A Statistical Inquiry Allying Frequency-Based Bird Song Characteristics with Satellite Imagery from Landsat 8 through Generalized Linear Regression and the Random Forest Machine Learning Algorithm

Thomas DeLillo ('21), Lili Gately ('21), David Guerra, Ph.D., Adam Albina, Ph.D. & Jay Pitocchelli, Ph.D.
Saint Anselm College, Department of Physics, Manchester, NH, USA

Introduction

In 2019, Jay Pitocchelli, Ph.D. trekked across North America and collected temporal and frequency-based data on bird songs characteristics of MacGillivray's Warbler *(Geothlypis tolmiei)*. The query posed through his data was whether bird songs from different locations across North America held any geographical correlations, or in other words, exhibited any regional dialects or "accents." Satellite imagery via Landsat 8 was tied to the bird song data through ArcGIS Pro, with each bird song location linked to a combination 7 bands. Each band is associated with a geographical feature and thus a mixture of these bands would generate a different geographical environment.

Generalized Linear Regression (GLR) and the Random Forest Machine Learning algorithm determined if there was a geographic environmental influence on the frequency-based bird song characteristics. The contrast between these two analyses presented a statistical inquiry through GLR's linearity and with the Random Forest's artificial intelligence direction.

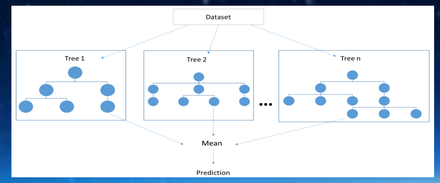
Generalized Linear Regression Equation

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots \beta_n X_n + \varepsilon$$

Dependent Variable · Coefficients · Explanatory Variables · Random Error Term/Residuals

Bird Song Characteristic = β_o + β_1(Band 1) + β_2(Band 2) etc...

Random Forest Machine Learning Algorithm

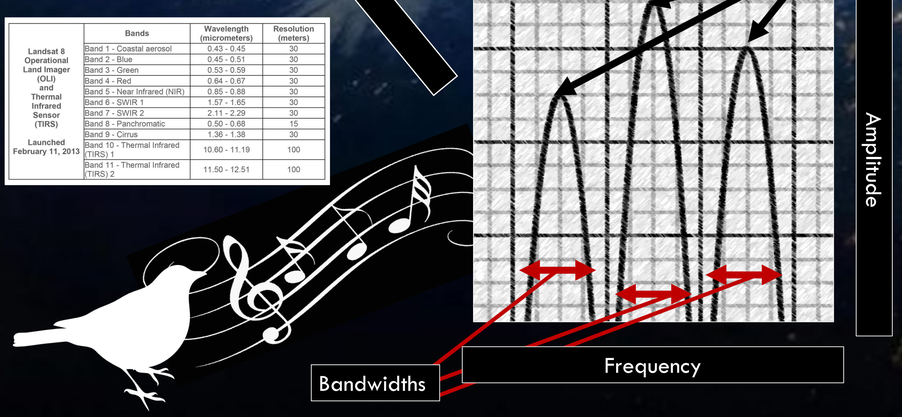


Landsat 8

Methods

DurAN	MinAN	MaxAN	BandwAN	PeakAN	DurBN	MinBN	MaxBN	BandwBN	PeakBN	DurCN	MinCN	MaxCN	BandwCN	PeakCN
-1.19	0.28	-0.28	-0.6	1.53	-1.28	0.18	-0.4	-0.7	1.39	-1.24	0.28	-0.02	-0.29	-1.01
-0.61	0.26	-0.56	-0.94	0.12	-0.69	0.2	-0.47	-0.82	0.22	-0.79	0.15	-0.19	-0.42	-0.48
0.51	-0.21	-0.46	-0.41	-0.14	0.22	-0.61	-0.48	-0.1	0.1	0.28	-0.69	-0.31	0.18	0.17
-1.07	-0.67	-0.15	0.38	-0.01	-1.33	-0.15	-0.54	-0.6	-1.07	-1.29	0.24	-0.25	-0.59	-0.88
-1.14	0.44	2.13	2.35	-0.01	-1.35	1.21	2.38	2.12	-0.03	-1.46	1.66	2.27	1.77	0.7
0.66	-0.59	-0.71	-0.4	-1.55	0.25	-0.48	-0.73	-0.55	-0.81	0.42	-0.44	-0.65	-0.54	-0.22
-1.57	2.44	2.25	0.79	0.76	-1.45	1.91	1.91	0.86	0.74	-1.61	2.09	2	0.97	0.57
-0.25	-1.04	-1.39	-0.9	-0.27	-0.44	-1.01	-1.71	-1.4	-1.07	-0.37	-1.64	-1.44	-0.57	-1.14
-1.19	-0.33	-0.68	-0.58	1.02	-1.33	-0.26	-0.89	-0.98	1.13	-1.11	-0.38	-0.17	0.11	0.83
0.22	-0.51	-0.1	0.31	-0.78	0.27	-0.48	-0.17	0.2	-0.55	0.28	-0.8	-0.12	0.57	-0.61
0.06	-0.33	-0.45	-0.3	-0.91	-0.07	-0.14	-0.42	-0.45	0.22	-0.05	-0.24	-0.5	-0.5	2.14
-1.38	-0.23	0.69	1.09	1.27	-1.48	-0.4	0.64	1.22	-0.03	-1.56	-0.25	0.74	1.32	0.3
-0.78	2.34	0.4	-1.5	1.78	-1.2	2.08	-0.19	-2.12	1.52	<Null>	<Null>	<Null>	<Null>	<Null>
-0.47	1.79	0.43	-0.97	0.37	-0.59	1.79	0.65	-0.72	0.35	-0.42	1.72	0.66	-0.64	0.43
0.97	-0.21	0.66	1.03	0.37	1.04	0.18	0.61	0.66	0.22	1.02	0.68	0.44	0.01	0.57
0.46	0.76	0.22	-0.38	0.89	0.52	1.13	0.33	-0.57	1.13	0.7	1.06	0.54	-0.2	0.3

