The Impacts of International Climate Policies on Global Temperature

Kevin Wu and Matthew Chen

University of Chicago

SOSC 13300

Abstract

In this paper, we will analyze global land surface temperature data from Berkeley Earth going back to 1850 to model the rate of climate change in the post-Industrial Revolution era. There is a general scientific consensus that in this era, climate change has significantly increased compared to past trends; however, whether this trend has continued on into the 21st century remains up for debate. This paper uses time-series techniques, particularly ARMA modeling and detrending techniques, to make conclusions about whether or not global climate policies beginning with the Kyoto Protocol have caused a statistically significant change in the rate at which global temperatures have risen over the 21st century.

Keywords: Kyoto Protocol, Paris Agreement, climate policies, global warming, climate change

1 Introduction

Climate change has been at the forefront of global issues for several decades. Research institutions, corporations, and governments worldwide are constantly working to reduce human-induced climate change, which has caused increasingly extreme weather patterns and has the potential to cause catastrophic damages across the globe. The first globally significant policy enacted to combat climate change was the Kyoto Protocol, which was signed in 1997 and officially enacted in 2005 (the second being the Paris Agreement of 2015). Our study aims to determine the effectiveness of international climate policies, which began with the Kyoto Protocol, on slowing the rate of global temperature increase.

To better understand the impacts of the Kyoto Protocol and its successors, we must begin by providing background on international climate policies. The United Nations Framework Convention on Climate Change (UNFCCC) was adopted in 1992 with the goal of "stabilizing greenhouse gas concentrations in the atmosphere at a level that will prevent dangerous human interference with the climate system, in a time frame which allows ecosystems to adapt naturally and enables sustainable development" (Leggett, 2020). Since 1995, countries have met annually at the Conference of Parties (COP) to assess progress in combating climate change (Zakerinia & Lin Lawell, 2019).

The Kyoto Protocol was the first subsidiary agreement of the UNFCCC, adopted at COP 3 in Kyoto, Japan, in December 1997. Participating nations, which included all developed countries besides the United States, agreed to a legally binding target of reducing greenhouse gas (GHG) emissions by 5.2% below their 1990 levels over an initial commitment period of five years lasting from 2008 to 2012 (Wigley, 1998). Upon the conclusion of the first period, the Doha Amendment to the Kyoto Protocol was made to initiate a second commitment period from 2012 to 2020 (Zakerinia & Lin Lawell, 2019). However, because developing countries such as China, India, and Brazil were exempt from the emissions targets of Kyoto and since many major emitters including the United States, Canada, Japan, and Russia showed a lack of commitment, the Doha Amendment was limited in scope to 15% of global CO2 emissions (Pagett, 2017).

The Kyoto Protocol was superseded by the Paris Agreement, the second subsidiary agreement of the UNFCCC, adopted at COP 21 in Paris, France in December 2015 (Leggett, 2020). The Paris Agreement required all 195 participating countries to contribute to a collective effort to keep the GHG-induced increase in temperature under 2° Celsius and to limit the global temperature increase to within 1.5° Celsius above the pre-industrial level. Due to the collective nature of the Paris Agreement, countries were allowed to implement their own climate policies (Nationally Determined Contributions, or NDCs), which resulted in a more flexible system than that of the Kyoto Protocol (Mehling, Metcalf, & Stavins, 2018). Continuing their lack of involvement in the Kyoto Protocol, the United States has had a limited contribution to the Paris Agreement. Former President Donald Trump pulled out of the agreement on June 1, 2017, ceasing all NDCs and financial contributions (Zhang, Dai, Lai, & Wang, 2017).

The effectiveness of the Kyoto Protocol and the Paris Agreement have been debatable, and previous studies have yielded conflicting results (Maamoun, 2019). Several studies (Aichele & Felbermayr, 2013; Grunewald & Martinez-Zarzoso, 2016; Maamoun, 2019) provide evidence that the Kyoto Protocol has influenced the reduction of emissions in participating parties. Another study (Almer & Winkler, 2017) has shown the opposite result; that the effect of the Kyoto Protocol on reducing emissions has been negligible. Due to the relative recentness of the Paris Agreement and the long-term nature of its goals, there has been little empirical evidence supporting or refuting its success, with most of the literature questioning and speculating its effectiveness (Raiser, Kornek, Flachsland, & Lamb, 2020; Clémençon, 2016).

A common theme in these studies is the focus on the reduction of GHG emissions as the metric of success. While we recognize that reducing emissions is a directly stated goal of the UNFCCC, we also claim that it is an intermediate step in reaching the final goal, which is to slow the rate of human-induced climate change. In this study, we will use the rate of change of global land surface temperatures as a metric for success rather than emissions reductions. While global surface temperatures are just one of many indicators of climate change, we use global surface temperature data for two main reasons: one, because the effects of temperature change on the environment have been closely studied, and two, because there is a large collection of raw data measuring global surface temperatures (Rohde et al., 2013).

The remainder of the paper is organized into five sections as follows. We discuss previously used methods and build our own theoretical framework for measuring the effectiveness of Kyoto in Section 2. We then describe the source of our data and provide some basic statistics and exploratory plots in Section 3. In Section 4, we first explain the mathematical framework of the ARMA time series model and how we employ the model in our forecasts and measurements. Then we detail several tests of significance and incorporate control variables, including sunspot cycles and natural climatic oscillations. We discuss the results of our method in Section 5 and conclude in Section 6.

2 Theoretical Framework

Back in 2016, the National Oceanic and Atmospheric Administration released data indicating that climate change had not slowed down over the 2000's. In response, an article published in Nature's climate change section did a breakdown of their temperature measurements over the past half-century to indicate that during the 2000's, there was indeed a trend of general slow-down in the rate of global warming, indicated by using average global mean surface temperature across many models. (Vaidyanathan, 2016)

However, in their analysis, they point to several key issues that arise when discussing global temperature change. First, they concede that simply choosing a 15-year time block out of a 50-year sample could easily be an arbitrary fluctuation, with no strong check for robustness. In particular, this period of "cool-down" could be part of a larger periodic trend dating back hundreds of years. Second, on a smaller-scale, there may be other confounding trends such as sunspot cycles or El Niño/La Niña climate cycles that additionally affect climate change. Ultimately, these sources of uncertainty present the biggest problem for any climate researcher. It is crucial to isolate as many external sources of additional change in order to determine whether current human-produced actions, such as GHG emissions rates, have positively or negatively impacted the rate of climate change. (Fyfe et al., 2016)

The "theory of change", then, that climate researchers attempt to answer, is twofold. First, do human-produced trends have a statistically significant impact on the rate of climate change? And secondly, have actions taken to change current trends, such as the Kyoto Protocol, successfully altered the rate at which climate change occurs? Most scientists have come to a consensus on the former, while it remains to be seen whether the latter is true. This paper, like many before it, will try to answer the latter question.

