Bird Sightings Across the United States

Bryce C. Turner

December 2022

Contents

1 Introduction

Birders are more than willing to travel to see their favorite bird. But, the trick is knowing when and where to travel. Most people need to plan trips months or even a year in advance to work around their vacation and school schedules. There are plenty of species-level and region-level population forecasts, but for an average birder a smaller geographic unit is necessary. The closest available resource, BirdCast's Bird Migration Radar, has two limitations; first, it is real-time and second, it is not available year-round [3].

To help address this gap, this paper will forecast monthly bird sightings by State[1]. This could allow travelers to pick the region that would best make use of their available time.

This work focuses on two different birds and two different forecasting models. The birds, the Bald Eagle and the Red Winged Black Bird, are distinct enough to merit interest from birders. They also differ on many dimensions. The Bald Eagle has been the subject of focused conservation efforts while the Red Winged Black Birds are "abundant and widespread" according to the National Audubon Society [4] [2]. The Bald Eagle is a large Raptor while the Red Winged Black Bird is a small Passerine.

The two forecasting models evaluated are an ARIMA model and a Structural Time Series model with local trend and a seasonal component. ARIMA is typically the first time series model that a user would attempt. The second model utilizes a maximum likelihood in order to adjust for noise at each observation. Given that noise is certainly present in the data, this method will attempt to extract the true value.

2 Data Description

2.1 Bird Sightings

Bird sighting data comes from eBird, a project of the Cornell Lab of Ornithology. It contains data about the sightings from its users such as the time, date, location, bird type, and number seen. While this is open-source and anyone is free to submit data into their database, eBird does conduct verification that the reporting is plausible based on a user's geography, time of year, and bird count[1]. The data is freely available for academic and other non-commercial purposes. To gain access, a user must have a free eBird account, complete a short questionnaire about the nature of their project and sign their acceptable use policy.

This data set is valuable because it provides more comprehensive coverage than would ever be possible by an individual research team. [2] There are, however, some limitations to the data set as well. The data sends to be very

[1] Certain species could be forecasted at the zip code level, but given the state of the current data this would have left many series without observations for a long period of time.

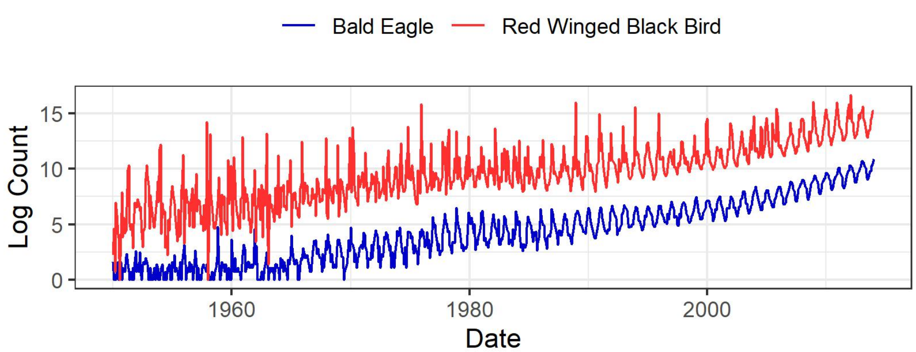
[2] It could conceivably cover the entirety of the planet, though it user base is likely focused in North America.

cyclical, as birders tend to be less active in the colder months up north. There is also a different level of effort that birders spend on each species.

This project will focus on bird sightings for two species: the Bald Eagle (*Haliaeetus leucocephalus*) and the Red Winged Black Bird (*Agelaius phoeniceus*). The Bald Eagle was chosen because it is both rare enough to ensure some level of effort was put in to sighting while being common enough across the states to ensure decent coverage. While the Red Winged Black Bird is significantly more common than a Bald Eagle, it's distinctive red patch seemed more likely to draw notice from birders.

The raw data is unique at the date-observer level. The Bald Eagle data is approximately 0.4 Gbs large and the Red Winged Black Bird data is approximately 1.2 Gbs large. To create more usable data, the data is aggregated to the Month-State counts. The data technically extends back to 1900, but the bird counts do not show consistent observations until the 1950s. Therefore, this data covers the period from 1950 to 2013. Figures 1 and 2 show the national counts and state-level counts from 4 large states, respectively. The variable log count is used on these plots to more easily compare the two species but it is not used in the analysis. The seasonality apparent in the national averages is also apparent at the state level, which spans diverse climates.

