ECE NTUA Machine Learning Project on Climate Change and CO2 Fossil-Fuel Emissions

Athanasios Pantos
Department of ECE
National Technical University of Athens
Athens, Greece
AM: 03400026

Abstract

We present a cross-dataset exploratory time series analysis on measurements of temperature, temperature anomalies, and CO2 emissions starting from 1750 up to 2015. Through time series decomposition of temperature into seasonal and trend components we verify the global warming phenomenon. In addition we find significant correlation between the temperature values and the extreme temperature events, as well as very strong significant correlations (~ 0.7) between global fossil fuel CO2 emissions and the global average temperature. Finally, we also perform a per country analysis of CO2 emissions. This report is accompanied by a zip file containing the source code.

Index Terms

global warming, climate change, fossil fuels, CO2 emissions, time series decomposition

I. Introduction

The IPCC (Intergovernmental Panel on Climate Change) special report of 2019 on Global Warming was especially concerning. According to the report, human induced warming has increased the average surface, air, and sea surface temperature above $1°$, compared to global temperature of pre-industrial levels, increasing at a rate of $0.2°$ per decade. According to the report, a world $1.5°$ larger than pre-industrial levels would have a catastrophic impact on life on earth: increased heat-related health impacts, fires, extreme weather events, ice meltdowns [1]–[3].

As a result, climate change is undoubtedly one of the most hot topics in the scientific community due to its large impact on the environment, and as a sequence to life on earth. Climate is mostly studied through the analysis of data as time series [3]–[6]. These can include time series of weather data such as precipitation, humidity, or time series of temperatures.

Motivated by the above, in this report we perform a cross-dataset analysis of time series data pertaining to climate change. We fit several different time series models on data of temperature, temperature anomalies, and CO2 emissions, studying not only their components, but also their intercorrelations.

II. Datasets

There is an abundance of datasets on climate change - for example, the google cloud platform [7] has a total of 28 datasets on climate change, out of a total of 212 datasets (13.2%), highlighting the importance of the field. One of the most important sources of data on climate is the National Oceanic and Atmospheric Administration (NOAA) of U.S.A., founded in 1970. Here, we study the trend of various different metrics, doing a cross-dataset analysis. The datasets we use are:

- Climate Change: Earth Surface Temperature Data [8] compiled by BerkeleyEarth. This is a curated meta-dataset, containing average temperature of land and ocean, by compiling a total of 16 pre-existing archives of data, starting in 1750. The dataset also includes the 95% confidence intervals.
- Fossil-Fuel CO2 Emissions [9] by the Carbon Dioxide Information Analysis Center (CDIAC)
- Global Temperature Anomalies [10] by the Met Office Hadley Centre.

III. Method

For each dataset, unless stated otherwise, we first identify missing values, and perform a nearest interpolation in order to fill these. Our main goal is identifying the trends as well as the seasonal components, by performing a time analysis on the above datasets. We consider three different methods, subsequently described.

