Prediction Models for Carbon Dioxide Emissions and the Atmosphere

Shou Hsing Shih and Chris P. Tsokos
Department of Mathematics and Statistics, University of South Florida
Tampa, Florida 33620

Abstract
The object of the present study is to develop statistical models for predicting the carbon dioxide emissions and the atmosphere in the United States. We used monthly emissions data from 1981 to 2003 that was collected by the Carbon Dioxide Information Analysis Center. For the carbon dioxide in the atmosphere, we used the data that was collected in Mauna Loa from 1965 to 2004 by the Scripps Institution of Oceanography. The developed statistical models take into consideration trends and seasonal effects. The quality of the prediction process is illustrated using the actual data.

Keywords – Time Series Forecasting, ARIMA, Multiplicative ARIMA, Carbon Dioxide, CO_2 Emissions, Atmospheric CO_2, Global Warming

1. INTRODUCTION

Global Warming is one of the most compelling and difficult problems facing our society. It is well understood that carbon dioxide, CO_2, along with temperature are the primary causes of global warming. The present study is concerned with developing analytical statistical models to predict CO_2. Jim Verhulst, Perspective Editor, St. Petersburg Times, writes, “Carbon dioxide is invisible- no color, no odor, no taste. It puts out fires, puts the fizz in seltzer, and is to plants what oxygen is to us. It is hard to think of it as a poison.” (Verhulst 2007). The United States is emitting approximately 5.91221 billion metric tons of carbon dioxide in the atmosphere, which makes us one of the World leaders. In addition to CO_2 in the atmosphere, we have CO_2 emissions that are related to gas, liquid, and solid fuels along with gas flares and cement production.

The aim of the present study is to develop two different statistical models for the carbon emissions and atmospheric carbon dioxide in the United States using historical data from the subject matter.

The CO_2 emissions data set that we used to develop the proposed model contains the monthly emissions data from 1981 to 2003. It was published by Carbon Dioxide Information Analysis Center (CDIAC), which is supported by the United States Department of Energy. The CDIAC is a well known organization, which responds to data and information requests from users worldwide investigating the greenhouse effect and global climate change. For detailed information, see (United States Environmental Protection Agency (EPA), 2004; Marland et al., 2003). A graphical presentation of the emissions data is given by Figure 1.1.

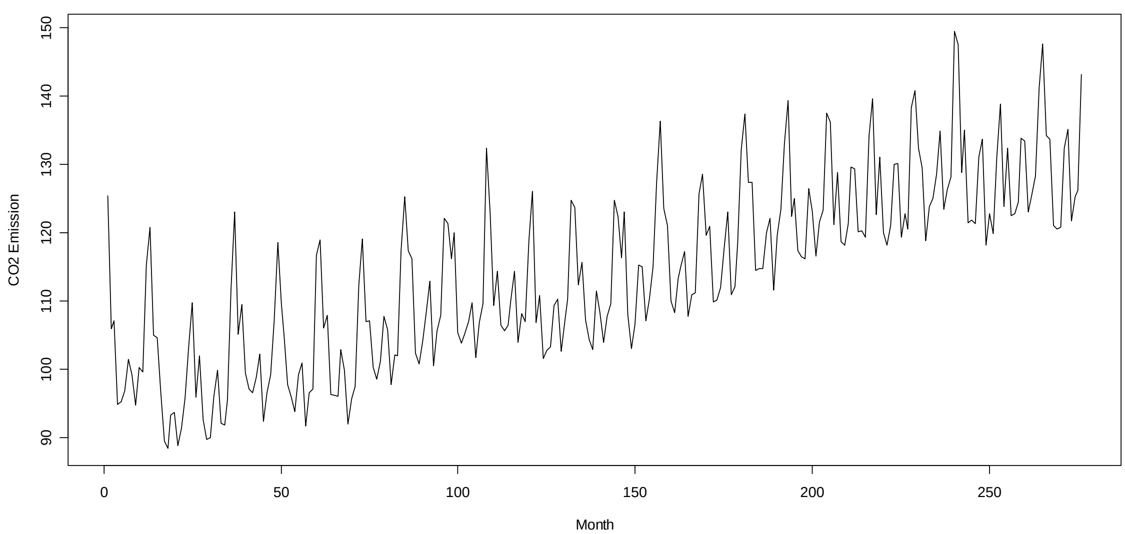

Figure 1.1 Time Series Plot on CO_2 Emission 1981-2003

The data set that we used to develop our second proposed model consists of monthly CO_2 concentrations in the atmosphere from 1958 to 2004. The data was collected in Mauna Loa by Carbon Dioxide Research Group, Scripps Institution of Oceanography, University of California. A map of geographical location of Mauna Loa is provided by Figure 1.2. At the earlier stage of our model building process, we spot several missing values in the early 1960s. To address this problem, we decided to use the data from 1965 to 2004, which is a period which contains no missing values. For additional information concerning the data set on CO_2 concentrations in the atmosphere, see (Bacastow, 1979; Bacastow & Keeling, 1981; Bacastow et al., 1980; Bacastow et al., 1985; Keeling, 1960; Keeling, 1984; Keeling, 1998; Keeling et al., 1976; Keeling et al., 1982; Keeling et al., 1989; Keeling et al., 1996; Keeling et al., 1995; Pales & Keeling, 1965; Keeling et al., 2002; Whorf & Keeling, 1998).

