Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/4/6/eaaq1084/DC1

Supplementary Materials for

Eavesdropping on the Arctic: Automated bioacoustics reveal dynamics in songbird breeding phenology

Ruth Y. Oliver, Daniel P. W. Ellis, Helen E. Chmura, Jesse S. Krause, Jonathan H. Pérez, Shannan K. Sweet, Laura Gough, John C. Wingfield, Natalie T. Boelman

Published 20 June 2018, *Sci. Adv.* **4**, eaaq1084 (2018)
DOI: 10.1126/sciadv.aaq1084

This PDF file includes:

Supplementary Text

Identifying songbird vocalizations and arrival date to breeding grounds

(1) Supervised classification. We selected the decision threshold with a resulting false positive rate of 30.3% and the true positive rate of 69.7%. The AUC was 0.78. The linear classifier described 65% of the variance ($R^2 = 0.65$) of the proportion of songbird vocalizations in recordings as determined by manual listening of the training dataset and had a root-mean-square-error of 0.19 (fig. S2). Discrepancies between the VAI and the listener scores may be caused by the sample size over which these values are calculated. In this case the VAI is the proportion of clips containing songbird vocalizations in a thirty-minute recording. In contrast, the listener scores are found only for 10 clips per recording (20 seconds).

Time series of the vocal activity index (VAI) for 2010-2014 at all sites show large daily and inter-annual variability in songbird vocalizations (Fig. 2 and S3-5).

Arrival date estimates were largely insensitive to the threshold used. Arrival dates estimated using thresholds of 30-70% of the maximum value in the VAI differed from the arrival dates estimated using a 50% threshold by less than a day on average (fig. S7).

(2) Unsupervised classification. The first five principal components, on average, explained 70% of the variance ($R^2 = 0.7$) in the VAI (fig. S2). The weighted sum of the first five principal components, using the linear model coefficients, and the VAI had a root-mean-square of 0.11 (fig. S2).

Time series of the weighted sum of the first five principal components for 2010-2014 at all sites show large daily and inter-annual variability in songbird vocalizations (Fig. 2 and S3-5).

The influence of environmental conditions and songbird phenology on the VAI

Statistically significant linear relationships ($p < 0.1$) were found between the VAI and environmental covariates for all twenty thirty-day recording periods as measured by a F-test. Snow cover, temperature, wind speed, atmospheric pressure, and precipitation were found to have statistically significant relationships with the VAI in the following number of cases: eight, eleven, six, seven, and one, respectively. The relationships between the VAI and environmental covariates over the thirty-day period, as measured the by the mean R^2 for a model type, were stronger for the significant multivariable models than the single variable models (multivariable: 0.52 +/- 0.06; snow cover: 0.33 +/- 0.07; temperature: 0.23 +/- 0.05; wind speed: 0.07 +/- 0.02; atmospheric pressure: 0.47 +/- 0.01; precipitation: 0.03+/- 0.01) (Fig. 3 and S6). In all cases mean R^2 values were higher, or equivalent, when considering the period prior to clutch initiation as compared to the entire thirty-day period. Conversely, mean R^2 values were lower when considering the period after clutch initiation, with the exception of models based on wind speed data (Fig. 3).