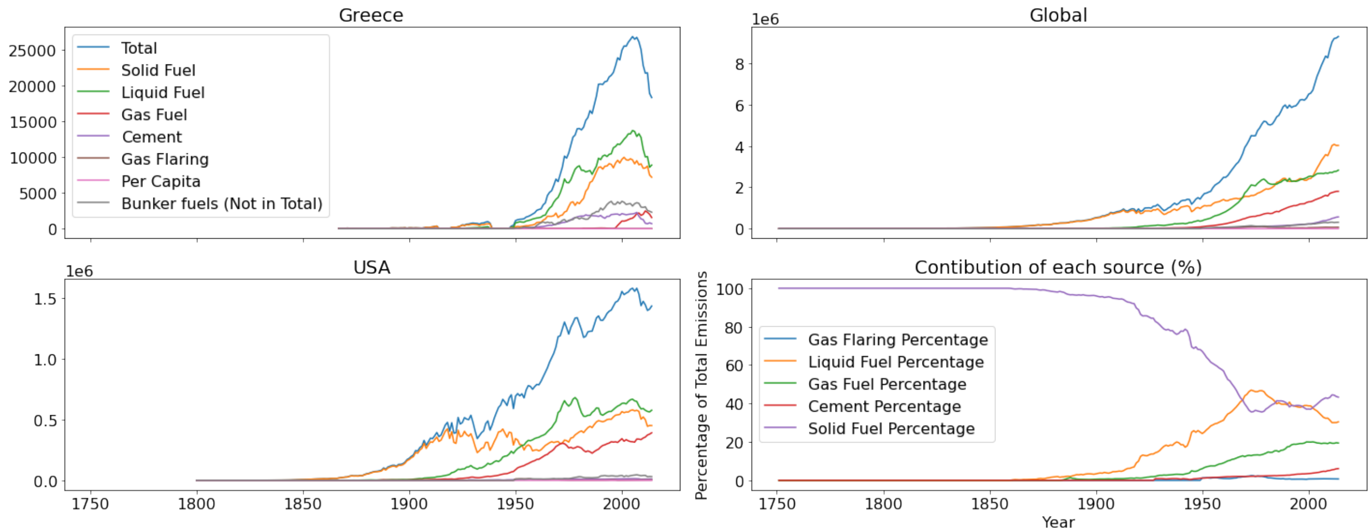


Figure 1: CO-2 Emissions per year for Greece, USA and Global, as well as the main source of emissions. Bottom right shows the percentage of each emission source against the total emissions.

A. *Seasonal Decomposition*

During this decomposition, the time series is modeled as:

$$y(t) = g(t) + s(t) + \epsilon_t \tag{1}$$

where $g(t)$ is the trend, $s(t)$ is the seasonality and ϵ_t is the residual error from fitting the model to a time series. This model is fit using the LOESS [11] method (locally estimated scatterplot smoothing). Using this decomposition, we can then deseasonalize the data (i.e., $y(t) - s(t)$), and we can fit an ARIMA model to do forecasting.

B. *Holt-Winters Forecast*

The Holt-Winter method [12] (triple exponential smoothing) is a generalization of Holt's exponential smoothing, designed to capture seasonality. In this method the time series is recursively modeled as

$$\begin{aligned}
s_0 &= x_0 \\
s_t &= \alpha\left(x_t - c_{t-L}\right) + (1-\alpha)\left(s_{t-1} + b_{t-1}\right) \\
b_t &= \beta\left(s_t - s_{t-1}\right) + (1-\beta)b_{t-1} \\
c_t &= \gamma\left(x_t - s_{t-1} - b_{t-1}\right) + (1-\gamma)c_{t-L} \\
F_{t+m} &= s_t + mb_t + c_{t-L+1+(m-1) \bmod L}
\end{aligned} \tag{2}$$

where x_t is the value of the time series at time t, b_t is the trend at time t, s_t is the smoothed value at time t, c_t is the seasonal value, α is the data smoothing factor, β is the trend smoothing factor, γ is the seasonal smoothing factor and F_{t+m} the estimated value of the time series at a future time $t + m$.

C. *Generalized Additive Model Decomposition and Forecasting*

In this method, the analyzed time series is split in three components like in the first method, using a generalized additive model (GAM) [13] and fast fitted using L-BGFS [14]. This is the method used in the widely used package Facebook Prophet [15] for forecasting at scale.

IV. Experimental Results

A. *Fossil-Fuel Co2 Emissions*

a) Global Emissions: We first explore the Fossil-Fuel CO2 Emissions dataset. The dataset includes the total CO2 emissions per country, as well as the CO2 emissions per category: Solid Fuel, Liquid Fuel, Gas Fuel, Cement, Gas Flaring. The final extra measurements provided are the per capita emissions, as well as the fuels from Bunkers, which are not accounted for in the total emissions. Based on these measurements we proceed to create a global estimate on the CO2 emissions, by aggregating

