Check for updates

CASE STUDY

Analysis of the stationarity and correlation of the global temperature and carbon dioxide time series [version 1; peer review: awaiting peer review]

Upul Rupassara (iD), Sarah Frantsvog, Ashley Holen, Karen Robinson

Mathematics and Computer Science, Minot State University, Minot, North Dakota, 58707, USA

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Background: The rapid and ongoing phenomenon of global warming has negatively
impacted both the Earth's environment and its inhabitants. Time series and regression analysis techniques play a significant role in weather forecasting and the interpretation of climate data. One of the key characteristics of time series analysis
is stationarity.

Methods: In this article, we explore how detrending and differencing techniques can be used to transform the time series of global temperature and carbon dioxide into stationary series. Regression models and goodness of fit tests were used to examine the relationship between carbon dioxide and data on global temperature. A cross-correlation time series model is also used to assess those time series' lagging
and leading characteristics.

Results: The study of data on global temperature anomalies indicates that detrending and differencing are helpful in transforming temperature time series into stationary time series. However, the first differencing and detrending methods do
not make the carbon dioxide time series stationary; instead, an alternate transformation is needed. Neither the carbon dioxide time series nor the global temperature time series lag or lead with regard to the cross-correlation function.

Conclusions: In this article, we looked into stationarity and some other topics associated with correlation in terms of data on CO2 and global temperature. Stationarity is one of the important properties to check before conducting a more thorough investigation of the time series. To transform a non-stationary time series into a stationary one, there are numerous techniques available. However, in this article, we just pay attention to detrending and differencing and how those methods perform with respect to time series data for global temperature and carbon dioxide.

Keywords

Time, Series, Temperature, Global, Stationary

This article is included in the Climate gateway.

Corresponding author: Upul Rupassara (upul.rupassara@minotstateu.edu)

Author roles: Rupassara U: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Frantsvog S**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Holen A**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Robinson K**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing

Introduction

Since the years 1850–1900, the rapid and relentless phenomenon of global warming has severely and negatively altered Earth's environment and its inhabitants. NASA describes global warming as "the long-term heating of Earth's surface observed since the pre-industrial period (between 1850 and 1900) due to human activities, primarily fossil fuel burning, which increases heat-trapping greenhouse gas levels in Earth's atmosphere" (NASA 2023). Carbon dioxide (CO_2) is naturally present in the atmosphere as part of the Earth's carbon cycle (the natural circulation of carbon among the atmosphere, oceans, soil, plants, and animals) (EPA 2023). Unlike oxygen or nitrogen (which make up most of our atmosphere), greenhouse gases absorb heat radiating from the Earth's surface and re-release it in all directions–including back toward the surface (Lindsey 2022). In order to preserve Earth's future, action must be taken now. The first step is tracking how atmospheric carbon dioxide emissions affect global warming rates.

The correlation between global temperature and carbon dioxide levels has been investigated by many researchers in the literature (Florides and Christodoulides 2009; Macedo and Madaleno 2023; Hereher 2016; Palmer et al. 2007; Woodward and Gray 1993).

In the article, Global warming and carbon dioxide through science, (Florides and Christodoulides 2009) revealed that there is no way to definitively say whether or not CO_2 directly impacts global warming and temperature increases. This study used three independent sets of data (collected from ice cores and chemistry) to perform an analysis. Through a specific regression analysis of their data, they found that the data stating that there is a correlation between CO_2 concentration and temperature relies heavily on specific choices of data. From this conclusion, they found that for both chemistry data and ice core data, "one cannot be positive that indeed such a correlation exists". Through further research into the topic, there is evidence that CO_2 is not inherently harmful to the environment, in addition to the fact that there is very little evidence linking CO_2 levels and temperature. In the end, they concluded, "It is our view that it is not yet sufficient, let alone rigorous, evidence that anthropogenic CO_2 increase is indeed the main factor contributing towards the global warming of the 20th century".

Evidence from a Maximum Entropy Approach (Macedo and Madaleno 2023) uses a statistical approach based on maximum entropy to conduct a study that supports the results from different techniques that found that CO_2 does, in fact, impact the increase in global temperature. This study, along with a recent review of emerging literature, led them to the conclusion that the impact of CO_2 on our atmosphere is trending toward a detrimental conclusion. They further state that if people continue emitting CO_2 into the atmosphere, it will have a detrimental impact on human, plant, and animal health in the long run.

A time series-related analysis was used in Hereher (2016) to study the global temperature trends of land surface temperatures and climate changes based on the temperature data in Egypt. This study reports serious climate change in Egypt by detecting variability of land surface temperature (LST) over the last decade at selected locations, with varied geomorphological characteristics and human stressors. The time series for the land surface temperatures were acquired from satellite images from 2003 to 2014, totaling 276 images. The analysis suggests that the LST in Egypt increases by 1.54 °C/decade. The variation of LST depends on latitude, geology, topography, and surface albedo. The dataset used in this analysis is named MODIS, from the NASA Land Processes Distributed Active Archive Center website. This analysis found that the time-series MODIS LST data proved sufficient for the short-term monitoring of land surface temperature variations in Egypt. It is noted that geology, topography, and surface albedo have significant impacts on the LST of Egypt. Further, it was revealed that between the years of 2003-2014, the LST of Egypt increased by 0.3–1.06 °C/decade, and for urban areas, the excess LST is higher at 1.54 °C/decade.

According to a new isothermal analysis by Palmer et al. (2007) in order to produce a more accurate depiction of the underlying warming. They proposed a new analysis of millions of ocean temperature profiles intended to filter out local dynamical changes. According to the authors' comments, it was a difficult analysis since oceans do not warm uniformly across the globe. They present decadal-scale analyses of the ocean's thermal state relative to a fixed isotherm. This new diagnostic is less prone to the influence of dynamical processes at both high and low frequencies, and the results present a more globally uniform picture of ocean warming. The limitation of the isothermal analysis is that high-latitude oceans were not included.

Using the Autoregressive Moving Average (ARMA) model, a study of time series data on global warming and trends is presented in Woodward and Gray (1993). They concluded that atmospheric greenhouse gases will affect the continued projection of the warming trend of global temperatures. Further, they stated that there is no conclusive evidence that this trend will continue. This is due to the difference between data on long-term trends and data on random trends with the length of the temperature series.

Data sources

Global temperature anomalies and carbon dioxide data were obtained from (NASA 2023) global climate change website. The temperature anomalies (in Celsius) and carbon dioxide (parts per million) were used from 1960 to 2022. The temperature anomalies are used instead of the absolute temperature data, as they accurately represent the temperature variability over larger areas. Further, they give a frame of reference that allows more meaningful comparisons between locations and more accurate calculations of temperature trends (NOAA 2023).

