Modeling Complex Phenotypes: Generalized Linear Models Using Spectrogram Predictors of Animal Communication Signals

Scott H. Holan, Christopher K. Wikle, Laura E. Sullivan-Beckers, and Reginald B. Cocroft

1Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A.
2School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, U.S.A.
3Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211, U.S.A.

SUMMARY. A major goal of evolutionary biology is to understand the dynamics of natural selection within populations. The strength and direction of selection can be described by regressing relative fitness measurements on organismal traits of ecological significance. However, many important evolutionary characteristics of organisms are complex, and have correspondingly complex relationships to fitness. Secondary sexual characteristics such as mating displays are prime examples of complex traits with important consequences for reproductive success. Typically, researchers atomize sexual traits such as mating signals into a set of measurements including pitch and duration, in order to include them in a statistical analysis. However, these researcher-defined measurements are unlikely to capture all of the relevant phenotypic variation, especially when the sources of selection are incompletely known. In order to accommodate this complexity we propose a Bayesian dimension-reduced spectrogram generalized linear model that directly incorporates representations of the entire phenotype (one-dimensional acoustic signal) into the model as a predictor while accounting for multiple sources of uncertainty. The first stage of dimension reduction is achieved by treating the spectrogram as an “image” and finding its corresponding empirical orthogonal functions. Subsequently, further dimension reduction is accomplished through model selection using stochastic search variable selection. Thus, the model we develop characterizes key aspects of the acoustic signal that influence sexual selection while alleviating the need to extract higher-level signal traits a priori. This facet of our approach is fundamental and has the potential to provide additional biological insight, as is illustrated in our analysis.

KEY WORDS: Acoustic; Bayesian model averaging; Classification; Empirical orthogonal functions; Functional data; Mating calls; Nonstationary; Stochastic search variable selection.

1. Introduction

A major task of evolutionary biology is to describe patterns of phenotypic variation and understand their evolutionary implications (Björklund, 2003). In particular, studies of phenotypic selection—how fitness covaries with variation in traits and trait combinations—estimate the strength and direction of selection acting in natural populations (Kingsolver et al., 2001). One challenge in characterizing selection is that many phenotypes of evolutionary interest are complex, consisting of multiple subordinate traits (Swallow and Garland, 2005). Current approaches to studying the fitness consequences of complex traits begin by atomizing the phenotype into a set of measurements chosen by the researcher. The choice of measurements is often to some extent arbitrary, particularly when the sources of selection are unknown. Furthermore, it may be difficult to determine when enough measurements have been made to capture all of the relevant trait variation (Michelle-Olds and Shaw, 1987). Because our ability to correctly infer the nature of phenotypic selection is contingent on adequate trait descriptions at this first stage of analysis, the characterization of complex phenotypes is an important problem within evolutionary biology.

The issues arising from phenotypic complexity are particularly acute in studies of sexual selection, which favors elaborate displays and other phenotypes used in mate acquisition (Andersson, 1994). It is increasingly clear that sexual selection on males is a multivariate process, influenced by multiple traits in combination (Gerhardt and Huber, 2002; Brooks et al., 2005; Bentsen et al., 2006). However, it is difficult to fully characterize the variation in traits such as acoustic signals; for example, one study of sexual selection in frogs used 15 variables to describe a one-half-second mating call, and even then the authors acknowledged that some aspects of the call were omitted (Ryan and Rand, 2003). Other studies narrow the focus to a small subset of signal traits thought to be important (e.g., Rodriguez, Ramaswamy, and Cocroft, 2006) or rely on qualitative descriptions such as the number of different signal elements in a repertoire (Searcy, 1992).

These studies exemplify a major obstacle in relating variation in complex phenotypes to outcomes such as female choice, survival, or other measures of fitness. That is, some of the phenotypic variation is discarded by the investigator prior to the analysis. This filtering increases the likelihood that important traits are omitted from the analysis, which complicates the interpretation of measures of selection (Lande and Arnold, 1983; Michelle-Olds and Shaw, 1987) and may be especially problematic in the case of multivariate sexual selection (Bentsen et al., 2006).
The type of model we propose, regression of scalar response (e.g., mating success) on signal predictors, can be thought of as signal regression (see Reiss and Ogden, 2007, and the references therein) and is often referred to as a functional data analysis (Ramsey and Silverman, 2005). Specifically, we develop a functional generalized linear model (GLM; Ramsey and Silverman, 2005) that is novel in several respects. First, our model induces a dimension-reduced representation of the spectrogram that is especially useful for characterizing nonstationary bioacoustic signals (mating calls). In addition, the spectrogram provides a highly interpretable time-frequency representation of the nonstationary signal (see Gröchenig, 2001, and references therein) that can be viewed as an image with its x-axis corresponding to time and y-axis corresponding to frequency. However, such a representation is high dimensional and difficult to incorporate into a complex statistical model. To reduce the dimension, we treat each spectrogram as an “image” and decompose the image using a Karhunen-Loève representation (Wikle and Cressie, 1999; Ramsey and Silverman, 2005).

