Predicting Microplastic Concentrations in the Ocean

Danielle Lamb

DSC680-T301 Applied Data Science

Professor Catherine Williams

Business Problem

The sluggish degrading process of plastic has caused immense amounts of plastic to end up in the world's oceans. This plastic has broken down into microscopic pieces that are ingested and inhaled by fish across all bodies of water. Farmed fish and wild caught fish have been found to have microplastic in their gills and gut, which has been proven to travel into parts of their body typically eaten by humans. Without the understanding of global microplastic impacts, the more detrimental the consequences could be in the future. This project aims to outline the current trend of microplastic concentrations in the surface ocean and predict future trends if plastic use continues with the frequency of today.

Background

Since the invention of plastic in 1907, the production of synthetic packaging materials has skyrocketed. No longer reliant on breakable packaging such as glass and paper, almost all manufactures turned to the more flexible polyethylene. However, the very thing that makes plastic protective, moldable, and durable, also makes it difficult to break down. Some plastics can take up to tens of thousands of years to fully degrade. So, after 114 years of production and no full degradation, a world full of plastic is left. Currently, in a single year "around 500 billion PET bottles are sold every year," (The Age of Plastic: From Parkesine to pollution, 2019) and only around 12.2% of plastic products were recycled in the United States in 2018 (EPA, n.d.). Globally recycling efforts could be much more efficient than current processes and more biodegradable packaging research could be done.

While each person making a personal goal to reduce and recycle plastic products, an individual level change is not what is needed. The leaders of plastic production in 2020 were companies such as Coca-Cola, Nestlé, Colgate-Palmolive, and Unilever. Many companies are starting to understand these impacts of plastics around the world and have led research into new packaging and the reduction of plastic all together. With the influence of these companies choosing to change methods, supply chain must also change with them to stay in business. As more businesses are pressured by consumers and other businesses, overall a cleaner way of creating and delivering goods can be reached. Only with these changes can the reduction of plastics into landfills and the ocean happen.

Data

Data for this project was found through the National Oceanic and Atmospheric Administration. Samples from surface waters in the North Atlantic Ocean and Caribbean sea were taken between 1986 and 2008. Date, latitude, longitude, pieces per km^2, and normalized pieces were logged. The normalized feature was empty, so the feature was removed. There are a total of 3737 non-zero rows. The data was recorded off the ship the SSV Corwith Cramer.

Methods

The statistical program R was used to analyze the data. To rule out missing data, a plot from the library DataExplorer was used. After, a histogram was used to analyze the distribution of the data. The pieces per kilometer squared is extremely right skewed. A boxplot of pieces per kilometer squared was made to further analyze any outliers. Outliers were removed beyond the extremes of the whiskers. Monthly averages were taken to even out observation time periods.

The time series was run through a Dickey-Fuller Test to test for stationarity. After, an auto-correlation function (ACF) and partial auto-correlation function (PACF) was plotted to compare present observations to past observations. To determine the status of heteroscedasticity, a Breusch-Pagan test was performed. A moving average was plotted over the original with an order of 13 to evaluate averages using observations 6 past and 6 future observations.

The model chosen for this project is the autoregressive integrated moving average model (ARIMA). The AR order was determined to be 3, the degree of difference 1, and the MA order zero. A coeftest was used to evaluate the effectiveness of the model. In addition, a Ljung-Box test was conducted to assess the dependence of residuals of the model. Values for one year and two years into the future were predicted. To rule out seasonality, an additional ARIMA model was made with a seasonality AR order of 3. There was no significant difference so the original model was kept.

Assumptions

The one major assumption of the ARIMA model is stationarity. An option for testing stationarity is the augmented Dickey-Fuller Test. Using this test on the data resulted in a p-value of 0.01, meaning that the data can say with strong confidence that they can reject the null hypothesis of non-stationary data.

Analysis

Initial summary of data indicates a total of 5772 rows of data. There were a total of 2035 rows with zero pieces per kilometer squared. These rows were removed. Boxplots indicated the possible presence of outliers because of the skewness (Figure 1), which was confirmed when looking at the the .stats method on the boxplot which pulls the outliers outside the tails. There was a total of 434 outliers based on the boxplot calculations (Figure 1). These outliers were removed.

