Determining the Significance of Global Warming on Earth's Temperatures

Kory Begley[1,*]

[1]Independent Researcher, Bachelor of Science in Data Analytics

Abstract

This study provides a comprehensive analysis of 300 years of global temperature data to examine the ongoing trend of global warming. Using data sourced from Berkeley Earth, statistical techniques such as linear regression and the Mann-Kendall test were applied to assess temperature changes from 1750 to 2024 at a 95% confident level. The results revealed that there is a statistically significant increase in global temperatures on Earth, with the fastest warming observed in the colder months, with slopes ranging from 0.006 - 0.008 degrees (C) per year, while the summer months calculated slopes from 0.004 - 0.005. The Mann-Kendall test determined that each month, (as well as by year), the global temperature has an increasing trend at 95% confidence. To contextualize these numbers, the underlying causes of global warming, also referred to as "the Greenhouse Effect", can be attributed to this rise in temperature. Through human activity, carbon and other gases are released into the atmosphere and trap the suns heat before escaping out into space. With more pollutants in the air, it becomes increasingly difficult for these gases to escape, and stay longer, becoming increasingly potent.

Key words: global warming; greenhouse gas; climate change; temperature

Introduction

The 5 warmest years on record (since the 1800s) have all occurred since 2016, with July 2024 being in contention to be the hottest month on record, with enough intense heat to make scrambled eggs on any unassuming Arizona sidewalk. This is a continuation of a trend humanity has been experiencing since the industrial revolution, aptly named global warming. This report intends to analyze 300 years of temperature data, gathered from Berkeley Earth using their Berkeley Averaging method through the statistical methods: linear regression and the Mann-Kendall test. In addition, the results of the analysis will be compared to institutions such as NASA and the NOAA, acting as a reassurance to the results gathered.

The aggregation process consisted of Python library Pandas, and the statistical analysis was performed through Python libraries sklearn, and pymannkendall, and visualized with MatPlotLib. As a result of the statistical analysis and observations of the what, how, and who of global warming, this report concluded that the increase in temperature over the last 300 years is statistically significant and has been steadily increasing over time. In retort, this report explored the root causes, such as the Greenhouse Effect, and humanities hand in rising carbon emissions since the industrial revolution. Highlighting the need for widespread change and cooperation with institutions like the NOAA and NRDC to combat our carbon emission processes so that the Earth can mitigate the catastrophes that come with rising temperatures.

A Warm Introduction

If you've ever seen those giant balloons that soar high in the atmosphere and wondered what exactly those do, well, those are weather balloons! They're one of many tools used to measure temperatures around the world; and these aren't a modern invention, meteorologists as far back as 1896 have been utilizing these devices, with French meteorologist Léon Teisserenc de Bort used weather balloons to find the top of the troposphere and stratosphere.[1] But now in the modern day, the Global Historical Climatology Network daily dataset, preserved by the NOAA is a major source for the up to date records and handling of data that relate to temperatures, wind speeds, and other climate conditions.[2]

The reason this practice is so important is because temperature influences variables all around us, heat waves, water availability, increased deforestation and habitat destruction, sea level increases, and all other variables are hosts of negative impacts on human, and animal lives.[3] This is also why it was imperative to start the effort of finding the temperatures and climate conditions of the past, to see how things have changed the last few hundred years. Scientists have discovered and been using a myriad of ways to get this data

that reinforces and expands our idea of the world we inherited, whether it be through ice cores, rock sediment, or coral reefs.[4]

This analysis will cover a roughly 274 year period, however, it was only until 1850 where the data gathered for this report covers the entire globe, whereas the century before that only captured land temperatures. This data remains relevant as it still captures the trends in temperatures and allows for a deeper look into the potential variables influencing this rise in temperature over time.

Data Sourcing and Aggregation

Berkeley Earth stands as the primary source of data for this analysis, using their Land Monthly Average Temperature dataset[5], this data was gathered using the Berkeley Averaging method. The Berkeley Averaging method is a mathematical framework to measure temperatures using an anomaly-based method from an accepted baseline (Jan 1951-Dec 1980), and when Berkeley Earth ran an analysis on this method, they found their results were within acceptable estimates made by NOAA, NASA, and the Hadley Center/Climate Research Unit at 95% confidence.[6] So this method was applied to their complete dataset, resulting in the set before us today, which is still updated and ran regularly.