In general, the empirical version of the Karhunen-Loève decomposition, often referred to as empirical orthogonal functions (EOFs) in spatial statistics, can be considered a form of functional principal component regression (FPCR; Ramsey and Silverman, 2005; Reiss and Ogden, 2007). However, rather than merely choosing the k-components in the customary manner, the components with the largest variance, we choose our components via a two-stage procedure. First, we achieve a substantial dimension reduction by considering a low-rank EOF representation. Then, given this representation the final model is selected through stochastic search variable selection (SSVS; George and McCulloch, 1993). In this sense, our variable selection approach is more similar, in principal, to functional partial least squares (FPLS), where the strategy is to choose components that are most relevant to predicting the outcome (see Reiss and Ogden, 2007, and references therein).

In an example related to our work, Wang, Ray, and Mallick (2007) develop a unified hierarchical model that encompasses wavelet-based function estimation and a logistic modeling framework. However, our approach differs from Wang et al. (2007) in several respects. First, prior to implementing our variable selection procedure, we apply a dimension reduction step in which we project the time-frequency decomposition onto a dimension-reduced set of EOFs (Wikle and Cressie, 1999; Banerjee, Carlin, and Gelfand, 2004, p. 257). Additionally, we include biologically relevant covariates and allow them to interact with the signal predictor (mating call) during the final variable selection stage. More importantly, the results of our approach provide an easily interpretable image that biologists can use to determine which aspects of the signal are important. Furthermore, the models we propose for classification take advantage of Bayesian model averaging (Hoeting et al., 1999). As a result, we reduce the underestimation of uncertainty at the model-selection stage as well as achieve better classification. Finally, when conducting classification, we present a distribution of probabilities of group membership, which provides a measure of uncertainty associated with group membership assignment. Of course, this aspect is a natural consequence of the Bayesian approach (Gelman et al., 2003); nonetheless, in the case of borderline group assignments, this aspect proves to be especially informative.

Our approach represents the first attempt at using dimension-reduced bioacoustic predictors for measuring sexual selection. While it is becoming more common to consider spectrogram representations for bioacoustic signals (see, for example, Valente et al., 2007, and the references therein), usually this is done on a subject-by-subject basis and the inclusion of spectrogram representations in a more general modeling framework is still lacking. Thus, our motivating application demonstrates a unique approach to animal communication modeling through the development of Bayesian hierarchical models with dimension-reduced spectrogram predictors.

The remainder of this article is organized as follows. Section 2 describes the motivating data—modeling mating success in an insect that uses acoustic signals to attract mates. Section 3 describes our statistical modeling approach for this problem including dimension-reduced spectrograms. Section 4 discusses specific modeling considerations and implementation details. The details of an extensive simulation are presented in Section 5. Section 6 provides results from an analysis of our motivating data using a time-frequency spectrogram GLM that classifies and characterizes insect mating success as a function of vibrational signaling and other relevant fitness covariates. A concluding discussion is provided in Section 7. Finally, supplementary material is included in the Web Appendix.

**2. Phenotypic Selection Application**

Measuring sexual selection on signals requires a study organism for which we can easily obtain both mating signals and a measure of reproductive success for a large number of individuals. The empirical system of signal evolution we analyze meets both of these criteria. The study animals are small plant-feeding insects known as treehoppers (Hemiptera: Membracidae). Males compete for mates using plant-borne vibrational signals (Sattman and Cocroft, 2003; Rodríguez, Sullivan, and Cocroft, 2004; Cocroft and McNett, 2006), a widespread communication modality in insects (Cocroft and Rodríguez, 2005). Females choose males on the basis of variation in their vibrational signals, and comparative evidence suggests that female mate choice is an important agent of signal evolution (Rodriguez et al., 2006). Furthermore, divergence in mating signals is important in speciation in the clade containing the study species, so understanding the nature of selection on signals is important for understanding the process of diversification in this group (Cocroft, Rodríguez, and Hunt, 2008).

The approach we propose is well suited for investigating selection in the species being studied (Enchenopa binotata Ptelea). The overall strategy is to consider recorded vibrational signals of males as predictors of mating success as measured in outdoor experimental populations; a typical signal is displayed in Figure 1. There are several aspects of the insect’s biology that are important for our analysis. First, there is significant repeatability of signal traits over the duration of the mating season (Sattman and Cocroft, 2003). Second, adults live for only one breeding season and females mate
only once, so the number of matings a male obtains during the breeding season is a reliable index of his lifetime reproductive success (Wood, 1993). In addition, individuals will readily colonize host plants in outdoor enclosures, allowing documentation of the mating history of marked individuals under conditions that closely mimic those of natural populations. This experimental design removes the confounding effect of correlations among environmental and phenotypic variation (Michelle-Olds and Shaw, 1987).