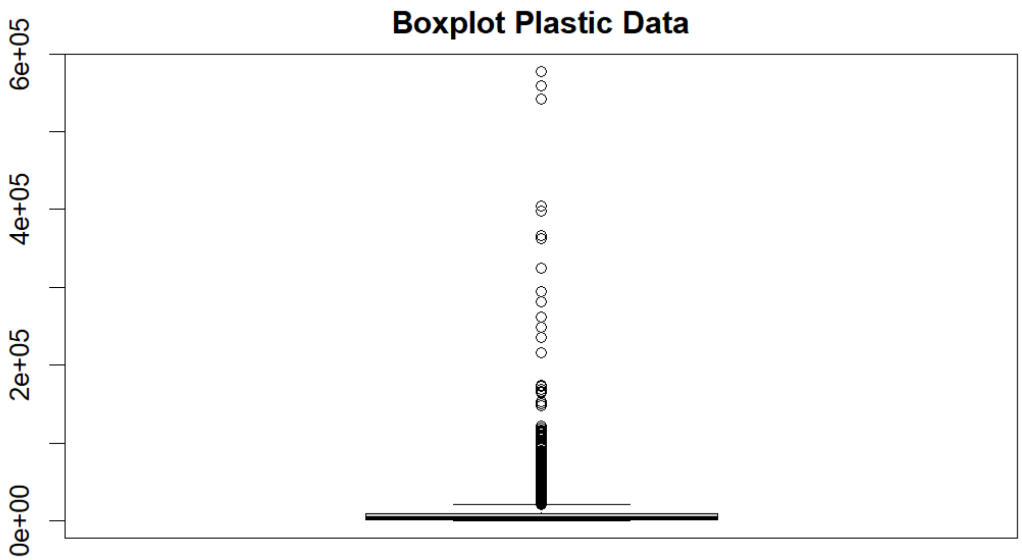

Figure 1: Boxplot of pieces of plastic per kilometer squared.

After the time scale was evened out using monthly averages, a total of 201 observations remained. Monthly observations from 1991 to 2008 were then determined to be stationary based on the augmented Dickey-Fuller test. To have a consistent lag order, monthly averages were taken. A general moving average was calculated and showed some sign of possible seasonality. This could be due to the organic nature of the ocean and its currents. In addition, river rates could change between seasons with greater input of water in the spring as the mountain snow melts.

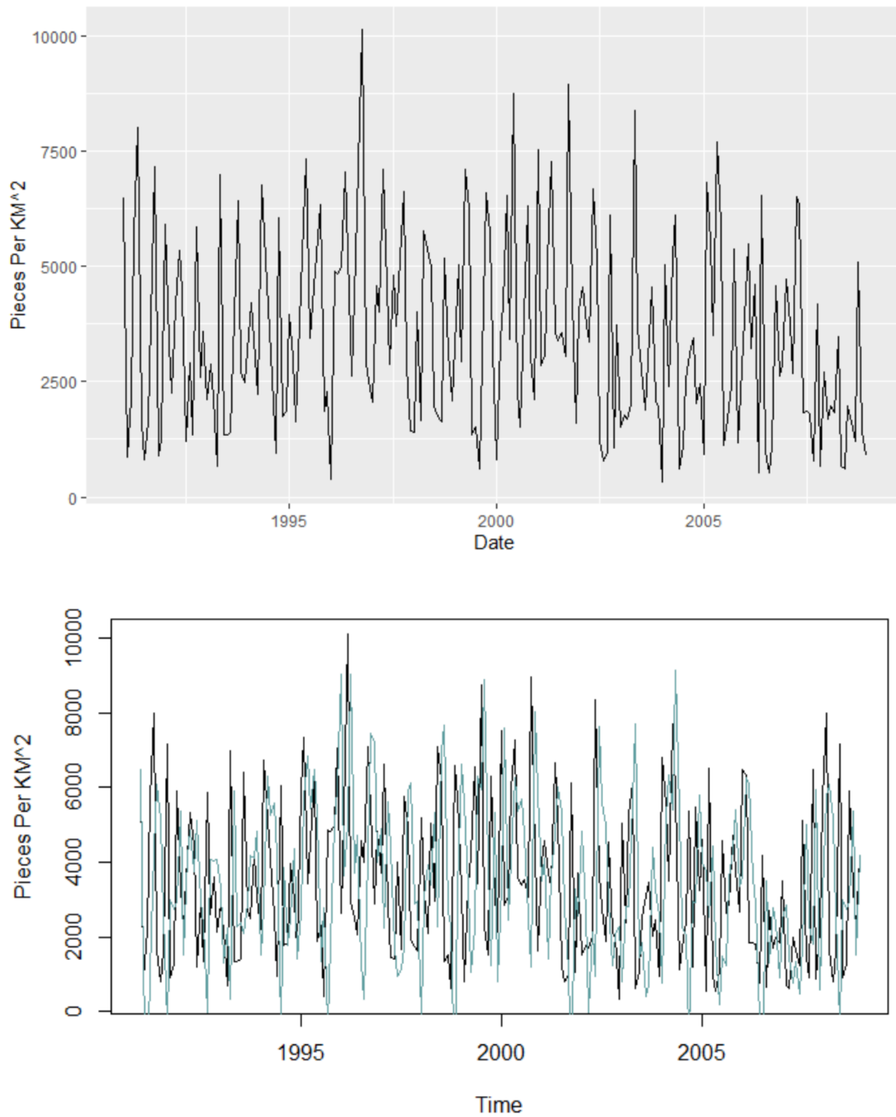

Figure 2: Line graph of date vs. pieces per kilometer squared upper. General trend is slightly positive and linear. Lower ARIMA model predictions overlayed on the actual data.