Figure 1: National Bird Sightings

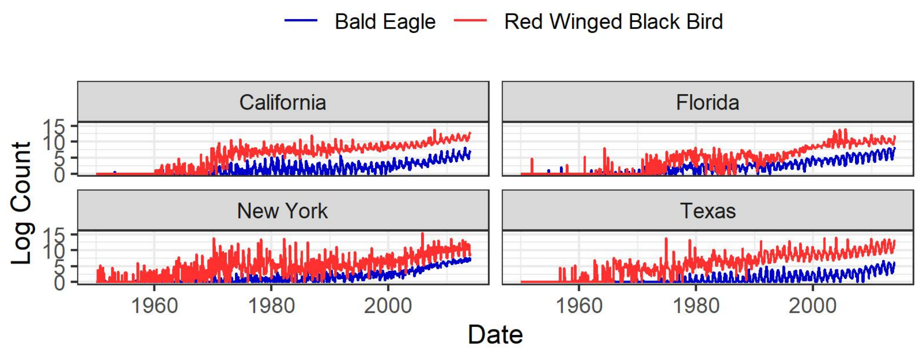


2.2 Weather Data

The raw weather data comes from the National Centers for Environmental Information, which falls under the National Oceanic and Atmospheric Administration (NOAA). State-level weather aggregates are available for the years 1895 to January 2014 at https://www.ncei.noaa.gov [5]. [3] This weather data only contains data on the "Lower 48" states, excluding Alaska and Hawaii. Thus, the

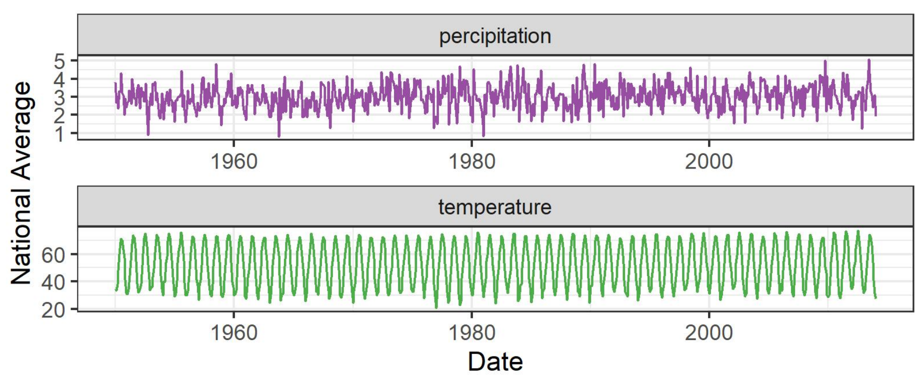
[3]The author spent much time attempting to get more recent aggregates online, but found NCEI's climate data search tool to be not researcher friendly.

Figure 2: Bird Sightings in Select States



entire analysis data will also exclude these states. For this analysis, the main weather events to track are average monthly temperature and average monthly precipitation. Figures 3 and 4 show these series nationally and in select states, respectively.

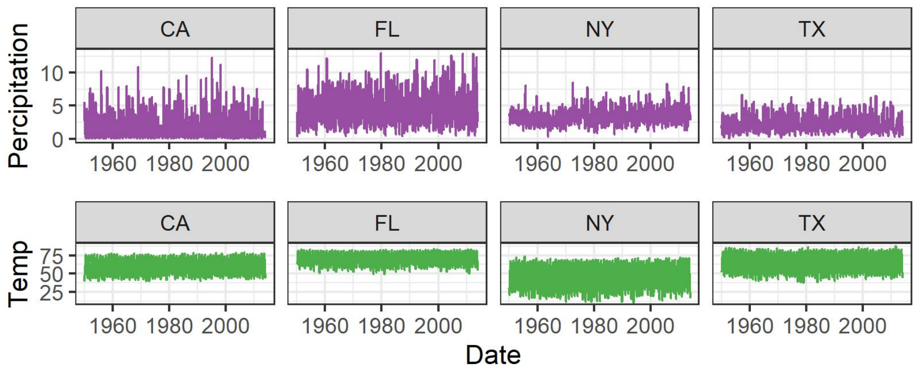
Figure 3: National Weather Averages



3 Data Exploration

The remainder of this work will focus on individual bird counts at the state-month level.