Our data come from a two-year experiment (though in our analysis we only consider 1 year) in which male *E. binotata* Ptelea had their signals recorded and mass measured prior to introducing them into outdoor cages one week preceding the start of mating. The sex ratio in the cages was 1:1, matching that of the source population (Sullivan-Beckers, 2008). Typically, only 5–10% of the females mated per day, and thus, over time there was an opportunity for individual males to attract and mate with multiple females.

Specifically, *N* = 160 males and females were marked and established in 10 outdoor enclosures, each simulating a natural population. Over the course of the experiment, individual mating behavior was recorded in 90-minute intervals during daylight hours (see Sullivan-Beckers, 2008, for a comprehensive discussion). Previous analysis of selection on these data, using standard GLM methodology (McCullagh and Nelder, 1989) on researcher-chosen measurements of male signals, identified several signal traits as targets of selection (Sullivan-Beckers, 2008).

Although the standard measurement approach used by Sullivan-Beckers (2008) has yielded insights into sexual selection on mating signals in *Enchenopa* treehoppers, these measurements are unlikely to capture all of the signal variation relevant to female choice. In fact, additional important variation is suggested by the examination of female responses to playback: although computer-generated signals based on the population average for five signal measurements reliably invoke responses, natural, recorded signals evoke more responses (R. Rodriguez and R. Cocroft, unpublished manuscript). Accordingly our approach incorporates the dimension-reduced time-frequency representation (discussed in Section 3) of the signal that frees the practitioner from choosing measurements a priori; thus we can potentially gain deeper insights into the nature of selection on the signal phenotype.
3. Statistical Modeling

3.1 Functional Signal Representation

Time-frequency analysis is an important component of signal analysis. In many cases, the global Fourier transform is of little interest when analyzing the spectrum of a long time series. In particular, nonstationary signals, such as bioacoustic signals, require a notion of frequency analysis that is local in time.

Our method takes advantage of the time-frequency representation of the signal. Specifically, we consider the spectrogram associated with each nonstationary signal. For example, Figure 1 displays a representative acoustic signal (mating call) and its corresponding spectrogram.

The spectrogram relies on a local time-frequency representation known as the short-time Fourier transform (STFT) and provides information about local properties of a signal $f$ (where, in this case, $f$ denotes the acoustic signal). Specifically, in order to produce a “local frequency spectrum,” $f$ is restricted to an interval and the Fourier transform is computed. However, the functions resulting from this segmentation are not periodic. Consequently, the Fourier transform will interpret these boundary jumps as discontinuities or as sharp variations in the signal, leading to large Fourier coefficients at high frequencies. These undesirable artifacts can be alleviated by introducing the concept of windowing. Rather than localizing the signal $f$ by a rectangular function (the case of no smoothing), the signal $f$ is localized using a smooth cutoff function as a “window,” which approaches one near the origin and decays toward zero at the boundaries. Several popular windows have been proposed for this purpose including the Hamming, Hanning, Bartlett, and Kaiser windows. See Feichtinger and Strohmer (1997) for further discussion. The window used in our analysis, the Hamming window, has the following form:

$$w_n = \begin{cases} 0.54 - 0.46 \cos(2\pi n/M); & 0 \leq n < M, \\ 0; & \text{otherwise}, \end{cases}$$

where $M$ is the length of the impulse–response function. The particular forms for the other windows mentioned and a comprehensive discussion of their usage can be found in Oppenheim, Schafer, and Buck (1999).

Let $g \neq 0$ denote a fixed window function. Then, for time $t$ and frequencies $\omega$ such that $x, \omega \in \mathbb{R}^d$, the STFT of a function $f$ with respect to $g$ is defined by

$$V_f(x, \omega) = \int_{\mathbb{R}^d} f(t) g(t-x) e^{-2\pi i \omega t} dt,$$

where $i = \sqrt{-1}$ and $g(-x) = \overline{g(x)}$ denote complex conjugation (cf. Definition 3.1.1, p. 37, Gröchenig, 2001). Note that in this context, (1) can be thought of as the Fourier transform of a segment of $f$ centered at time $x$ evaluated at frequency $\omega$ (assuming $g$ is compactly supported). Further, let $g \in L^2(\mathbb{R}^d)$ denote a window function such that $\|g\|_2 = 1$. Then, the spectrogram of $f$ with respect to $g$ is defined by $\text{SPEC}_g f(x, \omega) = |V_f(x, \omega)|^2$ (cf. Definition 4.1.1, p. 60, Gröchenig, 2001).

In our biological problem of interest (see Section 2), the time-frequency representation of the signal is extremely high dimensional. For example, the spectrogram representation shown in Figure 1 consists of 87 time points by 129 frequencies and thus produces 11,223 possible time-frequency covariates. Although it is possible to reduce the dimension by downsampling the signal in time and/or by considering fewer frequencies, such techniques might potentially eliminate critical signal components. However, the time-frequency “pixels” are not independent, as they result from covariance present in the original signal and from a smoothing procedure that uses overlapping windows. In fact, the spectrogram presents coherent “patterns” that should be accounted for. These patterns are analogous to what ones sees in spatial images.