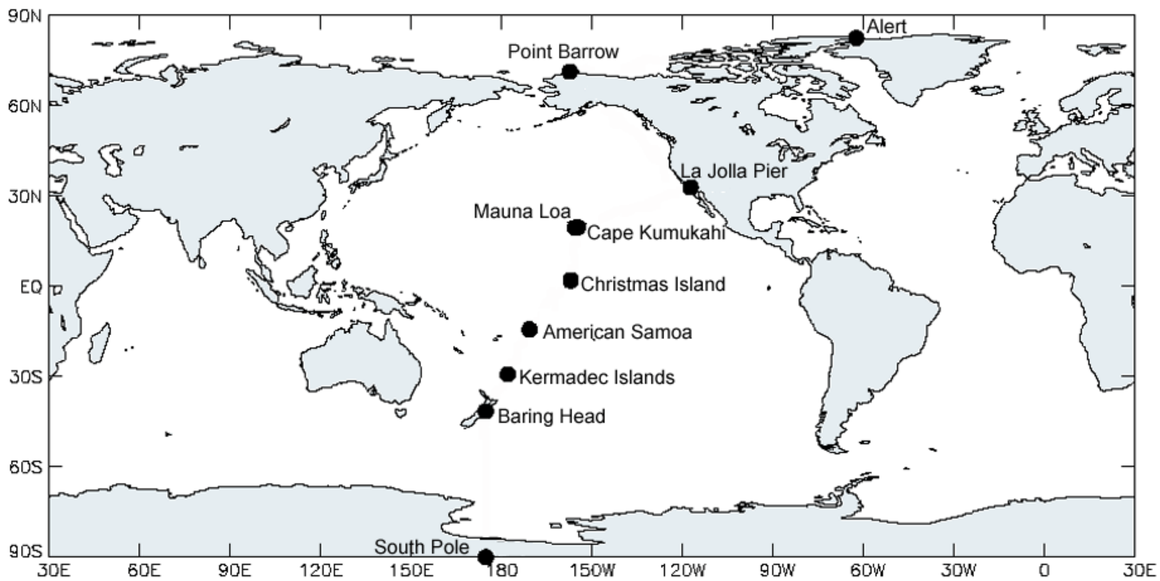

Figure 1.2 Geographical Location of Mauna Loa

A plot of the actual CO_2 concentration in the atmosphere is given by Figure 1.3. provides a visual presentation of the time series plot of CO_2 concentrations in the atmosphere.

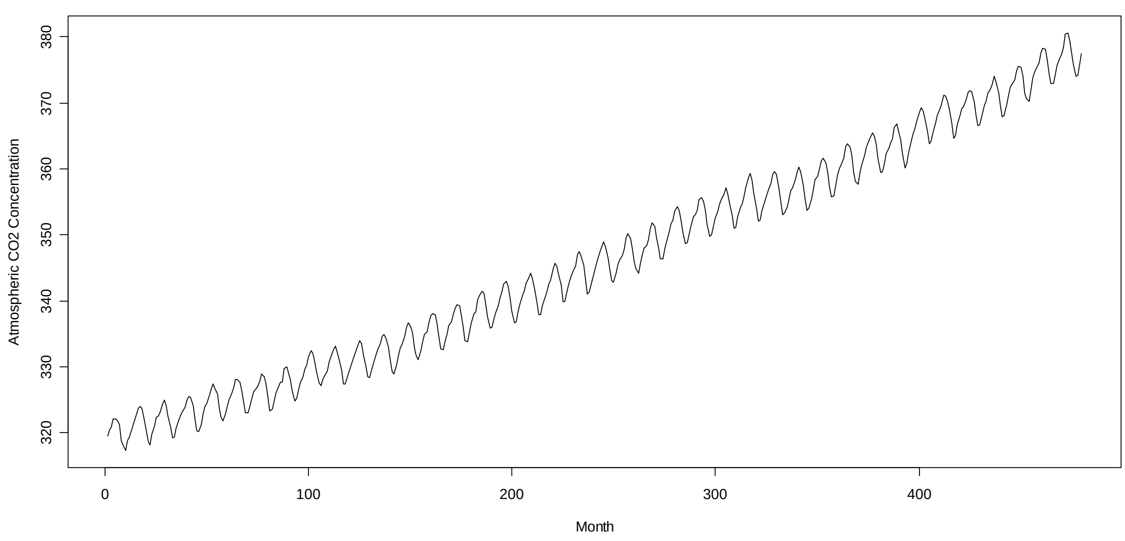

Figure 1.3 Time Series Plot for Monthly CO_2 in the Atmosphere 1965-2004

2. ANALYTICAL PROCEDURE

The multiplicative seasonal autoregressive integrated moving average, ARIMA model is defined by