In general, for any climate-based research, the dependent variable would be temperature and the independent variable would be time. The job of the researcher then is to model temperature as a function of time, find the rate of increase, and use this to predict future temperature. The causal mechanism, rather than being a continuous action, rather is a discrete classifier. We classify the time points by pre- and post-Kyoto Protocol, attempting to test whether the Kyoto Protocol caused a decrease in the rate of warming.

To test our theory, we will first build a model that determines the overall rate of climate change using monthly averages of global land surface temperature data from 1850 to 1999 (We choose 1999, two years after Kyoto, to allow time for the effects of Kyoto to take place). We will then create a separate model, using data from 2000 to 2020 and determine whether this model is significantly different from the 1850 - 1999 model. We will then use the 1850 - 1999 model to predict global temperature from 2000 to 2020 and compare our predictions to the 2000 - 2020 model. This process will be done with two separate datasets: first, a dataset measuring global average surface temperature, and second, a dataset measuring the average surface temperature specifically inside the contiguous United States.

Previous studies have employed time-series models to forecast temperature change (Hecke, 2010; Ye, Yang, Van Ranst, & Tang, 2013; Kumar, Puri, Selokar, Bokre, & Talhar, 2020). Our base models will likely be very similar to many of these, as time-series predictions are nothing new in climate change. Where we hope to improve on previous literature, however, comes in two ways.

First, while prediction is nothing new, we have yet to find studies that attempt to utilize the United States as a control variable the way we do. Most models are more concerned with doing model calibration to determine whether their time-series models are a good predictor of future data. We want to take this one step further and make a causality argument by demonstrating that in the control group (United States), the rate of global warming was not decreased the same way it was outside of the United States. This would help demonstrate that the decrease we (hypothetically) measure in climate change over the past two decades was indeed the cause of foreign policy in non-U.S. developed nations.

The reason previous studies have not looked at control variable countries is likely that while there has clearly been research about time-series prediction and Kyoto Protocol independently, none of the time-series researchers have considered analyzing the significance of a landmark event. Additionally, many of the previous empirical studies done with time-series modeling are at least five years old. We use the most up-to-date temperature data to build upon previous models which only forecast temperatures. This may seem small, but considering how recent the events we are measuring are, an additional five years of climate data can be critical to finding statistical significance.

3 Data and Descriptive Statistics

To analyze rates of global temperature change over time, we require data that captures global average land surface temperatures accurately and consistently. Ideally, the global average land surface temperature would be computed by taking the average of air temperature measurements recorded by a infinite amount of equivalent weather stations at every location. However, since this situation does not exist, finding the ideal average is impossible, and estimation is required.

We found a data set collected and maintained by Berkeley Earth, an independent U.S. non-profit organization affiliated with the Lawrence Berkeley National Laboratory. According to Berkeley Earth, their current archive contains data from over 39,000 unique weather stations, combining 1.6 billion temperature reports from 16 preexisting archives. They estimate global average land surface temperature by using available data to take a land-area weighted average, as opposed to a station average (Rohde et al., 2013).

Berkeley Earth developed a novel averaging process for estimating the average Earth surface temperature, with the goals of "increasing the size of the data set used to study global climate change, bringing different statistical techniques to bear on the problem with a goal of minimizing uncertainties in the resulting averages, and reanalyzing systematic effects, including data selection bias, urban heat island effects, and the limitations of poor station siting" (Rohde et al., 2013).

Their data is meticulously filtered and cleaned through a series of steps, including eliminating duplicate records and flagging values more extreme than the 99.9% threshold for normal climate variation in each record and region. More information on their averaging process can be found in their paper (Rohde et al., 2013). Data can be further filtered by hemisphere, country, and major city. Because the U.S. was the only major developed country to abstain from Kyoto, we will look at contiguous U.S. temperature averages alongside global averages as a baseline metric.

The data found on the Berkeley Earth website reports temperatures as anomalies, in Celsius, relative to the January 1951 to December 1980 temperature averages, which is 8.60°C for global and 11.36°C for the contiguous United States. We will add these averages to the anomaly values in the original global and U.S. data, respectively, to obtain the absolute temperature values. We will include both original data in text format and our modified data in CSV format in our supplementary materials. Additionally, for each month, Berkeley Earth reports the one-year, five-year, ten-year, and twenty-year moving averages centered about that month, rounding down if the center is in between months. For example, the annual average from January to December 1950 is reported at June 1950. We will use the following variables in our study:

1. Year: starts in 1750, but only require 1850 - 2020
2. Month: 1 (January) - 12 (December)
3. Global_Avg_1: global one-year average land temperature in Celsius
4. Global_Avg_Unc_1: the 95% confidence interval around the average, accounting for statistical and spatial under-sampling effects
5. Global_Avg_10: global ten-year average land temperature in Celsius
6. Global_Avg_Unc_10: the 95% confidence interval around the ten-year average, accounting for statistical and spatial under-sampling effects

The contiguous U.S. versions of items 3 - 6 will also be used. Ten-year moving average variables (items 5 - 6) will be used for descriptive statistics, but the models will use one-year moving averages (items 3 - 4).

Period	1850-1997	1997-2020	1850-2020
Global	8.42 (0.35)	9.61 (0.23)	8.56 (0.53)
Contiguous U.S.	11.16 (0.48)	12.23 (0.48)	11.30 (0.60)

Table 1: Average temperatures (°C); standard deviations in parentheses

We will begin exploring the data by providing some basic statistics, using the one-year moving average columns. Across our period of interest (1850 - 2020), the average global temperature was 8.56°C and the average U.S. temperature was 11.30°C. Pre-Kyoto (1850 - 1997), the average global temperature was 8.42°C and the average U.S. temperature was 11.16°C. Meanwhile, post-Kyoto (1998 - 2020), the average global temperature was 9.61°C and the average U.S. temperature was 12.23°C. There has been a drastic increase in average temperatures, marked by a 1.19°C increase in average global temperature and 1.07°C increase in average U.S. temperature. Averages and standard deviations are summarized in Table 1.

Next, we will consider mean rates of temperature change using the ten-year moving average columns. Mean rates of temperature change have increased exponentially both worldwide and in the U.S., as seen in 2. Since the Kyoto Protocol was signed in 1997, the mean rate of temperature

Since	1850	1900	1950	1997
Global	1.09	1.26	1.82	2.84
Contiguous U.S.	0.98	1.23	1.42	2.56

Table 2: Mean rates of change (°C/Century)

change has more than doubled compared to mean rate of temperature change since 1850, from 1.09°C/Century to 2.84°C/Century worldwide and from 0.98°C/Century to 2.56°C/Century in the United States.