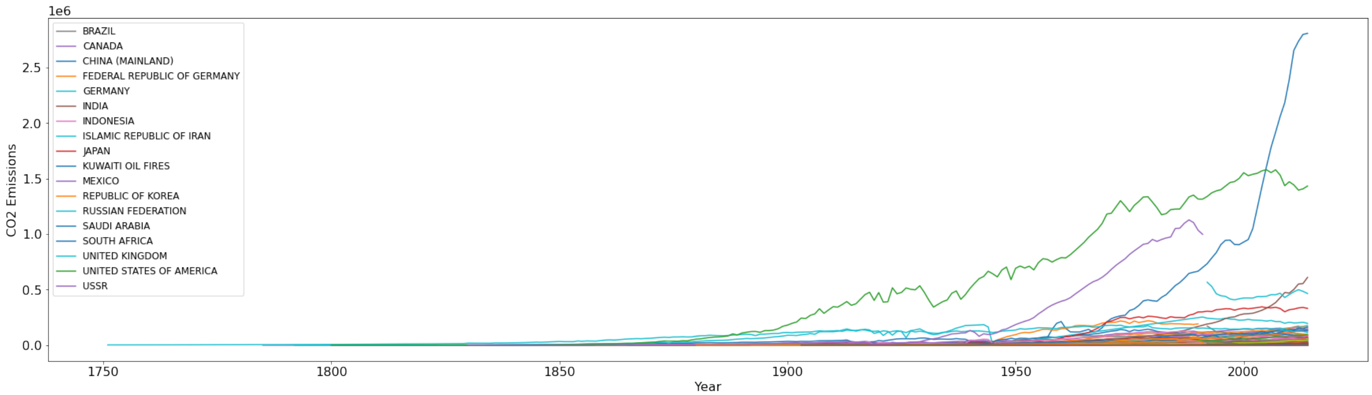


Figure 2: CO2 Emission timelines by country. Legend includes only countries where their last measured value was above 0.1e-6 million tones.

the emissions per year across all countries, ignoring missing values. In Figure 1 we plot the emissions of Greece, USA, as well as the global CO2 emissions estimates

We immediately observe a rapid increase in the global CO-2 emissions, while the time series for USA seem to attempt to stabilize after 1950s. In Greece, around 2005 there was a peak of fuels, and after that the emissions rapidly drop. This could be due to the economic crisis, stricter regulations on the emissions by the EU, or both. It is also interesting to study the percentage of each different emission source, against the total emissions. We create these features and plot them in the same Figure 1 (bottom right).

We see that initially all emissions were based on solid fuel, however especially after the start of the 20th century, the percentage of solid fuel source of emissions rapidly declined, and liquid and gas fuels started increasing. The same goes for cement, while gas flaring seems to generally have a minuscule percentage of total emissions, with larger values during the 1970s.

b) Per Country Emissions: We next explore the question, “Which country emits the most CO2?”. We show a representative plot in Figure 2. We can see that China, USA, India, Japan, and Russia, are the main drivers of emissions, with Japan and Russia maintaining a relative stable value. We also see how the *Kuwaiti oil fires* event [16] was so important that impacted the emissions on a global scale. During this event, a total of 605 to 732 oil wells and oil lakes were set to fire in 1991 during the Persian Gulf War.

B. Global Temperatures

a) Identifying the trend in climate change of the globe: . Subsequently, we proceed to perform an analysis of the global climate, using the *Climate Change: Earth Surface Temperature Data*. In Figure 3 we can see the global average temperature from 1750 to 2015, along with the uncertainty of these temperatures, in red. It is obvious that the uncertainty in these estimations gets reduced, since measurement instruments get more precise, and more organizations gather temperature data. Continuing, we select data from 1880 onwards due to the small uncertainty in these temperatures, resulting later in a more robust model fit of temperatures.