Below is the baseline temperatures (in Celsius) for each month that each row of data references (depending on the month):

Jan	Feb	Mar	Apr	May	Jun
2.57	3.19	5.29	8.29	11.27	13.41
0.06	0.06	0.06	0.06	0.06	0.06

Jul	Aug	Sep	Oct	Nov	Dec
14.29	13.83	12.05	9.20	6.06	3.61
0.06	0.06	0.06	0.06	0.07	0.07

The bottom value (that ranges from .06 - .07 is the margin of error (in Celsius) for each month, for simplicity, this study did not take into account the variation and took the temperatures at face value. It is worth mentioning though that the temperatures measured from this method do have a margin.

Since there was no easy way to download the data directly, the method used to import the data was through the Requests library in Python, loading it into a csv and then making a dataframe out of the delimited and formatted dataset. Below is a basic synopsis of the aggregation techniques performed on the dataset for ease of use and practicality.

Figures 1 - 6 cover the aggregation process for this study:

Fig. 1: Load data using python requests

Fig. 2: Read into a csv and check the first few rows

Fig. 3: Drop unnecessary columns and nulls

Fig. 4: Function to make the anomaly-based data in a Celsius format

Fig. 5: Another function but for Fahrenheit

```
In [85]:   #Doing a couple test cases for random years

           temps_2023 = globalTemps[globalTemps['Year'] == 2023]
           print(temps_2023)
                 Year  Month  Anomaly  Unc.  Celsius  Fahrenheit
           3276  2023    1    1.230   0.061    3.80      38.84
           3277  2023    2    1.428   0.080    4.62      40.32
           3278  2023    3    1.941   0.067    7.23      45.01
           3279  2023    4    1.227   0.094    9.52      49.14
           3280  2023    5    1.043   0.077   12.31      54.16
           3281  2023    6    1.315   0.107   14.72      58.50
           3282  2023    7    1.502   0.065   15.79      60.42
           3283  2023    8    1.646   0.092   15.48      59.86
           3284  2023    9    2.151   0.090   14.20      57.56
           3285  2023   10    2.092   0.097   11.29      52.32
           3286  2023   11    2.073   0.150    8.13      46.63
           3287  2023   12    2.003   0.137    5.61      42.10

In [86]:   temps_1895 = globalTemps[globalTemps['Year'] == 1895]
           print(temps_1895)
                 Year  Month  Anomaly  Unc.  Celsius  Fahrenheit
           1740  1895    1   -1.158   0.443    1.41      34.54
           1741  1895    2   -1.039   0.437    2.15      35.87
           1742  1895    3   -0.563   0.317    4.73      40.51
           1743  1895    4   -0.028   0.361    8.26      46.87
           1744  1895    5   -0.282   0.358   10.99      51.78
           1745  1895    6   -0.309   0.276   13.10      55.58
           1746  1895    7   -0.115   0.340   14.18      57.52
           1747  1895    8   -0.261   0.361   13.57      56.43
           1748  1895    9   -0.221   0.296   11.83      53.29
           1749  1895   10   -0.069   0.227    9.13      48.43
           1750  1895   11   -0.202   0.308    5.86      42.55
           1751  1895   12   -0.090   0.353    3.52      38.34
```

Fig. 6: Verify data on random chosen years

How Temperatures Have Changed Over 300 Years

The first statistical analysis performed was a linear regression analysis for each month, this would allow us to see the change in global temperatures overtime and give a sense of how temperatures will continue to increase.[7] Figures 7 - 18 demonstrate the change over time:
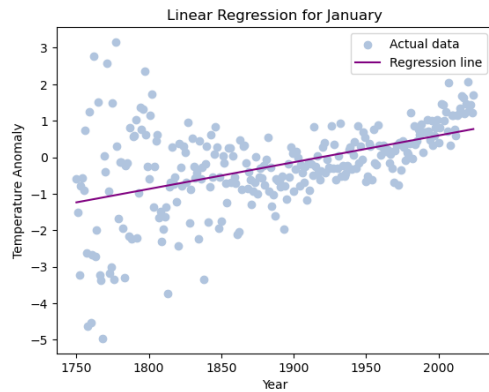
Fig. 7: Linear regression for January's temperatures
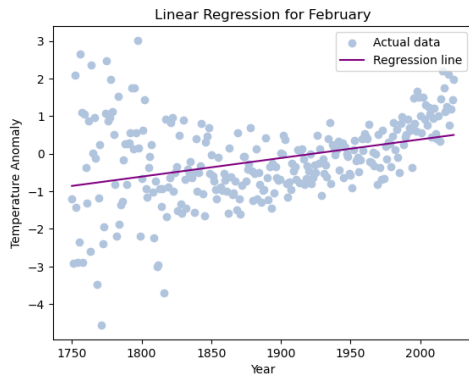
Fig. 8: Linear regression for February's temperatures

