29 Apr 2022

Predicting annual surface water area using climate and waterfowl data in Missouri's Lower Grand River watershed

Zachary J. Loken, Louisiana State University, School of Renewable Natural Resources, Baton Rouge, LA 70808, USA

Correspondence: Zachary J. Loken, Louisiana State University, School of Renewable Natural Resources, Baton Rouge, LA 70808, USA. Email: zloken1@lsu.edu

ABSTRACT

Waterfowl provide ecological, societal, and economic significance worldwide. Because of the values these birds provide, their populations have been monitored in North America for nearly a century. Recent concerns emerging about decreased abundances of ducks in places they have traditionally existed in large congregations have been related through research to warming winter temperatures and decreased surface water area (SWA). This study explores the association between relative abundance trends, habitat metrics, and winter climatic variables for a portion of Missouri's Lower Grand River watershed. Outcomes from this work indicate the potential for assessing annual surface water trends as a function of various habitat metrics, relative duck abundance, and climatic variables.

KEY WORDS dabbling ducks, diving ducks, eBird, Google Earth Engine, Mississippi flyway, Missouri, Spectral Mixture Analysis, winter.

This manuscript has been prepared in accordance with the Journal of Wildlife Management standards.

INTRODUCTION

Waterfowl are ecologically, societally, and economically significant worldwide (Grado et al. 2001, 2011; Green and Elmberg 2014; Guerry et al. 2015; Notaro et al. 2016). For example, an estimated 1.5 million citizens in the United States spent approximately 15 million hours hunting waterfowl in 2011. In addition, they contributed nearly $3 billion to the national economy through purchases of equipment and spending associated with travel, food, and lodging (U.S. Fish and Wildlife Service [USFWS] 2015). Because of the values provided by these birds, their populations have been systematically monitored in North America annually on breeding and non-breeding grounds since 1935—most notably through the Midwinter Waterfowl Survey (MWS; Stott and Olson 1972, Eggeman and Johnson 1989, Heusmann 1999, Pearse et al. 2008, Masto et al. 2020).

Concerns have recently emerged regarding observed decreased abundances of ducks using traditional southern wintering grounds in the Mississippi flyway (Moorman 2020). Research suggests these decreased abundances of ducks may be related to winter warming and related climatic phenomena (Meehan et al., 2021) that result in decreased SWA on a landscape. Thus, the objective of the current study was to explore the association between relative abundance trends, habitat metrics, and winter climatic variables. I tested the hypothesis that habitat and climatic water balance variables can accurately predict the SWA present on a habitat patch—a metric commonly associated with site selection and habitat use of waterfowl. Understanding these relationships will become increasingly important as water scarcity continues to restructure the timing and availability of wetland resources (Donnelly et al., 2022).

STUDY AREA

The region of interest includes Missouri's Lower Grand River watershed, residing within the Central Rivers floodplain (NAWMP's Geographies of Greatest Continental Significance to North American Waterfowl, 2012). The study area (Appendix A, Figure 1), located approximately 80 miles northeast of Kansas City, MO, spans over 82,000 acres and includes aquatic, cultivated, forested, grassland, riparian, wet meadow, and wetland habitat types (Table 1). In addition, this area has a high concentration of wetland restoration easements (WREs; 75 easements totaling 10,806 acres) that are managed to support the energetic needs of waterfowl; as such, it provides essential habitat for waterfowl in the Mississippi flyway during winter migration.

Habitat Class	Frequency	Sum of Acres
Aquatic	376	702.210855
Cultivated	1134	30728.32081
Forest - Hardmast	917	5010.36894
Forest - Riparian	2985	18662.29665
Grand River	137	1215.628684
Grassland	1765	6389.797436
Levee	803	1386.062737
Rotational Wetland	31	659.854642
Water Bodies	326	3152.040627
Wet Meadow	58	1690.688663
Wetland	1274	12901.85831

Table 1. Summary of patch frequency and total sum of acres for each habitat class found in the study area.

