Forecasting Mauna Loa Carbon Dioxide Emissions from Average Monthly Data

Stephen Yan,* Matthew Evans,† Michelle Floyd,‡ and Shelby Tanous§

(Dated: c. April 2016)

Mauna Loa, Hawaii is home to some of the most extensive CO_2 data in the world. With readings dating back to 1958, the Mauna Loa Observatory has been a reliable source for carbon dioxide readings and an innovative site in climate change research. Using monthly mean data from the past decade, our goal is to create a model for predicting carbon dioxide monthly means in the coming months, particularly the month of April. Our work will not only specify a model and predict these values, but will also explore the scientific implications and justifications of this model and prediction.

I. MODEL CHOICE

We chose the ARIMA model because of its prevalence in academia that dealt with carbon dioxide emission modeling. Scholarly examples include Chris Tsokos and Shou Shih's statistical analysis of Mauna Loa's CO_2 emissions. With more and more emphasis being put on climate change and predicting the future, scientists and statisticians are working in unison to discover the effects of our increased use of fossil fuels and natural resource consumption as it relates to carbon dioxide output and environmental effects.

Chris Tsokos and Shou Shih at the University of South Florida performed a similar statistical analysis of Mauna Loas CO_2 emissions and took into account trends and seasonal effects in order to forecast future values in 2008. The aim of his study was to develop two statistical models for the carbon emissions and atmospheric Carbon Dioxide. For our study we are only trying to predict the mean of CO_2 values in Mauna Loa for April, but we are still able to learn valuable takeaways from his other model in an attempt to improve our model. Their study used monthly emissions, similar to what we are using in our project. However, his data was from 1981 to 2003 and our data is more recent and runs until February 2016.

It is important to create an intial visualization of the CO_2 data. We accomplish this in a time series analysis package, *ITSM*. From the time series plot of CO_2 concentrations, upward trend and seasonality are evident.

Tsokos and Shih used the multiplicative seasonal autoregressive integrated moving average model (`ARIMA`) which is represented by $\texttt{ARIMA}(p, d, q) * (P, D, Q)$. For this model p is the order of the autoregressive process, d is the order of regular differencing, q is the order of the moving average process, P is the order of the seasonal autoregressive process, D is the order of the seasonal differencing, Q is the order of the seasonal moving average process, and s is the seasonal period. For our project we used an $\texttt{ARIMA}(p, d, q)$ model. We were able to remove the seasonality and trend by using `ITSM`.

The article discusses how it is important to have a good methodology when developing the model. First we need to determine the seasonal sub-index s as 12 because after looking at the time series plot, the peaks and troughs are during the same time of the year each year. This shows the time series is behaving as a periodic function with a cycle of 12 months. We then tested the stationarity of the process. We removed the seasonality and trend through the use of differencing and classical estimation. For this time series, we chose to use differencing to remove the trend and seasonality. By removing the trend and seasonality, the model becomes stationary. This also gave us the order of differencing. We can then use Akaikes information criterion (AIC) to select an appropriate model.

To evaluate the quality of the model, Tsoko and Shih used the residual model $r_t = x_t - \bar{x}$ where x_t and $\bar{x}$ are the observed and predicted values and evaluated his model by observing the mean, variance, standard deviation, and standard error of the residuals. Afterwards they found the model dealing with Mauna Loa could be best categorized as a $\texttt{ARIMA}(2, 1, 0) * (2, 1, 1)_{12}$ process (Shih 6).[1] They were able to forecast future recordings of CO_2 and when the values were later released and realized, they could compare their estimate to the actual value by using residuals. The University of South Floridas model is a good model because the residuals are small and are isolated around the zero axis, something we hope for our model as well (Shih 10). Overall, Tsoko and Shihs predictions are encouraging, and their model was helpful as we evaluated the time series and searched for a model in order to forecast the CO_2 values in April 2016

Springer Texts Introduction to Time Series and Forecasting gives a well thought out approach to creating an `ARIMA` model with a time series.[2] First, we differenced the data with $lag = 12$ in order to eliminate the seasonality, since the data was monthly. Next, we subtracted the mean from our model to ensure a stationary series to create a model. Using the Autofit Estimation tool in `ITSM`, orders for `AR` and `MA` were chosen based on AICC. The `ARMA` model obtained through autofit estimation was:

* sjyan@cs.unc.edu
† matthewe@live.unc.edu
‡ mtfloyd@live.unc.edu
§ tanous@live.unc.edu

[1] Shih, Shou Hsing, and Chris P. Tsokos. "Prediction Models for Carbon Dioxide Emissions and Atmosphere." The International Journal Neural 16 (2008). Web.