Fig. 9: Linear regression for March's temperatures
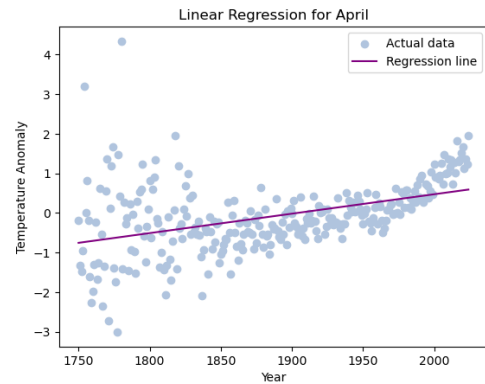
Fig. 10: Linear regression for April's temperatures
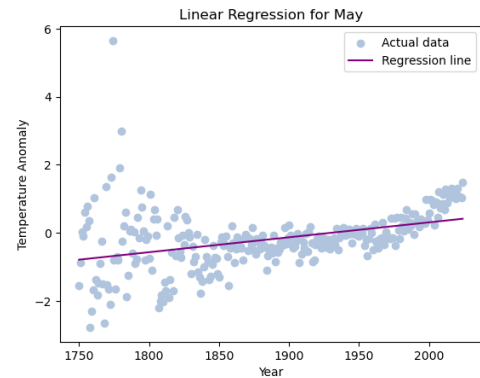
Fig. 11: Linear regression for May's temperatures
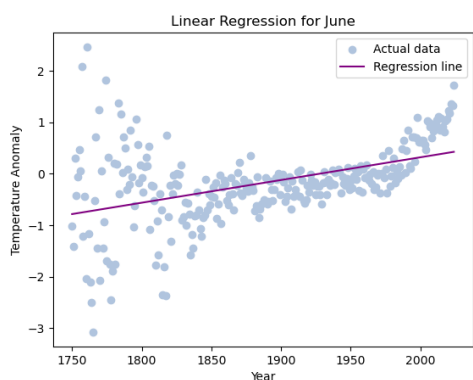
Fig. 12: Linear regression for June's temperatures

Fig. 14: Linear regression for August's temperatures
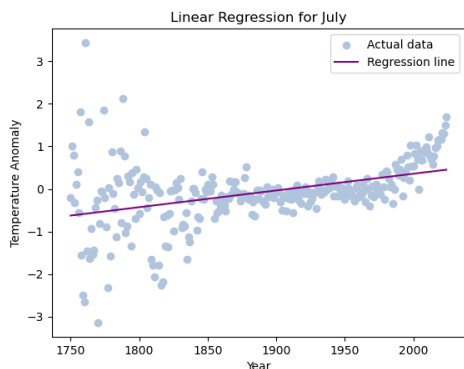
Fig. 13: Linear regression for July's temperatures
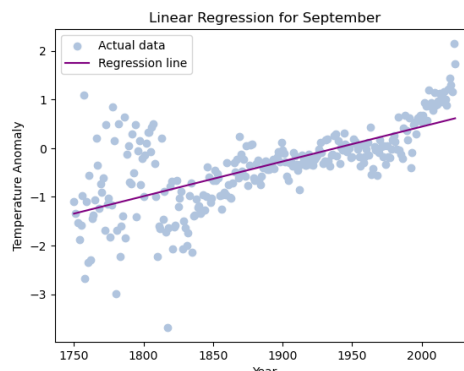
Fig. 15: Linear regression for September's temperatures

The linear regression slopes from the above figures are the following (rounded):