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D x_t = \theta_q(B)\Gamma_Q(B^s)\varepsilon_t \tag{2.1}$$

where *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 the subindex *s* refers to the seasonal period. We shall denote the subject model by ARIMA $(p,d,q)\times(P,D,Q)_s$, and $\phi_p(B),\theta_q(B),\Phi_P(B^s),\Gamma_q(B^s)$ defined as follows:

$$\phi_p(B) = (1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$$

$$\theta_q(B) = (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$$

$$\Phi_P(B^s) = 1-\Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$$

and

$$\Gamma_Q(B^s) = 1-\Gamma_1 B^s - \Gamma_2 B^{2s} - \ldots - \Gamma_Q B^{Qs}.$$

The order of the multiplicative ARIMA model determines the structure of the model, and it is essential to have a good methodology in terms of developing the forecasting model. In the present study, we start with addressing the issue of the seasonal subindex *s*. After we examine the original data sets, shown by Figure 1.1 and 1.3, we note that both the monthly CO_2 emissions and the atmospheric CO_2 behave as a periodic function with a cycle of 12 months. Hence, we let the seasonal subindex $s=12$.

In time series analysis, one cannot proceed with a model building procedure without confirming the stationarity of a given stochastic realization. Thus, we test the overall stationarity of the series by using the method introduced by Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y in 1992, (Kwiatkowski et al., 1992).

Once the order of the differencing is identified, it is common for one ARIMA $(p,d,q)\times(P,D,Q)_s$ model that we have several sets of (p,q,P,Q) that are all adequately representing a given set of time series. Akaike's information criterion, AIC, (Akaike, 1974), is used to select the appropriate model. That is, we select the model of the, (p,q,P,Q), that produces the smallest AIC.

Another important aspect of the model selection process is to determine the seasonal differencing, *D*, that results in minimum AIC. For convenience, given below a brief summary of step-by-step procedure that we follow to develop the subject models:

- Determine the seasonal period *s*.
- Check for stationarity of the given time series $\{x_t\}$ by determining the order of differencing *d*, where $d = 0,1,2,...$ according to KPSS test, until we achieve stationarity.
- Deciding the order m of the process, for our case, we let $m = 5$ where $p + q + P + Q = m$.
- After (d, m) being selected, listing all possible configurations of (p, q, P, Q) for $p + q + P + Q \le m$.
- For each set of (p, q, P, Q), estimates the parameters for each model, that is, $\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q, \Phi_1, \Phi_2, ..., \Phi_P, \Gamma_1, \Gamma_2, ..., \Gamma_Q$.
- Compute the AIC for each model, and choose the one with smallest AIC.
- After (p, d, q, P, Q) is selected, we determine the seasonal differencing filter by selecting the smaller AIC between the model with $D = 0$ and $D = 1$.
- Our final model will have identified the order of (p, d, q, P, D, Q).

In order to evaluate the quality of the models we use several criteria. The residuals of the model, $r_t = x_t - \hat{x}_t$, where x_t and $\hat{x}_t$ are the actual value and predicted value, respectively. Mean of the residuals, $\bar{r} = \frac{\sum_{t=1}^{n} r_t}{n}$; Variance of the residuals, $S_r^2 = \frac{\sum_{t=1}^{n}(r_t - \bar{r})}{n-1}$; Standard Deviation of the residuals, $S_r = \sqrt{S_r^2}$; Standard Error of the residuals, $SE = S_r^2 / \sqrt{n}$; and Mean Square Error, $MSE = \frac{\sum_{t=1}^{n} r_t^2}{n}$.

3. DEVELOPMENT OF THE FORECASTING MODELS

CO_2 emissions data that is shown by Figure 1.1, contains a small upward trend, and the data repeats its pattern every 12 months. Thus, we let the seasonal subindex $S = 12$. Follow by the step-by-step procedure as we described above, we found the model that best characterizes the monthly emissions of the United States is an ARIMA$(1,1,2)\times(1,1,1)_{12}$ process, analytically given by

$$(1-\Phi_1 B^{12})(1-\phi_1 B)(1-B)(1-B^{12})x_t = (1-\theta_1 B-\theta_2 B^2)(1-\Gamma_1 B^{12})\varepsilon_t \tag{3.1}$$