Figure 1 and Figure 2 show the trend in global and U.S. average land surface temperatures between 1850 and 2020. Each data point represents one month, at which both one-year and ten-year moving averages centered at that month are plotted (hence why the ten-year average line ends earlier than the one-year average line). The 95% confidence intervals around the means are also plotted. Due to more prominent statistical and spatial under-sampling effects, the confidence intervals are wider in earlier years than more recent years.

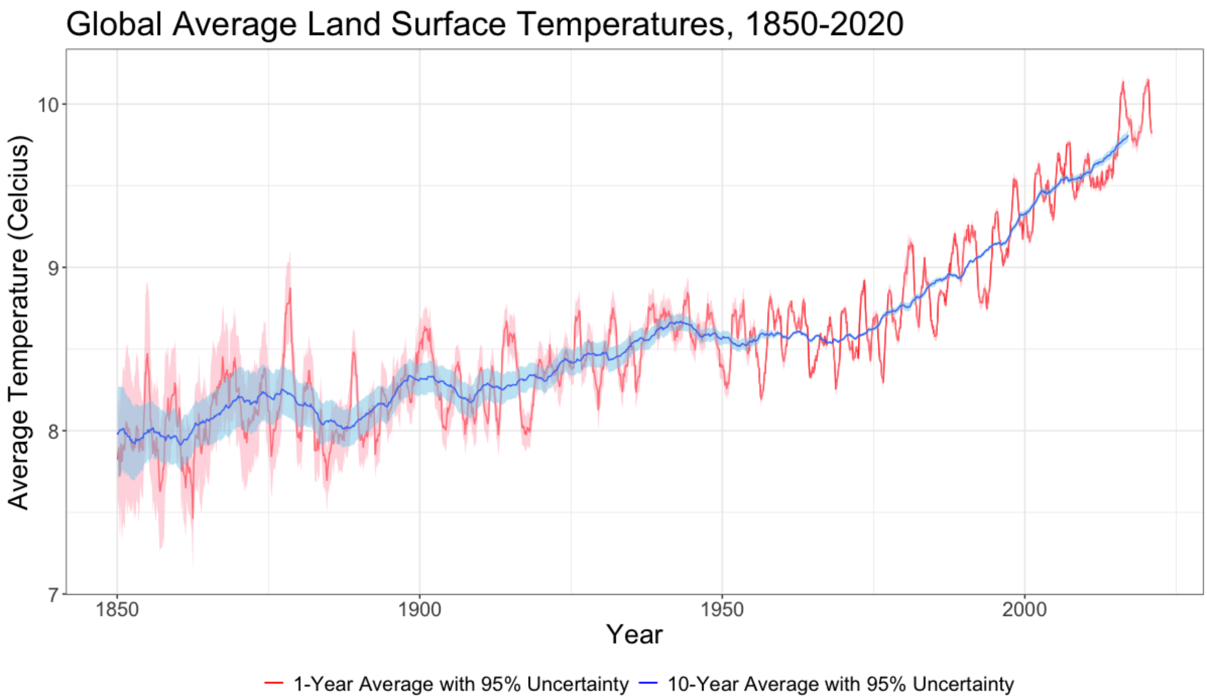

Figure 1: Global average land surface temperatures from 1850 - 2020. Note that the y-axis starts from 7°C in order to better see the trends.

Looking at the smoother ten-year average lines, there appears to be a consistent increase in average temperature. The mean rates of temperature change appear to be higher in Figure 1 than Figure 2, an observation supported by the values in Table 2. Both graphs appear to be

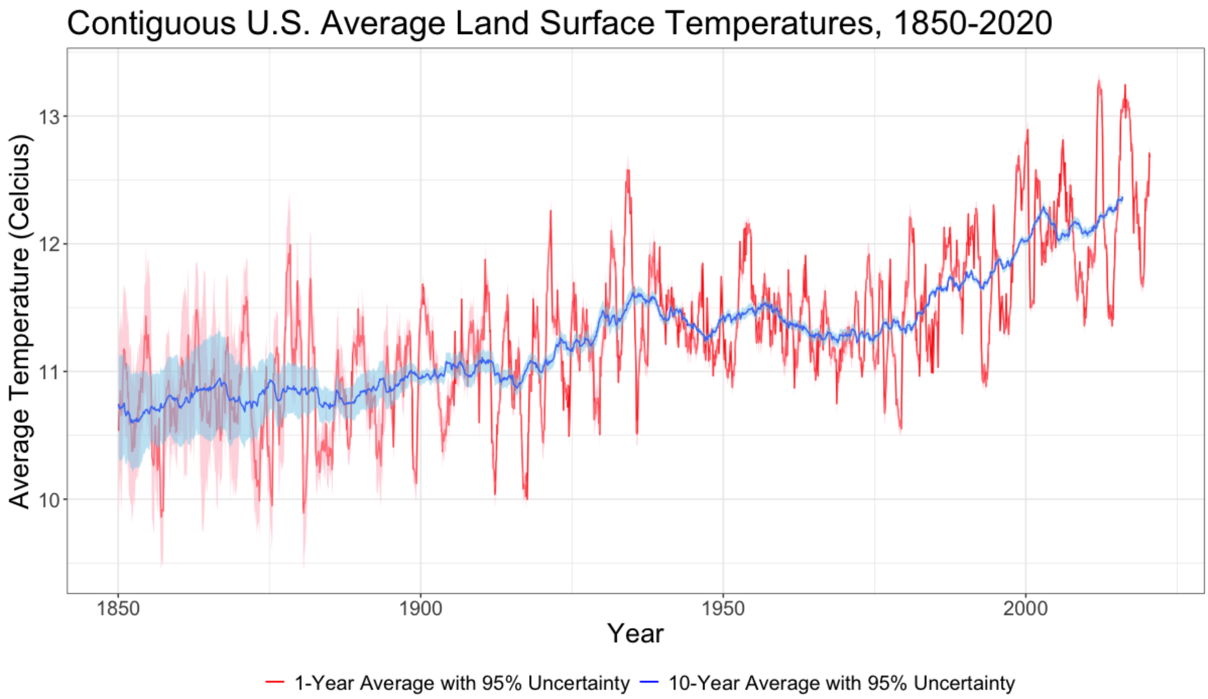

Figure 2: Contiguous U.S. average land surface temperatures from 1850 - 2020. Note that the y-axis starts from 9°C in order to better see the trends.

increasing at an increasing rate, which is also supported by Table 2.

While it is clear that there has been an increase in both mean temperature and mean rate of temperature change pre- and post- Kyoto Protocol, we cannot conclude the Kyoto Protocol has expedited global warming because these statistics offer no evidence of what would have happened in the absence of the Kyoto Protocol. Furthermore, we cannot rely only on mean rates of temperature change in determining the effectiveness of international climate policies because each data point is reliant on previous data points. Instead, we need to apply time-series techniques, which we will do through an ARMA time-series model and compare predicted vs. actual rates of change as will be described in Section 4.