To address the dependence in the spectrogram, we treat each spectrogram as an “image” and decompose it as a Karhunen-Loève representation (Ramsey and Silverman, 2005). In practice, we vectorize the spectrogram image and find its EOFs as is common in spatial statistics (Wikle and Cressie, 1999). Let $w_{s,i}$ denote the vectorized spectrogram. Then,

$$w_{s,i} = \Psi \tilde{x}_{s,i} = \sum_{j=1}^{N} \psi_j x_{s,ij},$$

where $\Psi = [\psi_1, \psi_2, \ldots, \psi_N]$ are the EOFs (eigenvectors from the empirical covariance matrix of the vectorized spectrograms), and $\tilde{x}_{s,i} = [x_{s,1}, x_{s,2}, \ldots, x_{s,N}]'$ are the corresponding spectral coefficients (i.e., principal components) given by $\tilde{x}_{s,i} = \tilde{\Psi} w_{s,i}$. Although in our application we make use of EOFs, other basis functions could be used in (2) as the critical aspect here is “lossless” dimension reduction.

It is important to note that although the EOFs can be viewed as spatial images analogous to spectrograms, in general, these images lack the interpretability associated with the original spectrogram. As with principal components, the lack of interpretability arises from imposing orthogonality constraints in the eigenvector decomposition (e.g., Mardia, Kent, and Bibby, 1979). The only EOF that does not suffer from this lack of interpretability is the first EOF. Even though it is often illuminating to examine the first EOF, our approach primarily focuses on interpreting the reconstructed signal that results from our modeling approach. This distinction is crucial as the reconstructed signal is highly interpretable and “filters” out spectral content not believed to be driving mating success.

3.2 Dimension-reduced Spectrogram GLMs

The basis functions in expansion (2) correspond to “spatial patterns” in the spectrogram. For example, Web Figure 6 shows two eigenvectors (EOFs) from the data set considered here. Note that these individual EOFs can be viewed as spatial “images” analogous to the spectrograms. The areas of relatively large magnitude in these images correspond to those portions of the spectrogram that load highly onto the given eigenvector. Thus, the associated expansion coefficients contain essential information that can be used as signal regressors in a GLM. As with principal components, the EOF coefficients are listed in a decreasing order of variance accounted for in the spectrogram. In this regard, there is a tremendous dimension reduction in considering the first $k$ ($k \ll N$) EOFs. For example, in the application considered in Section 2 we see that the first 5 EOFs account for 51% of the variation in the spectrogram whereas the first 50 coefficients account for 86%.
Although the time-frequency decomposition that we have introduced provides dimension reduction in terms of accounting for variation, there is no reason that the coefficients associated with EOFs that account for more variance should be the important coefficients in terms of measuring sexual selection on signals. Thus, we require a method of variable selection capable of identifying the important EOFs in terms of predicting mating success. We therefore propose a Bayesian hierarchical model and employ Bayesian SSVS as presented in George and McCulloch (1993, 1997).

In general, suppose we have a binary response variable \( Y_i = \{0, 1\} \) (where, in our case, 1 denotes mating success and 0 denotes no mating success) and covariates \( x_i \) for the \( i \)th subject. Additionally, as in Albert and Chib (1993), define a continuous latent variable \( Z_i \) such that if \( Z_i \) is positive then \( Y_i = 1 \), otherwise \( Y_i = 0 \), where \( Z_i \) is related to the covariates \( x_i \) by a normal regression model. In other words, let

\[
Y_i = \begin{cases} 
1; & \text{if } Z_i > 0, \\
0; & \text{if } Z_i \leq 0.
\end{cases} \tag{3}
\]

with \( x_i \) and \( \beta \) both \( p \times 1 \). Note, this is equivalent to a probit model on a Bernoulli response (Albert and Chib, 1993). Although we provide explicit details in terms of a GLM having a probit link, straightforward modifications allow for the use of other link functions.

Next, we specify a SSVS prior for the components of \( \beta \). Specifically, for \( j = 1, \ldots, p \) we have

\[
\beta_j \mid \gamma_j \sim \gamma_j N(0, c_j \tau_j^2) + (1 - \gamma_j) N(0, \tau_j^2) \tag{4}
\]

where \( \gamma_j \sim \text{Bern}(\pi_j) \),

where \( c_j, \tau_j, \) and \( \pi_j \) are all fixed hyperparameters and \( \gamma_j = 1 \) indicates that the \( j \)th variable is included in the model.