Statistical time series review

This section provides an overview of the definitions and test formulas that we use in this article to examine the time series models. Time series analysis can be used to examine the outcomes of either a planned or unplanned intervention as well as to better understand the underlying naturalistic process and the pattern of change over time (Velicer and Molenaar 2012). In time series analysis, regularity is essential, and if time series data exhibits irregular behavior over time, establishing meaningful conclusions about the underlying causes will be challenging. This regular behavior is explained by the concept of stationarity. In general, to satisfy the conditions of stationarity, the time series must satisfy some strong conditions, and those conditions are too strong for most of the applications. Therefore, most time series seen in real life are analyzed using a weaker version of stationarity.

The weakly stationary time series should satisfy the following conditions (Shumway and Stoffer 2017).

    (1) The expected value (mean) of the time series is constant and does not depend on time.

    (2) The autocovariance function does not depend on the actual time value but only depends on the time value through the time difference.

If two time series are taken into account simultaneously, such as x_t and y_t, they are said to be jointly stationary if the individual time series are stationary and the cross-covariance function solely depends on the lag value.

To evaluate the model's goodness of fit, we utilize Akaike's Information Criterion (AIC), the bias-corrected Akaike's Information Criterion (AICc), and Bayesian Information Criterion (BIC). Those criteria measure the goodness of fit of the models by balancing the number of parameters of the model and the error of the fit. If k is the number of parameters of the model, n is the number of observations, and SSE is the sum squared error of the fit, then AIC is given by

$$\text{AIC} = \ln(\text{SSE}/n) + \frac{n + 2k}{n}. \tag{1}$$

The AICc is given by

$$\text{AICc} = \ln(\text{SSE}/n) + \frac{n + k}{n - k - 2}. \tag{2}$$

With the Bayesian corrected term, the BIC takes the form

$$\text{BIC} = \ln(\text{SSE}/n) + \frac{k \ln(n)}{n}. \tag{3}$$

More details about those information criteria can be found in Akaike (1974, 1973, 1969) and Schwarz (1978).

Detrending and differencing are the two fundamental methods for transforming a non-stationary time series into a stationary one. The independence of the residuals can be visually demonstrated by using the residual plots. The autocorrelation plots and cross-correlation plots can be used the explain the nature of the correlation at different lag values. If $\mu_{x,t}$ and $\mu_{y,t}$ are the means of the time series x_t and y_t respectively, the autocovariance function of x_t is defined by

$$\text{cov}(x_{t+h}, x_t) = E\left[\left(x_{t+h} - \mu_{x,t+h}\right)\left(x_t - \mu_{x,t}\right)\right], \tag{4}$$

where h represents the time shift or lag value. The cross-covariance function for x_t and y_t is given by

$$\text{cov}(x_{t+h}, y_t) = E\left[\left(x_{t+h} - \mu_{x,t}\right)\left(y_t - \mu_{y,t}\right)\right] \tag{5}$$

The autocorrelation and cross-correlation functions are normalized versions of the above formulas. For a given time series x_t, and y_t the autocorrelation (ACF) and cross-correlation (CCF) functions are given by

$$\mathrm{ACF}(x_s, x_t) = \frac{\mathrm{cov}(x_s, x_t)}{\sqrt{\mathrm{cov}(x_s, x_s)\mathrm{cov}(x_t, x_t)}}, \tag{6}$$

and

$$\mathrm{CCF}(x_s, y_t) = \frac{\mathrm{cov}(x_s, y_t)}{\sqrt{\mathrm{cov}(x_s, x_s)\mathrm{cov}(y_t, y_t)}}, \tag{7}$$

respectively.

Because they depend on the locations of time points s and t, the autocovariance and cross-covariance functions can change during the course of the series. If the autocovariance function depends on the separation $h = |t - s|$ rather than the points where the time series are situated, we are able to analyze sample time series data when there is only one series available.

Analysis of the global temperature and carbon dioxide time series

We use the global temperature anomalies from 1960 to 2022, obtained from NASA (2023). As demonstrated in Figure 1 most of the planet is warming (yellow, orange, and red). Only a few locations, most of them in the southern hemisphere oceans, cooled over this time period. According to the fact highlighted in Lindsey and Dahlman (2023) earth's temperature has risen by an average of 0.14° Fahrenheit per decade since 1880. The temperature anomaly variability (in Celsius), including pre-industrial time, is shown in Figure 2. The trend is shown by the red dashed line while smoothing splines with the smoothing parameter value *spar* is 0.5, is demonstrated by the blue spline.

The trend and irregular behavior of the temperature series require a modification to make it stationary before additional analysis can be done to uncover its other properties. The stationary time series are easy to study using established principles created in the time series literature because of their predictable long-term behavior.

We investigate detrending and differencing techniques to transform the original time series into stationary time series.
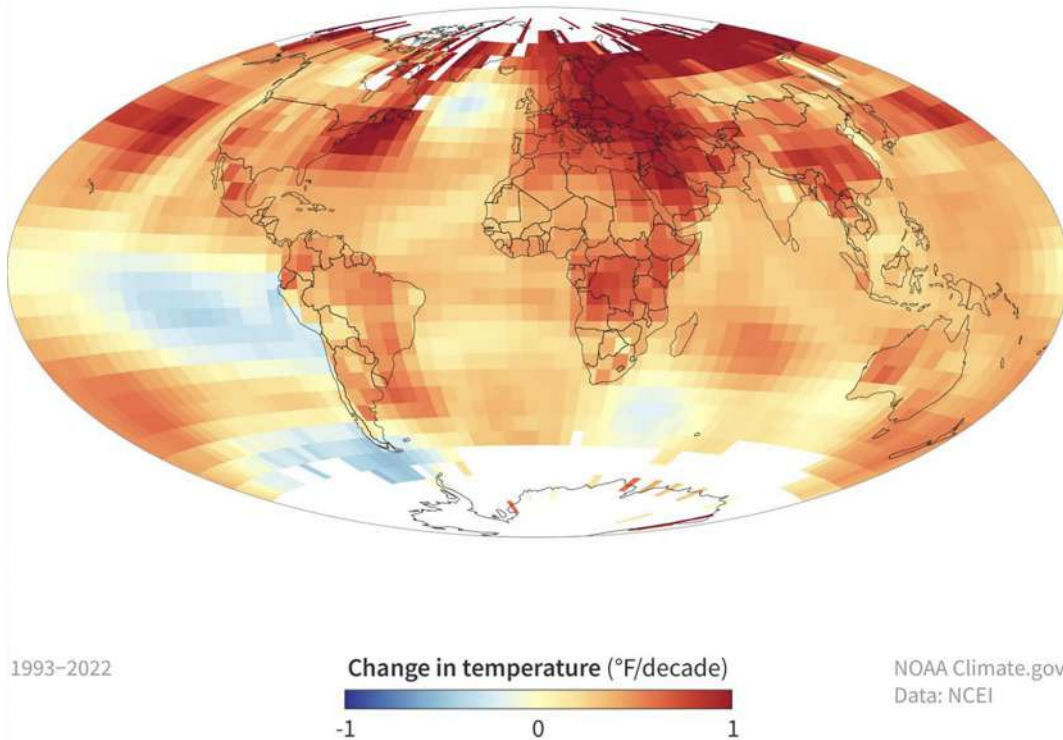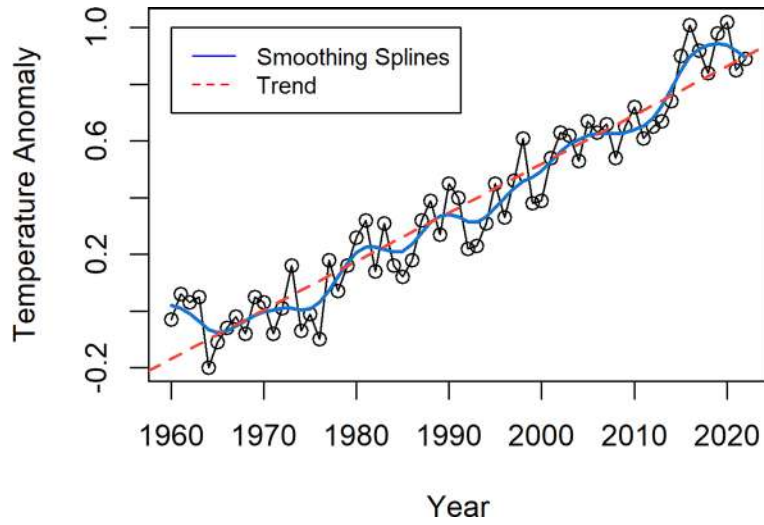