4 Empirical Method

(Preface: The explanations of time-series modeling come from (Shumway & Stoffer, 2016).)

Linear regressions on time-based data make statistical significance an almost impossible proposition when testing hypotheses. With such a large amount of variance on year-to-year time-based dependencies, and the fact that said variances scale together for each subsequent time-point, almost no coefficient we cast will be statistically significant. Thus we need another framework on which to work with time-dependent data.

The solution, then, it to construct a model where each term can be traced back to prior time-points, rather then trying to find an rate of change and fit all points to a line. Time-based effects are influenced by multiple periods of previous time to varying degrees. Thus, a fitting model would be to write y_t in terms of multiple past years could be

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} \cdots + a_n y_{t-n} + w_t + \epsilon_t,$$

where w_t is a time-dependent term indicating expected additional drift.

When turning this into a testable model, we will make use of a concept called the backshift operator. Define B to be an operator on any time-related variable y_t such that $By_t = y_{t-1}$. Then we can write

$$y_t = a_1 By_t + a_2 B^2 y_t \cdots a_n B^n y_t + w_t.$$

This is useful since we can now isolate w_t in terms of y_t. Moving everything over to one side, we get $y_t - (a_1 By_t + a_2 B^2 y_t \cdots a_n B^n y_t) = w_t$. We can then write our backshift operator as a polynomial expression in terms of B to get $\phi(B)y_t = w_t$. Ultimately, any autoregressive model analysis we perform will be attempting to determine what the best-fitting model of $\phi(B)$ is.

We can extend this same logic, however, to the collection of drift terms. It's possible that whatever drift change added at time $t-1$ still affects the drift at time t, in which case we could write our time-series model as $\phi(B)y_t = w_t + a_1 w_{t-1}$. Using the same logic as before, we see we can similarly create a backshift operator to determine how many significant drift terms are relevant and get a model of

$$\phi(B)y_t = \theta(B)w_t + \epsilon_t.$$

This is known as the autoregressive moving average model, or ARMA. ARMA regressions try to determine the polynomials ϕ, θ that best model this result.

We will take annual average data of all four datasets (pre- & post-1999, U.S & global) and make our time-series model using each annual measurement. To choose the best fitting model, we make use of something called the auto-correlation function. Similar to correlation and covariance in normal linear models, auto-correlation and auto-covariance are measures of covariance within time series variables. In a time series, this is defined by

$$\gamma(t_1, t_2) = E[(x_{t_1} - E[x_{t_1}])(x_{t_2} - E[x_{t_2}])].$$

(Shumway & Stoffer, 2016).

Similarly, we get auto-correlation from this auto-covariance by dividing the above term by $\sigma_{t_1}\sigma_{t_2}$. When working with data, we can directly compute the sample covariance and standard deviation across our data to calculate the "lag" auto-correlation, which is a measure of the correlation between a time t and previous times $t-k$ across all time points t. As a normal threshold test, we usually use $\rho(k) = 0.2$ as the measure for when the lag k has a statistically significant correlation with its previous terms. By calculating the auto-correlation function for our data, we get a rough idea of where to start by picking the statistically significant y_{t-k} terms, and can thus select a starting model size.

From there, we must optimize to find the ideal model, as using auto-correlation is not enough to pick the best model. When trying to find a model with an unknown number of parameters, in this case being the number of previous time points we have to account for in our model, we have to balance model size and accuracy. Of course, if we created a massive model where we incorporate every data point, then we can come up with a perfectly fitting function. However, this level of over-fitting would render our model useless. Similarly, a linear model of just time vs temperature would have way too much variance as discovered earlier. When selecting the model of best fit, we start by utilizing the Akaike Information Criterion and Bayesian Information Criterion. These are two statistics defined as $2k - \log(L)$ and $\log(n)k - \log(L)$ respectively, where n is the number of data points inputted, k is the number of parameters chosen, and L is the variance of the best-fitting model of size k. Using these two criterion, we analyze many different models of different sizes to select the number of parameters that fits best among all four models (Shumway & Stoffer, 2016).

Once this is complete, our models will have the following traits:

1. A "best-fit" polynomial denoting the coefficient of each lagging term

2. A "best-fit" polynomial denoting the coefficient of each moving average, as well as an approximate moving average term

3. An expected standard deviation term indicating how much variance we can expect in our

model

4. p-values and standard errors for each coefficient

To give an example of what that result looks like, our model could resemble the following:

$$(1 - 0.9B + 0.3B^2 + 0.74B^3)y_t = (1 + 0.9B + 0.7B^2)w_t.$$

Each of the coefficients for B will come with a standard error and p-value the same way R produces those statistics for a linear regression. We can then empirically calculate w_t to see what the drift term is for each state, to see the expected rate of temperature change.

Finally, once we have done this, we need a testable implication. If our theory of change is true, we should observe the following four things:

1. There is a statistically significant difference in the model created with data from 1850 - 1999 global temperatures vs 2000 - 2020 global temperatures.

2. When we use the data from 1850 - 1999 to predict temperature from 2000 - 2020, the predictions yield data that is outside the expected variance we'd expect within our model of global temperatures that includes 2000 - 2020 temperature data.

3. There is a statistically significant difference between the rate of temperature change using the model generated from 2000 - 2020 global temperatures and 2000 - 2020 U.S. specific temperatures.

4. *There is no statistically significant change in the rate of temperature change generated using data from 1850 - 1999 U.S. temperatures vs. 2000 - 2020 U.S. temperatures. (This last one isn't necessarily a consequence of our theory of change, as it could very well be the case that other factors have caused U.S. temperature rates to stabilize. However, if this statement was proven true, this would, in conjunction with the other statements, indeed support our theory that the Kyoto Protocol was a significant causal factor).

To measure the first claim, we will compare the coefficients of the two models and determine whether they fall within each other's standard deviation ranges. For example, the B_1 coefficient of one model might be 0.9 with standard deviation 0.08 and the other might be 0.7. In this instance, our z-score of $(0.7 - 0.9)/0.08 = 2.5$ would indicate that there is indeed a statistically significant difference between our variables. We generalize this test over all coefficients to see whether there is a statistically significant difference between the two models using a 2-sample t-test (Tunggawan, 2018). If so, we can reject the null hypothesis that the rate of warming has not been altered post-Kyoto Protocol. Measuring the fourth claim would require the same techniques, but on U.S. temperature data instead of global temperature data.

To measure the second claim, we will do a test of mean-squared error between the 2000 - 2020 model and the predicted outcomes based on the 1850 - 1999 model. We will compare this to an expected sum of squares based on the sample variance we gathered within the model. If the total variance between the predictions from the old model and the expected values of the new model is statistically significant, using methods like F-test for analysis of variance, then we can reject the null hypothesis that the new model's results can be explained as simply variance. In other words, this will indicate that there is indeed a significant difference between our collected data.