METHODS

Waterfowl observation data

The eBird Basic Data set (EBD) from the Cornell Laboratory of Ornithology was used to align annual waterfowl abundance with habitat and climate trends. The Auk package (Strimas-Mackey et al., 2018) was used to extract regional EBD count and presence data for 34 species in the order *Anseriformes* from 1984 to 2020. Most waterfowl observations used in the analysis were acquired post-2015 due to the relatively recent release of eBird. The extracted EBD observation data were filtered to only include observations for 34 species in the order *Anseriformes* (Table 2) common to Missouri recorded between October 1st and December 31st—a general window of when waterfowl are wintering in this region. The final filtered dataset contained 120,962 observations.

Table 2. Waterfowl species, by scientific name

Aix sponsa	*Aythya valisineria*	*Mareca strepera*
Anas acuta	*Branta canadensis*	*Melanitta americana*
Anas crecca	*Branta hutchinsii*	*Melanitta deglandi*
Anas platyrhynchos	*Bucephala albeola*	*Melanitta perspicillata*
Anas rubripes	*Bucephala clangula*	*Mergus merganser*
Anser albifrons	*Clangula hyemalis*	*Mergus serrator*
Anser caerulescens	*Cygnus buccinator*	*Oxyura jamaicensis*
Anser rossii	*Cygnus columbianus*	*Spatula clypeata*
Aythya affinis	*Dendrocygna autumnalis*	*Spatula cyanoptera*
Aythya americana	*Dendrocygna bicolor*	*Spatula discors*
Aythya collaris	*Lophodytes cucullatus*	
Aythya marila	*Mareca americana*	

Table 2. Waterfowl species included in the filtered EBD. These species are common to Missouri during winter migration (October 1 – December 31).

The filtered eBird dataset was then subset only to include relevant variables. Furthermore, a new column containing the observation year was created by isolating the year from the observation date. This data set was then converted into an ESRI shapefile using longitude and latitude as the (x, y) coordinates. Once converted to a shapefile, the EBD data could be read into R as a simple feature (SF) object—a class type that handles spatial features. In addition to loading the EBD dataset, the study area parcel boundaries were loaded into an SF object. Finally, both SF objects were transformed into the same coordinate reference system (EPSG: 3857) and then plotted onto a map to ensure correct projection (Figure 1).

Surface water area

SWA was derived by running a sub-pixel classification—spectral mixture analysis (SMA)—on Landsat 5 Thematic Mapper, Landsat 7 Enhanced Thematic Mapper Plus, and Landsat 8 Operational Land Imager satellite imagery in Google Earth Engine (GEE; Donnelly et al., 2019). SWA was classified annually for each parcel in the study area from 1984 to 2020. After running the SMA, each parcel's annual SWA was converted into hectares and exported as a CSV to be imported into RStudio.

Once in R, this dataset was converted to a narrow format data frame with one column containing the year and another containing annual hectares of surface water for each of the 9,806 parcels in the study area. The SWA data was then grouped by habitat class to sum the total hectares of surface water annually for each habitat type present. Lastly, the average SWA was calculated for each habitat type in a given year. It was necessary to group and summarize SWA as outlined above to join this data with the other datasets used for analysis; the SWA outputs from GEE are not georeferenced and therefore could not be spatially joined with the other datasets.

Climate data

TerraClimate is a monthly climate and climatic water balance dataset spanning back to 1958 (Abatzoglou et al., 2018). This data was extracted from the study area annually from 1984 to 2020 using GEE. Climate and climatic water balance variables extracted include actual evapotranspiration (mm), climate water deficit (mm), reference evapotranspiration (mm), precipitation accumulation (mm), runoff (mm), soil moisture (mm), snow water equivalent (mm), wind-speed at 10m (m/s), minimum temperature (Celsius), and maximum temperature (Celsius). Minimum and maximum temperatures were extracted as mean temperatures from October 1 to December 31, while all other climate and climatic water balance variables were extracted as total sums. After extracting the TerraClimate data using GEE, it was imported into R to be merged with the other datasets.