Figure 4: National Weather Averages



3.1 Stationarity

Each state-bird series was tested for stationarity via the Augmented Dickey-Fuller Test. Up to 12 lags were included in the test. The series were tested for trend using the type="trend" option in R's ur.df() function from the library urca. Table 1 shows the results from the 48 individual stationarity tests. Nearly all of the Bald Eagle series are stationary at the 1% level. However, just over half of the Red Winged Black Bird series are stationary at any level.

Table 1: Results from Augmented Dickey-Fuller Test for Stationarity

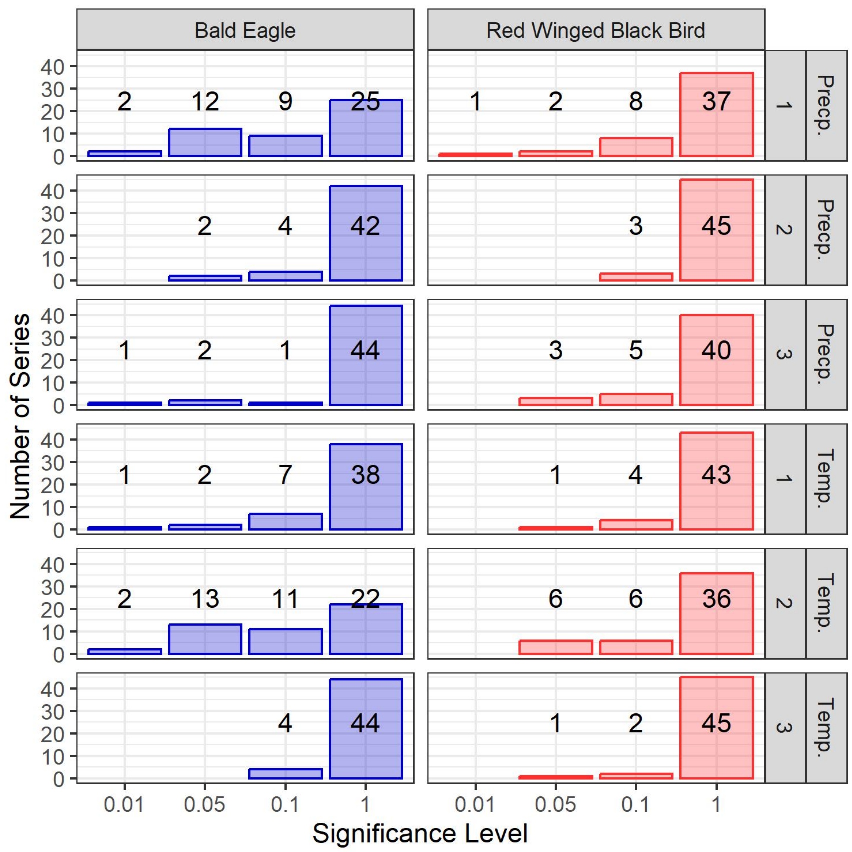
Statistic	1%	5%	10%	Not Significant
			Bald Eagle	
ϕ	43	0	0	5
τ	34	6	2	6
			Red Winged Black Bird	
ϕ	22	4	2	20
τ	23	2	1	22

Series were tested with lags up to 12 months.

3.2 Cointegration

The series were also tested for cointegration with their local weather data, described above. This test was conducted using covariance matrix returned from the NeweyWest function in R's sandwich library. Each series was first-differenced if it was determined to be non-stationary in the above results. The first 3 lags for both Temperature and Precipitation were evaluated, and the results are displayed in Figure 5. The left column shows the results for the Bald Eagle series and the right column shows the results for the Red Winged Black Bird series. The first three rows show the results for the first three lagged values of the Precipitation series, and the last three rows show the same for the Temperature series. All of the Bald Eagle series show no evidence of cointegration with their local weather. While this is largely the pattern with the Red Wing Black Bird series, about 10 series do show some level of cointegration with the weather. Given that weather does not appear to be a particularly useful series, the following forecast exercises will focus on univariate models.

