	63.5	60	61
Kstar	62.5	66.5	**70.5**	69.5	69.0
BayesNet	53.5	**56.5**	55.0	51.5	51.0
Simple Logistic	58.0	61.5	**63.0**	43.0	**63.0**
Decision Table	49.5	**55**	53.5	51.0	45.5
Random Forest	69.0	**74.0**	69.5	68.5	66.5
Logit Boost	64.5	65.5	**69.5**	64.0	59.5
Rotation Forest	70.5	**74.5**	70.5	72.5	70.5

the feature extraction methods are underperforming compared to the proposed DTW based constructions.

In Table IV we report the recognition rates with varying levels of down sampling and conversion to amplitudes. Each column shows the results of different levels of down sampling (varying the number of bins within which averaging takes place). The column headers indicate the length of the feature vector after down sampling occurs (e.g. 25 indicates a feature vector of length 25) and the *Raw* column shows the results when applying the various classifiers on the original data. Bold fonts indicate best performance of a classifier within specific feature extraction.

Table V presents results after FFT has been performed on either the original signal ("All") or in segmented versions (bins) with varying levels of down sampling and conversion to frequency. Again, the column headers indicate the length of the feature vector after down sampling occurs. The final alternative feature extraction in Table VI considers the aforementioned "global features" of the signals. Combining these features with FFT based extractions or amplitude information does not result in statistically significant improvements.

Table VII shows the performance of the proposed method and the best results for each classifier across feature extraction approaches for the primary dataset. The best window size for the DTW_{global} was found by grid-search and cross validation to be 450 with an associated bandwidth of 0.4. All methods perform best with the DTW_{global} feature construction and the adopted VBpMKL method achieves above 97% average recognition rate on the 5 class problem with the additional probabilistic benefits for prior knowledge inclusion, uncertainty quantification and data fusion. Considering the aforementioned recognition levels of human experts this is a significant step towards the development of automated flight call detection systems.

The second dataset that we consider has flight calls from a greater variety of species and hence poses a larger multi-class problem. In fact the inter and intra family structure of the flight calls can be exploited in a hierarchical manner that is ongoing work. In Table VIII we report preliminary recognition rates of VBpMKL with the proposed DTW_{global} kernel. We achieve an average of 74% accuracy which, considering the 42 different classes of flight calls in the data, is a very good classification level and a promising benchmark to improve upon.

VII. DISCUSSION AND FUTURE DIRECTIONS

The immediate extensions of this work are to deal with the streaming nature of the flight call sensor data and to integrate spatio-temporal and ecological (GIS) information into the model. Scalability becomes an important aspect of an online system, and sparsity can reduce the computational complexity of kernel based methods via only retaining a few representative samples (e.g. relevance vectors) [33]. In the larger picture, interesting computational problems come into play as (near) optimal sensor placements for flight calls becomes an exciting direction for research that can couple predictive models of species distributions to flight call detection; and human observation locations (e.g. eBird: http://ebird.org) to acoustic sensor placements.

The development of a successful feature construction and classification methodology for flight call detection is an important step towards the overall goal of understanding bird

TABLE VII
BEST RESULTS FOR EACH CLASSIFIER (BASELINE = 20%)

Classifier	*Feature Extraction Method*	10×10CV %
J48	$\text{DTW}_{\text{global}}$	87.1 ± 1.14
Kstar	$\text{DTW}_{\text{global}}$	**96.6 ± 0.65**
BayesNet	$\text{DTW}_{\text{global}}$	93.2 ± 0.27
Simple Logistic	$\text{DTW}_{\text{global}}$	94.9 ± 0.55
Decision Table	$\text{DTW}_{\text{global}}$	72.8 ± 3.82
Random Forest	$\text{DTW}_{\text{global}}$	93.2 ± 0.84
Logit Boost	$\text{DTW}_{\text{global}}$	91.7 ± 1.64
Rotation Forest	$\text{DTW}_{\text{global}}$	94.5 ± 1.06
$\text{SVM}^{\text{multiclass}}$	$\text{DTW}_{\text{global}}$ Kernel	95 ± 0.43
VBpMKL	$\text{DTW}_{\text{global}}$ Kernel	**97.6 ± 0.68**

TABLE VIII
VBPMKL & $\text{DTW}_{\text{GLOBAL}}$ ON AUXILIARY DATA SET (BASELINE = 3.4%)

Window Size	Bandwidth	5CV %
450	0.4	72.45 ± 12.39
450	0.05	74.07 ± 13.64
650	0.01	74.49 ± 12.50

migration. Increasing automation of the flight call analysis work flow for detection and classification will allow for the processing and reporting of increasing amounts of audio data in increasingly rapid and efficient analyses. This, in turn, will lead to more efficient use of trained experts' time in interpreting the acoustic record; and it follows that more timely and relevant analysis and interpretation of nocturnal migration is possible.

The proposed method, and specifically the $\text{DTW}_{\text{global}}$ kernel construction, appears to be generalizable to a larger class of sequence-based problems. The flexibility of being able to use different scoring functions within DTW routines is a further benefit for other applications where different scoring approaches can prove to be more beneficial.

VIII. COMPUTATIONAL COMPLEXITY

The DTW routine has complexity $\mathcal{O}\left(|\mathbf{x}_i||\mathbf{x}_j|\right)$ and the classifier a dominant term of $\mathcal{O}\left(CN^3\right)$ where $C, N, \mathbf{x}_i$ are the number of classes, samples and the length of sequence i. Both complexities can be improved via sparsity (sparsity inducing priors or regularization) and faster implementations.