Data analysis

Random forest analysis of the multivariate dataset was used to determine the most important predictor variables for average annual SWA. Following an approach outlined by Donnelly et al. (2021), variables were measured annually from 1988 to 2020 as 5-year running means beginning in 1984. Normalizing variable estimates as 5-year running means filtered out the short-term variability in factors influencing hydrologic conditions (Rajagopalan and Lall, 1998). After running the random forest analysis, confidence intervals at the 95% confidence level ($\alpha = 0.05$) were calculated and plotted as boxplots to assess relative variable importance (VIMP; Appendix A, Figure 2).

Based on the random forest results, a best subsets regression using Akaike's Information Criterion (AIC) scores was used to determine which combination of the most important variables resulted in a linear regression model with the highest predictor capability (Table 3). The final multiple linear regression equation included the variables of the best subsets regression model (i.e., lowest AIC). After running the regression analysis, the distribution of residuals was plotted to confirm that the data is normally distributed (Appendix A, Figure 3). Homoscedasticity was confirmed by plotting the residuals as a function of the fitted data (Appendix A, Figure 4).

Predictor variables with a *p*-value greater than 0.05 (Table 4) were removed from the model to test if AIC scores decreased. If model performance did not improve by removing variables with a *p*-value greater than 0.05, they were added back into the final model equation.

RESULTS

Based on the random forest results, the most important predictor variables for average annual SWA (Appendix A, Figure 2) were cover type (median = 137.2), actual evapotranspiration (median = 53.6), minimum winter temperature (median = 33.5), runoff (median = 32.9), precipitation accumulation (median = 22.2), and soil moisture (median = 59.9). An AIC best subsets regression was used to determine which variable combinations produced the top-performing model; the model with the lowest AIC score (Table 3) has the highest predictor capabilities. The top model included six variables: cover type (seven factor levels), actual evapotranspiration, precipitation accumulation, runoff, soil moisture, and minimum temperature.

Model	**Predictor Variables**	**R^2**	**Adjusted R^2**	**AIC**
1	Cover Type	0.5625	0.5586	4054.9191
2	Cover Type, Actual Evapotranspiration	0.6806	0.6772	3875.3613
3	Cover Type, Actual Evapotranspiration, Minimum Winter Temperature	0.7706	0.7678	3686.2727
4	Cover Type, Actual Evapotranspiration, Minimum Winter Temperature, Runoff	0.783	0.7799	3656.371
5	Cover Type, Actual Evapotranspiration, Minimum Winter Temperature, Runoff, Precipitation Accumulation	0.7946	0.7913	3626.6959
6*	Cover Type, Actual Evapotranspiration, Minimum Winter Temperature, Runoff, Precipitation Accumulation, Soil Moisture	0.7959	0.7923	3624.7964
7	Cover Type, Actual Evapotranspiration, Minimum Winter Temperature, Runoff, Precipitation Accumulation, Soil Moisture, Duck Count by Species	0.796	0.792	3626.6489

Table 3. AIC best subsets regression outputs summary table used to perform variable selection for the multiple linear regression final model. An Asterix (*) denotes the final model that was chosen.

The best subset of predictor variables determined through an AIC best subsets regression was then analyzed using multiple linear regression (Equation 1, Table 3).

$$y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \in_i$$

Equation 1. The multiple linear regression equation. α is the intercept, β_j is the mean parameter for each of the ten additional groups, subscript *j* indicates each of the levels within β, and ε_i is the error term.

Coefficient	Term	Mean	Std. error	t-value	p (> |t|)
Cover Type: Cultivated (Intercept)	α	-15.344706	2.233524	-6.87	1.70E-11 *
Cover Type: Riparian	β_1	-3.629988	1.078431	-3.366	0.000814 *
Cover Type: Grassland	β_2	-14.453487	2.300878	-6.282	6.68E-10 *
Cover Type: Levee	β_3	-13.509068	1.049341	-12.874	< 2e-16 *
Cover Type: Wet Meadow	β_4	7.704455	1.023052	7.531	2.00E-13 *
Cover Type: Wetland	β_5	15.657402	0.669438	23.389	< 2e-16 *
Actual Evapotranspiration	β_6	0.3118877	0.017506	17.816	< 2e-16 *
Precipitation Accumulation	β_7	-0.057581	0.009598	-5.999	3.53E-09 *
Runoff	β_8	0.089235	0.012791	6.976	8.49E-12 *
Soil Moisture	β_9	0.013811	0.00705	1.959	0.050591
Minimum Temperature (°C)	β_{10}	-3.831444	0.277242	-13.82	< 2e-16 *

Table 4. Outputs summary from final multiple linear regression model. Variables that are statistically significant at the 95% confidence level ($\alpha = 0.05$) are indicated using an Asterix (*).

