THE IMPACT OF GREENHOUSE GAS EMISSIONS ON MOURNING DOVE ABUNDANCE

Tan, Manfred. Tri, Noah. Wiley, Michael

STAT 5291 Final Project
May 9, 2025

Abstract
This study explores the potential association between greenhouse gas emissions and the abundance of the Mourning Dove, a prevalent North American migratory bird, using data from 1990-2019 for a selection of U.S. states. A multiple linear regression model, incorporating state-level controls and time trends, was employed to analyze relationships between various greenhouse gas emissions, including carbon dioxide, fluorinated gases, methane, nitrous oxide, and land use and forestry carbon stock changes, to Mourning Dove abundance. The model explained a modest portion of the variance in dove abundance (approximately 25%). Carbon dioxide showed statistically significant positive associations with dove abundance, while significant state-level differences were also observed; other emission types did not show clear linear associations. This paper discusses these findings and attempts to interpret the results in the context of the complex ecological dynamics. This paper also acknowledges notable model limitations and methodological challenges inherent in such observational ecological studies, underscoring the exploratory nature of these results. Consequently, this research emphasizes the need for continued investigation, potentially with refined data and more advanced modeling techniques, to further model these environmental relationships.

1. Introduction

The steady rise in greenhouse gas emissions over the past few decades has sparked significant concerns over their ecological consequences. Among these consequences are changes in migration patterns in various bird species, connecting greenhouse emissions to behavioral changes. Increased temperatures, habitat degradation, and altered food availability driven by elevated emission levels are among the key stressors affecting these populations. In this paper, we focus on a specific species, Zenaida macroura, colloquially known as the Mourning Dove. It is a prevalent North American migratory bird, known for its adaptability and widespread distribution. The question we aim to answer is if there exists an association between the volume of different types of gas emissions within a U.S. state and the Mourning Dove's population levels in that state. That is, is the volume of annual gas emissions in a state correlated with the abundance of Mourning Doves in that state? We aim to explore these questions using multiple linear regression to quantify potential relationships between emissions data and population trends over time.

Mourning Doves can be found throughout the continental United States, extending down to central Mexico and up to southern Canada. Some Mourning Doves remain in the warm climate of Texas and its neighboring states, as well as northern Mexico year round. However, most birds breed in southern Canada and northern U.S. states, particularly in the upper Midwest and upper Mountain West. In the winter, they migrate south to the southern U.S. region and northern Mexico. This broad range and migratory behavior make the species an ideal candidate for studying the potential impact of environmental changes across different locations. Shifts in their migratory timing, range boundaries, or relative abundance may signal larger ecological shifts linked to climate stressors. While imperfect, statistics on the abundance of Mourning Doves in a given area and their broader migration patterns typically rely on standardized bird counts, which record individual birds perceived by sight or sound during regular survey periods (**Sauer**; **Link**; **Kendall & Dolton**; **2010**). This data, although subject to observer bias and environmental variability, currently offers one of the most consistent long term datasets available for monitoring population-level trends in bird species.

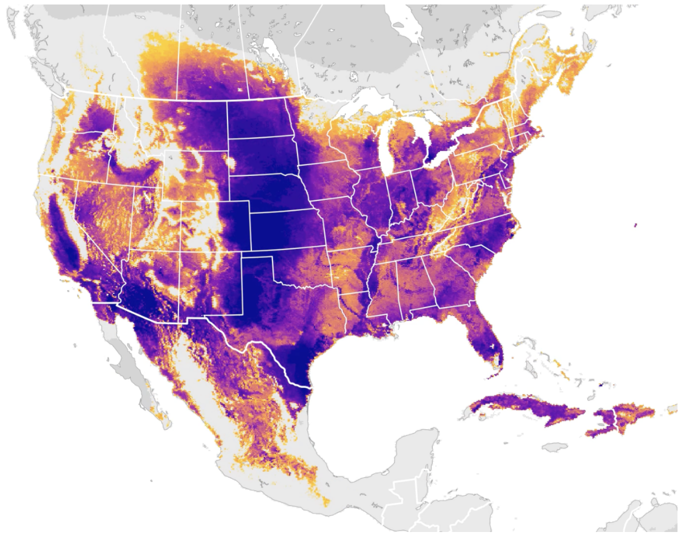

Figure 1. Mourning Dove Abundance Heatmap During August, 2024 (EBird Data)