Figure 5: Cointegration Test Results



4 Forecasts

4.1 ARIMA Models

The first forecasting exercise attempted to fit an ARIMA model for each series. Using R's auto.arima() function, each series was tested for all possible ARIMA models and the best result was returned [4]. The model was trained on the data from January 1950 to December 2012. The resulting model was then forecast out 13 months, until the end of the data. Two statistics were recorded: (1) The Root Mean Squared Error (RMSE) from the In-Sample and, (2) the RMSE from the forecasted 13 months. The results of this process are shown in Figure 10 in the Appendix.

4.2 Structural Models

The second forecasting exercise attempted to fit a Strucutral Time Series model on the data. This was done with R's StructTS() model using the "BSM" option. This option fits a local trend with a seasonal component. Given the clear seasonality event in Figure 2, this seemed the best option.

4.3 Comparison

The "level" of the RMSE was slightly lower in the Structural models. Therefore, the two models were mainly compared on the increase of the RMSE when applied to their respective training sets. This metric is used because it highlights whether a given model is over-fitting the training data and is therefore unlikely to be consistent over time. Figures 6 and 7 show the relative increase in a gives series' RMSE. The results are displayed as a histogram.

For the Bald Eagle series, the structural model results in a larger outlier but also results in many more series close to 0. For the Red Winged Black Birds, there is not a clear difference but the Structural Model appears to move more series away from zero. Therefore, it seems that the Structural Model is the better choice when forecasting Bald Eagles while it is approximately a tie for the Red Winged Black Bird series.

4.4 Forecasts: March 2023

Using the Structural Model described above, each state receives a forecast for each bird. Given the large differences between states, the logged values of these counts are displayed. The results are displayed in Figures 8 and 9. March 2023 was chosen because this is a common time for Spring Break. It is also before BirdRadar turns back on.

From these maps, it appears that the states with the highest Bald Eagle sightings are Illinois, Kansas, and Florida. States with the highest Red Winged

[4] It is possible to conducted an exhaustive search by setting the option "stepwise" to FALSE.

Figure 6: Percentage Increase in RMSE when Forecasting - ARIMA

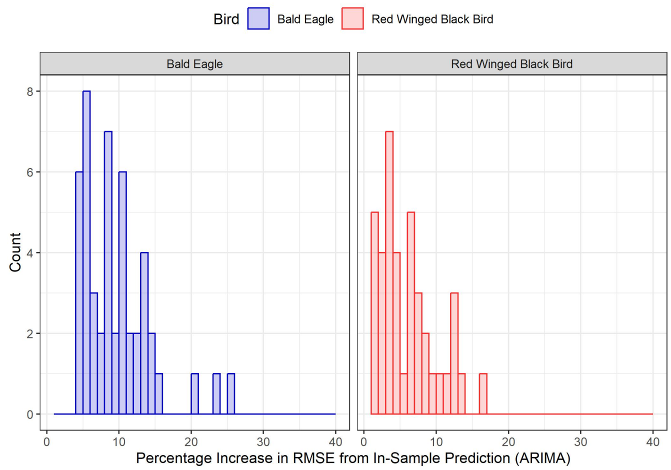
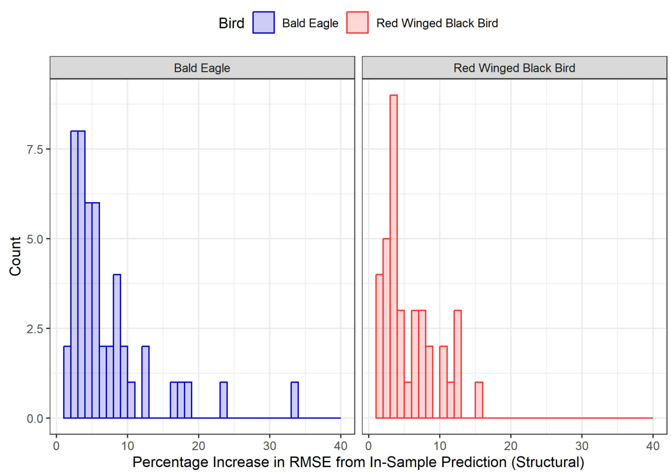


Figure 7: Percentage Increase in RMSE when Forecasting - Strctural Model



Black Bird counts are Kansas and South Carolina. Therefore, a very strong candidate for a planned bird watching trip would be Kansas.