To implement the SSVS sampler, several choices of hyperparameters must be made. Based on the guidance given by George and McCulloch (1993), we want \( \tau_j \) to be small (yet positive) so that when \( \gamma_j = 0 \) it would be reasonable to estimate \( \beta_j \) by something close to zero. Similarly, we would like \( c_j \) to be large (and greater than 1) so that we would expect a nonzero \( \beta_j \) in the model when \( \gamma_j = 1 \). Note that \( \pi_j \) can be viewed as the prior probability that \( \beta_j \) should be in the model. George and McCulloch (1997) discuss various approaches for selecting these parameters.

In our problem \( x_i \equiv [x_{i,1}, \ldots, x_{i,m}]' \) and \( \beta = (\beta_1, \ldots, \beta_m)' \) where the subscripts “b” and “s” correspond to biological and signal covariates, respectively, and \( x_{i,s} = [x_{i,1}, \ldots, x_{i,2}, \ldots, x_{i,a}]' \). In this case, \( x_{i,b} = \Psi w_{i,b} \) where \( w_{i,b} \) is the vectorized spectrogram and \( \Psi = [\Psi_1, \ldots, \Psi_s] \) are the first \( k \) EOFs (\( k \leq N \)) obtained from \( \hat{W} = \hat{\Psi} \hat{\Lambda} \hat{\Psi} \). It should be noted that in our context the spectrogram has not been estimated within the Bayesian framework. Estimation of the spectrogram within the Bayesian framework would make the approach fully Bayesian but this aspect currently requires further development and investigation.

The model, as it is proposed here, yields conjugate full conditionals for all of the parameters. Therefore, estimation proceeds by a straightforward application of Gibbs sampling (see, for example, Gelman et al., 2003). The exact forms for all of the full conditionals can be found in Web Appendix A.

4. Model Implementation

Initially, each of the \( N = 137 \) signals was of length 37,912 after registration. Although the original experiment consisted of \( N = 160 \) insects, males that did not live through the first day were removed from the analysis, see Sullivan-Beckers (2008) for a comprehensive discussion. These signals were downsampled every eighth observation (creating series of length 4739) and the spectrogram calculation was based on this downsampled series. There was no appreciable difference between the original signal and its downsampled equivalent since the measurement process oversamples the signal in this application. Specifically, the highest frequency in the signal is less than 4–5 kHz while the sampling rate was 44.1 kHz.

The spectrogram for each signal (subject) was calculated using the spectrogram (STFT) function in the Matlab® signal processing toolbox. By default, this procedure uses a moving Hamming window, a popular choice in speech processing (Oppenheim et al., 1999), chosen to be of length 64 in our case. The resulting power spectral density is then converted to decibels (dB) by taking 10 times the base-10 log of the modulus of the power spectral density and EOFs are calculated.

We conducted a sensitivity study on the downsampling and window length (window lengths of 16, 32, 64, and 128 were considered) in the spectrogram calculation and did not find any appreciable difference with regards to parameter estimation, model selection, or classification when considering no downsampling and downsampling up to every eighth observation using different window widths. Thus, we are comfortable that our results are robust to these specifications.

In addition to the signal covariates, we considered several nonsignal-related biological covariates for possible inclusion in the model. These variables were selected based on previous analyses (Sullivan-Beckers, 2008). In particular, we included variables related to the survival of the insect (number of days that the male survived during the experiment) and the insect’s weight (just prior to the start of the experiment). We standardized these variables and also included a quadratic term for each.

The modeling was split into two parts. First, the SSVS was performed where the possible variables include the first 50 EOF coefficients as well as the four biological covariates. Note that for sensitivity purposes we included different number of EOF components (25, 50, 100) and found no appreciable differences in the results. The results of an extensive sensitivity analysis can be found in Web Appendix B.

Several choices of hyperparameters must be made to implement the SSVS sampler (see Section 3). Since we are interested in classification, we felt it would be expedient to consider various possible values for these hyperparameters that improved classification. Thus, we performed sensitivity analyses, considering the factorial combinations when \( \tau = (0.1, 1, 1, 10), c = (0.1, 1, 10, 100), \pi = (0.2, .5, .8) \) (see Web Appendix B). Judging the classification results for the “model-averaged” SSVS and the “best” model (see the second part of the analysis discussed below), the parameters \( \tau = 0.1, c = 10, \pi = 0.5 \) yield good results overall. Although these
parameters could be tuned further, our experience is that such
tuning leads to minimal classification gains in our application.
The SSVS sampler results are based on 10,000 iterations with
a 1000 iteration burn-in.

In the first part of the analysis, we kept track of the classification
rate on the holdout sample based on mean probability
\(p_i\) over all iterations of the SSVS Markov chain Monte Carlo
(MCMC) run. The classification threshold was simply set to 1 (success)
if this probability was greater than or equal to 0.5, and was set to 0 (not success) if this probability was less than
0.5. This classification represents a “model averaging” over all
possible covariates, and accounts for their relative importance
through the stochastic search procedure.