First, we perform some exploration. An Augmented Dickey-Fuller test results in -2.026 statistic and p-value 0.275, meaning that the series is non-stationary. Next, we check for the existence of seasonal trends in the temperatures. For that, we plot the autocorrelation function in Figure 4, which strongly verifies the seasonal component of temperatures, as well as the existence of a trend. Furthermore, the relationship between the two seems to be additive, since the seasonal fluctuations do not seem to increase (only the trend increases).

We next proceed to fit the LOESS and LGBFS seasonal decomposition models on the temperature time series, which results in a mean residual error of 0.17 and 0.25 respectively, and plot the components in Figures 5 and 6.

We immediately observe the increasing nature of the trend in both models. The seasonal fluctuations are also depicted, showing how the temperature fluctuates around the year.

b) Forecast: Next, we employ all three described models (ARIMA(1,1,1) with deseasonalized series, Holt-Winters, and LGBFS) to do forecast into the future of global temperatures. Figure 7 shows the results. We also note the residual fit of Holt-Winters is 0.24. In the same figure we also show the average *yearly* temperatures, so that we get a better visual estimate. All models predict a rapid continuous temperature increase.

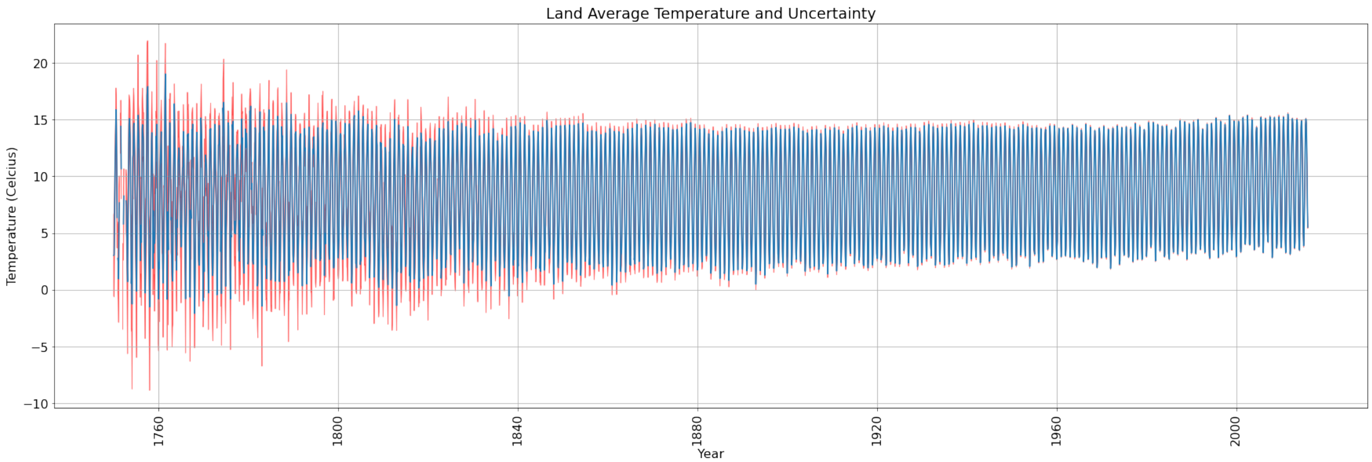


Figure 3: Global Average Land Temperature and uncertainty from 1750 to 2015.

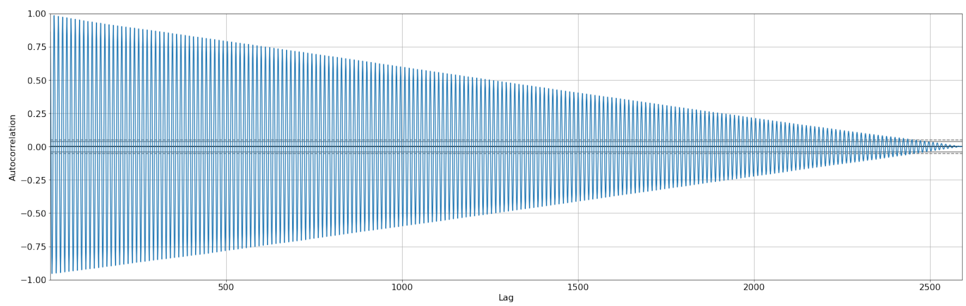


Figure 4: Autocorrelation of temperature.