Using the model coefficients (Table 4), the final linear regression equation (Equation 2) can be written as:

$$y_i = -15.345 - 3.630\beta_1 - 14.453\beta_2 - 13.509\beta_3 + 7.704\beta_4 + 15.657\beta_5 + 0.311\beta_6 - 0.058\beta_7 + 0.089\beta_8 + 0.014\beta_9 - 3.831\beta_{10} + \in_i$$

Equation 2. The final linear regression model with intercept and mean parameters defined. Y_i is the response, α is the intercept (control group), β_1 is the change in mean response for riparian habitat, β_2 is the change in mean response for grassland habitat, β_3 is the change in mean response for levee habitat, β_4 is the change in mean response for wet meadow habitat, β_5 is the change in mean response for wetland habitat, β_6 is the change in mean response for actual evapotranspiration, β_7 is the change in mean response for precipitation accumulation, β_8 is the change in mean response for runoff, β_9 is the change in mean response for soil moisture, β_{10} is the change in mean response for minimum winter temperature, and ε_i is the error term.

The minimum residual value of the final model was -11.434, the median was -2.038, and the maximum was 17.403. Based on the residuals, which are roughly symmetrical and somewhat centered around 0 (Figure 3), it can be determined that the model's predictions are slightly off by a median value of -2.038. Therefore, a multiple linear regression model is suitable for this data. Moreover, the adjusted R2 value of 0.7923 indicates that the independent variables in the model can predict 79.23% of the variance in the measure of average SWA. Lastly, the residual standard error (RSE) value is 5.534, corresponding to a 32.6% error rate (found by dividing the RSE by the mean response variable).

DISCUSSION

Habitat type, actual evapotranspiration, precipitation accumulation, runoff, and minimum temperature are significant ($p < 0.05$) predictors of annual average SWA for this region. Total SWA on a landscape is associated with site selection and habitat use of waterfowl and therefore is becoming an increasingly important metric for assessing waterfowl habitat suitability. We now live in an age where it is possible to use widely available and publicly accessible data (e.g., eBird, TerraClimate, Landsat) to predict suitable landscapes for waterfowl. As a result, habitat managers can better target conservation endeavors to provide sufficient waterfowl habitat as wetland ecosystems continue to change. Conversely, the economic, societal, and ecological effects will be omnipresent if suitable waterfowl habitat availability continues decreasing at unprecedented rates.

REFERENCES

Abatzoglou, J. T., S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch. 2018. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. Scientific Data 5:170191

Donnelly, J. P., D. E. Naugle, D. P. Collins, B. D. Dugger, B. W. Allred, J. D. Tack, et al. 2019. Synchronizing conservation to seasonal wetland hydrology and waterbird migration in semi-arid landscapes. Ecosphere 10:1-12.

Donnelly, J. P., J. N. Moore, M. L. Casazza, and S. P. Coons. 2022. Functional wetland loss drives emerging risks to waterbird migration networks. Frontiers in Ecology and Evolution 10:844278.

Donnelly, J. P., S. L. King, J. Knetter, J. H. Gammonley, V. J. Dreitz, B. A. Grisham, et al. 2021. Migration efficiency sustains connectivity across agroecological networks supporting sandhill crane migration. Ecosphere 12:3543.

Eggeman, D. R., and F. A. Johnson. 1989. Variation in effort and methodology for the Midwinter Waterfowl Inventory in the Atlantic Flyway. Wildlife Society Bulletin 17:227-233.

Grado, S. C., K. M. Hunt, C. P. Hutt, X. T. Santos, and R. M. Kaminski. 2011. Economic impacts of waterfowl hunting derived from a state-based mail survey. Human Dimensions of Wildlife 16:100-113.