The second part of the analysis was a classification based
on the variables selected in the SSVS. That is, we took the top
\(n_v\) variables based on the posterior mean of \(\gamma_i\) (the prob-
ability of a variable being in the model). Although this does not
guarantee that the chosen variables are necessarily “opt-
timal,” it does provide a good basis for the selection of vari-
bables for the model. For sensitivity evaluation, the number
\(n_v\) of selected variables was chosen to be 7 and 17. These
variables were then used in a Bayesian hierarchical probit
regression and the classification results and posterior distri-

5. Synthetic Signal Simulation Study

To illustrate the effectiveness of our approach, we conducted
a simulation study by creating a large sample of synthetic
signals whose traits have a defined relationship to fitness.
The synthetic signals were similar to those of the Enchenopa
treehoppers studied by Sullivan-Beckers (2008), except that
the synthetic signals consisted only of a frequency-modulated
whine, without the initial or terminal pulses present in natu-
ral insect signals. The amplitude envelope of all signals con-
sisted of a gradual rise in amplitude, reaching a peak at \(t =
80\%\) of the total signal duration, followed by a gradual fall
(Figure 2 (top)). Signals varied in beginning frequency, end-
ing frequency, and duration. The mean and standard devi-
ation were defined for both winners and losers (fitness = 1
vs. 0, respectively), and then a sample of winning and los-
ing signals were generated using a custom-written program
in Matlab\textsuperscript{®}. These signals were then analyzed in the same
manner as the natural insect signals (see Section 4 above;
note that aside from a fitness value of 0 or 1, there were no
biological covariates for the simulated signals).

We simulated two scenarios. In the first, individuals with
higher fitness had signals that were shorter and lower-pitched
than those with lower fitness. In the second, individuals with
high and low fitness had the same mean values, but individu-
als with values closer to the mean had higher fitness (this sim-
ulated the action of female preferences for trait values close to
the population mean, resulting in stabilizing selection). The
first scenario was chosen to illustrate a hypothetical relation-
ship between signal variation and fitness, while the second was
chosen because stabilizing selection on signal frequency is ex-
pected based on female preferences for male signal frequency
(Rodríguez et al., 2006).

In the first scenario, we obtained a perfect classification for
both the “model-averaged” approach and for the “best” five-
variable model. Similarly, in the second scenario, we achieved
a near-perfect classification for the “model-averaged” ap-
proach: 00 = 99\% (the case of correctly classifying 0—no
success) and 11 = 100\% (the case of correctly classifying
1—success). However, for the “best” five-variable model we
obtained: 00 = 72\% and 11 = 91\%. In both cases, the dif-
ferences between more-fit and less-fit individuals were clearly
recovered by the analysis, and in particular by the difference
spectrogram (the mean spectrogram for the winners minus
the mean spectrogram for the losers). A precise definition for
how this quantity is calculated can be found in Web Appendix
C. In the first simulation, the difference spectrogram shows a
region of high values indicating a shorter, lower-pitched signal,
with a nearby region of low values indicating a longer,
higher-pitched signal. In other words, the relationship be-
tween fitness and phenotypic variation in the simulated data
set (short, low-pitched signals are preferred) was revealed in
the difference spectrogram in a straightforward, easily inter-
pretiable way (Web Appendix Figure 2).

In the second simulation, the difference spectrogram shows a
region of high values indicating an average-frequency sig-
ual, with surrounding regions of low values indicating signals
with frequencies above or below the mean (Figure 2; see Web
Figure 5 for a color version). This analysis thus clearly re-
covers the signature of stabilizing selection present in the
simulated data set. An extensive summary of the simulation
results along with all associated (color) spectrograms can be
found in Web Appendix C.

6. Enchenopa Treehopper Results

Table 1 shows the top 10 variables chosen as most important
(based on the posterior probability of inclusion in the model,
\(i.e., \gamma_i\)) for the 7- and 17-variable SSVS. In addition, the
percentage of cross-validation model runs for which these vari-
ables were chosen is listed. First, we note that of the biological
covariates, survival (and its quadratic) is always selected, but
the weight variables are not favored. In addition, EOF 2, 13,
18, 8, 21, and 48 are in the top 10 most selected variables
for both the 7- and 17-variable models. It is interesting to
note that EOF 49 appears in the 7-variable model 62\% of
the time while in the 17-variable model it appears 79.6\% of
the time. However, since the 10th most selected variable in
the 17-variable model occurs 98.2\% of the time this variable
does not make the list of the top 10 variables.

In the case of the signal variables, it is interesting to ex-
amine the average spectrogram based on just the most highly
selected EOFs. Specifically, we consider the difference in the
average-reduced spectrograms for those subjects who had
Figure 2. Representative simulated mating signal (top), and image (bottom) representing the mean reduced spectrogram of the “winners” (= mated/high fitness), minus that for the “losers” (= unmated/low fitness) in the simulated data set. In this simulation, signals of winners had frequencies close to the mean, while those of losers had frequencies farther from the mean, with other variables held constant. This pattern of stabilizing selection is indicated by the band of high values (lighter tones) representing the average winner signal, surrounded above and below by bands of low values (darker tones) representing loser signals.