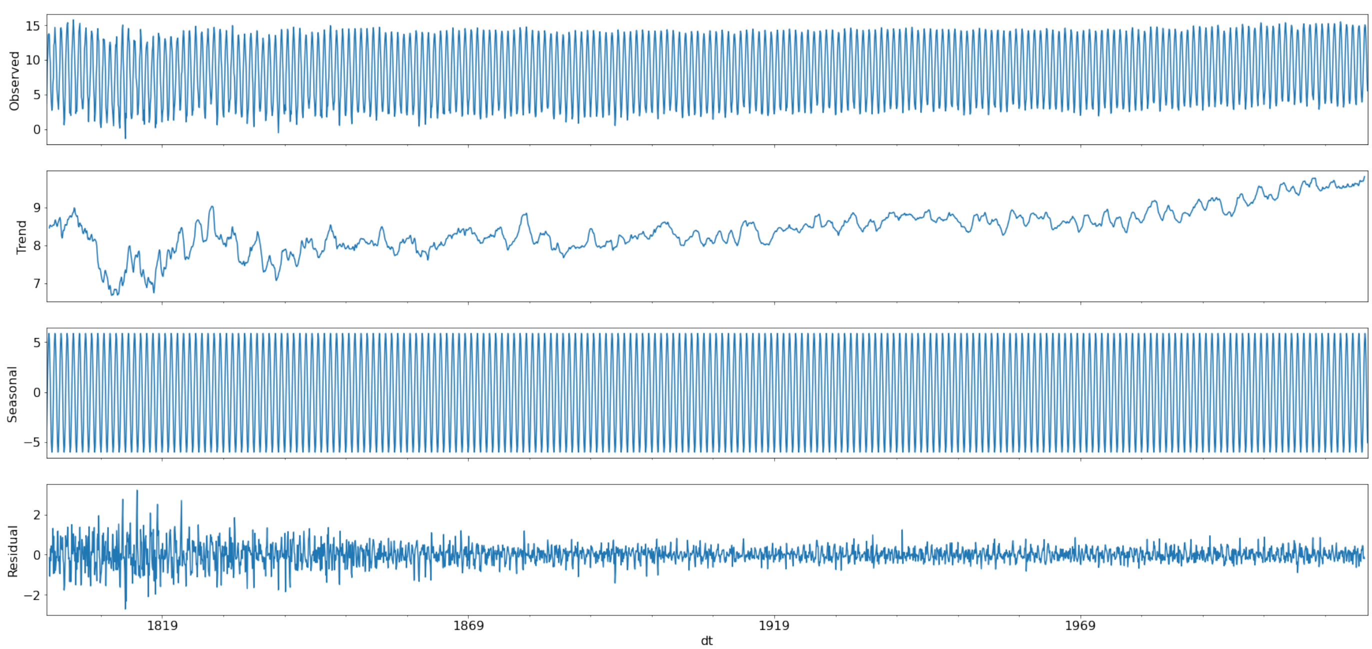


Figure 5: Components of the fitted time-series decomposition model using LOESS on the average global temperature per year time series from *Climate Change: Earth Surface Temperature Data*. From top to bottom are a) the observed time series, b) the trend of the decomposed model, c) the yearly seasonal component, and d) the residual.