Measuring the third claim requires us to think about cross-sample trends. We can look at two different results to make our argument - the predicted results of each dataset in terms of the other model, and the difference in rate of moving average change. For the former, we will try and predict the global temperature in terms of the U.S.-specific model, and vice versa, using prior data points to calibrate this model. To be clearer, what we'd be doing is taking the model generated from 2000 - 2020 on U.S. temperatures and using it to predict the global temperatures from 2000 - 2020, and then looking at the mean squared error. Similar to the first case, the variance being unexplained would indicate that there is a significant difference in how global temperatures and U.S. specific temperatures are changing.

In the latter case, we'd be measuring the coefficients of $B_1, B_2, \cdots B_n w_t$ and analyzing whether those are statistically significant. (We cannot do this with the raw temperature rates of change like we can within the same dataset since there's a scale difference). This allows us to measure if the difference in change rate between global temperature and U.S. temperature is statistically significant. If it is, then we can argue that global temperatures are stabilizing significantly more than U.S. temperatures, indicating that trends outside the U.S. are causing a statistically significant impact on global temperature.

Unfortunately, we cannot completely prove causality with a topic like climate change since the scope of possible factors is so large, and we cannot locally isolate an environment to test hypotheses. However, there are a few things we can do in order to try and argue causality. The main argument we will make is that utilizing the United States as a control group gives us a way to argue that the Kyoto Protocol was indeed a starting point for a large-scale change in warming trends. Since the U.S. was the last major developed nation to engage in climate change control (having only rejoined the Paris Agreement in 2021), and the only one not in Kyoto, it serves as a baseline metric to look at global progress. The Berkeley Earth data which we will be using is based on many independent sources collecting temperature points and averaging them, so we essentially have psuedo-random sampling of temperature in two sample locations (Rohde et al., 2013). Assuming all sources are collected independently and with similar methods, we have independent random data, where our primary isolated variable being that the United States has not taken the same global measures other developed nations have.

This is, of course, not perfect, as there are many other policy decisions besides Kyoto and other UNFCCC policies that can also affect climate change. Additionally, it is difficult to know to what extent U.S. temperatures might also depend on global temperature trends. So it is possible that U.S. temperatures might also stabilize at a similar rate to global temperatures due to it being a small part of global land surface area. If this becomes an issue, something we can consider is comparing the U.S. vs E.U. temperature trends as a control, to see if this gives us a stronger conclusion. Either way, this at least gives us a reasonable way to split datasets in a way that gives us a counterfactual, something that is otherwise impossible in climate changed research. As mentioned in Section 2, most time-series models on climate change have primarily been used for prediction and not in a setting testing for a potential causal mechanism. This project, while it may not make a perfect causality argument, attempts to isolate a significant variable in a way that most climate research has not attempted yet.

A separate issue we have to consider altogether is the task of de-trending. We can control for the existence of alternative mechanisms that impact global climate temperature, such as sunspot cycles, El Niño, La Niña, and other climate oscillation patterns. There have already been some papers exploring this exact subject. (Jones & Bradley, 1992) claimed that there exist both global seasonal warming trends from the 17th century onwards and an overall increasing warming trend over the past century. They analyzed warming rate changes in dozens of cities and used these long sample sizes to de-trend patterns in temperature and precipitation and make their conclusions. Additionally, (Benner, 1999) analyzed other potential confounding factors, both in seasonal temperature trends such as El Niño-Southern Oscillation patterns, and correlation with sunspot trends in solar irradience. Benner claimed that no apparent relationship to the El Niño-Southern Oscillation, but that there was indeed a strong correlation with sunspot trends.

To analyze these trends, we can use a test called periodogram analysis, which looks at the time-series data in terms of its frequency domain (i.e. how often certain data points recur) and analyzes how strong of a trend data points have with each other. From there, we can isolate seasonal trends like El Niño and determine whether or not they have a significant impact on climate change trends over time. If so, we can use a method called detrending to difference our coefficients in a way that gets rid of these trends, similar to how diff-in-diff methods work. When doing so, we get detrended time series on which we can isolate the linear effect the last decades have had on global warming overall. We will also do an analysis looking at the correlation rate in the cycles of sunspots and of climate change. (The best way to think of this is that we're creating two separate time-dependent data models and looking at correlation coefficients between them.)

Finally, using ARMA models relies on an assumption of constant variance. In time-series analysis, we can usually verify this assumption by looking at a residual diagram and determining whether or not it has constant slope. We expect this assumption will be satisfied, as past studies in time-series data on climate modeling have been able to get satisfactory predictions using ARMA models (Ye et al., 2013). If this does fail, however, which is possible given some of our models will be built using new data points and on shorter periods of time, we will include changes, utilizing time-series models for non-constant variance, that will be discussed as needed in future reports.

5 Results and Discussion

First, we start by analyzing the periodic cycles found within global climate data. What we are doing here is applying a transformation to our data to see the likelihood that our data contains a periodic term x_t such that it returns recurrently at each given time x_{t+n}. Each point in the x-axis corresponds to the frequency domain of a period in $\frac{1}{t}$ months.

As Figure 3 shows, there is no significant periodic cycle at any level in either the global temperature cycle or the U.S.-specific cycle. No terms stand out, and our highest period is simply $t = 170$, i.e., the entire dataset. Thus we can confidently say that there is not a periodic term within our dataset. This allows us to proceed with an AR(12) assumption, only including the obvious monthly period in our data.

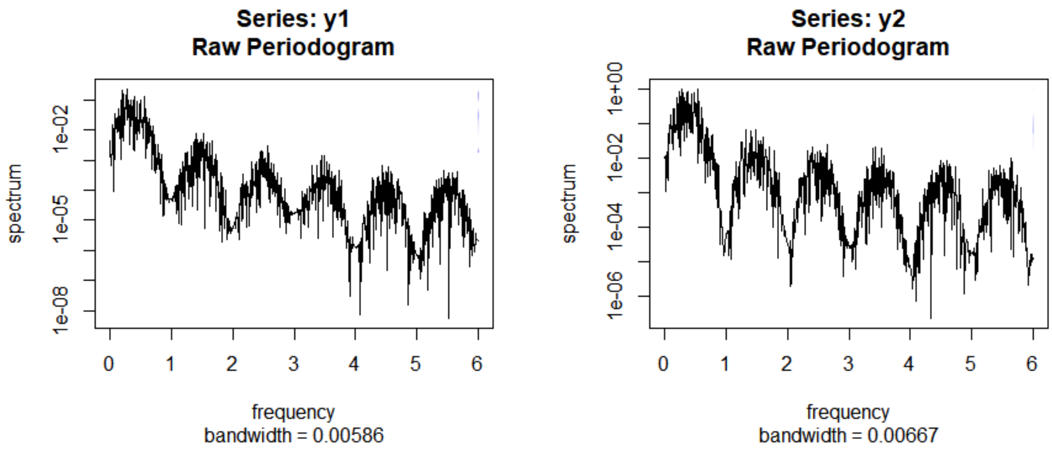

Figure 3: Periodograms showing periodic cycles for global (left) and U.S. (right).