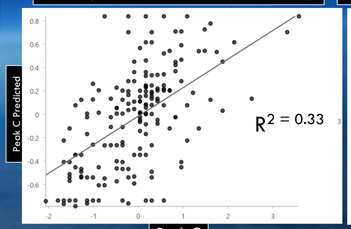
Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Launched February 11, 2013	Bands	Wavelength (micrometers)	Resolution (meters)
	Band 1 - Coastal aerosol	0.43 - 0.45	30
	Band 2 - Blue	0.45 - 0.51	30
	Band 3 - Green	0.53 - 0.59	30
	Band 4 - Red	0.64 - 0.67	30
	Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
	Band 6 - SWIR 1	1.57 - 1.65	30
	Band 7 - SWIR 2	2.11 - 2.29	30
	Band 8 - Panchromatic	0.50 - 0.68	15
	Band 9 - Cirrus	1.36 - 1.38	30
	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100
	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100



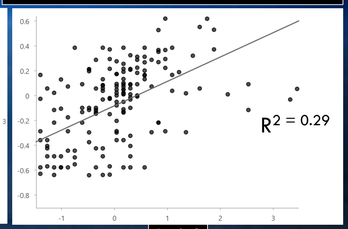
Generalized Linear Regression ArcGIS Model

Random Forest ArcGIS Model

Relationship between Peak C and Peak C (Train + Predicted)



Relationship between Peak C and Peak C (No Band 2 (Train + Predicted))



Bandwidth A Random Forest

Frequency-Based Bird Song Characteristic	R^2 Value
Bandwidth A Trained	0.38
Bandwidth A Trained + Predicted	0.46
Bandwidth A Trained + Predicted (No Bands 5 or 7)	0.45

Results

Generalized Linear Regression Peak C

Variables	Coef	StdError	t_Stat	Prob
b1_MOS_2019_CB	0.413938	0.266052	1.555853	0.121581
b2_MOS_2019_CB	-0.755285	0.376294	-2.007166	0.046298
b3_MOS_2019_CB	0.169052	0.380788	0.443953	0.657634
b4_MOS_2019_CB	0.034378	0.028738	1.196251	0.233245
b5_MOS_2019_CB	-0.055286	0.153744	-0.359599	0.719588
b6_MOS_2019_CB	0.001453	0.003631	0.400229	0.689484
b7_MOS_2019_CB	0.106579	0.288358	0.369608	0.71213
Intercept	4198.761011	665.054517	6.313409	0

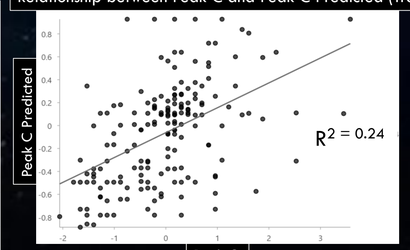
Random Forest Variable of Importance Table Peak C

Variables	Importance
b6_MOS_2019_CB	0.36675
b4_MOS_2019_CB	0.15243
b2_MOS_2019_CB	0.131363
b1_MOS_2019_CB	0.109404
b7_MOS_2019_CB	0.097324
b5_MOS_2019_CB	0.078707
b3_MOS_2019_CB	0.064023

Generalized Linear Regression Significant Bands

Frequency-Based Bird Song Characteristic	Statistically Significant Bands (P<0.05)
Bandwidth A	Bands 5 and 7
Bandwidth B	Bands 1, 4, 5 and 7
Bandwidth C	Bands 1 and 7
Peak A	No Bands
Peak B	No Bands
Peak C	Band 2

Relationship between Peak C and Peak C Predicted (Train)



Conclusion

The results of the GLR exhibited some correlation between frequency-based bird song characteristics and bands 1-7, specifically bandwidth B which displayed correlation concerning bands 1, 4, 5 and 7. However, the results of the remaining 5 frequency-based bird song characteristics were not consistent enough to establish a bird song dialect linked to a geographic environmental region, particularly with peaks A and B holding no correlation with any bands.

The Random Forest Machine Learning Algorithm's results also did not establish a bird song dialect, with the R^2 values between 0.24 and slightly more than 0.5. Nevertheless, Random Forests conducted with the removal of only GLR's statistically significant bands showed almost no change, thus demonstrating that the Random Forest Machine Learning algorithm is robust to the absence of significant data.

Although these results did not support a geographic environmental bird song dialect, this study's methods and procedures are applicable to other datasets to potentially yield more suggestive results or even to unveil noteworthy associations.