Jan	Feb	Mar	Apr	May	Jun
0.0073	0.0050	0.0071	0.0049	0.0044	0.0044

Jul	Aug	Sep	Oct	Nov	Dec
0.0039	0.0047	0.0071	0.0083	0.0066	0.0088

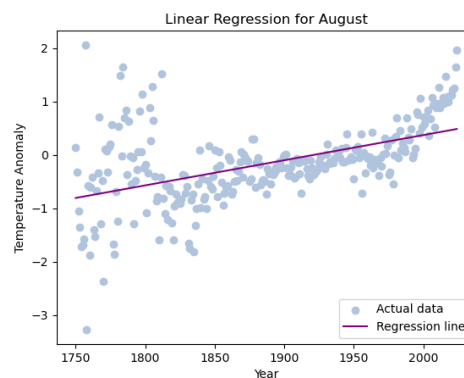Interestingly, the colder months have the highest rate of change, which is, according to the NOAA is due to a term coined as "Arctic Amplification", stating: "...dominated by the fact that the arctic is the fastest-warming large region on the plane...is driven by a handful of factors; the largest of these is the retreat of seasonal snow and ice"[8] The reason this makes the cold climates warm faster is because it creates higher minimums (in temperature) and longer stretches of time where the temperature remains warmer then it previously was. And this isn't isolated to just the poles, conditions in America are susceptible to the exemplified reasons,

with NOAA explaining: "you may have noticed that this year's extensive warmth in the western United States is driven by extremes in minimum (overnight low) temperatures, even more so than maximum (afternoon high) temperatures..."[8] The NOAA found that the winter months were rising at about twice the rate of the summer months, which can be reflected here in our slopes, with October, November and December reaching up to .008, while the summer months stay in the .004 range.

In addition to the slopes, the R-squared values were calculated (listed below, rounded), which represents the line of best fit and how accurate the actual measured data varies from our dependent variable.[9]

Jan	Feb	Mar	Apr	May	Jun
0.23	0.13	0.28	0.19	0.16	0.21

Jul	Aug	Sep	Oct	Nov	Dec
0.17	0.26	0.46	0.45	0.28	0.33

Table 1.

Fig. 16: Linear regression for October's temperatures

Fig. 17: Linear regression for November's temperatures
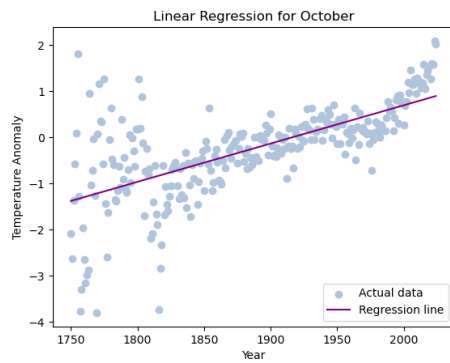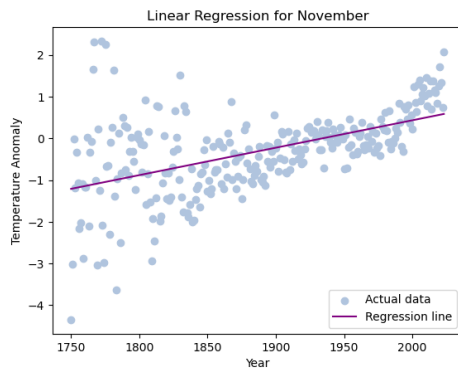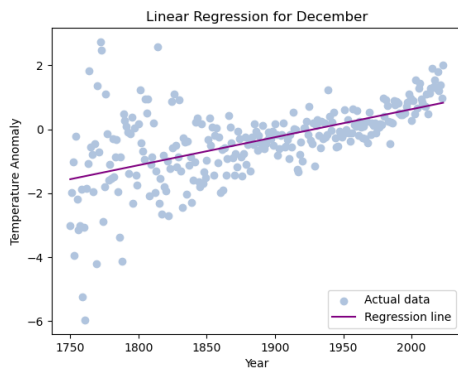