On the other hand, the CO_2 in the atmosphere data has a more obvious upward trend, and the shape of its pattern is almost identical every year, as shown by Figure 1.3. Thus, we also set the seasonal period $S = 12$. We have identified that the model that best described the monthly CO_2 concentrations in the atmosphere is an ARIMA$(2,1,0)\times(2,1,1)_{12}$ process, analytically given by

$$(1-\Phi_1 B^{12}-\Phi_2 B^{24})(1-\phi_1 B-\phi_2 B^2)(1-B)(1-B^{12})x_t = (1-\Gamma_1 B^{12})\varepsilon_t \tag{3.2}$$

The actual statistical estimated models for (3.1) and (3.2) with the appropriate estimate of the weights are given by

$$\begin{aligned}\hat{CO}_{2_E} = {} & 1.5203x_{t-1} - 0.5203x_{t-2} + 1.0049x_{t-12} - 1.527749x_{t-13} + 0.5228495x_{t-14} \\ & - 0.0049x_{t-24} + 0.007449x_{t-25} - 0.002549x_{t-26} - 0.9988\varepsilon_{t-1} + 0.1234\varepsilon_{t-2} \\ & - 0.8523\varepsilon_{t-12} + 0.8512772\varepsilon_{t-13} - 0.10517\varepsilon_{t-14}\end{aligned} \tag{3.3}$$

and

$$\begin{aligned}\hat{CO}_{2_A} = {} & 0.6887x_{t-1} + 0.1989x_{t-2} + 0.1124x_{t-3} + 1.0759x_{t-12} - 0.74097x_{t-13} - \\ & 0.213997x_{t-14} - 0.12093x_{t-15} - 0.0683x_{t-24} + 0.047038x_{t-25} + \\ & 0.013585x_{t-26} + 0.00768x_{t-27} - 0.0076x_{t-36} + 0.005234x_{t-37} + \\ & 0.0015116x_{t-38} + 0.00085x_{t-39} - 0.8787\varepsilon_{t-12}\end{aligned} \tag{3.4}$$

for emissions, $\hat{CO}_{2_E}$, and atmosphere, $\hat{CO}_{2_A}$, respectively. We shall proceed to evaluate these models and illustrate the quality of both models in the next section.

4. EVALUATION OF THE PROPOSED MODELS

The proposed statistical model, (3.3), for CO_2 emissions in the United States is used to forecast the last one hundred recordings. A graphical comparison of the actual CO_2 and the predicted are given by Figure 4.1.

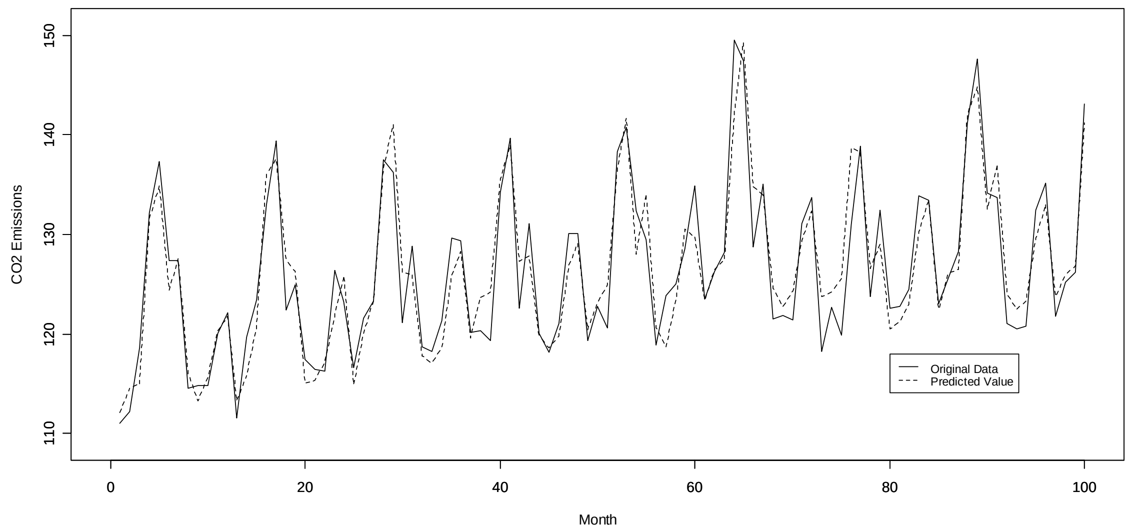

Figure 4.1 Monthly CO_2 Emissions VS. Forecast Values for the Last 100 Observations

As can be observed, the overall quality of the model is good. We proceed to calculate the residuals estimates, $r_t = x_t - \hat{x}_t$, and the results are graphically presented below by Figure 4.2