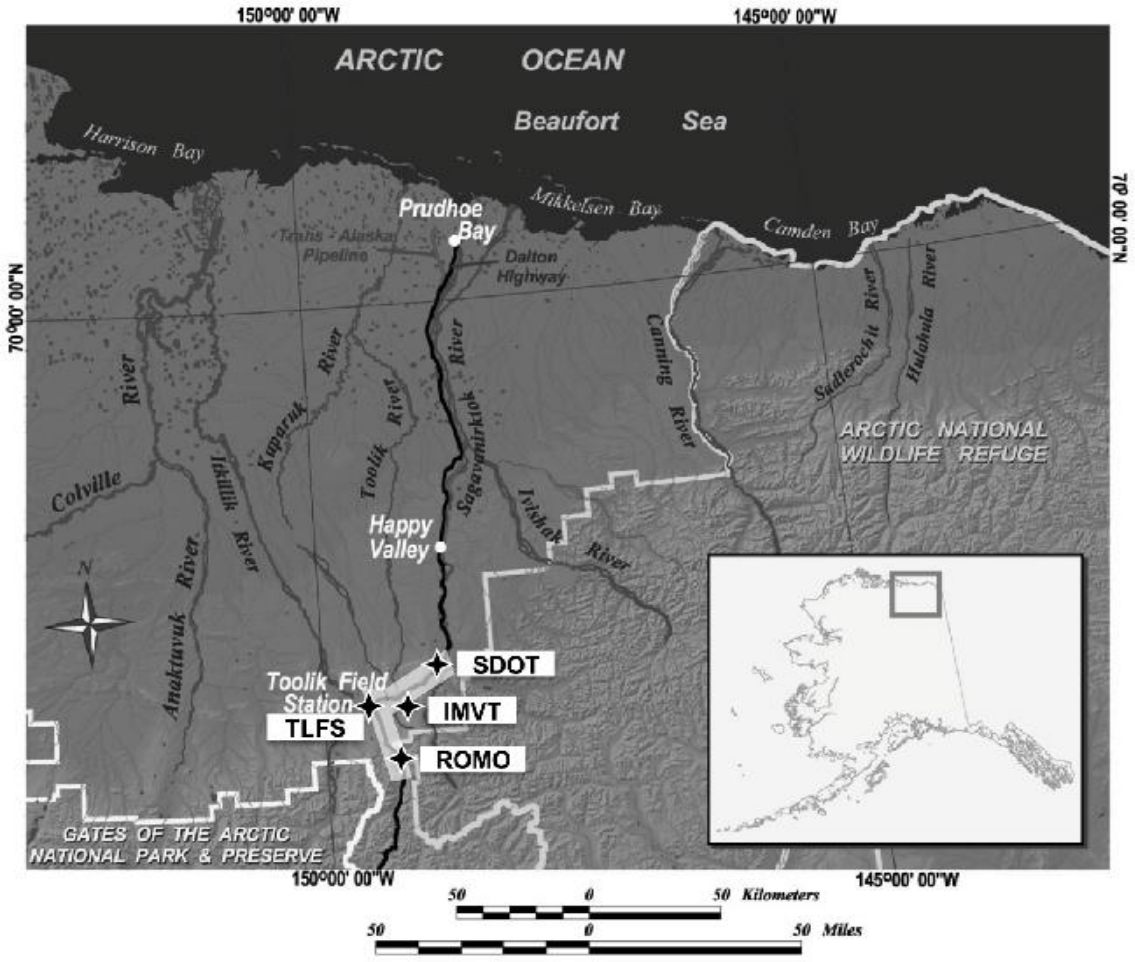

fig. S1. Map of Alaska (inset) and TLFS with approximate locations of acoustic recording units.