Mourning Doves exhibit consistent migratory patterns, often returning to the same location to breed. In some cases, they even reuse the same nest (**Noble Research Institute**). Accordingly, there has been attention devoted to studying and discovering the conditions that cause Mourning Doves to divert from their expected migratory patterns. One recent study concentrated in North Dakota found that Mourning Dove abundance decreases in regions with larger land areas of harvested corn and soybeans, while a large land area of woodlands was associated with higher Mourning Dove abundance (**Dinges**; **Szymanski**; **& Parent**; **2022**). Building on this line of inquiry, our analysis investigates whether broader environmental pressures, in this case greenhouse gas emissions, are similarly associated with regional Mourning Dove abundance patterns.

2. Data Collection & Data Description

This study integrates two primary datasets: one detailing state-level greenhouse gas emissions and another quantifying Mourning Dove abundance. Both datasets were processed to yield annual totals for each selected U.S. state, forming the basis for our analysis over the period 1990-2019.

2.1 Gas Emissions Data Collection

State-level greenhouse gas emissions data were obtained from the U.S. Environmental Protection Agency's (EPA) Greenhouse Gas Inventory Data Explorer, an interactive online tool that allows the exploration and download of emissions data across various sectors and geographic regions. The specific dataset utilized corresponds to the information available through the "All Sectors" and "All GHGs" view, accessible via the portal. To systematically retrieve the data for each U.S. state relevant to this study, a custom data collection script was programmed to iterate through each state and collect and format the relevant emissions data. The script is attached in the relevant code files, allowing users to reproduce the analysis with different sets of data as needed. The collected datasets provided state-level emissions, measured in Million Metric Tons of Carbon Dioxide Equivalent ($MMTCO_2e$), for the analysis period of 1990-2019. The datasets included emissions for carbon dioxide (CO_2), fluorinated gases, net carbon stock changes attributable to Land Use, Land-Use Change, and Forestry (LULUCF), methane (CH_4), and nitrous oxide (N_2O). These datasets were subsequently aggregated to derive total annual emissions for each specified gas type, which were further separated by state.

2.2 Mourning Dove Data Collection

Data for the Mourning Dove was found through the North American Breeding Bird Survey (BBS). The BBS is a large scale initiative monitoring the trends of North American bird populations. The BBS has recorded data since 1966 and is one of the most comprehensive long term bird tracking programs. The BBS tracks birds by having avian identification personnel stationed across different stations along the survey routes. They conduct measurements during the height of the avian breeding season and use this data to estimate population trends and abundances at different geographical scales. The sighting data for the Mourning Dove was sourced from the BBS, collecting every station sighting through the years 1990 - 2019.

2.3 Description and Methods

The gas emissions dataset contains the MMTCO2e (millions of tons of carbon dioxide equivalent) greenhouse gasses released by each state annually from 1990-2019 broken down by gas type (carbon dioxide, fluorinated gases, land use and forestry carbon stock change, and nitrous oxide). On the other hand, the mourning dove dataset is more granular as it provides mourning dove abundance data on each route (as explained earlier) within a state annually. The consistent collection methodology of mourning dove abundance data allowed us to disregard the route and sum the abundance for each year in a state to match the gas emissions data. We chose to only use the years 1990-2019 as COVID-19 disrupted the consistent collection of mourning dove abundance data in 2020 and beyond. Below, Figure 2 details a general outline of the total Morning Dove sightings, labeled by the variable 'SpeciesTotal'.