Grado, S. C., R. M., Kaminski, I. A. Munn, and T. A. Tullos. 2001. Economic impacts of waterfowl hunting on public lands and at private lodges in the Mississippi Delta. Wildlife Society Bulletin 29:846-855.

Guerry, A. D., S. Polasky, J. Lubchenco, R. Chaplin-Kramer, G. C. Daily, R. Griffin, M. Ruckelshaus, I. J. Bateman, A. Duraiappah, T. Elmqvist, et al. 2015. Natural capital and ecosystem services informing decisions: from promise to practice. Proceedings of the National Academy of Sciences 112:7348-7355.

Heusmann, H. W. 1999. Let's get rid of the Midwinter Waterfowl Inventory in the Atlantic Flyway. Wildlife Society Bulletin 27: 559-565.

Masto, N. M., R. M. Kaminski, P. D. Gerard, B. E. Ross, M. R. Kneece, and G. L. Wilkerson. 2020. Aerial strip-transcet surveys: indexing autumn-winter waterbird abundance and distribution in South Carolina. Journal of Southeastern Association of Fish and Wildlife Agencies 8:89-100.

Meehan, T. D., R. M. Kaminski, G. S. LeBaron, N. L. Michel, B. L. Bateman, and C. B. Wilsey. 2021. Half-century winter duck abundance and temperature trends in the Mississippi and Atlantic Flyways. The Journal of Wildlife Management 1-10.

Moorman, T. 2020. Are waterfowl migrations changing? Ducks Unlimited Magazine (March/April 2020):46-51.

North American Waterfowl Management Plan [NAWMP]. 2012. Process for developing the 2012 NAWMP map – geographies of greatest continental significance to north American waterfowl. U.S. Fish and Wildlife Service, USA; Ducks Unlimited, Inc., USA; North Dakota Game and Fish Department, USA; Ducks Unlimited Canada, Canada.

Notaro, M., M. Schummer, Y. Zhong, S. Vavrus, L. Van Den Elsen, J. Coluccy, and C. Hoving. 2016. Projected influences of changes in weather severity on autumn-winter distributions of dabbling ducks in the Mississippi and Atlantic flyways during the twenty first century. PLOS ONE 11:30167506.

Pearse, A. T., S. J. Dinsmore, R. M. Kaminski, and K. J. Reinecke. 2008. Evaluation of an aerial survey to estimate abundance of wintering ducks in Mississippi. Journal of Wildlife Management 72:1413-1419.

Rajagopalan, B., and U. Lall. 1998. Interannual variability in western US precipitation. Journal of Hydrology 210:51-67.

Stott, R. S., and D. P. Olson. 1972. An evaluation of waterfowl surveys on the New Hampshire coastline. Journal of Wildlife Management 36:468-477.

Strimas-Mackey, M., E. Miller, and W. Hochachka. 2018. Auk: eBird data extraction and processing with AWK. R package version 0.33. Vienna: R Core Team.

U.S. Fish and Wildlife Service [USFWS]. 2019. Economic impact of waterfowl hunting in the United States: addendum to the 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. USFWS, Falls Church, Virginia, USA.