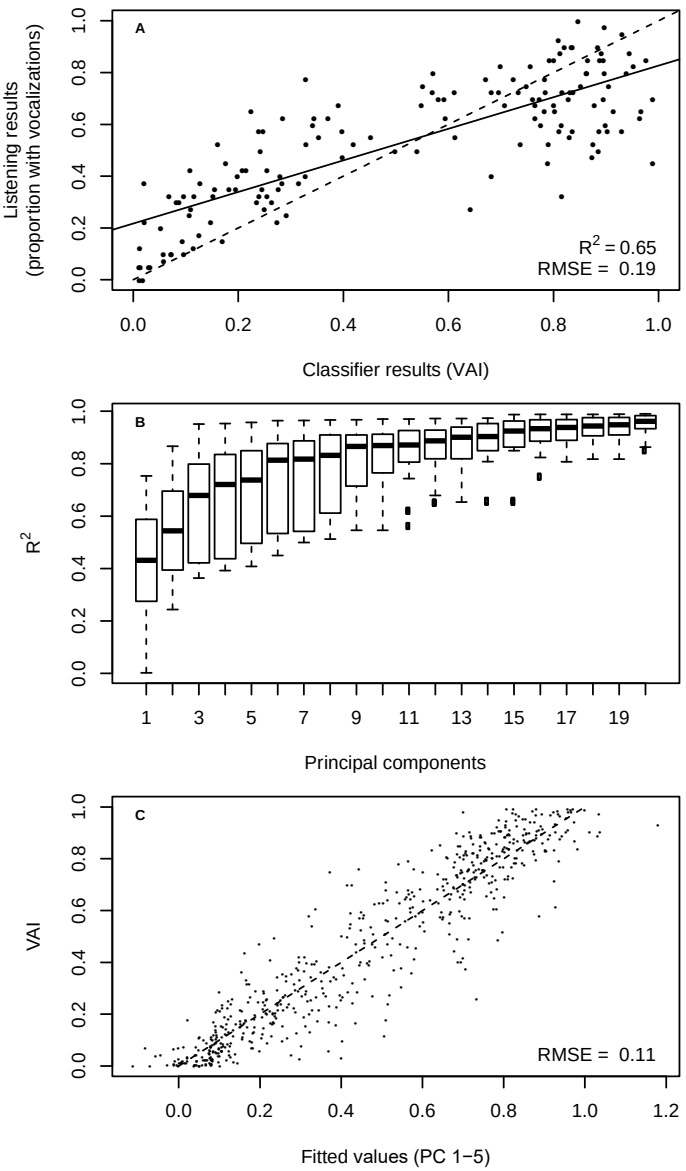

fig. S2. Performance of supervised and unsupervised classification approaches. (**A**) Performance of classification of acoustic data based on presence/absence of songbird vocalizations compared to listener scores. Each point represents the proportion of clips in a thirty-minute recording containing songbird vocalizations as determined by the linear classifier (VAI) and listener scores. The dashed line represents the one-to-one showing perfect agreement between the VAI and listener scores. The solid line is the least squares regression line. (**B**) Proportion of variance in the VAI explained (R^2) by the top principal components (from the unsupervised approach), measured by a linear regression, as a function of the number of components used. Linear models were built for each thirty-day recording period independently and principal components were added in succession. Mean R^2 values represent means across all study sites and years. (**C**) Comparison of the VAI to the weighted sum of the first five principal components fit by a multivariable linear model (B).