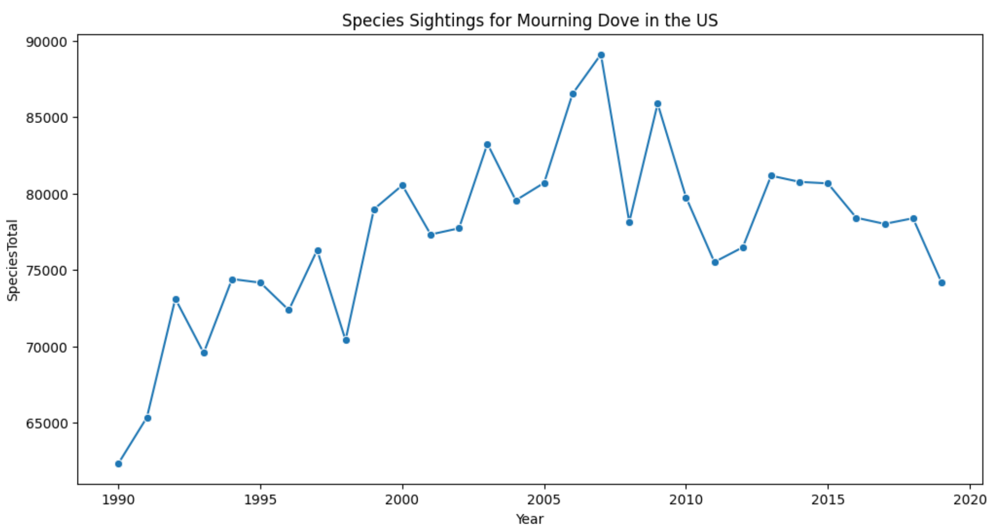

Figure 2. Mourning Dove Abundance from 1990 to 2019

Although we have the data for all 50 US states as well as some southern Canadian provinces, we chose to select a subset of states for a more focused and specific study. To ensure we had a subset of states that varied in terms of climate, gas emissions, populations, and habitats, we selected the following five states: Alabama, Arizona, Ohio, Oklahoma, and Wisconsin. Of these five states, Ohio and Wisconsin are prominent summer breeding states for mourning doves; on the other hand, Alabama, Arizona, and Oklahoma are prominent climates for mourning doves to migrate to for winter or stay year-round. The states are also spread out far from east-to-west. We also wanted a variety of greenhouse emission levels, as nearby states may have their emission levels impacted by neighboring states. Figure 4 below plots the nitrous oxide emissions for each state annually in our time frame, revealing the largely dissimilar nitrous oxide emissions for each state.