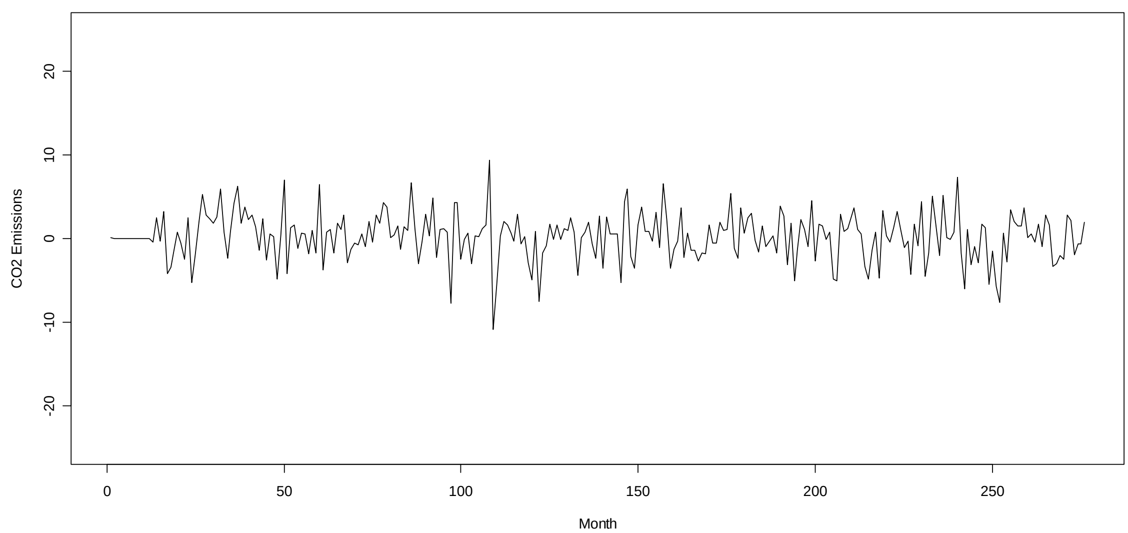

Figure 4.2 Residuals Plot for CO_2 Emissions

We observe that the residuals are quite small and isolating around the zero axis as expected. Thus, the proposed model is capable of forecasting the CO_2 emissions accurately in the United States.

The mean of the residuals, $\bar{r}$, the variance, $S_r^{\;2}$, the standard deviation, S_r, standard error, *SE*, and the mean square error, *MSE*, are presented below by Table 4.1.

Table 4.1 CO_2 Emissions

$\bar{r}$	$S_r^{\;2}$	S_r	*SE*	*MSE*
0.2339641	8.055668	2.838251	0.1708426	8.08122

We observe that all evaluation criteria support the quality of the proposed forecasting model for CO_2 emissions.

We now proceed to further evaluate model (3.3) hiding the last 12 months of the CO_2 recordings and re-estimating the coefficients of the model (3.3). Having restructured the model (3.3) we proceed to estimate the hidden recordings. For example, we used the first 264 observations $\{x_1, x_2, ..., x_{264}\}$ to forecast $\hat{x}_{265}$. Then we use the observations $\{x_1, x_2, ..., x_{265}\}$ to forecast $\hat{x}_{266}$, and continue this process until we obtain the forecasting values of the last 12 observations, that is, $\{\hat{x}_{265}, \hat{x}_{266}, ..., \hat{x}_{276}\}$. Table 4.2, gives the actual, forecasting and residual data for the subject 12 months.

Table 4.2 CO_2 Emissions Forecast

	Original Values	Forecast Values	Residuals
January 2003	147.6298	145.2361	2.3937
February 2003	134.1716	132.6554	1.5162
March 2003	133.6979	137.3912	-3.6933
April 2003	121.0047	124.5518	-3.5471
May 2003	120.4789	122.4091	-1.9302
June 2003	120.7394	123.101	-2.3616
July 2003	132.4187	129.3481	3.0706
August 2003	135.1314	132.787	2.3444
September 2003	121.7753	123.8295	-2.0542
October 2003	125.2487	125.9811	-0.7324

November 2003	126.2127	126.812	-0.5993
December 2003	143.1509	141.1834	1.9675

Note the closeness between the original and forecast values. A graphical presentation of the results given in Table 4.2 is given below by Figure 4.3.