Fig. 18: Linear regression for December's temperatures

These values are low-medium which works against the regression model representing the data accurately, leading to these models not being the most effective if we were to use them to predict future temperatures, but if we were to omit the 1750-1850 range our R-squared values would shoot up and more accurately fit the data. This is due in part to the high variance of data from before the standards were created, and the data only measuring land temperatures until 1850. So while linear regression doesn't tell the whole story of statistical significance and accuracy, we now have a basic understanding of how the temperature anomalies have

changed overtime in comparison to the baseline. In order to test for statistical significance, a Mann-Kendall test was performed on each month, this resulted in a trend statement: (Increasing, Decreasing, No Trend), a P-value, and an S-Statistic. A Mann-Kendall test was chosen due to its independence from non-normal distribution of data, and simplicity in measuring the significance of the trends in data at a 95% confidence level. With a null hypothesis of "Temperatures haven't increased universally over the 300 year period", we can perform the test to affirm/reject the trend we saw in the linear regression.

```python
# Perform a Mann-Kendall test for each month

month_results = {}
for month in range(1, 13):  # Loops through months (1 to 12)
    month_data = globalTemps[globalTemps['Month'] == month]['Anomaly']  # Filter the data for the month
    result = mk.original_test(month_data)  # Perform the Mann-Kendall test
    month_results[month] = result

# Display
for month, result in month_results.items():
    print(f"{month}: Trend={result.trend}, p-value={result.p}, S={result.s}")
```

Fig. 19: Code to perform the Mann-Kendall test

Below is the results of the Mann-Kendall test performed on the data by month:

Jan	Feb	Mar
Trend = Increasing	Trend = Increasing	Trend = Increasing
p-value=0.0	p-value=2.1449e-25	p-value=0.0
S=14214.0	S=11188.0	S=15676.0

Apr	May	Jun
Trend = Increasing	Trend = Increasing	Trend = Increasing
p-value=0.0	p-value=0.0	p-value=0.0
S=14912.0	S=14484.0	S=14271.0

Jul	Aug	Sep
Trend = Increasing	Trend = Increasing	Trend = Increasing
p-value=0.0	p-value=0.0	p-value=0.0
S=12850.0	S=16552.0	S=20353.0

Oct	Nov	Dec
Trend = Increasing	Trend = Increasing	Trend = Increasing
p-value=0.0	p-value=0.0	p-value=0.0
S=21211.0	S=16021.0	S=17932.0

As can be seen from the results, there's a universal relationship in the rise in global temperatures and time, this is affirmed by the P-value, being far less than the alpha value chosen in confidence for this test ($p < 0.05$); for the sake of simplicity in analysis, the P-values measured are essentially equal to 0.01, even though they are actually less. Applied to the context of this analysis, this means that we can reject our null hypothesis, reassuring the trends observed in the linear regression analysis that global temperatures have increased over the 300 year period of data we have.

The significance of represents that observations obtained later in time tend to be larger than observations made earlier, and the larger the value of S, (the larger the variance between the earlier data from the later data). Directly contributing to the trend observed from the P-value, and linear regression analysis, as time passes, the global temperature increases more, and starts to increase at larger intervals (as can be noticed in the most recent points in figures 7 - 18, with the last 5 years being consistently

the hottest 5 years, bar the data pre-1750 which didn't take into account ocean temperatures).

Looking at the data from a monthly perspective isn't the only way to measure these trends; however, we can also look at the yearly anomalies and get a more conventional sense of how the Earth has been changing over time. The process to measure significance for the yearly data is the same as the process for the monthly data, performing a linear regression analysis, and affirming the hypothesis with a Mann-Kendall test to test statistical significance. In addition, we can look at other institutions' measurements and compare how our measure of anomalies fits within their methods, as our data was gathered using the Berkeley Averaging method, facilities like NASA or the NOAA use different methods to capture their temperature data.

Similar to the monthly data, a linear regression of the yearly anomalies was plotted and fitted with a line of best fit (image below), and continues to reassure the trends initially observed.
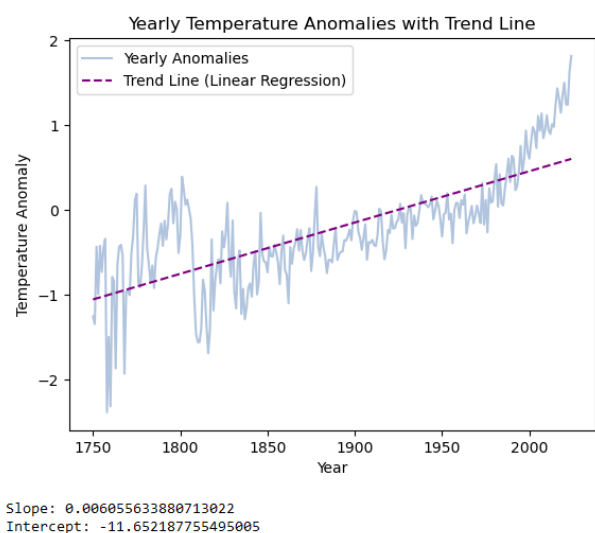