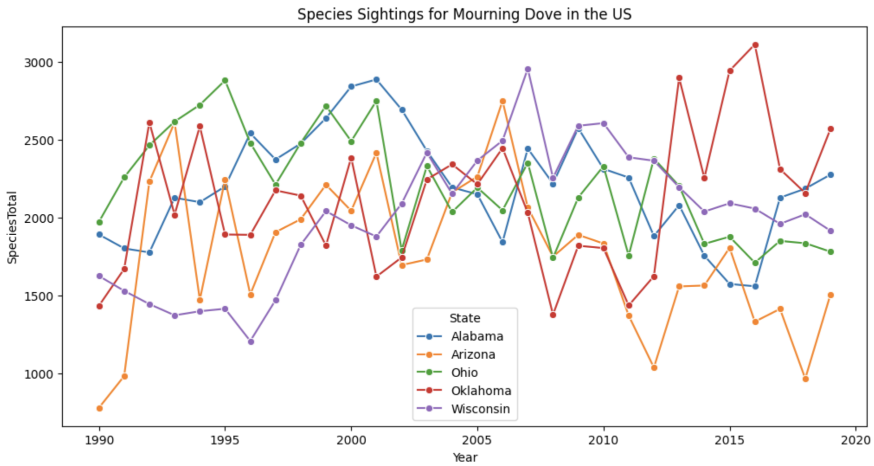

Figure 3. Mourning Dove Abundance by state

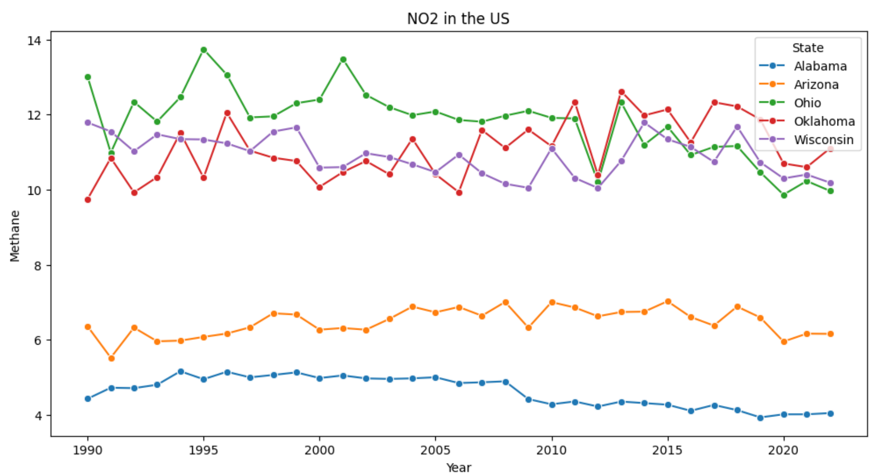

Figure 4. Annual nitrous oxide emissions (in MMTCO2e) for each state in our selected subset

Additionally, these five states had reliable data with no missing values. When combined, the distribution of data was roughly normal, as documented in Figures 5 and 6.

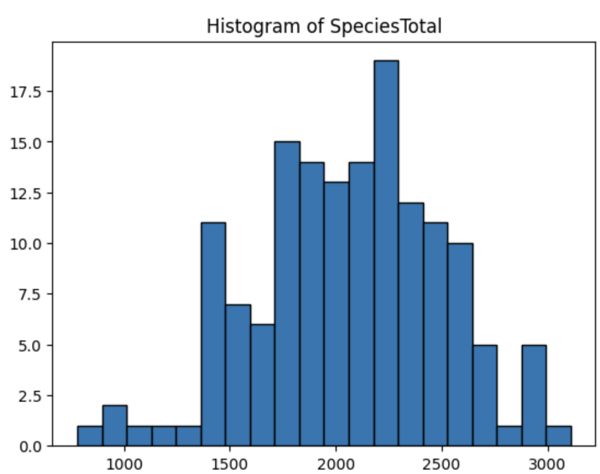

Figure 5. Histogram of Mourning Dove Total Abundance

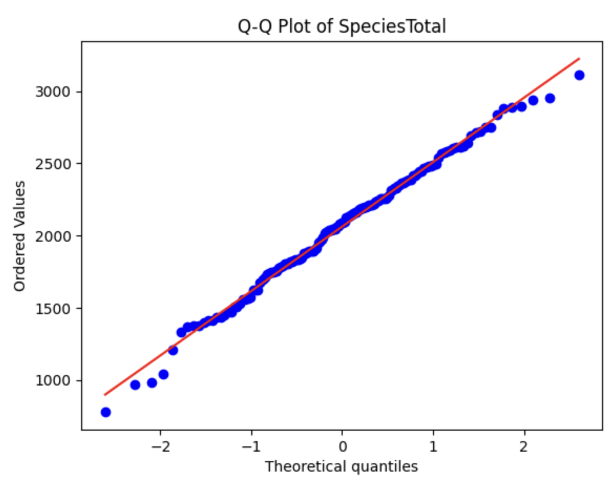

Figure 6. Q-Q Plot for normality

We combined the datasets such that we now had one unified dataset that contains the sightings and emissions for each state by year. Using this dataset, we had 165 data points (rows) to use for our analysis.

3. Statistical Model(s)

Our unified dataset contains annual Mourning Dove abundance estimates and the corresponding volumes of greenhouse gas emissions, specifically carbon dioxide (CO_2), fluorinated gases, net carbon stock changes from land use and forestry (LULUCF), methane (CH_4), and nitrous oxide (N_2O), for each of the five selected U.S. states from 1990 to 2019. The central objective of this analysis is to explore whether variation in greenhouse gas emissions is statistically associated with changes in Mourning Dove abundance over time.

To investigate this relationship, we employed a multiple linear regression model where the dependent variable is SpeciesTotal, representing Mourning Dove abundance. The independent variables include annual emissions of five types of greenhouse gases, along with fixed effects for U.S. states to control for time-invariant state-level differences. The model statement is specified as follows:

$$\text{SpeciesTotal}_{it} = \beta_0 + \beta_1 \text{CO}_{2_{it}} + \beta_2 \text{Fluorinated}_{it} + \beta_3 \text{LULUCF}_{it} + \beta_4 \text{CH}_{4_{it}} + \beta_5 \text{N}_2\text{O}_{it} + \beta_6 \text{Year}_t + \gamma_i + \epsilon_{it} \quad (1)$$