Figure 1. Recent temperature trend (1993-2022). Climate.gov Media: https://www.climate.gov/media/15022.

Figure 2. Temperature anomaly from 1960 to 2022. The red dashed line shows the trend and the blue line shows the temperature variability with smoothing spline (spar = 0.5).

Stationarity in the temperature time series

The trend stationary model is the form of non-stationary series that is easiest to work with. As seen in Figure 2, the temperature series exhibits stationary behavior around the trend line (trend stationary). The stationary component of the time series might thus be filtered by just removing the trend. This trend stationary model can be written as

$$T_t = \mu_t + T_t^s \tag{8}$$

where μ_t is the trend and T_t^s is the stationary component. The equation for the trend line can be used to detrend the data, and it can be estimated by ordinary least squares regression.

$$\widehat{\mu}_t = -33.82 + 0.02t \tag{9}$$

Using global temperature anomaly data from 1960 to 2022, the estimated trend line equation can be obtained as follows:

Figure 2 shows the superimposed estimated trend line (*red*). To get the detrended series, we simply subtract $\widehat{\mu}_t$ from T_t,

$$\widehat{T}_t^s = T_t + 33.82 - 0.02t \tag{10}$$

Because of the error of the estimated model in Eq. 10, $\widehat{T}_t^s$ may not be independent and identically distributed. In fact, our main goal is investigating the behavior of this stationary component $\widehat{T}_t^s$. According to the graphical representation in Figure 3a, detrending eliminates the original temperature time series trend. One of the main requirements of a stationary time series is the maintenance of a constant mean across the specified time period, which the elimination of trends helps to achieve. Figure 4a and Figure 4b show the autocorrelation function for the original and detrended global temperature anomaly time series respectively. The majority of the autocorrelation values for the detrended series are inside the 95% confidence band, indicating a significant improvement in the independence of lag-related correlation.

Differencing is the other technique we are utilizing here and in time series literature this is known as *integrating*. Differencing simply means subtracting past values from the current value. Figure 3b shows the differenced global temperature anomaly time series. This series also shows a similar pattern to that observed in the detrended series.

However, instead of treating drift as a fixed model, it can be modeled using a stochastic component. We define the stochastic drift model

$$\mu_t = \alpha + \mu_{t-1} + w_t \tag{11}$$

where w_t is the independent identically distributed random variable (called white noise or Gaussian process), which follows the normal distribution with mean 0 and fixed variance (say) σ^2. If w_t is independent of T_t^s, differencing the model in Eq. 8 and substituting Eq. 11, we get

a. Detrended temperature series

b. Differenced temperature series
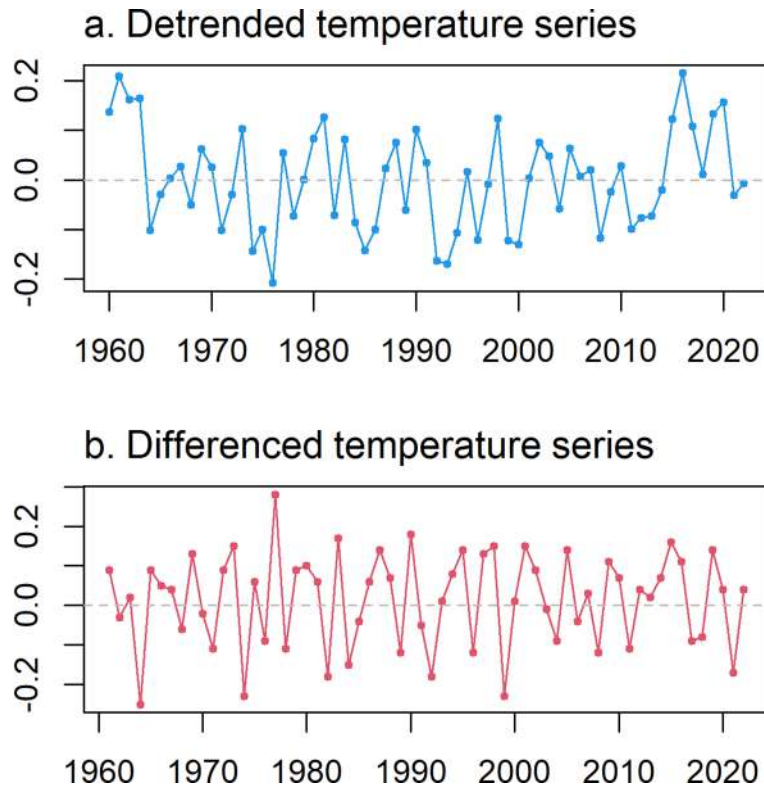
Figure 3. Detrended and differenced global temperature series from 1960 to 2022.

$$T_t - T_{t-1} = \alpha + w_t + T_t^s - T_{t-1}^s \tag{12}$$

Since T_t^s is the stationary component of the global temperature time series T_t, we can show that the difference $T_t^s - T_{t-1}^s$ is also stationary. For that define, $Z_t = T_t^s - T_{t-1}^s$. Then $E(Z_t)$ does not depend on time t since T_t^s is stationary for $1960 \leq t < 2022$.