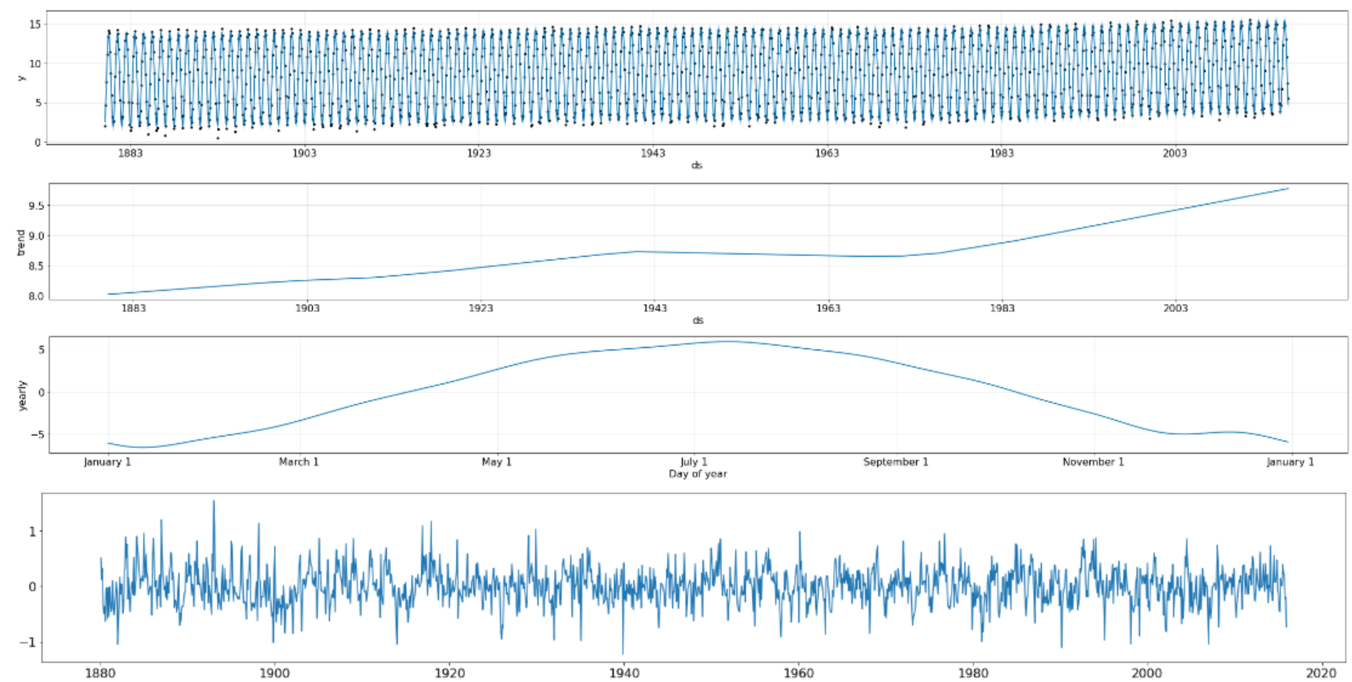
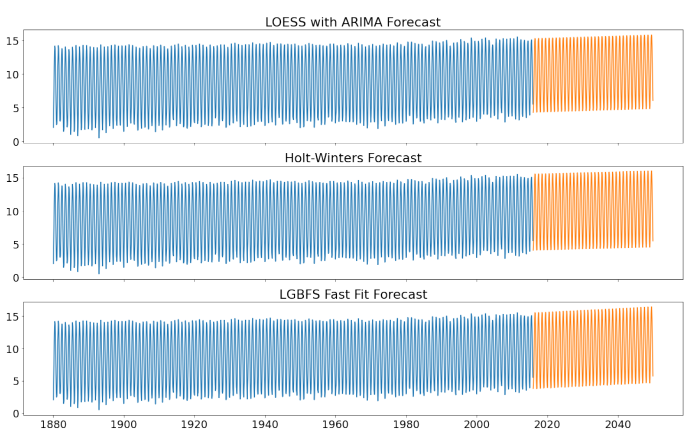
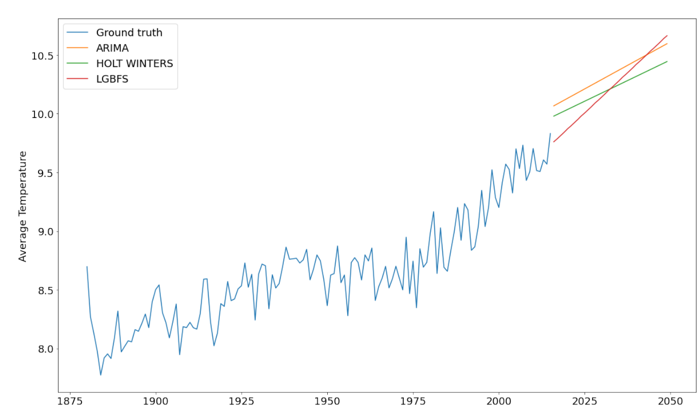