- SpeciesTotal_{it} is the Mourning Dove abundance in state i and year t,
- $\text{CO}_{2_{it}}$ denotes carbon dioxide emissions,
- Fluorinated_{it} denotes fluorinated gas emissions,
- LULUCF_{it} denotes net carbon stock changes from land use and forestry,
- $\text{CH}_{4_{it}}$ denotes methane emissions,
- N_2O_{it} denotes nitrous oxide emissions,
- Year_t captures the linear effect of time,
- γ_i captures state-level fixed effects,
- ϵ_{it} is the error term.

This model statement allows us to examine how variation in greenhouse gas emissions may be associated with changes in Mourning Dove abundance over time, while controlling for unobserved differences between states. In the next section, we present the regression results and interpret the findings in the context of our exploratory research objectives.

4. Results

Using Python StatsModels to perform regression, we evaluated the relationship between greenhouse gas emissions, state level variation, year, and Mourning Dove abundance across selected US states. The results are displayed in Figure 7. The model explained approximately 25% of the variance in bird sightings ($R^2 = 0.251$, Adj. $R^2 = 0.198$), suggesting a modest degree of explanatory power. This is unsurprising, as ecological variables and models tend to have many sources of variation that are extremely difficult to pinpoint.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           SpeciesTotal   R-squared:                       0.251
Model:                            OLS   Adj. R-squared:                  0.198
Method:                 Least Squares   F-statistic:                     4.670
Date:                Fri, 09 May 2025   Prob (F-statistic):           9.63e-06
Time:                        01:16:40   Log-Likelihood:                -1104.5
No. Observations:                 150   AIC:                             2231.
Df Residuals:                     139   BIC:                             2264.
Df Model:                          10
Covariance Type:            nonrobust
===============================================================================================================
                                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------------
Intercept                           -5553.0577   1.26e+04     -0.440      0.661   -3.05e+04    1.94e+04
C(State)[T.Arizona]                  -439.6260   1103.383     -0.398      0.691   -2621.210    1741.958
C(State)[T.Ohio]                    -1284.0583    623.496     -2.059      0.041   -2516.820     -51.297
C(State)[T.Oklahoma]                 -466.7222    646.269     -0.722      0.471   -1744.511     811.066
C(State)[T.Wisconsin]                  13.7502    629.216      0.022      0.983   -1230.321    1257.822
Carbon_dioxide                          7.9537      2.430      3.273      0.001       3.149      12.759
Fluorinated_gases                      39.5574     22.262      1.777      0.078      -4.459      83.574
Land_use_and_forestry_carbon_stock_change   14.9173     17.996      0.829      0.409     -20.664      50.499
Methane                                24.6262     18.886      1.304      0.194     -12.715      61.968
Nitrous_oxide                         -32.9332     69.374     -0.475      0.636    -170.097     104.231
Year                                    3.4614      5.917      0.585      0.560      -8.238      15.161
==============================================================================
Omnibus:                        0.607   Durbin-Watson:                   1.163
Prob(Omnibus):                  0.738   Jarque-Bera (JB):                0.612
Skew:                           0.150   Prob(JB):                        0.737
Kurtosis:                       2.910   Cond. No.                     7.86e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 7.86e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 7. Regression Results using StatsModels

Among the predictors, two variables were statistically significant at the 5% alpha level. The state indicator for Ohio showed a significant negative association with species counts ($\beta = -1284.05$, $p = 0.041$), indicating that, holding everything constant, Ohio recorded approximately 1284 fewer sightings compared to the reference state of Alabama. For the purposes of this research paper, we will not focus on the state-level fixed effects, as our primary research question lies in understanding the relationship between greenhouse gas emissions and bird populations. Carbon dioxide emissions exhibited a positive relationship with bird sightings ($\beta = 7.95$, $p = 0.001$), implying that increased emissions correspond to higher reported sightings, although causality cannot be inferred from this association. Additionally, fluorinated gases were almost significant at the 5% level and were also positively associated with sightings ($\beta = 39.55$, $p = 0.078$), though with wider confidence intervals, suggesting greater variability in the estimates.