5 Conclusion

This work set out to develop a methodology for predicting bird sightings across the United States. The granularity of the data set allowed for a long historical series across the lower 48 United States. It was surprising that there was such a weak correlation between Weather and Bird Sightings, but that could be due

Figure 8: Forecasted Bald Eagle Sightings: March 2023

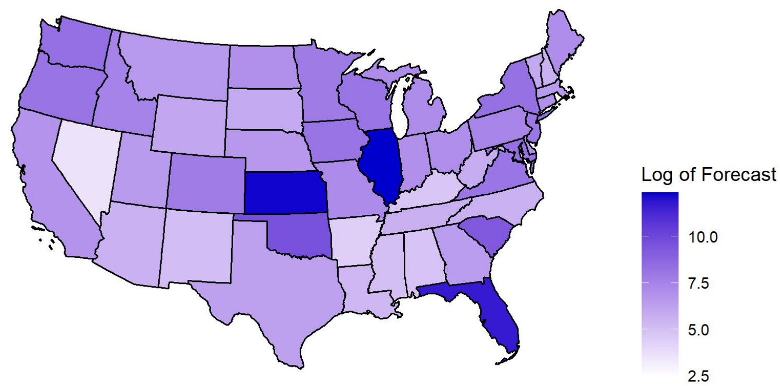
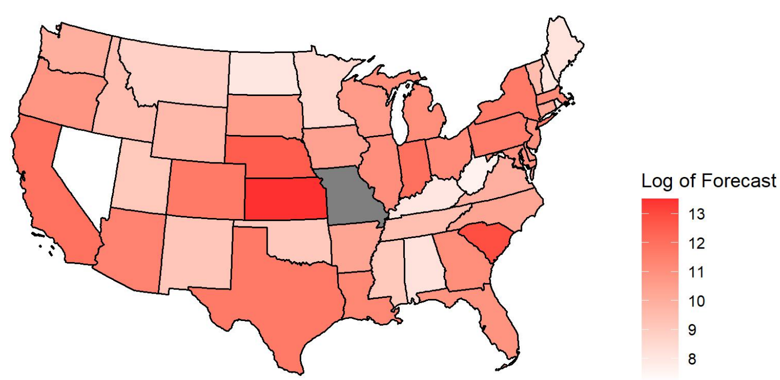


Figure 9: Forecasted Red Winged Black Bird Sightings: March 2023



to the fact that the core user base of eBird is slightly more resistant to weather than the average person. It could also be that the frequency ad lag of the weather data was incorrect. It seems more likely that bird sightings would be affected by the weather on the day of the sighting than the weather over the past three months. However, to prove this more granular weather data would be needed.

Given that the data is freely accessible, this methodology could be expanded to study different species and include Canada and Mexico. Another logical next step would be to further shrink the geographic unit, to either a County or Zip

Code level. This would likely lead to a significantly higher number of zeros in a given series, but may provide even more useful information to Bird Watchers.

While there is certainly further work that could be done in this area, this paper does highlight the usefulness of a data set to Bird Watchers.

6 References

[1] eBird: A citizen-based bird observation network in the biological sciences. 2009. URL: https://science.ebird.org.

[2] *Bald Eagle*. National Audubon Society. Dec. 2022. URL: https://www.audubon.org/field-guide/bird/bald-eagle.

[3] *Migration Dashboard*. BirdCast. Dec. 2022. URL: https://birdcast.info/migration-tools/live-migration-maps/.

[4] *Red Winged Black Bird*. National Audubon Society. Dec. 2022. URL: https://www.audubon.org/field-guide/bird/red-winged-blackbird.

[5] *Time Bias Corrected Statewide-Regional-National Tempurature-Precipitation*. National Centers for Environmental Information. Dec. 2022. URL: https://www.ncei.noaa.gov/pub/data/cirs/drd/.