[2] Brockwell, Peter J., and Richard A. Davis. Introduction to Time Series and Forecasting. New York: Springer, 2002. Print.

$$(1-B)X_t - 0.2758X_{t-1} + 0.7166X_{t-2} - 0.6421X_{t-3}$$

$$= Z_t + 0.3258Z_{t-1} + 0.9997Z_{t-2}$$

with White-Noise Variance $= 0.14813$

After applying a $p = 3$ and $q = 2$ to an `ARMA` model, along with including the fractionally integrated aspect of the model, it was time to run an Autofit Estimation again to find the fractionally integrated `ARMA` model (`ARIMA`). A new p and q was decided as $(0, 0)$, which gave our model an estimation of

$$(1-B)^{.4828}X_t = Z_t$$

II. RESULTS, ANALYSIS, & FORECASTING

After discovering our model, we had a couple choices for forecasting available in `ITSM`, but not a stable one for `ARIMA`. `ARMA` forecasting was the logical choice given our `ARIMA` model. Using the model as the basis for forecasting, we predicted the next 2 monthly means for CO_2 emissions in Mauna Loa with an estimate of **ppm=406.13** in the month of April.

On inspection of the ACF of the residuals of the model, we see that most correlations fall within the threshold, suggesting that the residuals are behaving like white noise.

FIG. 1. ACF of Residuals of model

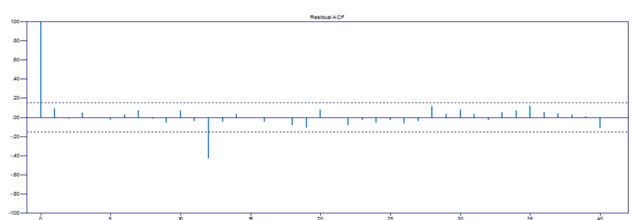
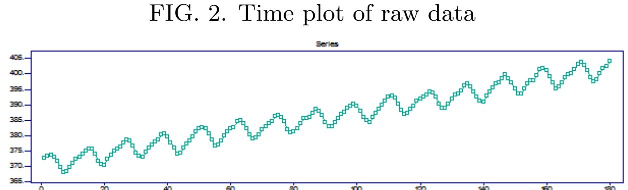
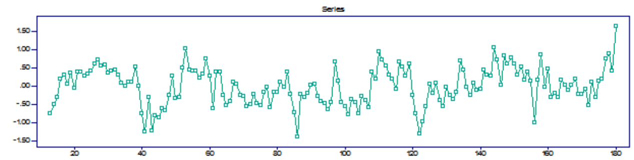


FIG. 2. Time plot of raw data



On inspection of the monthly mean emission data over the course of 2001 to present day, we obtain the time plot Fig. 1 as our ground truth. We observe seasonality and a linear trend in our data and proceed to follow our model by stationarizing, the first step in consolidating an appropriate forecast model. An important part of `ARIMA` and `ARMA` is to first difference the data with seasonal period of 12 months, $(1 - B^{12})X_T$

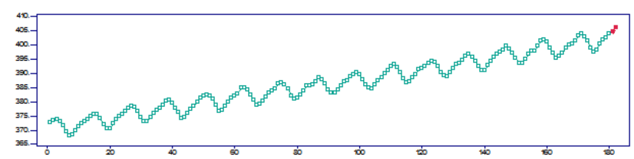
FIG. 3. Differenced Data



Next we forecast the next two monthly emission means and construct a 95% confidence interval using the `ARMA` model through autofit estimation using maximum likelihood.

We obtain the interval $[405.29, 406.97]$ PPM.

FIG. 4. `ARIMA`$(0, 0.4828, 0)$ forecasted March/April with 95% Confidence



We also construct a 99% confidence interval using this same model and obtain the interval $[405.03, 407.23]$ PPM ($\hat{\mu} = 406.13$ PPM) and conclude our prediction as 406.13 PPM for the monthly mean emissions for April.

FIG. 5. `ARIMA`$(0, 0.4828, 0)$ forecasted March/April with 99% Confidence

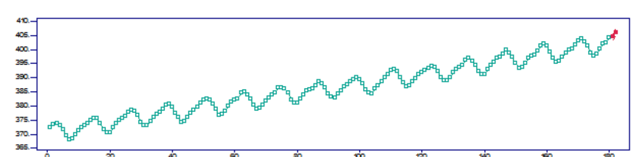


FIG. 6. Zoomed 95% Confidence Prediction

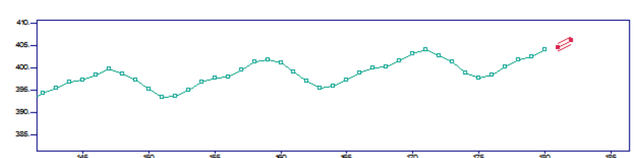
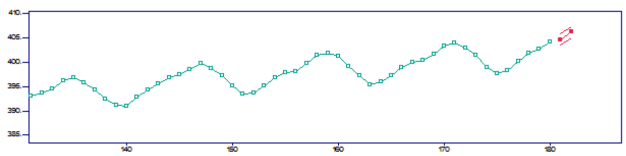