The positive correlations observed between carbon dioxide and fluorinated gases with bird sightings may initially seem counterintuitive, as these emissions are typically associated with environmental degradation. However, these findings do not necessarily imply a causal benefit of emissions on wildlife populations. Rather, they may reflect underlying patterns of wildlife or human activity. For example, the higher emissions may cause some impact on wildlife habitats while not directly impacting the wildlife itself. Certain areas can sometimes support higher wildlife abundance or density due to resource availability or fewer natural predators. Additionally, it may reflect the resilience of the Mourning Dove as a species, showing they are more adaptable than other vulnerable bird populations. Thus, the observed associations may be indicative of other effects rather than direct ecological benefit.

Other predictors, including methane, nitrous oxide, and land use carbon stock change, did not reach statistical significance. The intercept and several other state indicators did not show statistically meaningful deviations from the baseline.

Taken together, these results highlight some regional and environmental factors that may correlate with Mourning Dove sightings, while also indicating the need for more robust modeling approaches, such as generalized linear models, to account for overdispersion and residual nonnormality observed in subsequent analyses. Many shortcomings of this model are discussed in the following section.

5. Model Validation and Generalizations

The Ordinary Least Squares (OLS) regression model is widely used to estimate the relationship between a dependent variable and multiple independent variables. For the OLS estimates to be considered the Best Linear Unbiased Estimates (BLUE) and for the associated inferential statistics (such as t-tests and F-tests) to be valid, several key assumptions must be met. These assumptions include linearity in parameters, no perfect multicollinearity among parameters, independence of error terms, homoscedasticity (constant variance) of error terms, normality of error terms, and other assumptions. When these assumptions are violated, the reliability of the coefficient estimates, their standard errors, and the overall model fit can be compromised, potentially leading to invalid conclusions. We now test these assumptions and detail any concerns and potential violations in the current model.

One critical assumption for the reliability of the OLS model is the normality of the error terms. This was visually inspected using a histogram of the residuals (Figure 8) and a Quantile-Quantile plot of the residuals against a theoretical normal distribution (Figure 9). The histogram of residuals exhibits a roughly symmetric, bell-shaped distribution centered around zero. The Q-Q plot also shows that the residuals closely follow the diagonal reference line. There do not appear to be any systematic deviations from the line, although the outlier points may indicate a deviation. While OLS estimates are unbiased even with non-normal errors, it has an impact on the standard errors, confidence intervals, and tests reported by the model.

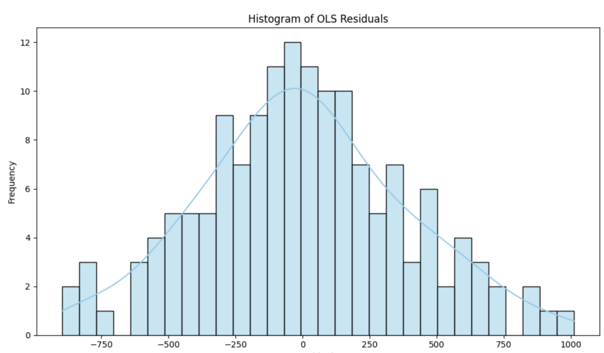

Figure 8. Histogram of Residuals

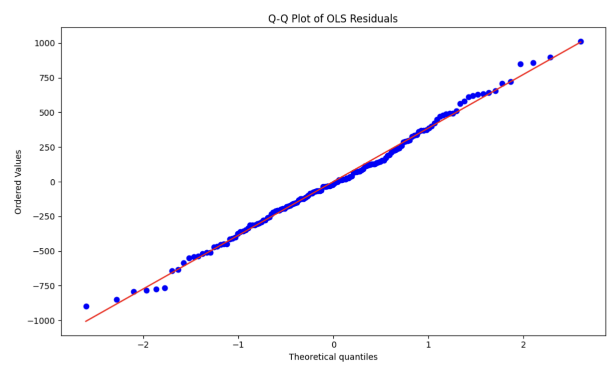

Figure 9. Q-Q Plot for normality

A limitation of this model is the potential impact of positive autocorrelation in the residuals, as suggested by a Durbin-Watson statistic of 1.163 found in Figure 7, which is below the standard value of 2. This indicates that the error terms may not be independent, a violation of a core OLS assumption. This can lead to an underestimation of the true variance of the coefficients. Consequently, this may result in inflated t-statistics and p-values, potentially causing variables such as Carbon dioxide ($p = 0.001$) to appear statistically significant when they might not be if the autocorrelation were appropriately addressed.