Another key requirement of a stationary time series is the independence of the autocovariance from the time. It can depend on the time difference, but not on the actual time value. If h represents the lag or shift value, the autocorrelation can be found as

$$\text{cov}(Z_{t+h}, Z_t) = 2\text{cov}\left(T_{t+h}^s, T_t^s\right) - \text{cov}\left(T_{t+h+1}^s, T_t^s\right) - \text{cov}\left(T_{t+h-1}^s, T_t^s\right) \tag{13}$$

Therefore, the autocorrelation of the differenced temperature time series does not depend on the actual time value, t rather on the time lag or time difference. In addition to the computation results, this can be further verified by the ACF plots for differenced temperature data. Figure 4a displays the autocorrelation for the initial temperature anomaly. The ACF value decreases as the lag value increases, providing strong evidence for the temporal reliance of the covariance. On the other hand, Figure 4b and Figure 4c show, the autocorrelation for detrended and differenced data respectively. In comparison to the original time series ACF plot, both graphs demonstrate a significant reduction in the time-dependent correlation. The blue horizontal lines represent the 95% confidence level.

Stationarity of the carbon dioxide time series

The atmospheric carbon dioxide level, measured in parts per million (*ppm*) during the period from 1960 to 2022 is shown in Figure 5. We use detrending and referencing techniques to analyze the stationarity behavior of the CO_2 time series. Figure 6a shows the detrended carbon dioxide series. Unlike in the temperature series, the detrending does not remove the trend of the original CO_2 series. Also, according to Figure 6b the first differenced series either does not show stationary behavior in terms of the trend component or it still shows some sort of trend. But the second differenced series in Figure 6c shows a significant improvement toward the stationarity. Therefore, although the first differencing does not work here, the second differencing transforms the CO_2 series into a stationary time series. There are different transformations
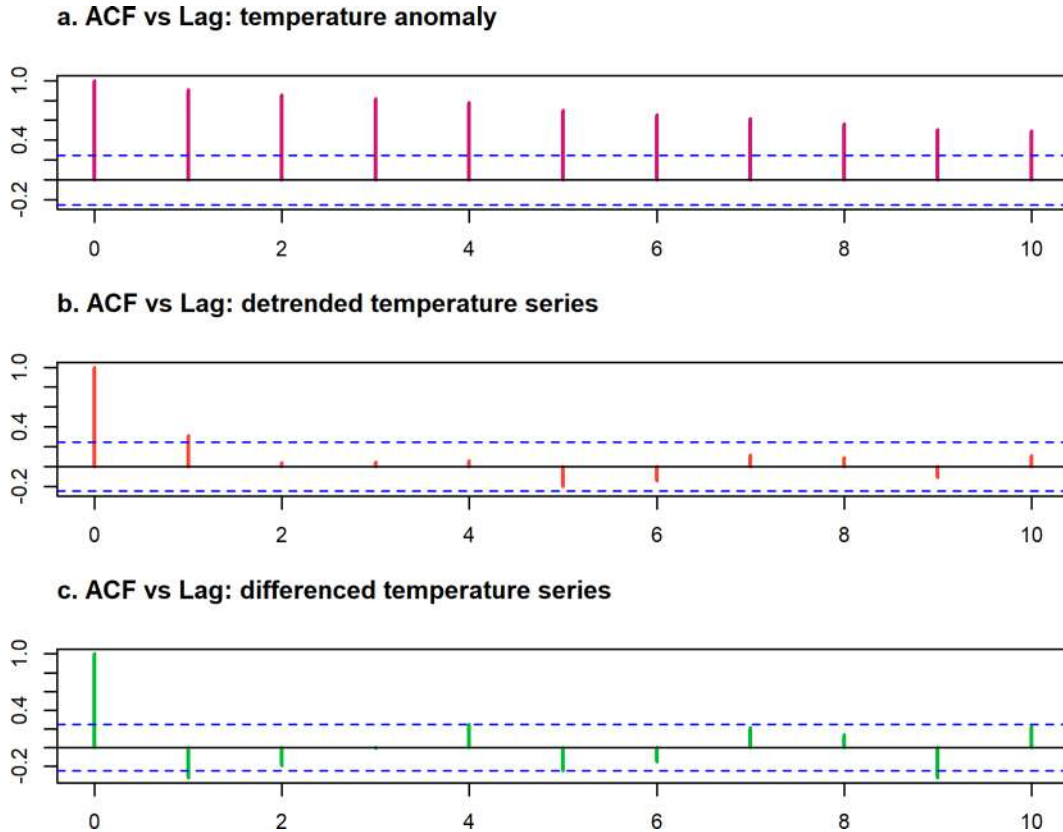
a. ACF vs Lag: temperature anomaly

b. ACF vs Lag: detrended temperature series

c. ACF vs Lag: differenced temperature series

Figure 4. ACF for a) Global temperature anomaly, b) Detrended, and c) Differenced time series.
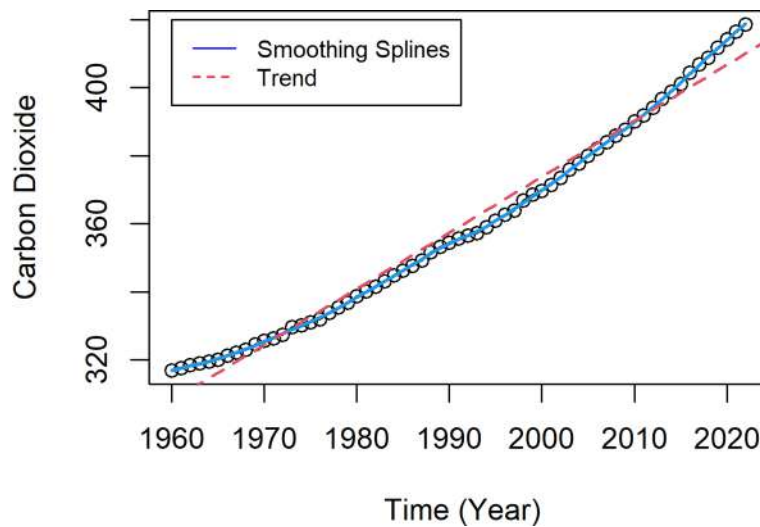

Figure 5. Carbon dioxide level (*ppm*) from 1960 to 2022. The red dashed line shows the trend (spar = 1) and the blue line shows the CO_2 variability with smoothing spline (spar = 0.5).

available in the time series literature to transform non-stationary time series into stationary form. In this article, we are not going to analyze other transformations, but one can try *log* or *Box-Cox* (Sakia 1992) family transformations.