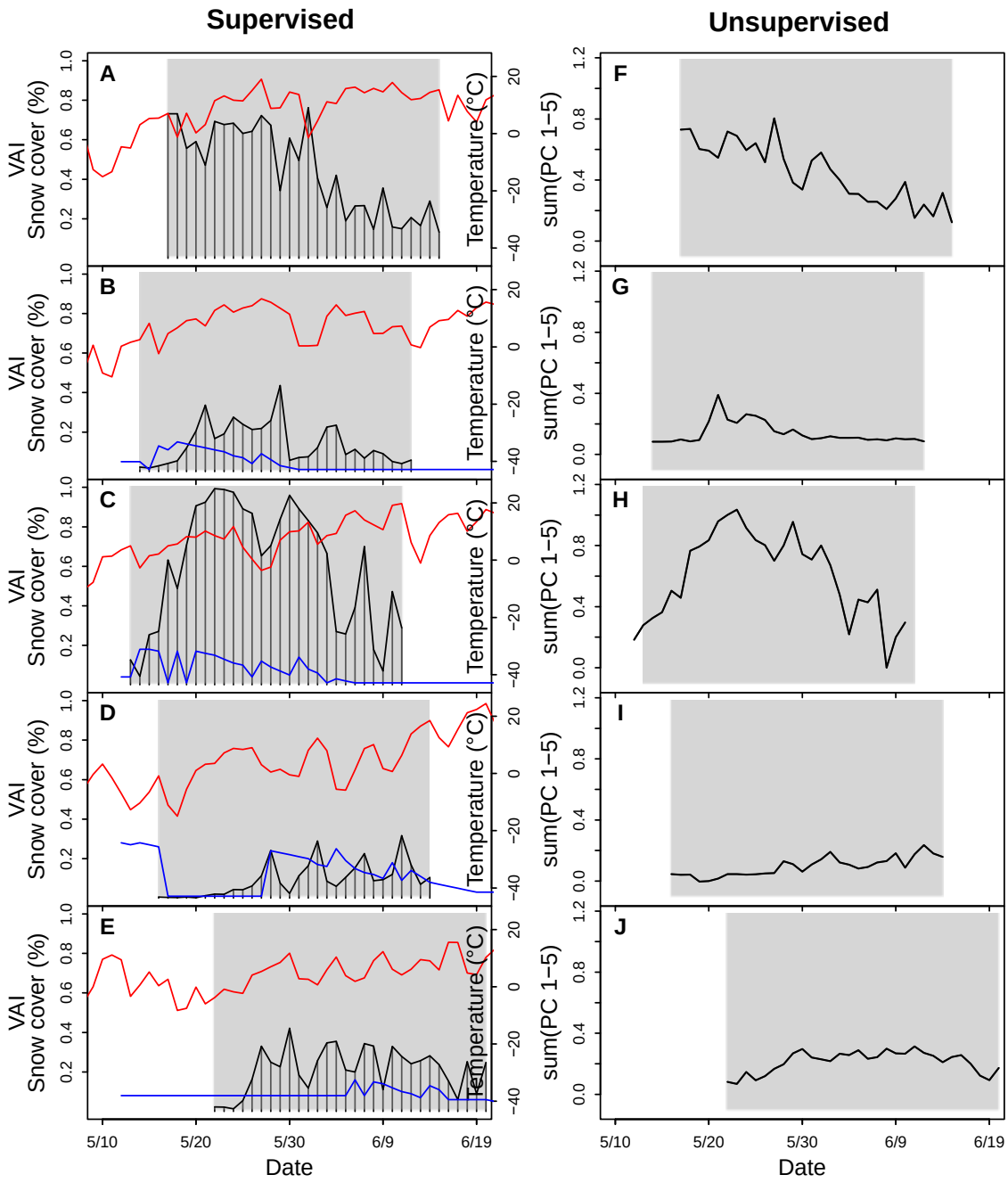

fig. S3. Songbird community vocal activity estimated by supervised and unsupervised approaches near IMVT. (**A**)-(**E**) Songbird daily Vocal Activity Index (VAI), snow cover (blue), and air temperature (red) near Imnaviat Creek (IMVT) between 2010-2014. (**F**)-(**J**) Weighted sums of the first five principal components at the same site and time. Grey boxes identify the available recording period for acoustic data.