Further analysis of the diagnostic plots reveals potential issues with heteroscedasticity (non-constant variance) in the residuals. Specifically, the Residuals vs. Fitted Values plot (Figure 10) shows a potentially non-random scatter of points around the zero line. The smoothed conditional mean (red line) should be linear about the zero line, but in actuality has curves, suggesting that the variance of the residuals may change across different levels of the fitted values. If heteroscedasticity is present, the standard errors reported by the model would be inefficient and biased, thereby affecting the reliability of the coefficients.

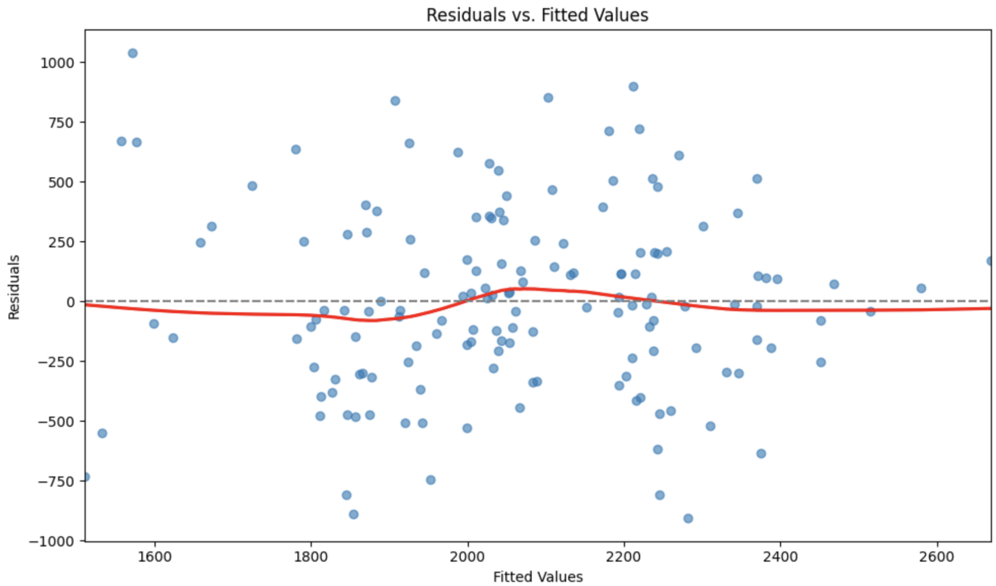

Figure 10. Residuals vs. Fitted Values

Additionally, the model exhibits indications of multicollinearity, highlighted by the high condition number of 786000. This suggests that some of the independent variables are correlated with one another. Multicollinearity may cause minor alterations in the input data to produce dramatically different estimated effects, reducing the model's generalizability to other data. Although multicollinearity does not bias the coefficient estimates themselves, it can affect their variances. This can lead to wider confidence intervals and can make it challenging to precisely determine the individual impact of each predictor on the total bird abundance. This change in variance could explain why variables such as Fluorinated Gases and Nitrous Oxide display relatively large coefficients but do not achieve statistical significance. A high condition number can also be a symptom of poor scaling among the independent variables. Because some emission values might be in thousands or millions, while other emission values might be in a smaller range, this disparity can artificially inflate the condition number due to the numerical properties involved in OLS estimation. Therefore, the observed multicollinearity might be partly due to scaling issues, in addition to any underlying correlations between the variables themselves. Standardizing the predictor variables could be a helpful step to help reduce scaling-based problems and investigate more fundamental structural multicollinearity.

Finally, a major concern is that the relationship between the greenhouse emissions and the Mourning Dove abundance is not linear in its parameters. We inherently assume that each unit increase in a given greenhouse gas emission results in a constant additive effect on Mourning Dove abundance. However, ecological systems are often have much more complex dynamics that may not be linear. For example, the impact of an emission might exhibit diminishing returns, or there could be thresholds beyond which it causes more dramatic change. If the underlying data does have non-linear relationships, the current linear model would be misleading. This misspecification could lead to biased and inconsistent coefficient estimates and potentially misleading statistical inferences. Figure 10 may be a result of non-linearity, suggesting that the linear model might not fully represent the complex relationship between emissions and Mourning Dove abundance.