Fig. 20: Linear regression for yearly temperature change

Found at the bottom of the figure is also the slope, which is (when rounded) 0.006, this could've also been found by finding the average of the slopes calculated from figures 7 - 18.

When compared to other institutions we can notice the similarity of the rise in temperature:[10]
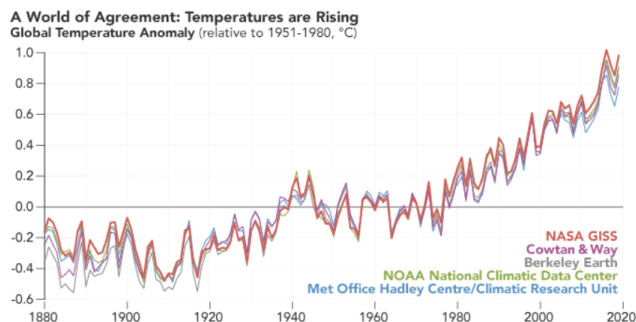

Fig. 21: Anomaly change in temperature by year as measured from NASA, NOAA, and others

The first observation to notice is that figure 21 only covers 1880-2020, but inferences can still be seen in the graph processed from the linear regression analysis. At about the 1980 mark we see the large shift in temperatures where the trend has a steeper slope, whereas the years previous were less extreme. We can also see similar peaks in years like 1940, and dips like 1910 and 1960. Another variable to consider is the baseline being used, which fortunately, the Berkeley Averaging method uses the same temperature baseline as in the above figure, giving us more consistency in data and accurate parallels to the observations we can make.

In addition to the linear regression, a Mann-Kendall test was performed on the yearly data, with the same null hypothesis of "Temperatures haven't increased universally over the 300 year period", resulting in similar conditions seen from the original monthly analysis:

Years (1750-2024)
Trend: increasing
p-value: 0.0
S: 21954.0

Which once again allows us to reject the null hypothesis at a 95% confidence, providing more evidence to the notion that temperatures are rising, and have been rising for the last few hundred years. The rate of which is statistically significant and allows us the ability to predict and measure the future temperatures at our current rates. Exemplified, we can take the slope of the yearly regression (0.006), and multiply it by 274 (# of years since 1750), which gives us 1.644, the total rise in temperature since 1750 in Celsius. This calculation is in line with estimates made by NASA, at a 1.1 - 1.36 degree Celsius increase since 1880. It's not unreasonable to predict that in the 130 year unobserved time-frame, an additional 0.3 - 0.5 degrees could be allocated. It's worth noting that this only takes into account NASAs estimates, and not more conservatives estimates like the UNs at 1.5. Even filtering out the years 1750 - 1850, the change in temperature remains around the same, even though the slope changes to 0.0096 (as 0.006 * 274 = 1.644; 0.0096 * 174 = 1.670).
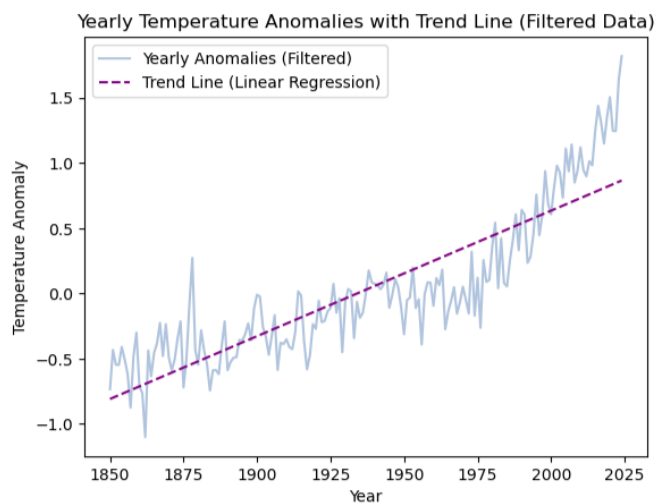

Fig. 22: Anomaly change in temperature by year (1850-2024)

What's Causing This?