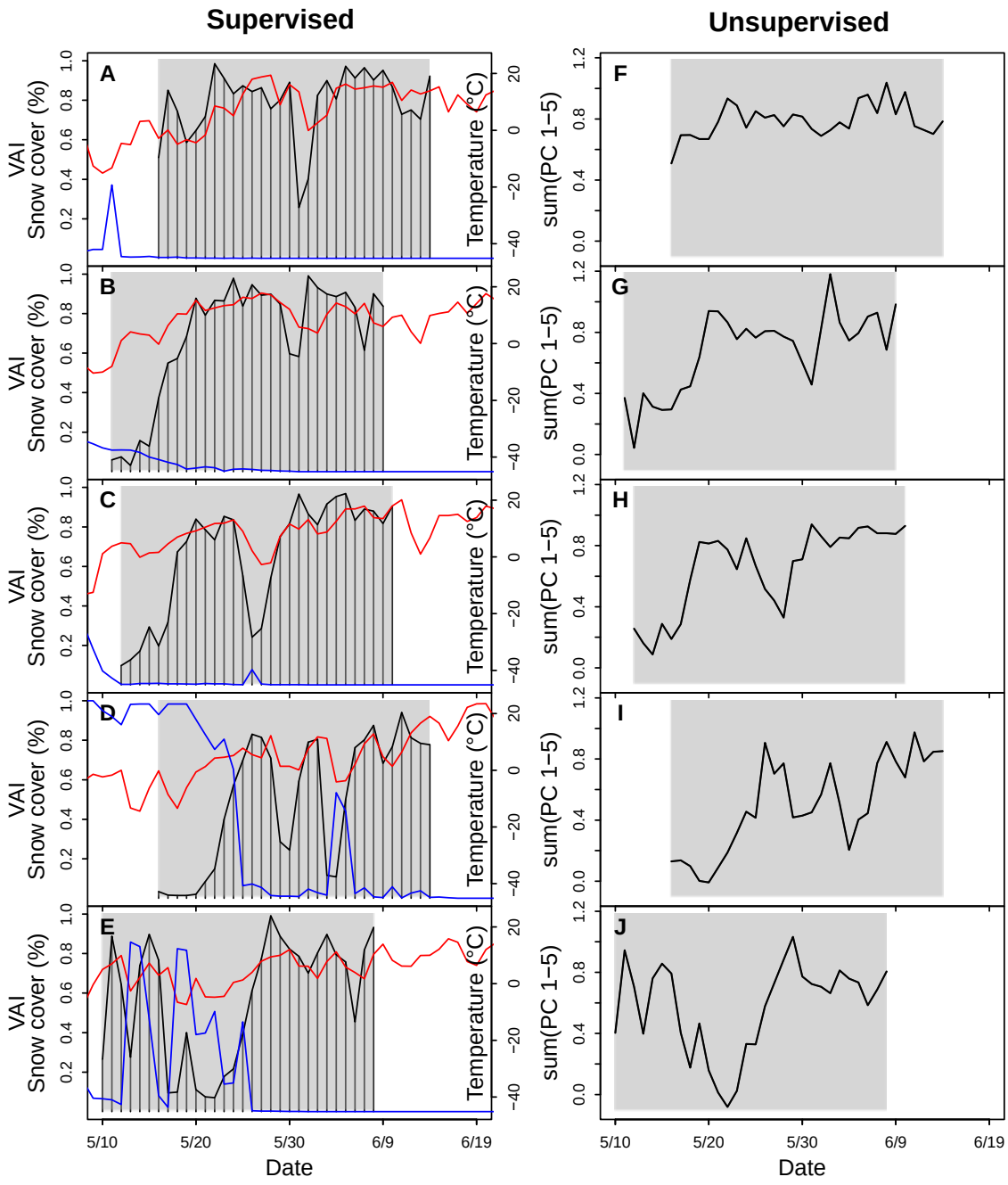

fig. S4. Songbird community vocal activity estimated by supervised and unsupervised approaches near ROMO. (**A**)-(**E**) Songbird daily Vocal Activity Index (VAI), snow cover (blue), and air temperature (red) near Roche Mountonee Creek (ROMO) between 2010-2014. (**F**)-(**J**) Weighted sums of the first five principal components at the same site and time. Grey boxes identify the available recording period for acoustic data.