Figure 7a shows the ACF of the original carbon dioxide series. Detrending and differencing (first difference) are unable to eliminate the autocorrelation of the CO_2 time series as they did in the temperature series. Figure 7b-c illustrates this by demonstrating how the height of the vertical lines decreases as the lag increases. It is clear that autocorrelation breaks a stationarity rule by leaning on the actual time value.
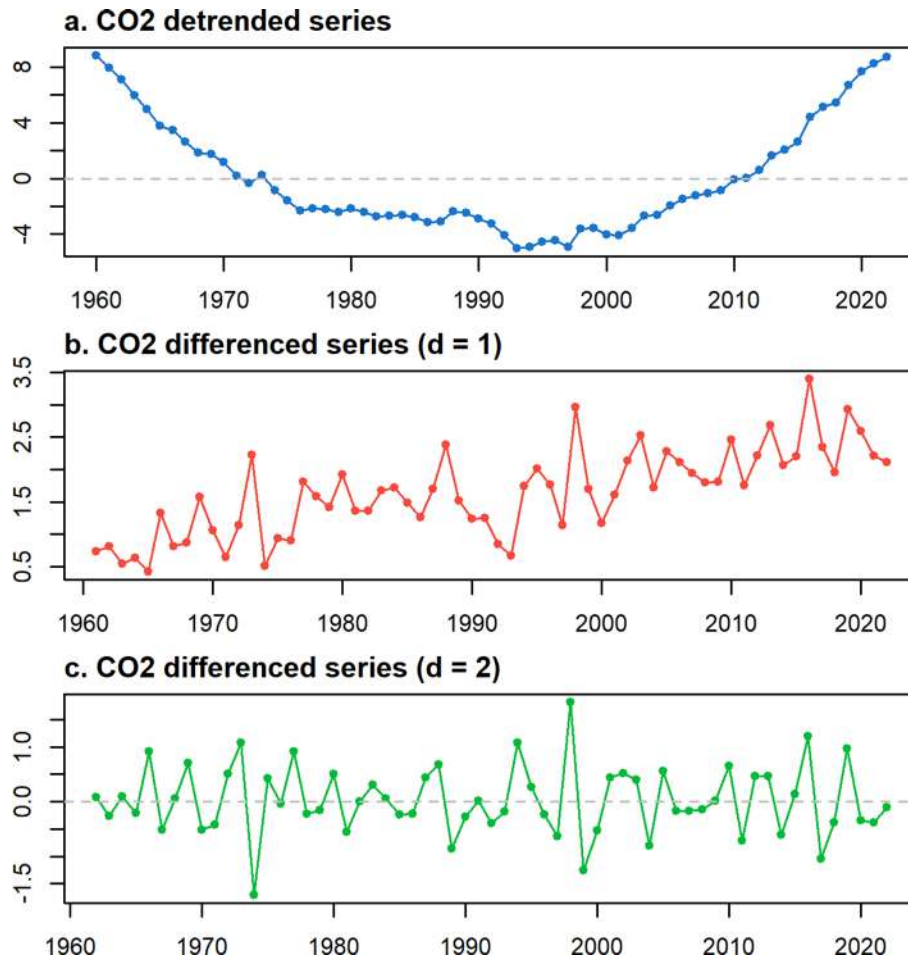
Figure 6. Detrended and Differenced CO$_2$ time series.

Higher order differencing

When first-order differencing fails to convert the original data into a stationary form, higher-order differencing may be necessary for the majority of time series applications. If a quadratic trend is present, second-order differencing is the optimal order to apply. In general, if the trend is linear, first-order differencing might be sufficient. If a non-stationary time series needs to be differenced d times in order to become stationary, it is said to be of order d integrated. The generalized form of higher-order differencing can be described by the backshift operator, say B. According to the exposition in Shumway and Stoffer (2017), this operator can be defined as

$$Bx_t = x_{t-1} \tag{14}$$

where x_t is any time series. Extending to the power k, we can get the generalized form of the Backshift operator

$$B^k x_t = x_{t-k} \tag{15}$$

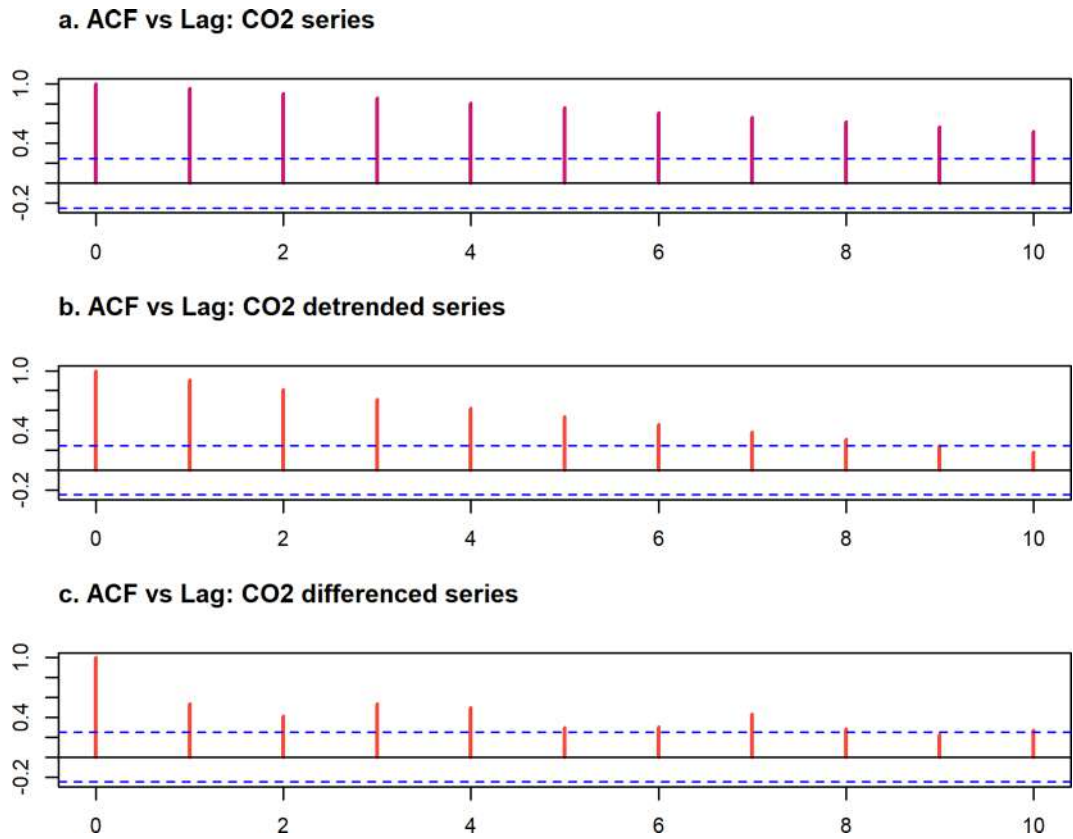If the difference is defined by

$$Dx_t = x_t - x_{t-1} \tag{16}$$

Then we can write,

$$Dx_t = (1 - B)x_t, \tag{17}$$

and further extending this iteration into the kth power

$$D^k = (1 - B)^k. \tag{18}$$

a. ACF vs Lag: CO2 series

b. ACF vs Lag: CO2 detrended series

c. ACF vs Lag: CO2 differenced series

Figure 7. ACF vs Lag: a) CO_2 (original) time series, b) CO_2 detrended time series, and c) CO_2 differenced (first differenced) time series.