IX. ACKNOWLEDGMENT

T.D. and C.G. acknowledge funding from NSF Expeditions in Computing grant on Computational Sustainability (Award Number 0832782). The authors would like to thank T.G. Dietterich and S. Kelling for feedback on drafts of the paper.

REFERENCES

[1] C. P. Gomes, "Computational sustainability," *The Bridge, National Academy of Engineering*, vol. 39, no. 4, 2009.

[2] T. G. Dietterich, "Machine learning in ecosystem informatics and sustainability," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence. Pasadena, Calif.: IJCAI*, 2009, pp. 8–13.

[3] D. Sheldon, M. Elmohamed, and D. Kozen, "Collective Inference on Markov Models for Modeling Bird Migration," in *NIPS*, vol. 20. MIT Press, 2007, pp. 1321–1328.

[4] J. Yu, W.-K. Wong, and R. Hutchinson, "Modeling experts and novices in citizen science data for species distribution modeling," in *ICDM*. IEEE, 2010.

[5] A. Farnsworth and R. W. Russell, "Monitoring flight calls of migrating birds from an oil platform in northern gulf of Mexico," *Journal of Field Ornithology*, vol. 78, pp. 279–289, 2007.

[6] S. Kelling, W. Hochachka, D. Fink, M. Riedewald, R. Caruana, G. Ballard, and G. Hooker, "Data-intensive science: a new paradigm for biodiversity studies," *BioScience*, vol. 59, no. 7, pp. 613–620, 2009.

[7] W. R. Evans and D. K. Mellinger, "Monitoring grassland birds in nocturnal migration," *Studies in Avian Biology*, vol. 19, no. 219-229, 1999.

[8] A. Farnsworth and I. J. Lovette, "Phylogenetic and ecological effects on interspecific variation in structurally simple avian vocalizations," *Biological Journal of the Linnean Society*, vol. 94, 2008.

[9] ——, "Evolution of nocturnal flight calls in migrating wood-warblers: apparent lack of morphological constraints," *Journal of Avian Biology*, vol. 36, pp. 337–347, 2005.

[10] T. Damoulas and M. A. Girolami, "Pattern recognition with a Bayesian kernel combination machine," *Pattern Recognition*, vol. 30, no. 1, pp. 46–54, 2009.

[11] H. Figueroa and D. Michael, "Opening the BARN, a bioacoustic resource network," Presented at the XXII Meeting of the International Bioacoustics Council, 2009.

[12] D. Mellinger, D. Gillespie, H. Figueroa, K. Stafford, and T. Yack, "Software for bioacoustic analysis of passive acoustic data." *The Journal of the Acoustical Society of America*, vol. 125, p. 2547, 2009.

[13] R. R. Graber and W. W. Cochran, "Evaluation of an aural record of nocturnal migration," *Wilson Bull.*, vol. 72, pp. 252–273, 1960.

[14] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoustic Speech Signal Processing*, vol. 26, pp. 43–49, 1978.

[15] M. Cuturi, J.-P. Vert, óystein Birkenes, and T. Matsui, "A kernel for time series based on global alignments," in *ICASSP*. IEEE, 2007.

[16] H. Shimodaira, K. ichi Noma, M. Nakai, and S. Sagayama, "Dynamic time-alignment kernel in support vector machine," in *NIPS*. MIT Press, 2002.

[17] H. Lei and B. Sun, "A study on the dynamic time warping in kernel machines," in *IEEE Conference on Signal-Image Techonologies and Internet-Based Systems*, 2007.

[18] T. Jebara, Y. Song, and K. Thadani, "Spectral clustering and embedding with hidden Markov models," in *ECML*. Springer, 2007, pp. 164–175.

[19] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in *NIPS*, 1999, pp. 487–493.

[20] D. Ellis, "Dynamic time warping (dtw) in matlab," 2003, available: http://www.ee.columbia.edu/ dpwe/resources/matlab/dtw/.

[21] L. Liao and W. S. Noble, "Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships," *Journal of Computational Biology*, vol. 10, no. 6, pp. 857–868, 2003.

[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[23] T. Joachims, "Multi-class support vector machine," Online, 2008, available: http://svmlight.joachims.org/svm_multiclass.html.

[24] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA.: Morgan Kaufmann Publishers, 1993.

[25] J. G. Cleary and L. E. Trigg, "K*: An instance-based learner using an entropic distance measure," *ICML*, pp. 108–114, 1995.

[26] R. R. Bouckaert, "Bayesian network classifiers in Weka for version 3-5-7," http://www.cs.waikato.ac.nz/ remco/weka.bn.pdf, May 2008.

[27] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," 2005.

[28] M. Sumner, E. Frank, and M. Hall, "Speeding up logistic model tree induction," in *PKDD*, 2005, pp. 675–683.

[29] R. Kohavi, "Power of decision tables," in *8th ECML*, 1995, pp. 174–189.

[30] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[31] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," Master's thesis, Stanford University, 1998.

[32] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Transactions PAMI*, vol. 28, no. 10, pp. 1619–1630, 2006.

[33] I. Psorakis, T. Damoulas, and M. A. Girolami, "Multiclass relevance vector machines: Sparsity and accuracy," *IEEE Transactions on Neural Networks*, 2010, in Press.