We now consider an ARMA(p,d,q) model with fractional differences.

$$\phi(B)\ (1-B)^d\ x_t = \theta(B)$$

First, we estimate the parameter d. For small ω, we can fit a linear regression model of $\log \hat{f}(\omega) \sim \log(|\omega|)$, and our expected slope would be $\approx -2d$. The fitted linear model results show that the coefficient $-2d \approx -0.057, -0.019$ respectively, which indicates a fractional differencing component near 0, so we can assume $d \approx 0$. This allows us to operate under the ARMA(p,q) assumptions discussed above.

For the sake of comparison, we will use the same ARMA(p,q) parameters for both global and U.S. based temperatures. The parameters will be chosen as those that best fit global trends, as even if another model would forecast U.S.-specific warming rates of change better, this would in fact provide stronger evidence for our claims that global warming trends are different between the U.S. and the rest of the world post-Kyoto protocol. In other words, assuming both models can fit the same parameters does not weaken our conclusion if we do find statistical significance.

Now, we want to select an ideal moving average size, or the q in our model. To do so, we start by generating the auto-correlation function and partial auto-correlation function plot. As a reminder, the auto-correlation function is a measure of how much each variable is correlated with the other variables, and the partial auto-correlation function is the conditional variance of each x_t with its prior lag terms.

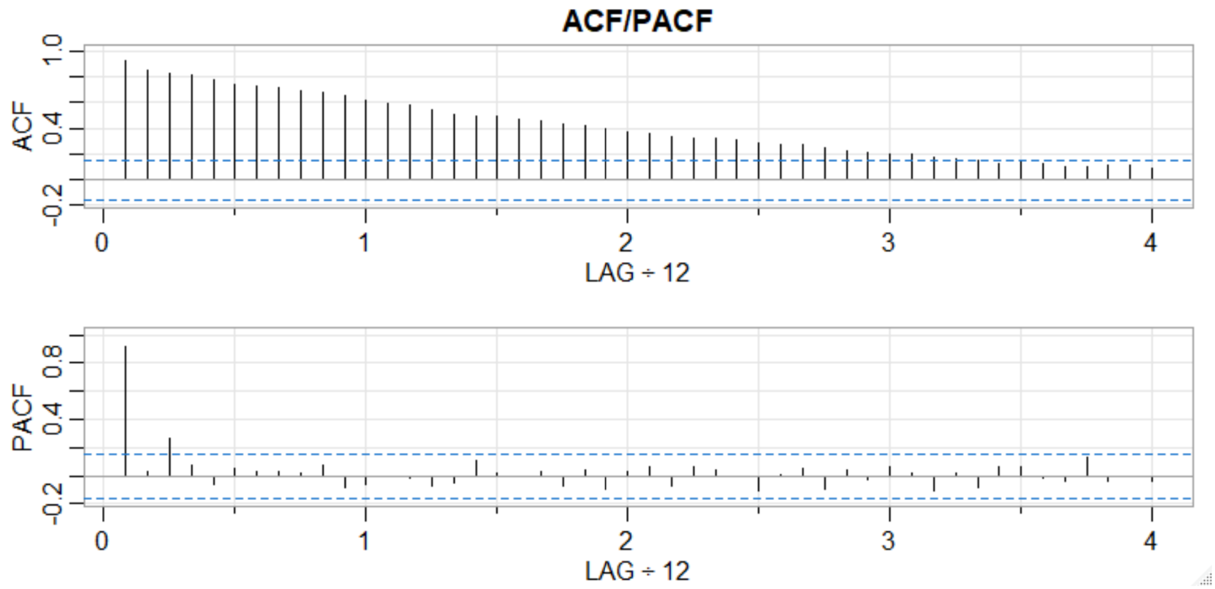

Figure 4: Auto-correlation and partial auto-correlation graphs.

These results show us that the trend of correlation drops off steadily over time, and appears to drop below 0.5 at 16 months back. To select our optimal model, we will use the Akaishe Information Criterion to balance between model size and accuracy for the given dataset by looking at models of many sizes surrounding our cutoff point. AIC is inversely proportional to log likelihood, and we want the smallest AIC, so we pick the highest log likelihood in below.

q in ARMA(12,q)	log-likelihood
10	4265.13
11	4407.69
12	4436.44
13	4436.03
14	4424.1
15	4435.49
16	4429.09
17	4420.65
18	4435.82

Table 3: AIC comparisons

We see that $q = 12$ ends up being optimal, so despite some extra lagged terms, we can actually proceed with the smaller ARMA(12,12) model. We then develop the models for each of the four cases. The resulting coefficients are in the four tables below.

The moving average coefficient is an interpretation of how much the moving average is affected by previous changes, not a statement on the rate of climate change itself. In other

Term	Coefficient	StdError	P-Value
ma1	1.3396	0.0372	$< 10^{-4}$
ma2	1.3455	0.0392	$< 10^{-4}$
ma3	1.3546	0.0409	$< 10^{-4}$
ma4	1.3631	0.0427	$< 10^{-4}$
ma5	1.3667	0.0447	$< 10^{-4}$
ma6	1.3672	0.0453	$< 10^{-4}$
ma7	1.3660	0.0439	$< 10^{-4}$
ma8	1.3588	0.0423	$< 10^{-4}$
ma9	1.3513	0.0410	$< 10^{-4}$
ma10	1.3382	0.0376	$< 10^{-4}$
ma11	1.3496	0.0387	$< 10^{-4}$
ma12	0.3550	0.0402	$< 10^{-4}$

Table 3: Global Model - 1850-1999

Term	Coefficient	StdError	P-Value
ma1	1.4156	0.1102	$< 10^{-4}$
ma2	1.3549	0.1223	$< 10^{-4}$
ma3	1.4245	0.1319	$< 10^{-4}$
ma4	1.4663	0.1414	$< 10^{-4}$
ma5	1.4681	0.1463	$< 10^{-4}$
ma6	1.4683	0.1502	$< 10^{-4}$
ma7	1.4678	0.1497	$< 10^{-4}$
ma8	1.4651	0.1471	$< 10^{-4}$
ma9	1.3851	0.1454	$< 10^{-4}$
ma10	1.3640	0.1228	$< 10^{-4}$
ma11	1.4579	0.1186	$< 10^{-4}$
ma12	0.5004	0.1183	$< 10^{-4}$