Various differencing orders are used to decrease the time series' unwanted deviations. These techniques are frequently used in ARIMA time series models.

Although $d = 1$ works for global temperature anomalies, the CO_2 time series required higher-order differencing. The milder quadratic pattern suggests second-order differencing, and the corresponding figure after the second difference is shown in Figure 6c. As demonstrated in the autocorrelation plot for the second differenced time series in Figure 8, progress toward stationarity can be seen as the correlation values at different lags decline in contrast to the first differenced auto-correlation plot.

Along with the graphical evidence of stationarity, the Augmented DickeyFuller (ADF) test (Dickey and Fuller 1981) can be utilized to determine whether or not the given time series is stationary. The ADF test results are shown in Table 1. The p-value for the detrended and differenced time series of the global temperature is less than 0.01. This allows us to reject the null hypothesis of the ADF test and accept the resulting time series as stationary within the corresponding
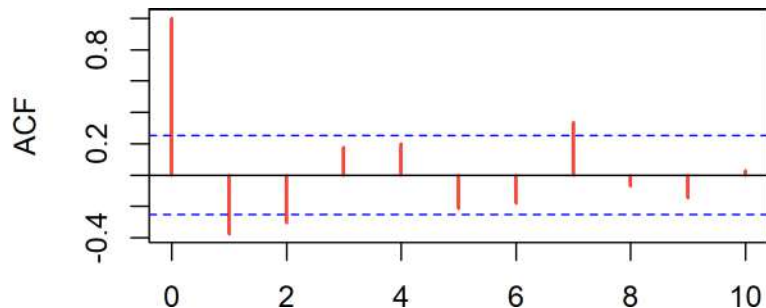
Figure 8. ACF plot: Carbon dioxide differenced series (d = 2).

Table 1. Dickey - Fuller test's p - values for detrended and differenced time series.

Time series/Transformation	Detrended	Differenced (d = 1)	Differenced (d = 2)
Global temperature anomaly	< 0.01	< 0.01	< 0.01
Carbon dioxide	0.99	0.07	< 0.01

significance level $\alpha = 0.05$. The hypothesis tests for detrended and first differenced CO_2 time series are not statistically significant at $\alpha = 0.05$, but the integrated order 2 series is stationary.

Models for correlation analysis

We analyze the basic global temperature trend model and temperature and carbon dioxide models. We compare two models using AIC, AICc, BIC, R-squared value, and sum squared error (SSE). The main objective of these models is to further analyze the effect of carbon dioxide on global temperatures. The basic trend model for temperature (model 1):

$$T = \beta_0 + \beta_1 t + w_t \tag{19}$$

where T is the global temperature anomaly (°C) and t is the time in years ($1960 \leq t \leq 2022$), w_t is the white noise and is assumed to be normally distributed with mean zero and fixed variance (say) σ^2.

The trend model (model 1) in Eq. 19 can be improved by introducing CO_2 as an explanatory variable.

Figure 9 shows the autocorrelation values for the carbon dioxide series at different lag values. Due to the considerable strength of the lag-related correlation in the second model, we adjust the carbon dioxide data for its mean, $\overline{C} = 358.9675$ (ppm). If C_t is the carbon dioxide level at time t, then we have model 2:
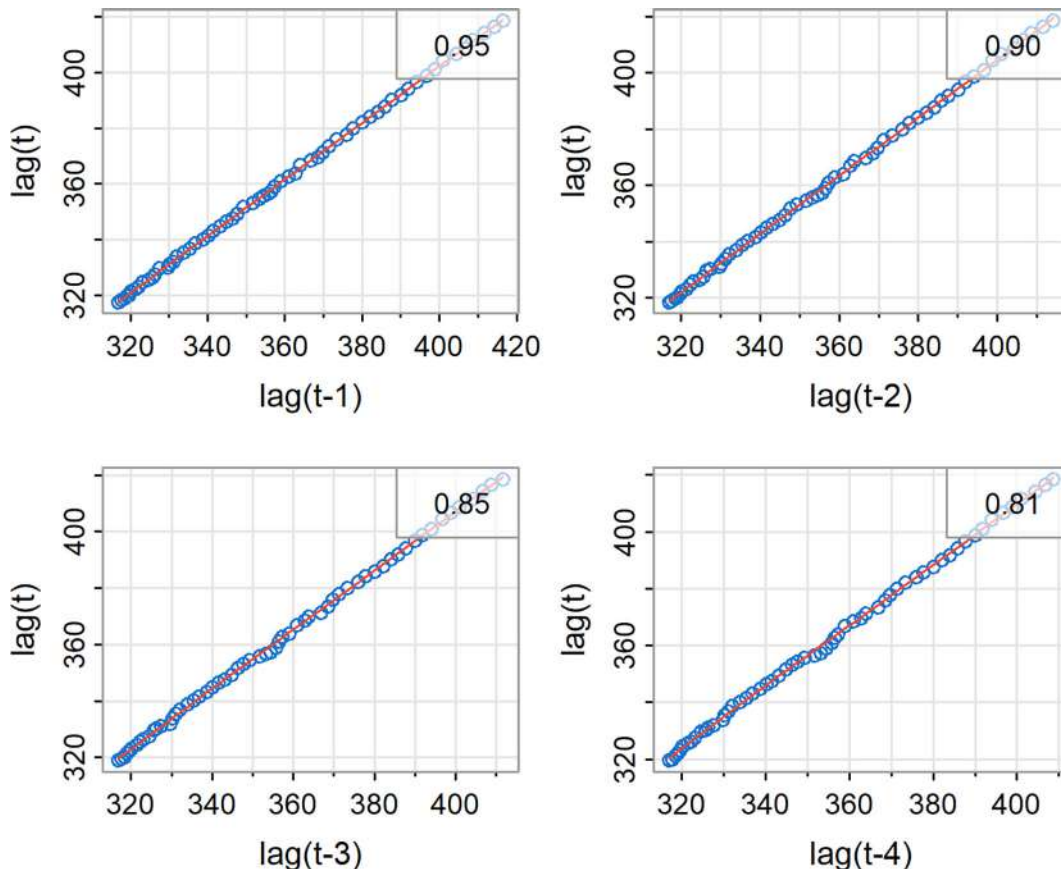
Figure 9. Scatter-plot matrix relating current CO_2 values (C_t), to past CO_2 values (C_{t-h}) for $h = 1, 2, 3, 4$. The values at the upper right corner are the sample autocorrelation at the corresponding lag values.