Table 1
Top 10 selected variables from the stochastic search variable selection cross-validation. The top 7 and 17 variables were chosen based on the posterior probability of the variable being in the model for each of the 500 cross-validation runs. The percentages shown are the percentages of those 500 runs in which this variable had one of the highest 7 and 17 posterior probabilities, respectively.

<table>
<thead>
<tr>
<th>Variable</th>
<th>% of models</th>
<th>Variable</th>
<th>% of models</th>
</tr>
</thead>
<tbody>
<tr>
<td>surv</td>
<td>100</td>
<td>surv</td>
<td>100</td>
</tr>
<tr>
<td>surv²</td>
<td>100</td>
<td>surv²</td>
<td>100</td>
</tr>
<tr>
<td>eof 2</td>
<td>100</td>
<td>eof 2</td>
<td>100</td>
</tr>
<tr>
<td>eof 13</td>
<td>100</td>
<td>eof 13</td>
<td>100</td>
</tr>
<tr>
<td>eof 18</td>
<td>100</td>
<td>eof 18</td>
<td>100</td>
</tr>
<tr>
<td>eof 8</td>
<td>87</td>
<td>eof 35</td>
<td>100</td>
</tr>
<tr>
<td>eof 49</td>
<td>62</td>
<td>eof 48</td>
<td>100</td>
</tr>
<tr>
<td>eof 21</td>
<td>39</td>
<td>eof 8</td>
<td>99.8</td>
</tr>
<tr>
<td>eof 48</td>
<td>3</td>
<td>eof 21</td>
<td>99.8</td>
</tr>
<tr>
<td>eof 5</td>
<td>2</td>
<td>eof 30</td>
<td>98.2</td>
</tr>
</tbody>
</table>

Several features are of note in this figure. First, winners had more energy in the broadband clicks at the start of the signal, as indicated by the region of relatively high intensity in the difference spectrogram in the region highlighted in box (a); this aspect of the signal has been ignored in previous studies but the current analysis strongly suggests that it is important for mating success. Second, if we examine the whine component of the signal, highlighted in box (b), winners had values closer to the mean, while losers had values farther from the
mean. This is qualitatively the same pattern as that shown in Figure 2 (bottom), though less marked than in the simulation, in which frequency was the only trait associated with fitness. Studies based on user-defined measurements also showed that more-fit individuals had frequencies closer to the average (Sullivan-Beckers, 2008). Finally, winners had more energy in the region highlighted in box (c), evidently reflecting the presence of more pulses at the end of the signal; a pattern also reflected in the studies based on user-defined measurements (Sullivan-Beckers, 2008). The current analysis thus illustrates both the strengths and weakness of the use of user-defined measurements: they can indeed identify important aspects of the signal, but important aspects of this complex trait were “filtered out” by the choice of measurements to include. In general, when studying animal communication signals, our approach provides an easily interpretable image that biologists can use to determine what aspects of the signal are important.

Table 2 shows the cross-validation classification results for the “model-averaged” SSVS probit model for 7 and 17 variables as well as the corresponding results from the “best” 7- and 17-variable models. These results are based on the three holdout samples over each of the 500 cross-validation runs (i.e., based on 1500 classifications). Note that the 7- and 17-variable models give similar results, with the 17-variable model performing better in the “model-averaged” classification and the 7-variable model performing better on the “best model” classification. We also note that it is apparently easier to classify mating failure (0) than mating success (1). For comparison, note that the within-sample classification probabilities (averaged over all cross-validation runs) are 85.6% for...
classifying from failure to failure, and 79.5% for classifying
from success to success. These results are for the 7-variable
model “model-averaged” scenario. (The other scenarios yield
similar results.)

We note that these results are based on a simple classifi-
cation rule with a cutoff at $p_{i} = 0.5$. One of the strengths of
the Bayesian approach is that one gets realistic measures of
the uncertainty in the probability of group membership. For
example, consider the posterior distribution of the classification
probabilities for subject 7 and 51. Both of these subjects
were successful maters, but both were classified as unsuccessful.
Figure 4 shows their posterior probabilities. In this case,
there is very little ambiguity concerning the classification of
subject 7, although there is a reasonable amount of mass at
probabilities above 0.5. For the most part, there is not enough
information in the selected biological and signal covariates to
correctly classify this subject the majority of the time. How-
ever, the posterior distribution for subject 51 is clearly more
ambiguous with regards to classification. In fact, the mode
of this distribution is greater than our cutoff of 0.5 (but the
mean and median are below 0.5).