Throughout this analysis we've referred to the trends being observed as just "global temperature increasing" but there is actually a term and field of science based on this phenomena alone, this is global warming. There's a broader term that reflects the multitude of changes humanity and Earth experience as a result of these climate conditions changing and becoming more drastic, this has been coined as climate change. But what exactly is global warming? According to the Natural Resources Defense Council, it can be explained as:

"Global warming occurs when carbon dioxide (CO_2) and other air pollutants collect in the atmosphere and absorb sunlight and solar radiation that have bounced off the earth's surface"[11]

With the knowledge of the **what**, we can explore the **how** and **who** behind global warming. Most importantly, is the **how**, learning what variables contribute to this scarily significant increase in global temperature is what will aid in the steps to mitigate the damage caused. The biggest contributor to climate change and global warming is the Greenhouse Effect, defined as the gas(es) produced that "absorbs infrared radiation from the Sun in the form of heat, which is circulated in the atmosphere and eventually lost to space."[12] (British Geological Survey) But during this process, these gases remain in the atmosphere and absorb the sun's heat, acting as a sort of insulation blanket on Earth. With the NRDC explaining, "Normally this radiation would escape into space, but these pollutants, which can last for years to centuries in the atmosphere, trap the heat and cause the planet to get hotter."[11]

So what are these greenhouse gases?

Greenhouse gases are the products of systems and processes around us we encounter in our day to day. Whether its carbon dioxide through burning fossil fuels or deforestation, methane through rice farming or livestock, or even nitrous oxide from fertilization production/use,[13] these seemingly everyday processes are primary contributors to global warming. In large part this started during the industrial revolution that took place around the world, and that is not to say that global warming was never around before our time, the Earth has had cycles of carbon emissions from nature, with the EPA explaining, "Over the last several hundred thousand years, carbon dioxide levels varied in tandem with the glacial cycles. During warm interglacial periods, carbon dioxide levels were higher. During cool glacial periods, carbon dioxide levels were lower."[14] However, the current cycle of global warming this past century is directly related to the exponential emission of greenhouse gases. The following graph from the EPA charted the 3 biggest contributors of the Greenhouse Effect: Carbon Dioxide (CO_2), Methane (CH_4), and Nitrous Oxide (N_2O) and their concentrations in the atmosphere over the last 2000 years.[14]

As depicted in figure 23, the point of time in which the gases have the highest concentration is around the point in history where the industrial revolution began. The Greenhouse Effect has had a more substantial impact on the health of the atmosphere and heat retention absorbed from the sun, than the natural processes, such as volcanoes that were the main contributors to the carbon cycle in the pre-industrial period.[14]

With all of this in mind, the **who** of this equation comes quite naturally, it's humanity. The methods used to make products the everyday person consumes, like oil or beef directly contribute to
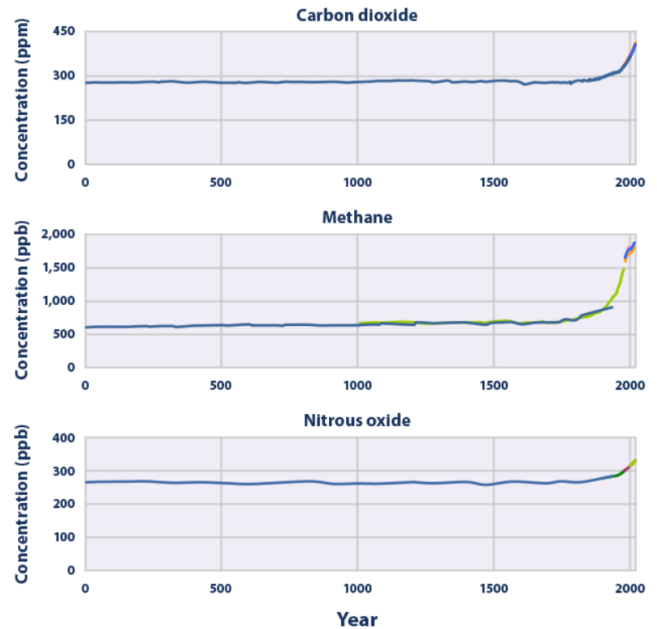