APPENDIX A

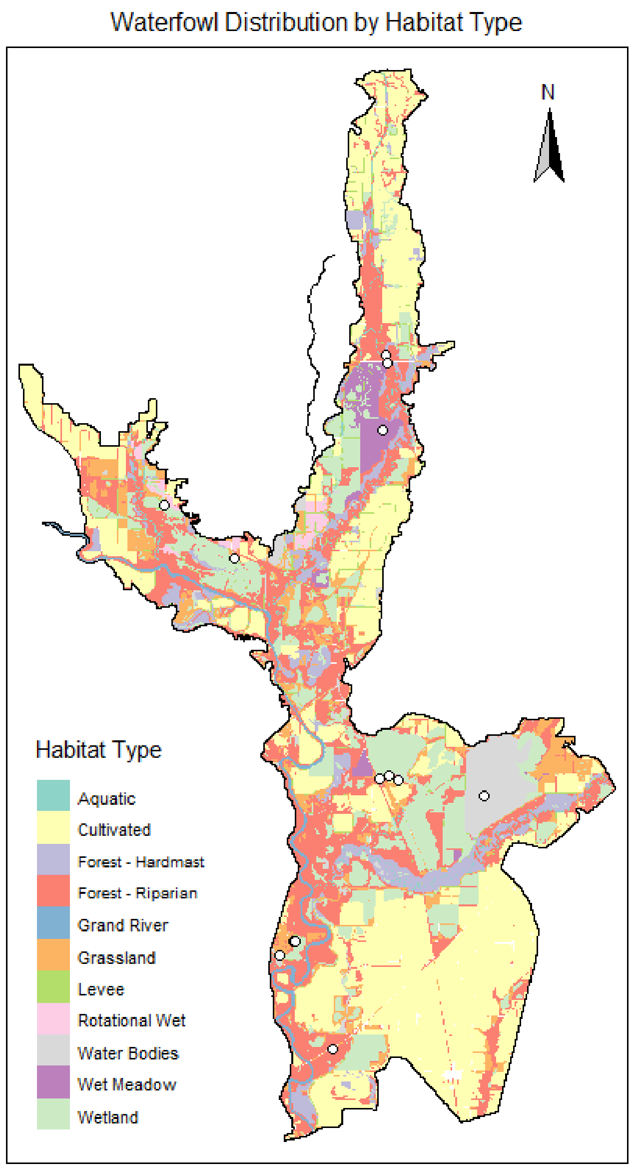

Figure 1. Study area situated within the Lower Grand River watershed, approximately 80 miles northeast of Kansas City, MO. Colors represent different habitat types in the study area. White circles are waterfowl observations in study area.

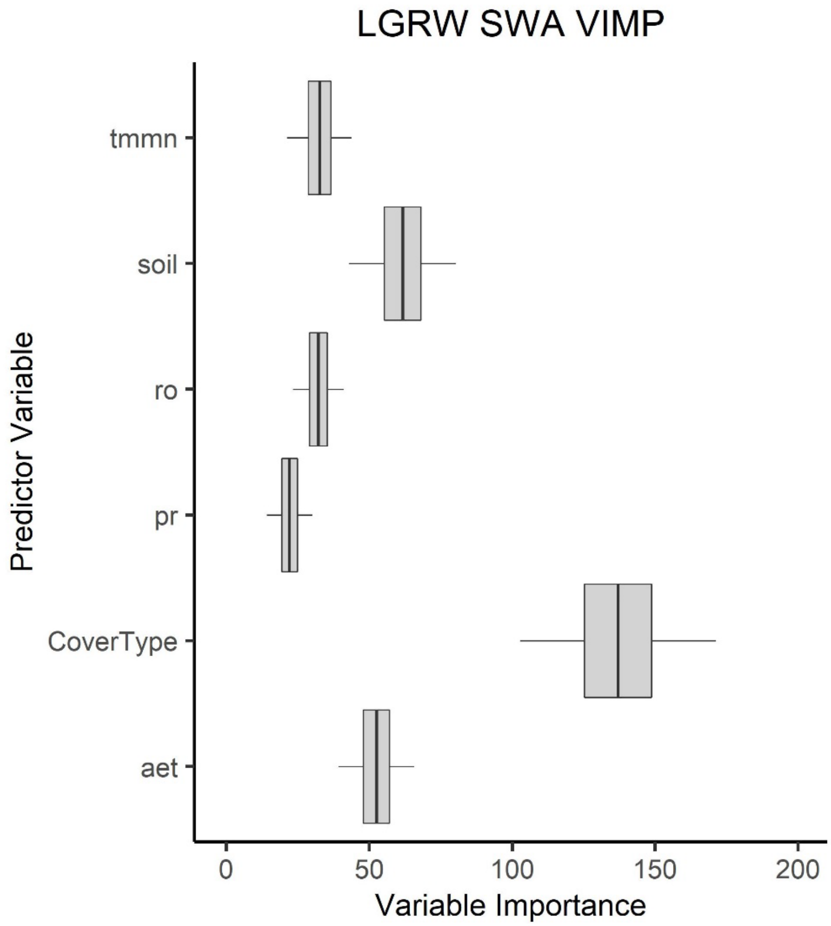

Figure 2. The horizontal black line indicates the median, top and bottom of the boxes represent the 25th and 75th interquartile range, and the whiskers extend to the most extreme data points which are no more than 1.5 times the interquartile range away from either edge of the box.