FIG. 7. Zoomed 99% Confidence Prediction



III. SCIENTIFIC JUSTIFICATION

Global concentration of carbon dioxide is a primary cause of global warming as CO_2 is responsible for about 63% of global warming attributed to greenhouse gases (Braasch).[3] These CO_2 levels result from both natural emissions (i.e. volcanoes), as well as human emissions. Although these levels vary throughout the year, they have significantly increased over time.

According to the seasonal trend in our model, atmospheric CO_2 levels dip the lowest during September and October, and they peak during May and June. Levels decline during the spring and summer months (from June to September) and rise during the fall and winter months (from October to May).

There is a scientific explanation for this trend. Decomposition occurs in the fall as microbes break down leaf litter and other dead plant material throughout the winter months. This decomposition process steadily increases atmospheric CO_2 levels, as microbes respire and produce CO_2. In the spring, as new plant growth begins to form and photosynthesis increases, plants draw CO_2 from the atmosphere, decreasing CO_2 levels into the summer months (Monroe).[4]

The Keeling Curve explains this pattern of a CO_2 decline during the months of photosynthesis and increase during the months of decomposition (Monroe). May marks this turning point from decomposition throughout the winter months to the photosynthesis process beginning in the spring months.

Although the northern and southern hemispheres maintain a reverse seasonal relationship, there is a much larger amount of land in the Northern Hemisphere (particularly concentrated in the dense forest areas of Siberia), which dominates the CO_2 cycle. Although photosynthesis also occurs in oceans (which comprises the majority of the Southern Hemisphere), marine photosynthesis has little effect on atmospheric CO_2, as very little CO_2 is actually released into the atmosphere (Monroe).

As you may notice, the seasonal trend from the data lags that of the mainland. So although Spring usually begins in March/April, it is not reflected in the data until May. It takes a few weeks for seasonal swings to circulate to Mauna Loas high latitude, thus creating the lag in the measurement at the observatory.

In addition to the seasonal trend, we consider the overall increasing trend, which is largely attributed to human emissions as a byproduct of societal development, technology, waste and fuel, and increase in population. As of March 2015, CO_2 levels reached 400 ppm (parts per million), measurably the highest concentration of CO_2 in over 800,000 years (Braasch). This measurement was taken from the average of all of NOAAs 40 CO_2 measuring sites. Scientists suggest that CO_2 levels need to be kept below 400-450 ppm to prevent irreversible damage and disastrous climate change effects (Braasch). Evidence of dangerously high CO_2 levels can already be seen through melting glaciers, rising sea levels, moving habitats, extinction of certain species, and sever weather catastrophes.

IV. CONCLUSIVE REMARKS & IMPROVEMENTS

We were able to develop a time series model taking into effect the trend and seasonality to predict Aprils estimate of Carbon Dioxide levels in Mauna Loa. We were able to represent the data using an `ARIMA` model by removing the upward trend and seasonality by differencing and then establishing the degrees of p for the autoregressive model and q for the moving average model through `ITSM`. We used observed recordings of CO_2 provided by the National Oceanic and Atmospheric Administration (NOAA) to develop our model. We then used `ARMA` to forecast Aprils levels using a lag of 2 (because our most recent data point is from February 2016) to forecast our prediction of 406.13 PPM. In `ITSM`, `ARIMA` forecasting is not readily available, but the `ARMA` forecasting tool was a great predictor for CO_2 in Mauna Loa.

In `ITSM`, `ARIMA` forecasting is not readily available - however there are tools available in R to perform this using the *forecast* package through `forecast.Arima()` and `auto.arima()`. We left it to the procedures in software to autofit our model after differencing our data and relied on the AICC model to trust that the model was appropriate. We could have further performed an Ljung-Box test as a portmanteau for verifying white-noise in the residuals of our models. We were limited by the API of `ITSM` - however the forecast results returned by an `ARMA` forecast were favorable. In the end, we could have been more robust in our procedure of selecting an appropriate model instead of a trial-and-error-esque approach.

[3] Braasch, Gary. "Mauna Loa Carbon Dioxide Levels." Mauna Loa Carbon Dioxide Levels. N.p., 2015. Web. 31 Mar. 2016.

[4] Monroe, Rob. "Why Does Atmospheric CO_2 Peak in May?" The Keeling Curve. N.p., 04 June 2013. Web. 01 Apr. 2016.