Fig. 23: Greenhouse gas concentrations over time as measured by the EPA (Environmental Protection Agency

the emissions of greenhouse gases that are causing this increase in temperatures.[13] While humanity at the level of you and me can only do so much, the real change has to occur at the corporate and national level. Pushes have been made in Europe, with the Paris Climate Agreement, which the United States rejoined January 20th, 2021, after its untimely departure in November 2016, which has committed to reducing carbon emissions by 2025. With the United States being the 2nd largest emitter of greenhouse gases in recent years (following China), but being the top contender for all time gas emissions,[11] local and state governments need to push for alternative methods of our current processes that are contributing to these record high carbon emissions and global warming.

Conclusion

The effects and consequences of climate change are ones that can be seen through your own eyes. With hotter summers, there's increased risk of heat exhaustion; with hotter winters there's less chance for snow to fall. Global temperatures are a complicated metric and this study barely scratches the surface, but as time goes on, as methodologies are developed and improved upon, these estimates and measurements are only getting more accurate. With enough support and effort, humanity can remain under the predicted catastrophic levels of global warming (estimated to be 2 degrees Celsius)[11] and rectify the last 300 years of global warming, returning to snow in the winter and less eggs cooked on the sidewalk in Arizona summers.

References

1. Education, UCAR Center for Science. "Center for Science Education." Exploring the Atmosphere with Weather Balloons

— Center for Science Education, 2014, `scied.ucar.edu/learning-zone/atmosphere/weather-balloons`.

2. Bolles, Dana. "How Do Scientists Measure Global Temperature? - NASA Science." NASA, NASA, Feb. 2024, `science.nasa.gov/climate-change/faq/where-do-global-temperature-data-come-from/`.

3. Buis, Alan. "A Degree of Concern: Why Global Temperatures Matter - NASA Science." NASA, NASA, 19 June 2019, `science.nasa.gov/earth/climate-change/vital-signs/a-degree-of-concern-why-global-temperatures-matter/`.

4. Education, UCAR Center for Science. "Center for Science Education." Investigating Past Climates — Center for Science Education, Accessed 28 Nov. 2024. `scied.ucar.edu/learning-zone/how-climate-works/investigating-past-climates`

5. Earth, Berkeley. "Data Overview." Berkeley Earth, 10 Nov. 2023, `berkeleyearth.org/data/`

6. Rohde, Robert, et al. "Berkeley Earth Temperature Averaging Process." Geoinformatics & Geostatistics: An Overview, Berkeley Earth, 5 Mar. 2013, `berkeleyearth.org/wp-content/uploads/2022/12/Methods-GIGS-1-103.pdf`

7. What Is Linear Regression?, IBM, 1 Nov. 2024, `www.ibm.com/topics/linear-regression`

8. Arndt, Deke. "Climate Change Rule of Thumb: Cold 'Things' Warming Faster than Warm Things." Climate Change Rule of Thumb: Cold "Things" Warming Faster than Warm Things, NOAA, 24 Nov. 2015, `https://bit.ly/3Vx907g`

9. Frost, Jim. "How to Interpret R-Squared in Regression Analysis." Statistics By Jim, Making Statistics Inclusive, 26 Feb. 2024, `statisticsbyjim.com/regression/interpret-r-squared-regression/`

10. "World of Change: Global Temperatures." NASA, NASA, 2024, `earthobservatory.nasa.gov/world-of-change/global-temperatures`

11. Amanda MacMillan, Jeff Turrentine. "Global Warming 101." Definition, Facts, Causes and Effects of Global Warming, NRDC, 7 Apr. 2021, `www.nrdc.org/stories/global-warming-101#warming`

12. "The Greenhouse Effect." Discovering Geology — Climate Change, BGS, 5 Apr. 2023, `www.bgs.ac.uk/discovering-geology/climate-change/how-does-the-greenhouse-effect-work`

13. Bolles, Dana. "Causes - NASA Science." Edited by SMD Content Editors, NASA, NASA, 2022, `science.nasa.gov/climate-change/causes/`

14. "Causes of Climate Change." EPA, Environmental Protection Agency, 12 Apr. 2024, `www.epa.gov/climatechange-science/causes-climate-change`