Table 4: Global Model - 2000-present

Term	Coefficient	StdError	P-Value
ma1	1.1642	0.0490	$< 10^{-4}$
ma2	1.1809	0.0507	$< 10^{-4}$
ma3	1.1763	0.0538	$< 10^{-4}$
ma4	1.1742	0.0543	$< 10^{-4}$
ma5	1.1709	0.0543	$< 10^{-4}$
ma6	1.1704	0.0543	$< 10^{-4}$
ma7	1.1735	0.0542	$< 10^{-4}$
ma8	1.1750	0.0540	$< 10^{-4}$
ma9	1.1824	0.0519	$< 10^{-4}$
ma10	1.1672	0.0499	$< 10^{-4}$
ma11	1.1623	0.0535	$< 10^{-4}$
ma12	0.1620	0.0540	0.0028

Table 5: US Model: 1850-1999

Term	Coefficient	StdError	P-Value
ma1	1.5230	0.1586	$< 10^{-4}$
ma2	1.5436	0.1790	$< 10^{-4}$
ma3	1.5729	0.1921	$< 10^{-4}$
ma4	1.5524	0.2005	$< 10^{-4}$
ma5	1.5516	0.2015	$< 10^{-4}$
ma6	1.5594	0.2059	$< 10^{-4}$
ma7	1.5443	0.2038	$< 10^{-4}$
ma8	1.5673	0.1968	$< 10^{-4}$
ma9	1.5631	0.2008	$< 10^{-4}$
ma10	1.5350	0.1796	$< 10^{-4}$
ma11	1.5202	0.1732	$< 10^{-4}$
ma12	0.5173	0.1688	0.0025

Table 6: US Model: 2000-present

words, it tells us that the rate change from t years ago has an influence on the current year multiplied by some a_t attached. A difference in coefficients tells us that the models are different, as the rates of change in climate are different, not a fact about the linear rate at which warming is going up or down.

Over the span of our cycle, the moving average component within global data from 1850-1999 (note that the x-axis is labelled as years after 1850) ended up being pretty similar for the first through eleventh lagged terms $w_t \approx 1.35$, with standard error ≈ 0.03, before the final lagged term drops off a lot. We see a similar story for the model generated over 2000 - 2020, with our average w_t coefficients over 2000 - 2020 was $w_t \approx 1.42$. All our differences in coefficients fall outside the 95 percent confidence interval, indicating there is a statistically significant change in the trend for global climate change.

By comparison, we look at our control group on just the U.S., as well as a second pseudo-control on Central England climate data. The moving average component within global data from 1850 - 1999 was $w_t \approx 1.17$, with standard error ≈ 0.05. Our average w_t coefficients over

2000 - 2020 was $w_t \approx 1.55$. (Again, the final term falls outside this pattern). This falls outside the 95% confidence interval, indicating there is a statistically significant change in the trend for global climate change. These conclusions tell us that in both the case of global data and U.S. specific data, we have significant reason to believe that the change rate of warming has been impacted in a statistically significant way post-Kyoto.

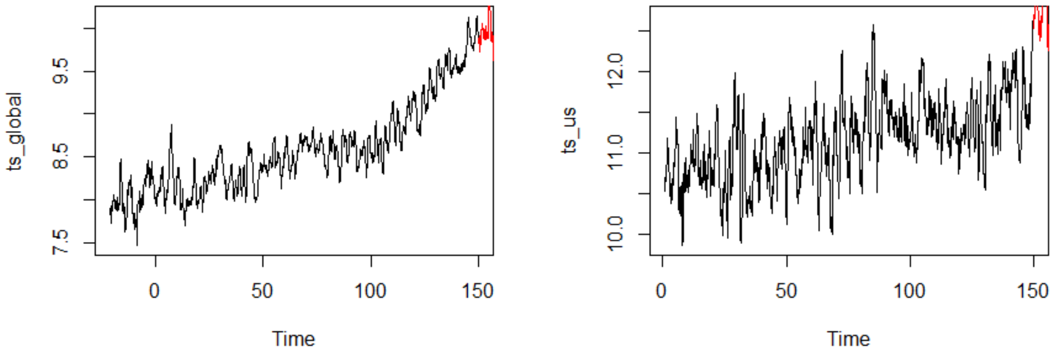

Figure 5: Predicted global (left) and U.S. (right) temperatures using the 1850 - 1999 models are highlighted in red.

Referring to Figure 5, we see that that the predicted results using the 1850 - 1999 model are higher than the predicted results using the 2000 - 2020 model. This indicates that the trends we observe with current data show a pattern of warming slowdown. Now, we want to look at residuals and our mean-squared error between our predictions and the actual results. We can use ARMA to forecast predictions for 2000 - 2020 temperature based on the model generated on the dataset from 1850 - 1999, and compare those to our real values. Additionally, we look at the residuals of the model we generated in order to ensure validity via the Box-Pierce test.

Model	Global	US
R-Squared	0.568	0.6635
P-Value	0.3639	0.1893

Table 3: R^2 and p-values for Global and U.S.

All of our p-values are above 0.05, so we fail to reject the null hypothesis that the residuals are indeed normally distributed, meaning our model works. We also calculated the mean-squared errors for the predicted results our model on 1850 - 1999 would have gotten for 2000-present data versus the actual data results. In the global case, we have a mean-squared error of 566.19, and in the U.S.-specific case, we have a mean-squared error of 534.16. In other words, past temperature trends are a worse predictor for current global temperatures than they are for past U.S. data. However, comparing the mean-squared error on its own is not sufficient to say that the two models have distinct when they are so similar. Thus we don't have a sufficient basis for arguing that the Kyoto protocol as a control was uniquely the reason for altered warming trends.

6 Concluding Remarks

Lastly, we will conclude by by addressing each of the four items is Section 4.

1. We found a significant different in the model created with data from 1850 - 1999 global temperatures vs 2000 - 2020 global temperatures.

2. For both global and U.S., when we used the data from 1850 - 1999 to predict temperature from 2000 - 2020, the predictions indeed yielded data outside the expected variance we'd expect within our model of global temperatures from 2000 - 2020. The predicted temperatures were significantly higher than the actual temperatures, indicating a decrease in the rate of temperature change from 2000 - 2020.

3. We did not find a statistically significant difference between the rate of temperature change using the model generated from 2000 - 2020 global temperatures and 2000 - 2020 U.S. specific temperatures. Thus, we cannot use the U.S. as a control. We will discuss the implications of this later.

4. *There is a statistically significant change in the rate of temperature change generated using data from 1850 - 1999 U.S. temperatures vs. 2000 - 2020 U.S. temperatures. This result aligns with items 1 and 3, since there was a significant difference in the rate of change of global temperatures pre- and post-Kyoto, and that there was not statistically significant difference between global and U.S. temperatures in the post-Kyoto model.