Figure 6: Components of the fast fitted time-series decomposition model using LGBFS on the average global temperature per year time series from *Climate Change: Earth Surface Temperature Data*. From top to bottom are a) the observed time series (in black dots) and the fitted model, b) the trend of the decomposed model, c) the yearly seasonal component (one period), and d) the residual.

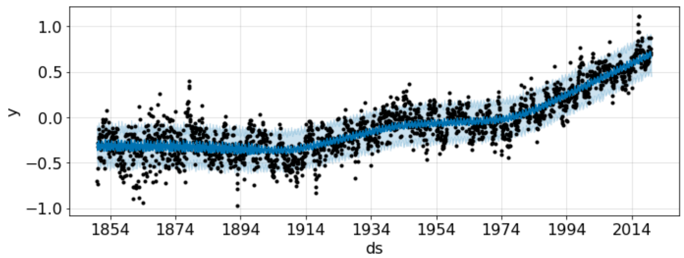


(a)

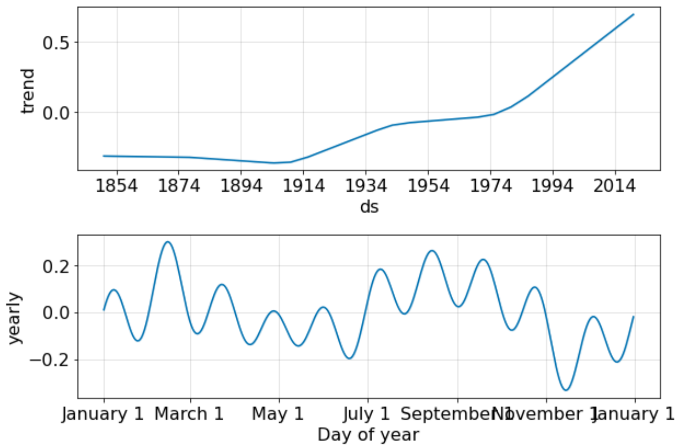


(b)

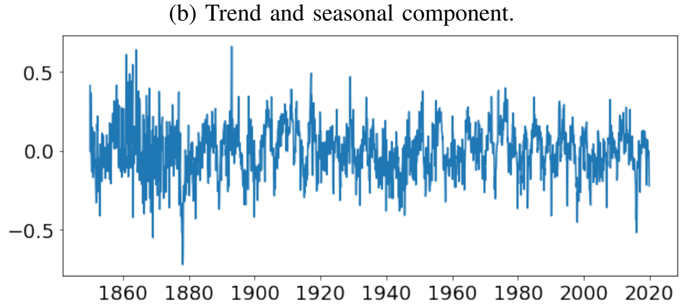
Figure 7: Monthly 7a and Yearly 7b predictions of average temperature into the future from LOESS with ARIMA, Holt-Winters, and LGBFS fast fit.



(a) Fitted model on the GTA dataset.



(b) Trend and seasonal component.



(c) Residual.

Figure 8: Components of the LGBFS fitted model on *Global Temperature Anomalies dataset*

C. Temperature anomalies

According to studies, a significant outcome of global warming is the more and more frequent appearance of extreme weather conditions, including temperature anomalies. We fit the fast (LGBFS) time series decomposition model and show the results in Figure 8.