7 Appendix

Figure 10: In-Sample and Forecast RMSE Results - ARIMA

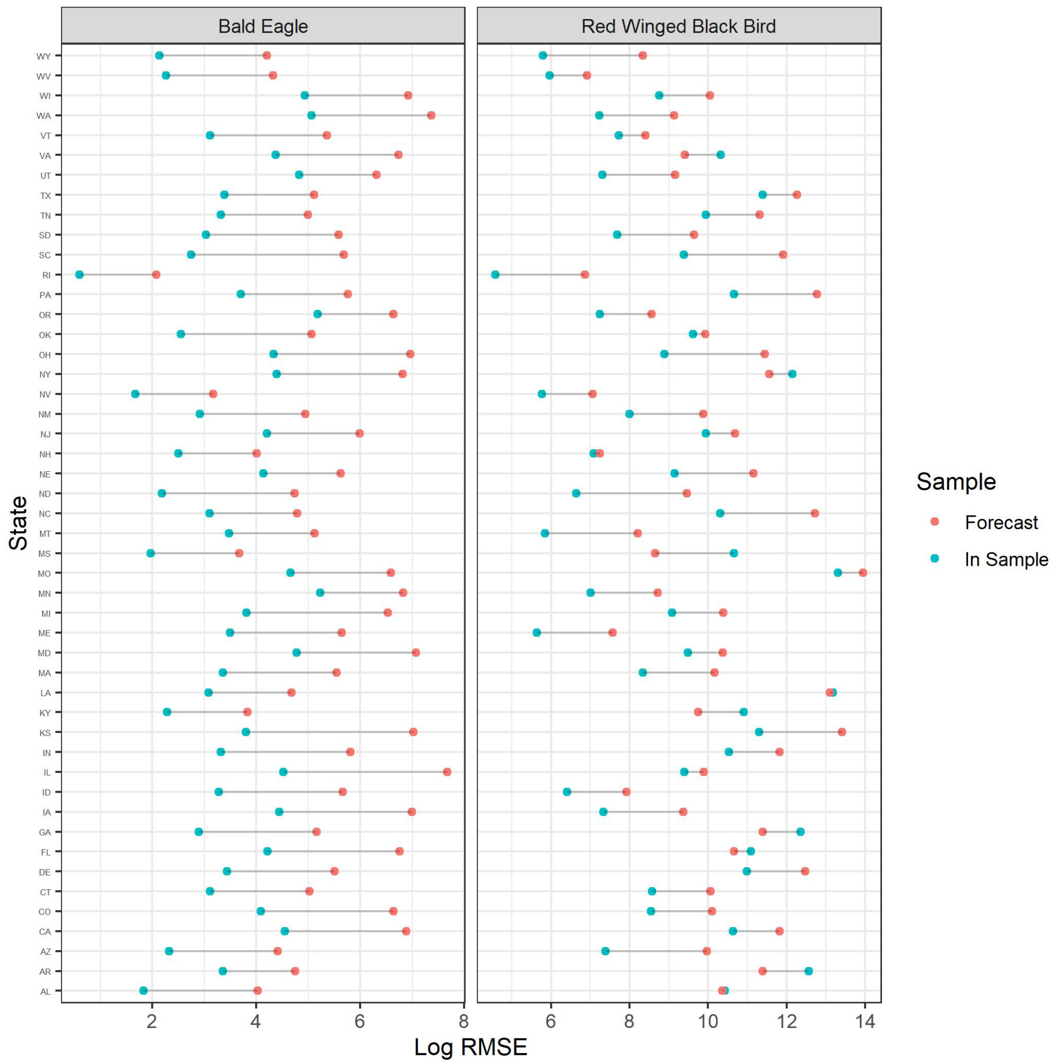


Figure 11: In-Sample and Forecast RMSE Results - Structural Model

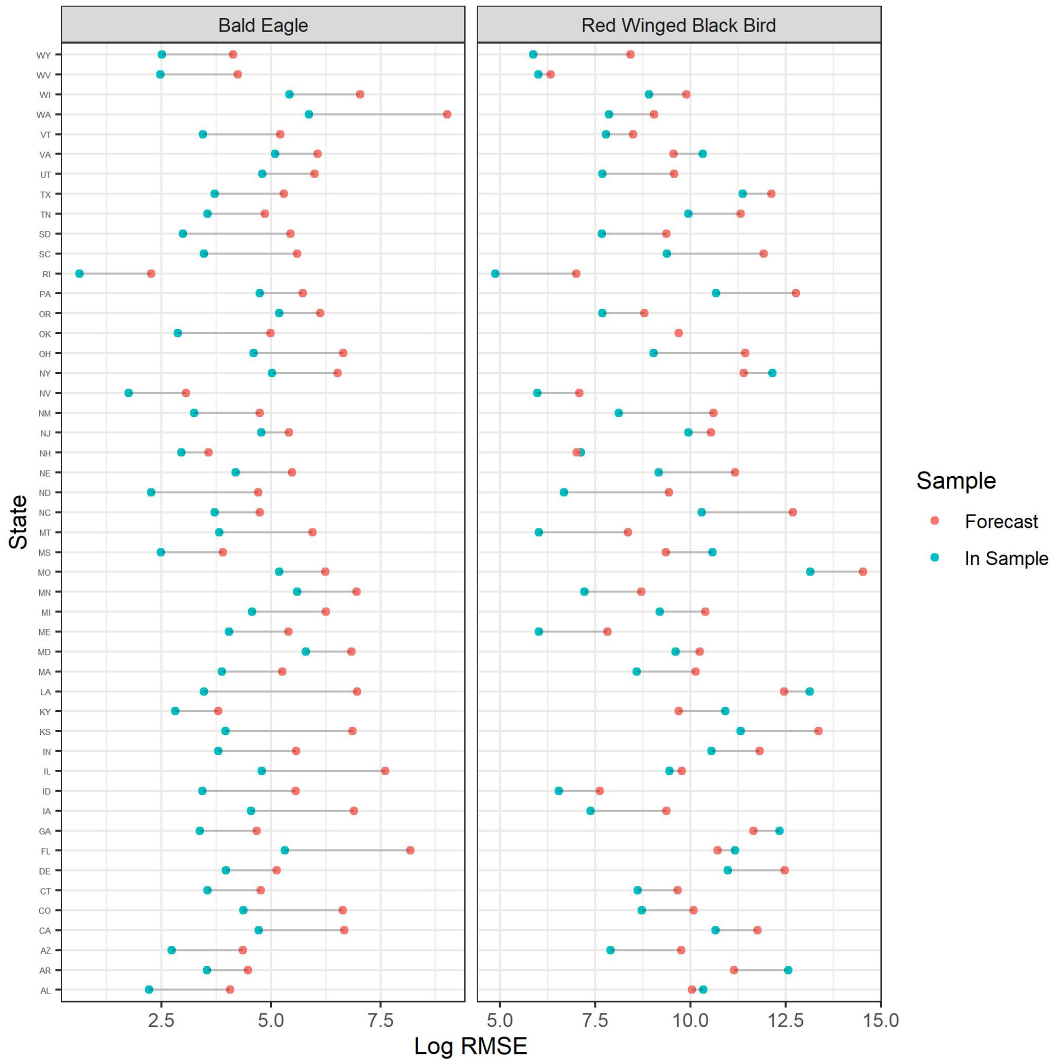


Table 2: Tabular Result of State-Level ARIMA Models

State	Bald Eagle		Red Wing Black Bird	
	In-Sample RMSE	Forecast RMSE	In-Sample RMSE	Forecast RMSE
AL	5.21	55.03	33500.06	31695.08
AR	27.59	113.71	286995.02	88279.39
AZ	9.21	81.75	1620.76	21413.05
CA	93.11	976.21	41310.48	136526.17
CO	58.38	763.44	5169.85	24400.02
CT	21.52	150.60	5261.44	23251.20
DE	29.95	244.55	58911.24	261918.60
FL	66.53	854.10	65910.16	42345.58
GA	17.14	172.45	231120.08	88811.33
IA	83.90	1091.46	1530.51	11739.04
ID	25.54	286.97	606.25	2754.26
IL	90.38	2138.32	12108.39	19853.81
IN	26.67	332.88	37197.02	135409.14
KS	43.88	1122.96	80787.04	668325.45
KY	8.77	44.98	54878.39	17113.20
LA	20.73	106.07	526365.48	487516.40
MA	27.83	253.53	4171.80	26035.61
MD	117.06	1176.74	13213.50	32064.81
ME	32.07	280.34	280.10	1940.53
MI	44.22	684.95	8833.91	32383.70
MN	185.28	920.20	1104.47	6108.59
MO	104.70	722.86	598117.87	1133770.78
MS	6.14	38.24	42739.05	5758.18
MT	31.28	166.65	347.50	3686.54
NC	21.18	118.36	29970.74	333867.52
ND	7.91	112.82	771.00	12770.46