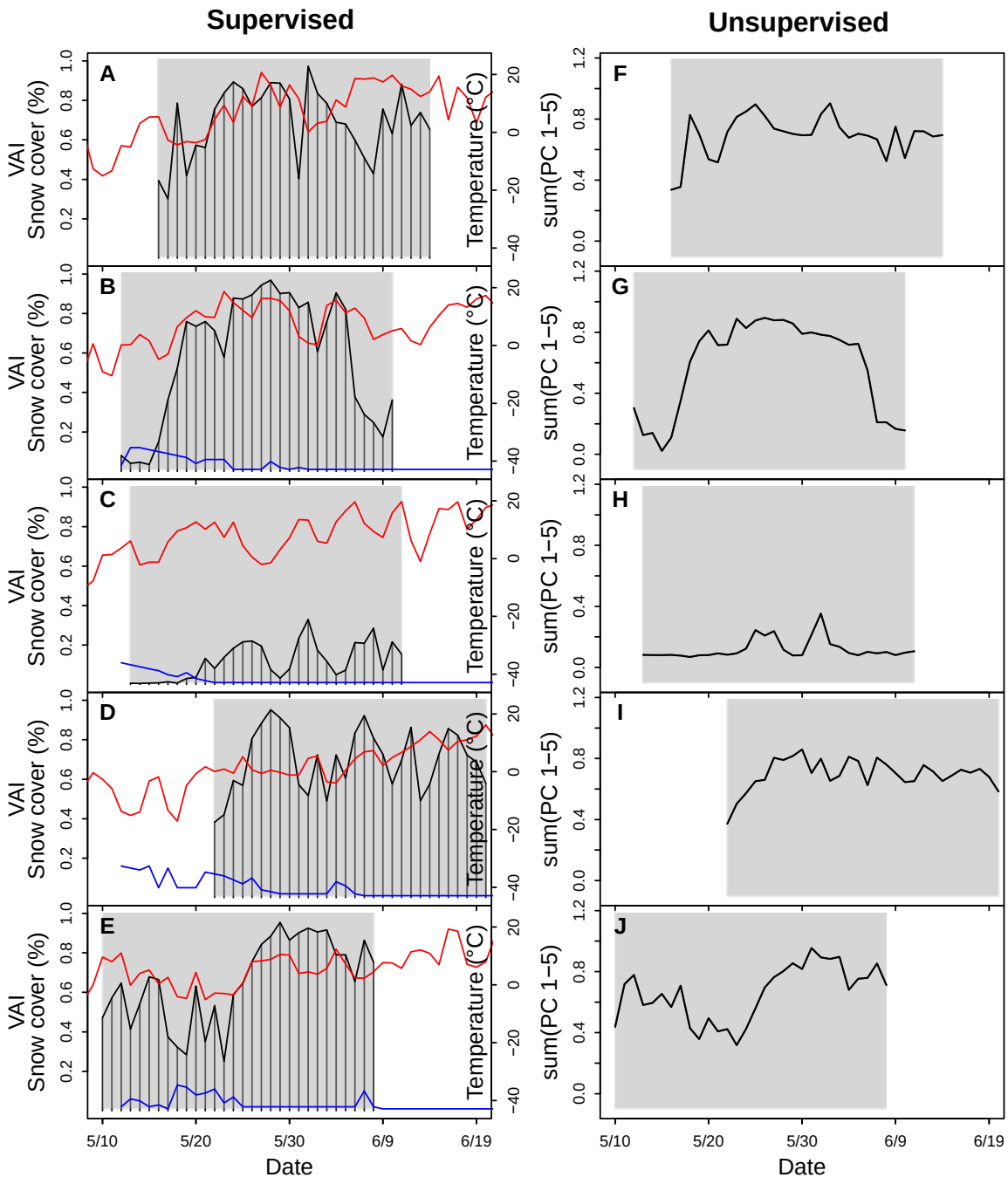

fig. S5. Songbird community vocal activity estimated by supervised and unsupervised approaches near SDOT. (**A**)-(**E**) Songbird daily Vocal Activity Index (VAI), snow cover (blue), and air temperature (red) near Sagavanirtok Department of Transportation (SDOT) between 2010-2014. (**F**)-(**J**) Weighted sums of the first five principal components at the same site and time. Grey boxes identify the available recording period for acoustic data.

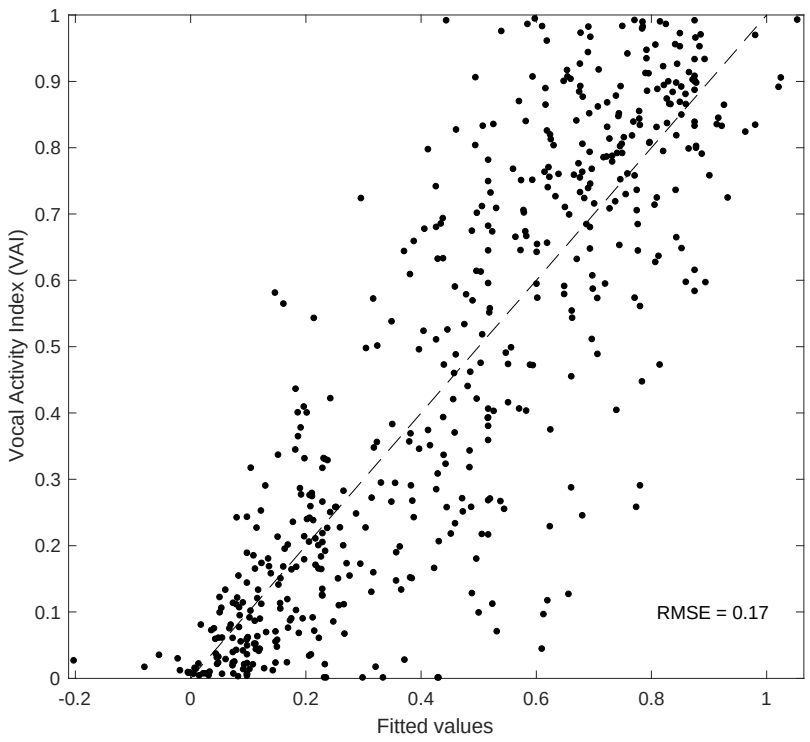

fig. S6. Comparison of the VAI to linear model predictions using only environmental covariates found to be statistically significant.