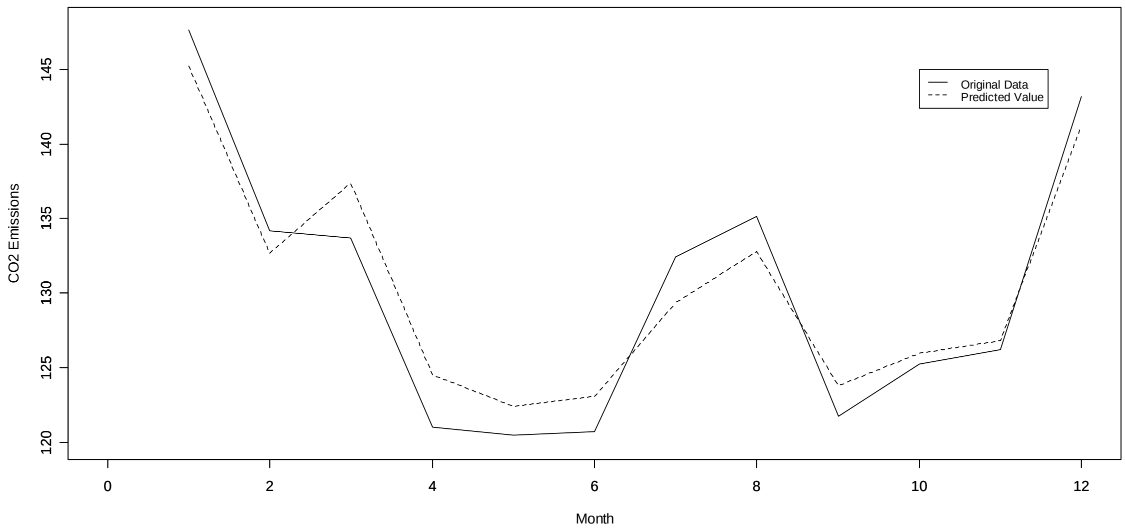

Figure 4.3 Monthly CO_2 Emission VS. Forecast Values for the Last 12 Observations

It can be observed that the forecasting values generated by our proposed model follows the pattern of the original series and attests to the accuracy of the model. Similar evaluation for model (3.4) for the atmospheric CO_2 is shown by Figure 4.4, that compares the actual recordings with those that are being estimated.

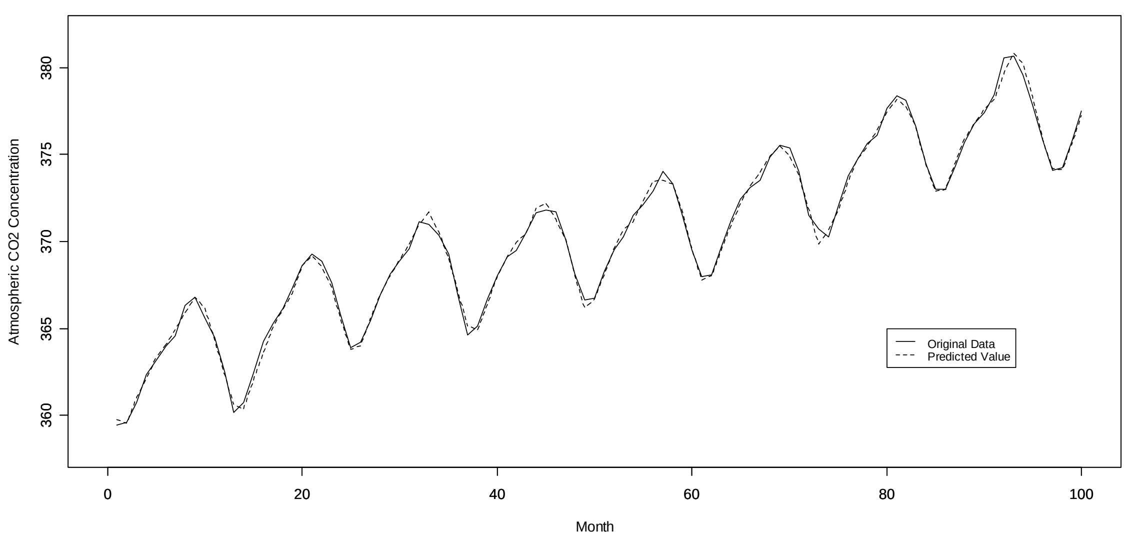

Figure 4.4 Monthly Atmospheric CO_2 VS. Forecast Values for the Last 100 Observations

Obviously, this graphical presentation attests to show the quality of the proposed model. A plot of the residuals is given by Figure 4.5 below.