NE	61.63	274.43	9382.23	69710.37
NH	11.15	53.92	1203.03	1400.90
NJ	66.24	395.49	20735.85	43935.62
NM	17.44	139.82	2987.30	19375.42
NV	4.34	22.91	321.66	1162.37
NY	80.09	906.47	189349.19	104239.59
OH	74.94	1054.88	7268.30	93156.45
OK	11.77	156.20	15109.73	20455.35
OR	175.78	764.95	1398.91	5256.19
PA	39.69	315.23	42654.04	352712.03
RI	0.82	6.97	97.17	953.08
SC	14.58	294.00	11883.17	148625.41
SD	19.81	264.75	2173.89	15356.96
TN	26.66	146.26	20683.79	81944.50
TX	28.46	164.51	88531.80	212522.52
UT	123.80	553.03	1484.01	9548.30
VA	77.96	837.48	30237.30	12127.91
VT	21.42	210.33	2275.83	4482.51
WA	156.34	1574.41	1382.82	9285.14
WI	138.40	1012.90	6342.66	23008.47
WV	8.64	74.44	390.16	1005.12
WY	7.49	66.25	329.39	4166.97

Table 3: Tabular Result of State-Level Structural Models

State	Bald Eagle		Red Wing Black Bird	
	In-Sample RMSE	Forecast RMSE	In-Sample RMSE	Forecast RMSE
AL	8.20	57.26	30556.11	22819.77
AR	33.21	86.31	285985.07	68577.47
AZ	14.39	76.89	2699.23	17183.88
CA	111.64	792.59	42209.32	127450.20
CO	77.99	758.21	6151.01	23675.44
CT	33.76	115.95	5459.92	15667.85
DE	51.78	167.71	58441.20	258316.61
FL	202.51	3579.17	70380.66	44722.26
GA	28.13	106.22	226977.37	115045.30
IA	92.80	988.37	1595.40	11612.05
ID	30.05	258.09	696.41	2043.26