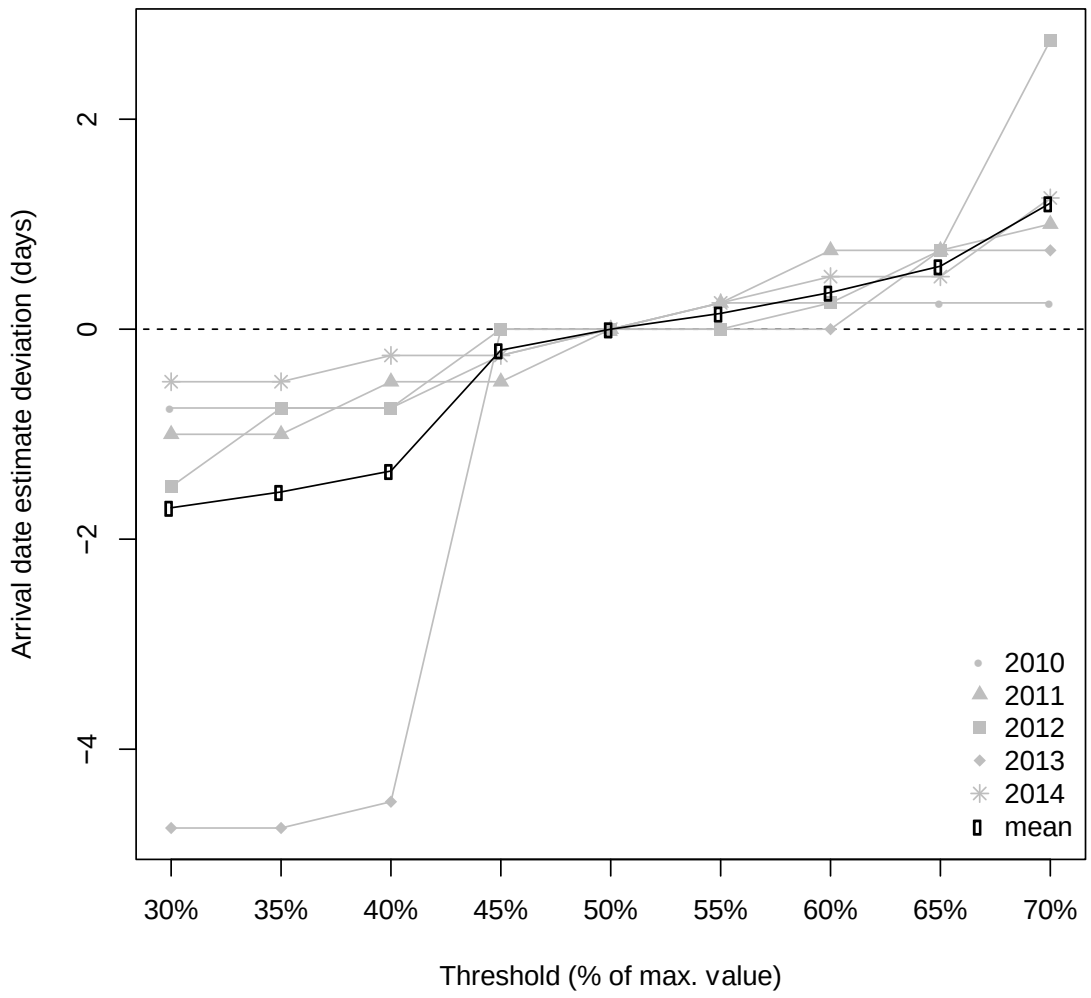

fig. S7. Threshold sensitivity of arrival date estimates from supervised approach. Sensitivity of arrival date estimates to various thresholds. Under the supervised approach, songbird arrival date estimated as the first date that had a VAI that exceeded 50% of the maximum value over the thirty-day recording period. Grey lines show the difference in the arrival date estimate for each year under a range of thresholds as compared to the estimate using a 50% threshold. The black shows the mean difference across years.