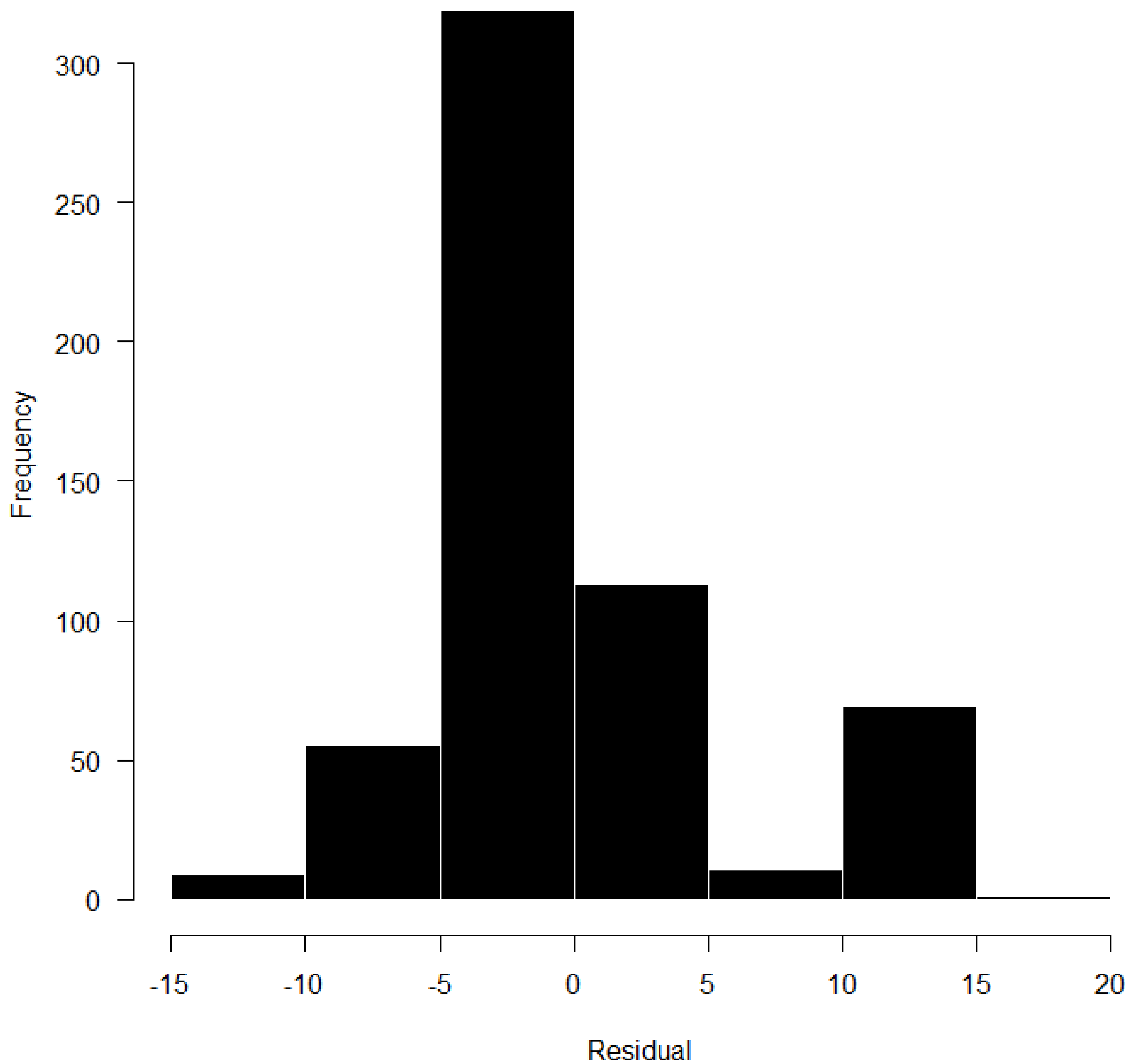

Figure 3. Histogram showing distribution of the residuals, which assume an approximately normal distribution.