IL	119.68	2003.12	12595.34	17470.38
IN	43.56	261.41	38258.06	134400.79
KS	51.19	946.85	82357.96	632303.97
KY	15.58	43.60	54589.60	16131.95
LA	31.10	1056.65	503203.96	256793.08
MA	47.10	191.17	5341.77	24984.11
MD	328.95	929.77	14828.23	27767.98
ME	55.74	219.57	407.67	2488.96
MI	95.68	514.67	9729.75	32514.32
MN	268.00	1042.38	1358.62	6076.58
MO	176.95	510.69	507724.36	2020326.32
MS	11.02	48.47	38951.17	11434.86
MT	44.29	382.67	412.53	4269.10
NC	39.76	113.20	29302.56	320362.10
ND	8.56	109.83	790.59	12552.12
NE	65.54	238.70	9581.71	70311.19
NH	18.05	34.72	1240.98	1117.05
NJ	118.43	223.20	20718.25	37720.79
NM	24.80	113.11	3351.30	40261.27
NV	4.76	20.22	393.26	1182.93
NY	151.29	674.43	188022.88	89073.15
OH	98.41	771.34	8365.78	92716.28
OK	16.67	146.10	16168.28	16071.61
OR	178.27	455.41	2165.23	6545.47
PA	113.16	305.65	42899.58	350938.87
RI	0.87	8.58	130.54	1102.06
SC	31.02	267.70	11808.42	148945.31
SD	18.94	229.44	2144.35	11634.52
TN	33.82	128.58	20716.92	81434.41
TX	39.87	196.62	86562.91	182458.25
UT	120.61	400.98	2161.34	14214.43
VA	161.74	430.58	30261.71	14105.54
VT	30.42	181.41	2393.58	4865.85
WA	349.89	8247.50	2595.68	8453.34
WI	226.16	1133.22	7432.90	19571.54
WV	10.92	68.16	406.66	562.31
WY	11.23	61.44	351.52	4527.00