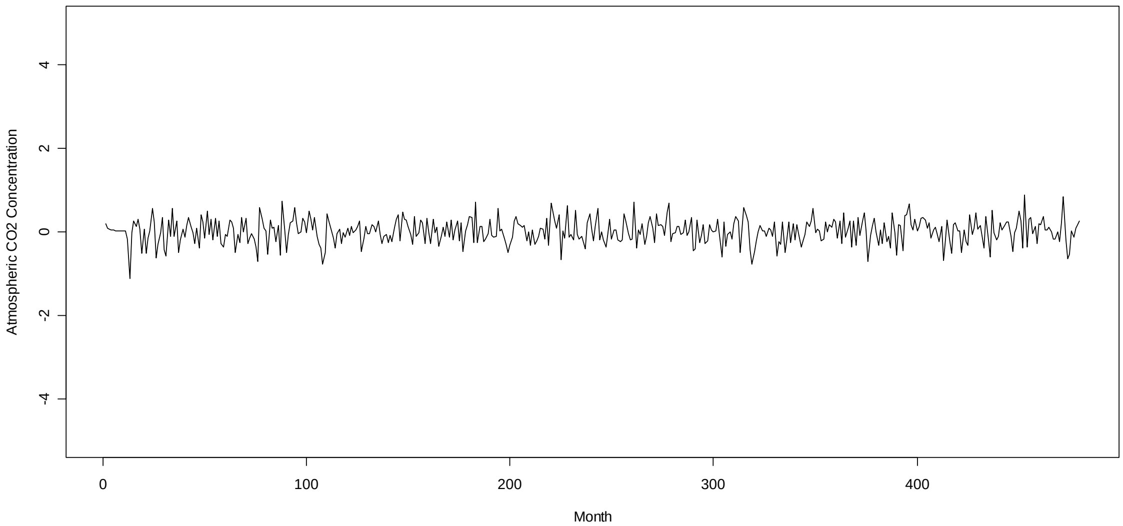

Figure 4.5 Residuals Plot for Atmospheric CO_2

The residuals of our proposed model are very small and isolating around the zero axis. It illustrated the quality of the model. The following Table 4.3 gives a basic evaluation statistics of the proposed model.

Table 4.3 CO_2 in the Atmosphere

$\bar{r}$	S_r^2	S_r	SE	MSE
0.01140137	0.08460756	0.2908738	0.01327651	0.08456128

These results also confirm the effectiveness of the proposed model for forecasting CO_2 in the atmosphere.

We shall use the same technique as we used in the previous application to illustrate the quality of our proposed model in terms of forecasting values in the future. Again, we hide the last 12 months of atmospheric CO_2 recordings, and try to predict them only using the information from the past. Table 4.4 gives the numerical comparison between the original series and the forecasting.

Table 4.4 CO_2 in the Atmosphere Forecast

	Original Values	Forecast Values	Residuals
January 2004	376.79	376.7963	-0.0063
February 2004	377.37	377.609	-0.239
March 2004	378.41	378.1837	0.2263
April 2004	380.52	379.6653	0.8547
May 2004	380.63	380.8268	-0.1968
June 2004	379.57	380.2339	-0.6639
July 2004	377.79	378.3489	-0.5589
August 2004	375.86	375.837	0.023
September 2004	374.06	374.1871	-0.1271
October 2004	374.24	374.1482	0.0918
November 2004	375.86	375.6897	0.1703
December 2004	377.48	377.2186	0.2614

The residuals that were calculated are shown by Table 4.4 are all very small, and a graphical presentation of the results is given below by Figure 4.6.

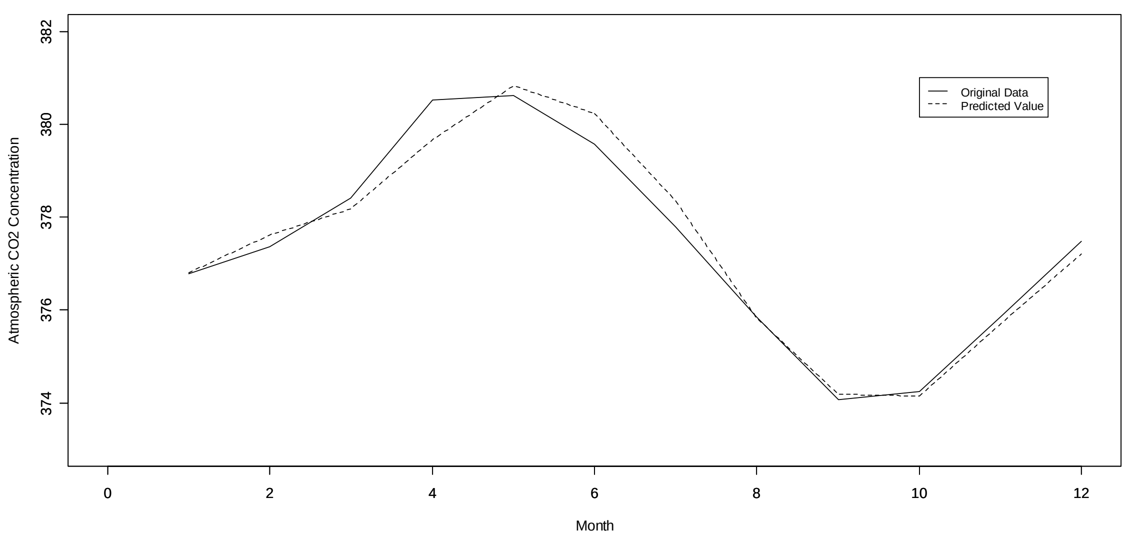

Figure 4.6 Monthly CO_2 in the Atmosphere VS. Our Predicted Values for the Last 12 Observations

Thus, we can conclude that the proposed model, (3.4), forecasts very well on the future behavior of CO_2 in the atmosphere.

5. CONCLUSION

We have developed two non-stationary time series models with trend and seasonal effects to predict future estimates of carbon dioxide emissions and that in the atmosphere. We use actual CO_2 recordings in both situations to develop the subject statistical models. The developed processes were evaluated to attest the degree of quality by using various statistical criteria. Finally, we tested the accuracy of the proposed models by predicting and analyzing the CO_2 emission and atmosphere for 12 months. The results are very encouraging.