While some predictions were overfitted, a majority of the values were accurately predicted. With this model future values can be predicted to estimate what the trend of microplastics in the ocean could be. Forecasted values predicted an increase in pieces per kilometer squared over the next year (Figure 3).

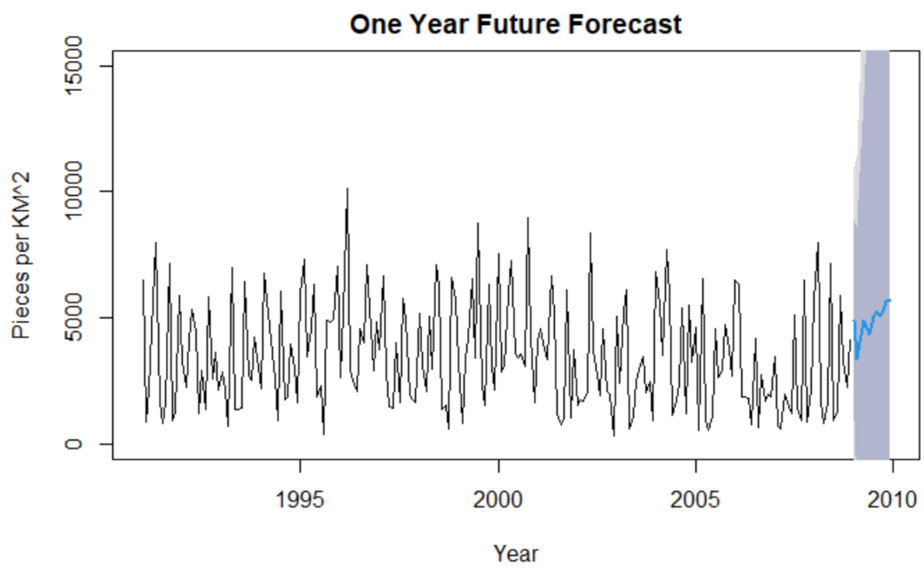

Figure 3: Prediction of pieces of plastic per kilometer squared. General increase in pieces is predicted.

Conclusion

This data describes a fairly steady level of microplastic levels with numerous outliers. Further analysis will be performed on whether these outliers are impactful in modeling versus unnecessary. Overall the trend is slightly positive but much less so than was thought. A fairly accurate model was created with the possibility of deployment to help understand the global plastic pollution trend.

Limitations

This model is based off data located along the east coast. If this model were to be applied globally, changes might have to be made to keep the effectiveness. The lack of knowledge on R-based time series analysis drew out the modeling process as more education and research was needed. Additional data collected in a more current time frame to continue predicting into further future dates.

Challenges

Reliable data that covered the information that was needed for this analysis was difficult to come by. Data cleaning also proved to lose much of that data. Adjusting the model to the correct hyperparameters was also challenging.

Future Uses

Knowing the changing amounts of microplastics will spark motivation for groups to improve their plastic use. In addition, new methods of safely breaking down plastic instead of throwing it by the wayside can help reduce the overall amount of plastics entering the landfills and potentially the world's waterways. It serves as a motivation to beat the trendline and enforce habits that cause the concentrations to drop below the proposed amount.

Implementation Plan

This current model uses only data from the North Atlantic Ocean. To be useful globally, data would have to be fit from numerous bodies of water around the world. Understanding the true origin of the plastics would be difficult, however, physical oceanography in the form of ocean currents and wind flow can predict a general pattern of how plastic flows around the globe. In addition, this is an every changing process, with different inputs and values daily. Better collection methods, potentially off buoys would allow for a more consistent collection across multiple locations.

Literature Cited

EPA. (n.d.). Facts and Figures about Materials, Waste and Recycling. Retrieved from https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling

Lavender Law, Kara; Morét-Ferguson, Skye; Maximenko, Nikolai; Proskurowski, Giora; Peacock, Emily; Hafner, Jan; Reddy, Christopher (2020). Plastic collected from SSV Corwith Cramer in North Atlantic Subtropical Gyre from 1986-10-15 to 2008-12-20 (NCEI Accession 0211007). NOAA National Centers for Environmental Information. Dataset. https://www.ncei.noaa.gov/archive/accession/0211007.

The Age of Plastic: From Parkesine to pollution. (2019, October 11). Retrieved from https://www.sciencemuseum.org.uk/objects-and-stories/chemistry/age-plastic-parkesine-pollution