Our results have shown that the rate of temperature increase has slowed since 2000, supporting the argument that climate change is indeed reversible. However, whether this slowdown is anthropogenic or naturally occurring remains in question. Because we did not find a statistically significant difference between the rate of temperature change using the model generated from

2000 - 2020 global temperatures and 2000 - 2020 U.S. specific temperatures, we cannot make the claim that the United States, as the only major country to largely ignore international policies in the past two decades, can act as a control environment in the absence of Kyoto and other policies. Thus, our reservations about possibility of using the U.S. as a control in Section 4 (we mentioned the possibility of U.S. temperatures stabilizing at a similar rate to global temperatures) were validated. Thus, we cannot infer causality, i.e. we cannot claim that international climate policies beginning with the Kyoto Protocol caused the slowdown of temperature increase.

Besides the lack of a control, there are several other limitations to our study that ought to be addressed. One, there are many alternative factors besides international climate policies affecting climate change that are difficult to measure and control for on a global scale, such as emissions from corporations and national and regional policies. Two, we did not collect the data ourselves, rather, we used Berkeley Earth's data, which uses its own averaging algorithm to consolidate millions of temperature records (Rohde et al., 2013). While we trust that their process provides generally accurate results, the fact that temperature data was collected from many (possibly correlated) data sources could result in higher correlations than should be expected. This is particularly true of instances where Berkeley earth is backdating their data collection process for data that was not collected in that exact year. Similar methods used in computing the backdated data points might result in the data points being correlated.

Despite the limitations of our study, we can conclude that trends in global temperature increase are reversible, which is a promising result. Combatting climate change is a difficult task that has and will continue to require a lot of research and resources, but our results show that this is not for naught; rather, it should encourage ourselves and other researchers to further. From the experience of performing this study, we have several ideas that research in the future could incorporate or pursue. One, models could be built with heavier machinery. There are more suitable, more computationally intensive models than ARMA, such as Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) and multivariate linear Gaussian, but we are limited by computing power. Two, studies using more data points, like weekly or even daily temperatures, may yield higher significance. Three, there are important measures of climate change besides land surface temperatures that can be considered in a similar manner, including rising ocean temperatures, rising sea levels, and decreasing snow coverage. Lastly, we believe that our idea of using a pseudo-control is valid in theory, but perhaps not on such a large scale as using U.S. temperatures as a control for global temperatures. For inferring causality, it will be more practical to conduct studies on a smaller scale, observing the effect of international policies locally. Since countries are generally given climate targets and not specific instructions on how to reach those targets, an example would be to select a country, determine how the international policy has manifested itself in national policies, and measure the effect of those national policies. In this case, it would be easier to find uncorrelated pseudo-controls, as well as consider other factors that may impact climate change such as major local industries and emissions from major corporations in that country.

References

Aichele, R., & Felbermayr, G. (2013). The effect of the Kyoto Protocol on carbon emissions. *Journal of Policy Analysis and Management*, *32*(4), 731-757.

Almer, C., & Winkler, R. (2017). Analyzing the effectiveness of international environmental policies: The case of the Kyoto Protocol. *Journal of Environmental Economics and Management*, *82*, 125-151.

Benner, T. C. (1999). Central England temperatures: long-term variability and teleconnections. *International Journal of Climatology*, *19*(4), 391-403.

Clémençon, R. (2016). The two sides of the Paris Climate Agreement: Dismal failure or historic breakthrough? *The Journal of Environment & Development*, *25*(1), 3-24.

Fyfe, J. C., Meehl, G. A., England, M. H., Mann, M. E., Santer, B. D., Flato, G. M., ... Swar, N. C. (2016). Commentary: Making sense of the early 2000s warming slowdown. *Nature*.

Grunewald, N., & Martinez-Zarzoso, I. (2016). Did the Kyoto Protocol fail? An evaluation of the effect of the Kyoto Protocol on CO_2 emissions. *Environment and Development Economics*, *21*(1), 1–22.

Hecke, T. V. (2010). Time series analysis to forecast temperature change..

Jones, P., & Bradley, R. (1992). Climatic variations in the longest instrumental records. *Climate Since A.D. 1500: Revised Edition*, 246-268.

Kumar, A., Puri, S., Selokar, P., Bokre, R., & Talhar, Y. (2020). Forecasting of temperature by using time series analysis. *International Journal of Research and Analytical Reviews (IJRAR)*, *7*(1), 547-551.

Leggett, J. A. (2020). The United Nations Framework Convention on Climate Change, the Kyoto Protocol, and the Paris Agreement: a summary. *Congressional Research Service*.

Maamoun, N. (2019). The Kyoto Protocol: Empirical evidence of a hidden success. *Journal of Environmental Economics and Management*, *95*, 227-256.

Mehling, M. A., Metcalf, G. E., & Stavins, R. N. (2018). Linking climate policies to advance global mitigation. *Science*, *359*(6379), 997-998.

Pagett, R. (2017). *Building global resilience in the aftermath of sustainable development.* Berlin: Springer International Publishing.

Raiser, K., Kornek, U., Flachsland, C., & Lamb, W. F. (2020). Is the Paris Agreement effective? a systematic map of the evidence. *IOP Publishing*, *15*(8).

Rohde, R., Curry, J., Groom, D., Jacobsen, R., Muller, R. A., Perlmutter, S., ... Wurtele, J. (2013). Berkeley Earth temperature averaging process. *Berkeley Earth*.

Shumway, R. H., & Stoffer, D. S. (2016). Time series analysis and its applications, 4th edition. *Springer Texts in Statistics*.

Tunggawan, E. (2018). *Time series analysis: validating effect of changes.* Retrieved from https://elvyna.github.io/2018/time-series-hypothesis-testing/

Vaidyanathan, G. (2016). *Did Global Warming Slow Down in the 2000s, or Not?* Retrieved 2022-05-02, from https://www.scientificamerican.com/article/did-global-warming-slow-down-in-the-2000s-or-not/

Wigley, T. M. L. (1998). The Kyoto Protocol: CO_2 CH_4 and climate implications. *Geophysical Research Letters*, *25*(13), 2285-2288.

Ye, L., Yang, G., Van Ranst, E., & Tang, H. (2013). Time-series modeling and prediction of global monthly absolute temperature for environmental decision making. *Advances in Atmospheric Sciences*, *30*(2), 382-396.

Zakerinia, S., & Lin Lawell, C.-Y. C. (2019). Climate change policy: dynamics, strategy, and the Kyoto Protocol [2019 Annual Meeting, July 21-23, Atlanta, Georgia].

Agricultural and Applied Economics Association.

Zhang, H.-B., Dai, H.-C., Lai, H.-X., & Wang, W.-T. (2017). U.S. withdrawal from the Paris Agreement: Reasons, impacts, and China's response. *Advances in Climate Change Research*, *8*(4), 220-225.