In the “user defined” approach implemented in Sullivan-
Beckers (2008) a multinomial logistic regression model was fit
to the data. The model considered used number of matings as
the response variable along with treatment, cage(treatment),
survival, weight, calls per bout, whine length, interval, pulse
rate, frequency, and the square of all variables except treat-
ment and cage(treatment) as covariates. The only signifi-
cant covariates in the model were frequency, survival, and
survival$^2$. A direct comparison of this model with our model
is not really meaningful as the specific models and approaches
differ. Therefore, to compare our approach to the “user de-
efined” model of Sullivan-Beckers (2008) we fit a maximum-
likelihood probit regression model using survival, survival$^2$,
and the above signal variables. The within-sample classifica-
tion rate was 84% for the 00 case and 63% for the 11 case.
Therefore, we achieve slightly better classification for the 00
case (85.6%) and substantially better classification for the 11
case (79.5%).

7. Discussion
Acoustic signals are subject to strong sexual selection for
many species. Because of their role in mate choice, diversifi-
cation in acoustic signals is important for the process of spe-
ciation (Coyne and Orr, 2004). Thus, they represent complex
phenotypes of great behavioral and evolutionary importance.
Such signals are typically nonstationary and their representa-
tion as “covariates” in GLMs has been based on a scient-
ist’s perception of important features. In order to effectively
and efficiently represent the signal in an objective manner,
we demonstrate a functional approach for characterizing the
very high-dimensional spectrogram in terms of a relatively
small number of EOFs. Further, we illustrate that variable
selection is critical for choosing the most important expan-
sion coefficients as covariates.

When treating nonstationary acoustic signals as if they
were images, the problem of registration is important. That
is, given variation in the duration of the signal or in the timing
of its components, how should the signals be aligned along the
time axis with respect to each other? Registrations that do
not reflect biological reality can potentially lead to diminished
inferential power. As in functional data analysis, one solution
to this problem is to properly register the signals according
to specific signal features or some other objectively chosen
criterion. In our analysis, we have chosen to rely on an align-
ment based on biological considerations determined by ex-
pert biologists. To validate that our final analysis was robust
to this registration we considered several registrations (sup-
plied by the expert biologists) and obtained qualitatively sim-
ilar results. The systematic registration of bioacoustic signals
within a hierarchical modeling framework remains an open
problem.

Although specific aspects of our model and inference could
be obtained using a classical approach based on maximum
likelihood estimation, appealing to the Bayesian paradigm in
this setting provides several distinct advantages. Specifically,
through an application of Bayesian model averaging we are
able to boost classification performance while providing accu-
rate measures of uncertainty of the estimated probabilities of
classification (i.e., the estimated probabilities of group mem-
bership). These measures of uncertainty are crucial when in-
ference on individual subjects is of interest. That is, we can
determine with statistical significance whether individual in-
sects are successful or unsuccessful maters.

Additionally, taking a Bayesian approach allows us to ef-
iciently perform variable selection in the case where we are
interested in a single explanatory model. In particular, the
SVS priors on the EOF coefficients are equally suited for an
analysis of this type. The Bayesian framework proposed here
can be easily extended to handle more complex stochastic
systems in which outcomes can be predicted from nonstation-
ary signals. Although our motivating application is concerned
with understanding evolutionary consequences of phenotypic
variation, the methods developed here provide firm founda-
tions for using bioacoustic signals in predictive animal social
behavior models.

An important facet to our approach is the differenced spec-
trogram (Figure 3). By taking the difference between the
mean spectrograms for the successful and unsuccessful maters
it is especially apparent what features in the signal are driving
mating success. These features may suggest further hypothe-
ses to explore, potentially leading to important discoveries. In
fact, as a result of our analysis, we confirmed that both the
pitch of the frequency sweep and the duration of the signal
are important for mating success. However, more importantly,
our analysis suggested that a higher-frequency region near the
beginning of the signal may also be important, an aspect of
the signal that has been ignored in previous studies.

Finally, the method employed in our analysis could also
provide the basis for a new approach to the study of animal
communication signals. The current approach is to measure
a set of signal features of unknown relevance, then exper-
imentally explore those features to discover which may be
important to the communicating animals. Using the method
proposed here, it would be possible to target the important
features of the signal from the start, as determined by the an-
imals’ responses to the signals rather than human perception.
Then, if desired, the differenced spectrogram could guide an
informed choice of signal features, whose importance could
then be tested experimentally.
Figure 4. Posterior classification probabilities for subject 7 (top) and subject 51 (bottom). Note, both of these subjects were successful maters, but both were classified as unsuccessful. There is very little ambiguity concerning the classification of subject 7 whereas for subject 51 there is clearly more ambiguity with regards to classification.
8. Supplementary Materials
The Web Appendices, Figures, and Tables referenced in Sections 3, 4, and 5 are available under the Paper Information link at the Biometrics website http://www.biometrics.tibs.org.

ACKNOWLEDGEMENTS

The authors would like to thank a co-editor, associate editor, and three anonymous referees for their insightful comments that helped substantially improve this article. Wikle acknowledges the support of National Science Foundation grant ATM-0434213 and Cocroft acknowledges the support of National Science Foundation grant IBS-0318326.

REFERENCES