REFERENCES

1. Akaike, H. (1974). A New Look at the Statistical Model Identification, IEEE Transactions on Automatic Control, AC-19, 716-723.

2. Bacastow, R.B. (1979). Dip in the atmospheric CO_2 level during the mid-1960s. Journal of Geophysical Research 80:3109-14.

3. Bacastow, R.B., and C.D. Keeling. (1981). Atmospheric carbon dioxide concentration and the observed airborne fraction. In B. Bolin (ed.), Carbon Cycle Modelling, SCOPE 16. John Wiley and Sons, New York.

4. Bacastow, R.B., J.A. Adams, Jr., C.D. Keeling, D.J. Moss, T.P. Whorf, and C.S. Wong. (1980). Atmospheric carbon dioxide, the Southern Oscillation, and the weak 1975 El Niño. Science 210:66-68.

5. Bacastow, R.B., C.D. Keeling, and T.P. Whorf. (1985). Seasonal amplitude increase in atmospheric CO_2 concentration at Mauna Loa, Hawaii, 1959-1982. Journal of Geophysical Research 90(D6):10529-40.

6. EPA (U.S. Environmental Protection Agency) (2004). Inventory or U.S. Greenhouse Gas Emissions and Sinks: 1990-2002, EPA 430-R-04-003, U.S. Environmental Protection Agency, Washington, D.C., 308 pp. plus annexes (291 pp.). Available electronically from: http://yosemite.epa.gov/oar/globalwarming.nsf/content/ResourceCenterPublicationsGHGEmissionsUSEmissionsInventory2003.html.

7. Keeling, C.D. (1960). The concentration and isotopic abundance of carbon dioxide in the atmosphere. Tellus 12:200-203.

8. Keeling, C.D. (1984). Atmospheric and oceanographic measurements needed for establishment of a data base for carbon dioxide from fossil fuels. In The Potential Effects of Carbon Dioxide-Induced Climatic Changes in Alaska. (Miscellaneous, etc.). The Proceedings of a Conference. Fairbanks, Alaska, April 7-8, 1982. School of Agriculture and Land Resources Management, University of Alaska, Fairbanks.

9. Keeling, C.D. (1998). Rewards and penalties of monitoring the earth. Annual Review of Energy and the Environment 23:25-82. Annual Reviews Inc., Palo Alto.

10. Keeling, C.D., R.B. Bacastow, A.E. Bainbridge, C.A. Ekdahl, Jr., P.R. Guenther, L.S. Waterman, and J.F.S. Chin. (1976). Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. Tellus 28(6):538-51.

11. Keeling, C.D., R.B. Bacastow, and T.P. Whorf. (1982). Measurements of the concentration of carbon dioxide at Mauna Loa Observatory, Hawaii. In W.C. Clark (ed.), Carbon Dioxide Review: 1982. Oxford University Press, New York.

12. Keeling, C.D., R.B. Bacastow, A.F. Carter, S.C. Piper, T.P. Whorf, M. Heimann, W.G. Mook, and H. Roeloffzen. (1989). A three-dimensional model of atmospheric CO_2 transport based on observed winds: 1. Analysis of observational data. In D.H. Peterson (ed.), Aspects of Climate Variability in the Pacific and the Western Americas. Geophysical Monograph 55:165-235.

13. Keeling, C.D., J.F.S. Chin, and T.P. Whorf. (1996). Increased activity of northern vegetation inferred from atmospheric CO_2 measurements. Nature 382: (6587) 146-49. MacMillan Magazines Ltd., London.

14. Keeling, C.D., T.P. Whorf, M. Wahlen, and J. van der Plicht. (1995). Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. Nature 375:666-670.

15. Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, & Y. Shin. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root., Journal of Econometrics, 54, 159-178.

16. Keeling, C.D., P.R. Guenther, G. Emanuele III, A. Bollenbacher, and D.J. Moss. (2002). Scripps Reference Gas Calibration System for Carbon Dioxide-in-Nitrogen and Carbon Dioxide-in-Air Standards: Revision of 1999 (with Addendum). SIO Reference Series No. 01-11.

17. Marland, G., T.A. Boden, and R. J. Andres (2003). Global, Regional, and National CO_2 Emissions. In Trends: A Compendium of Data on Global Change. Carbon

Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, USA.

18. Pales, J.C., and C.D. Keeling. (1965). The concentration of atmospheric carbon dioxide in Hawaii. Journal of Geophysical Research 24:6053-76.

19. Verhulst, J. (April 22, 2007). “Feeling The Heat”. St. Petersburg Times. St. Petersburg, Florida.

20. Whorf, T.P., and C.D. Keeling. (1998). Rising carbon. New Scientist 157:(2124) 54-54. New Scientist Publ Expediting Inc., Elmont.