Table 2. Summary statistics for the global temperature models.

Model	k	SSE	df	R^2	AIC	AICc	BIC
Model 1	2	0.632	61	0.905	-3.539	-3.501	-4.471
Model 2	3	0.495	60	0.924	-3.750	-3.707	-4.648

$$T = \beta'_0 + \beta'_1 t + \beta'_2 C' + w_t \tag{20}$$

where $C' = C_t - \overline{C}, 1960 \le t \le 2022$, is the adjusted carbon dioxide level (in *ppm*). If k is the number of parameters of the model and *df* is the degree of freedom, Table 2 shows the summary statistics for the two models.

The value k yielding the minimum AIC, AICc, and BIC specifies the best model. We note that the second model is substantially better than the first model. Model 2, which includes CO_2 accounting for 92.4% of the variability (R^2 value of model 2), which was 90.5% (R^2 value of model 1) without that. Also, it gives the best value for AIC and BIC. In addition, we can notice that AIC and AICc are nearly equal. To calculate those values, we used the formulas given in the equations 1, 2, and 3. Without using the aforementioned formulas, one can utilize the regression model summaries to obtain these values, although there might be a few minor variations from the corresponding values shown in Table 2.

Further, the trend model (model 1) can be compared to the model with carbon dioxide (model 2) with the null hypothesis, $H_0 : \beta_3 = 0$. The corresponding F statistic:

$$F = \frac{(\text{SSE}_r - \text{SSE})/(q-r)}{\text{SSE}/(n-q-1)} \tag{21}$$

where SSE_r is the sum squared error of the reduced model, and q and r are the numbers of predictor variables in the full and reduced models, respectively. When $q = 2$, $r = 1$, and $n = 63$

$$F = \frac{(0.632 - 0.495)/1}{0.495/60} = 16.606, \tag{22}$$

which exceeds $F_{1,60}(0.001) = 11.973$. Hence, model 2 is a better prediction model compared to model 1.

Further for the purpose of predicting the global temperature the prediction model can be given by,

$$\widehat{T}_t = 4.943 - 0.002t + 0.012(C_t - 358.968) \tag{23}$$

where $\widehat{T}_t$ is the estimated temperature (anomaly) at time t. A negative (but very small) weight is present as the time coefficient. But a relatively larger constant (β_0) value mitigates the effect of the negative weight. The positive weight of the carbon dioxide indicates a positive contribution to the global temperature when $C_t > 358.968$ (*ppm*). Because of this, whether carbon dioxide has a positive or negative effect on the rise in global temperatures depends on its relative value to the average annual carbon dioxide level (in this study, the average was computed throughout the years from 1960 to 2022).

Model assumption

This section examines the validity of model 2 that we previously presented. Here, we go into further detail about the regression model's primary assumptions, such as linearity, residual normalcy, homoscedasticity, and independence of the residual errors.

As shown in Figure 10 "Residual vs. Fitted" subplot, the horizontal line without any obvious patterns is a sign of a solid linear relationship. In Figure 10, the "Normal Q-Q plot" is used to determine whether the residuals are distributed normally. It's ideal if the residual points fall along the dashed straight line, which is not perfectly satisfied in this case. A "Scale Location" subplot is utilized to verify the homoscedasticity (homogeneity of variance) of the residuals. Homoscedasticity can be detected by a horizontal line with evenly spaced points, which is satisfied except for a few outliers as shown in Figure 10, "Scale-Location" subplot. A few influential observations can be identified according to Figure 10, "Residual vs. Leverage" subplot. According to the model diagnostic analysis, since model 2 satisfies the fundamental assumptions, it can be considered a viable model to explain the relationship between temperature anomaly and CO_2 across time.
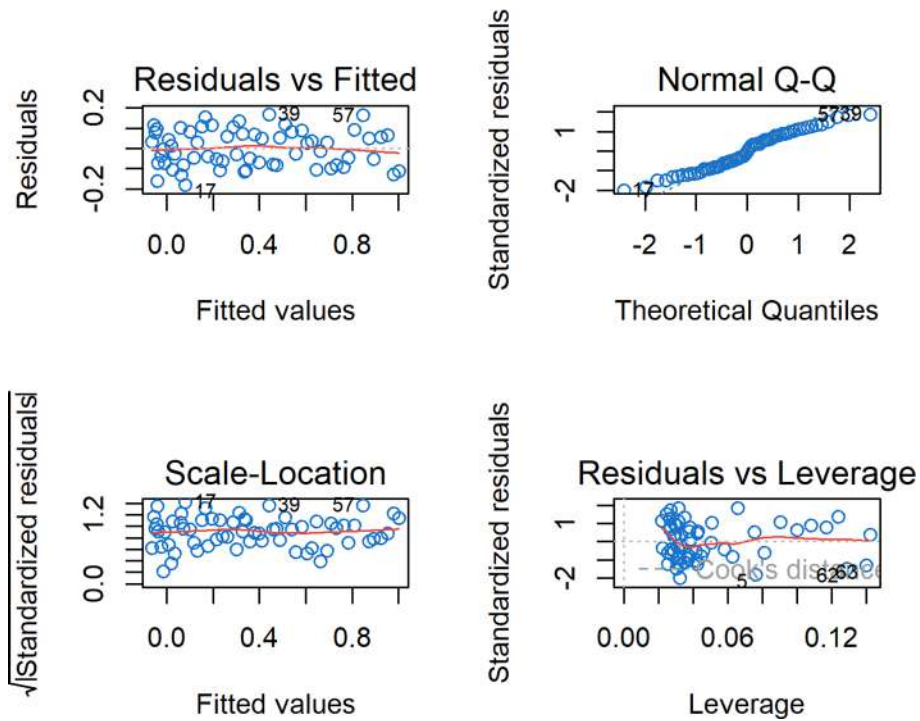
Figure 10. Residual vs Fitted, Normal Q-Q plot, Scale Location plot, and Residual vs Leverage plot.
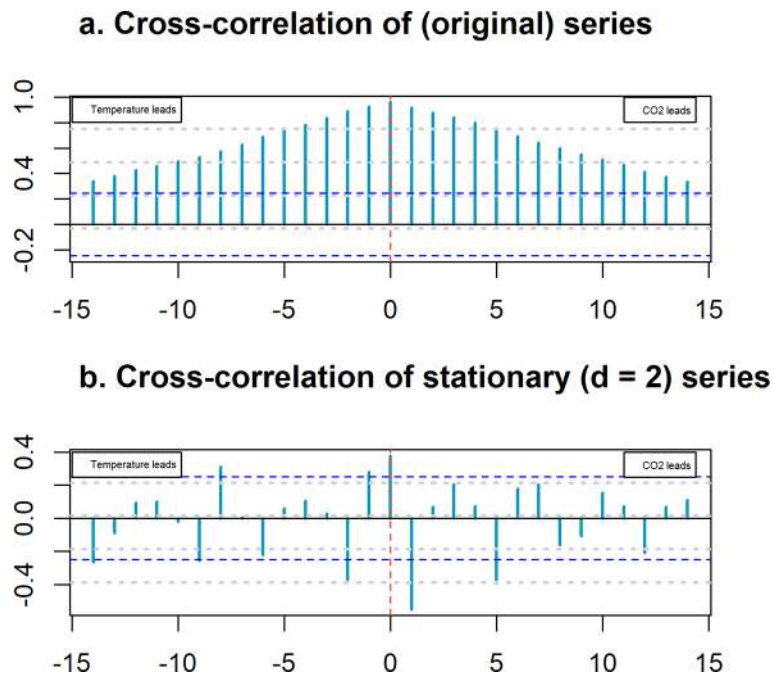
Figure 11. Cross correlation of global temperature anomaly and carbon dioxide time series from 1960 to 2022: a) Original, b) Stationary (d = 2).