In the future, many improvements could enhance the model's validity and the reliability in inference. For autocorrelation, employing methods such as Generalized Least Squares (GLS) with corrections for the error terms could be beneficial. For heteroscedasticity, recalculating standard errors using robust standard errors may be appropriate. To address multicollinearity, more effort on variable selection and experiment setup could prove helpful, along with using techniques such as Principal Component Regression (PCR). Finally, investigating alternative models and including potential non-linear relationships or interaction terms might improve the model's robustness.

6. Conclusion

This study dove into exploratory analysis in order to investigate the potential association between greenhouse gas emissions and the abundance of Mourning Doves across five selected U.S. states from 1990 to 2019. Utilizing a multiple linear regression model, we sought to determine if variations in carbon dioxide, fluorinated gases, land use and forestry carbon stock change, methane, and nitrous oxide emissions correlated with changes in Mourning Dove populations, while controlling for state-level fixed effects and year.

Our findings indicate a complex and not entirely intuitive picture. The model demonstrated modest explanatory power suggesting that the selected greenhouse gas emissions, along with state and year effects, account for about a quarter of the observed variance in Mourning Dove abundance. Notably, carbon dioxide emissions showed a statistically significant positive association with Mourning Dove sightings, and fluorinated gas emissions also exhibited a positive, although marginally nonsignificant, relationship. Conversely, some state-level effects, such as the negative coefficient for Ohio compared to Alabama, were significant, while others did not show statistically significant associations with dove abundance in this model. The positive correlation with certain emissions is counterintuitive, and as discussed, likely reflects complex underlying ecological factors or correlations with other unmeasured variables rather than a beneficial impact on the birds.

The limitations of this study, as detailed in the model validation section, must be carefully considered when interpreting the results. Issues with the OLS assumptions such as positive autocorrelation in the residuals, heteroscedasticity, multicollinearity, and the inherent assumption of linear relationships, all affect the generalizability of our findings. There may also exist more localized effects where the impact of emissions on habitats and dove populations could be more pronounced. The availability of local level Mourning Dove data contrasts with the broader scale of state-level emissions data, presenting a challenge for analysis.

Future research should aim to address these limitations. Better data collection and methodology is a crucial first step towards answering questions of this caliber. Moreover, exploring non-linear relationships and interaction terms could provide much deeper insights into the complex relationships between emissions and wildlife.

In conclusion, while this study intially identified some statistically significant associations between certain greenhouse gas emissions and Mourning Dove abundance at the state level, the relationships are not straightforward and the model faces several limitations. The findings underscore the complexity of assessing environmental impacts on widespread and adaptable species like the Mourning Dove and highlight the need for continued research, data collection, and modeling approaches to draw more definitive conclusions. This analysis may serve as an initial step for future ecological studies in this domain.

Appendix

References

Dinges, A. J., Szymanski, M. L., & Parent, C. J. (2022). Effects of weather and landscape use on mourning dove population trends in North Dakota. *Wildlife Society Bulletin*, *46*, e1346. https://doi.org/10.1002/wsb.1346

Noble Research Institute. (2022, September 1). Resource management aids mourning dove migration. *Noble Research Institute*. Retrieved May 9, 2025, from https://www.noble.org/regenerative-agriculture/resource-management-aids-mourning-dove-migration/

Sauer, J. R., Link, W. A., Kendall, W. L., & Dolton, D. D. (2010). Comparative Analysis of Mourning Dove Population Change in North America. *The Journal of Wildlife Management*, *74*(5), 1059–1069. http://www.jstor.org/stable/40665180

eBird Status and Trends, Cornell Lab of Ornithology. (2022). *Mourning Dove (Zenaida macroura) Weekly Abundance Map* (Data version: eBird Status and Trends 2024). Retrieved May 9, 2025, from https://science.ebird.org/en/status-and-trends/species/moudov/abundance-map-weekly