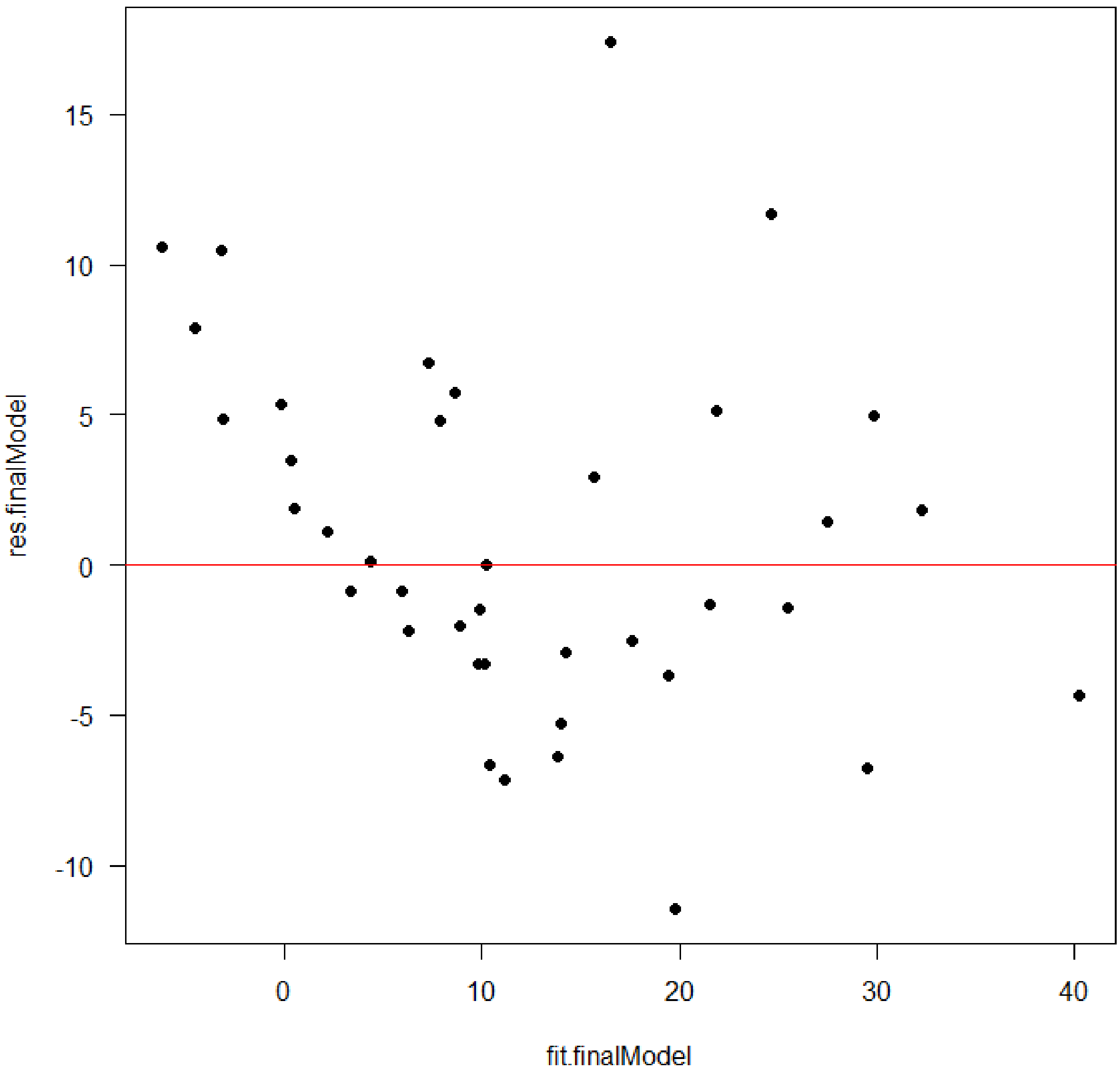

Figure 4. Scatter plot showing the residuals as a function of the fitted data. The width of the scatter as fitted values increase is roughly the same, which indicates homoscedasticity.