Cross correlation: Global temperature anomaly vs carbon dioxide

Analyzing the potential leading and lagging relationships between the global temperature and carbon dioxide series is another intriguing investigation. Leading and lagging relationships might be helpful when one time series is used to predict another. Let T_t and C_t represent the global temperature and CO_2 time series respectively. Consider the model of the form

$$T_t = \beta_t^l C_{t-l} + w_t \tag{24}$$

where β_t^l is a real value that depends on t or l. C_t *lead* T_t if $l > 0$, and *lag* T_t if $l < 0$. If w_t is uncorrelated with C_t time series, the cross-correlation is given by

$$\operatorname{cov}(T_s, C_t) = \operatorname{cov}\left(\beta_t^l C_{s-l} + w_s, C_t\right) = \operatorname{cov}\left(\beta_t^l C_{s-l}, C_t\right), t > 0, s > 0. \tag{25}$$

According to the graphical demonstration of the cross-correlation function in Figure 11a for the original data, no time series leads or lags other series since peak shows occur at $l = 0$. When viewed in relation to the zero-lag value, the cross-correlation is approximately symmetric. However, those original series are not stationary. We take integration order 2 into consideration when we analyze the behavior of the cross-correlation functions of the (weakly) stationary CO_2 and global temperature anomaly series. Although the first integrated temperature series is stationary, we use the differencing order 2 ($d = 2$) to transform both series in order to tackle the dimension issues. According to Figure 11b most dominant cross-correlation for stationary series occurs at $l = 0$ and $l = 1$. Those integrated series are jointly stationary if the cross-correlation or cross-covariance function is a function only on the lag value.

Discussion
According to the summary statistics criterion in Table 2, the impact of CO_2 on the global temperature cannot be negligible. The models we've examined can be improved by adding additional greenhouse gases and other factors that could have an impact on the global temperature. Although carbon dioxide is not the only greenhouse gas that can affect the temperature, it occupies a significant amount of space compared to other greenhouse gases.

Data availability
The data used in this analysis is available on the NASA global climate change website [NASA(2023)].

Source data
This project contains the following underlying data:

- Carbon dioxide data; https://climate.nasa.gov/vital-signs/carbon-dioxide

- Temperature data; https://climate.nasa.gov/vital-signs/global-temperature

Software availability
We used R/R Studio as the statistical program for our computations and data visualizations.

- Archived source code available from: DOI: 10.5281/zenodo.8234016

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC BY 4.0).

References

Akaike H: **A new look at the statistical model identification.** *IEEE Trans. Autom. Control.* 1974; **19**(6): 716–723.
Publisher Full Text

Akaike H **Information Theory and an Extension of the Maximum Likelihood Principle.** Petrov BN, Csaki F, editors. *Proceedings of the 2nd International Symposium on Information Theory.* Budapest: Akademiai Kiado; 1973; pp. 267–281.

Akaike H: **Fitting autoregressive models for prediction.** *Ann. Inst. Stat. Math.* 1969; **21**: 243–247.
Publisher Full Text

Dickey DA, Fuller WA: **Likelihood ratio statistics for autoregressive time series with a unit root.** *Econometrica.* 1981; **49**: 1057–1072.
Publisher Full Text

[EPA] Environmental Protection Agency: **Overview of Greenhouse Gases.** 2023. Last updated on April 13, 2023.
Reference Source

Florides GA, Christodoulides P: **Global warming and carbon dioxide through sciences.** *Environ. Int.* 2009; **35**(2): pp. 390–401. 0160-4120.
Publisher Full Text

Hereher ME: **Time series trends of land surface temperatures in Egypt: a signal for global warming.** *Environ. Earth Sci.* 2016; **75**: 1218.
Publisher Full Text

Lindsey R, Dahlman L: **Climate Change: Global temperature. Science and information for a climate-smart nation.** 2023. January 18, 2023.
Reference Source | **Reference Source**

Lindsey R: **Climate Change: Atmospheric Carbon Dioxide. Science and information for a climate-smart nation.** 2022. June 23, 2022.
Reference Source | **Reference Source**

Macedo P, Madaleno M: **Global Temperature and Carbon Dioxide Nexus: Evidence from a Maximum Entropy Approach.** *Energies.* 2023; **16**(16): 277.
Publisher Full Text

NASA: **Global climate change, Vital signs of the planet.** 2023. last updated on April 20.
Reference Source

[NOAA] National centers for environmental information: **National oceanic and atmospheric administration.** 2023.
Reference Source

Palmer MD, Haines K, Tett SFB, *et al.*: **Isolating the signal of ocean global warming.** *Geophys. Res. Lett.* 2007; **34**: L23610.
Publisher Full Text

Sakia RM: **The Box-Cox Transformation Technique: A Review.** *Journal of the Royal Statistical Society. Series D (The Statistician).* 1992; **41**(2): 169–178.
Publisher Full Text

Schwarz G: **Estimating the Dimension of a Model.** *Ann. Stat.* 1978; **6**: 461–464.
Publisher Full Text

Shumway RH, Stoffer DS: *Time Series Analysis and Its Applications With R Examples.* 4th ed. Springer; 2017. 978-3-319-52451-1.
Publisher Full Text

Velicer WF, Molenaar PC: **Time Series Analysis for Psychological Research.** *Handbook of Psychology.* 2nd ed. Weiner I, Schinka JA, Velicer WF, editors. 2012.
Publisher Full Text

Woodward WA, Gray HL: **Global Warming and the Problem of Testing for Trend in Time Series Data.** *J. Clim.* 1993; **6**: 953–962.
Publisher Full Text

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000 Research