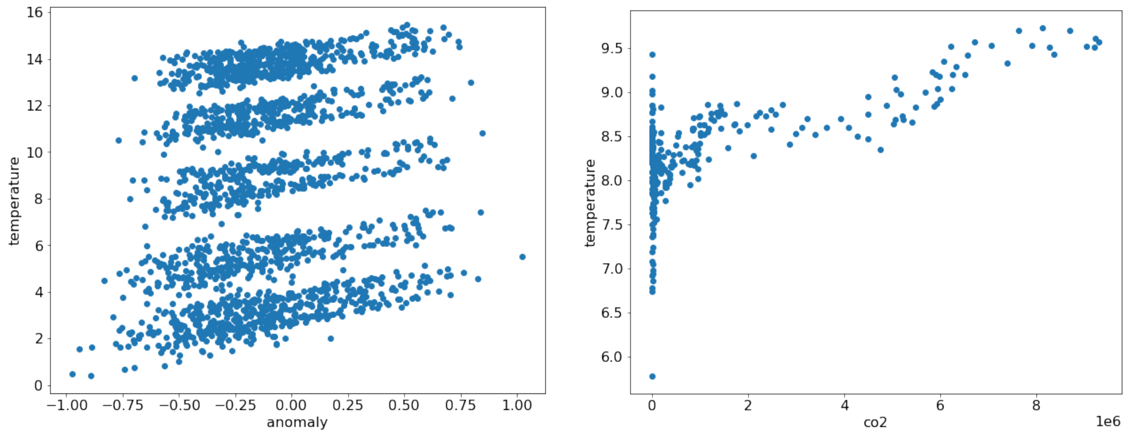
According to this decomposition, we can see that the temperature anomalies do not have any seasonal component (we have also verified this fact with the autocorrelation plot), but it does have a very strong increasing trend.

D. Cross-dataset examinations

a) Correlations between Global Average Temperature and Temperature Anomalies: In this subsection, we will explore if and how temperature extremes (i.e. anomalies) correlate with the global average temperature. To do that, we merge the Earth Surface Temperature data with the Global Temperature Anomalies, and calculate the correlation between the anomaly value, and the average land temperature. The results are shown in the Table I:

	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	1992	0.181697	[0.14, 0.22]	0.033014	0.032042	3.020619e-16	8.723e+12	1.0

Table I: Correlation between average land temperature and temperature anomalies.



(a) (b)

Figure 9: Scatter plots of temperature anomalies against average temperature 9a and CO2 emissions against average temperature 9b

.

We find an extremely significant albeit small correlation between the two, with a pvalue of $\sim 1e-16$ and bayesian factor of $8.7e+12$, leaving no doubt on the relationship between the two. The found correlation coefficient is 0.182, with a CI of $[0.14, 0.22]$. We however note the *Correlation does not imply causation principle* [17]; this correlation does not necessarily imply a causal relationship between the two - there could by other unknown underlying factors. A scatter plot showing the average temperature against the anomaly values is shown in 9a.

b) Correlations between Global Average Temperature and CO-2 Emissions: Doing the same for Earth Surface Temperature data with the CO-2 emissions dataset we find an **extremely strong and significant correlation** (0.698 with a CI of $[0.63, 0.75]$ with $p = 7e-40$ and BF $3.63e+36$) as can be shown in Table II. A scatter plot showing the average temperature against CO2 emissions (increase) values is shown in 9b.

	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	264	0.697752	[0.63, 0.75]	0.486859	0.482926	7.727123e-40	3.628e+36	1.0

Table II: Correlation between average land temperature and CO2 emissions.

Finally, we also explore the question "Do the emissions of a country correlate with the temperature increase of this specific country?". To answer this, for each country we calculate the temperature increase using the first and last temperature measurements, as well as 1) the sum of its emissions over time, and 2) the increase in its emissions (last measurement versus first measurement). Subsequently, we calculate the pearson correlation between the two, and show it in Table III.

	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	175	0.096262	[-0.05, 0.24]	0.009266	-0.002254	0.205068	0.21	0.245484
pearson	175	0.160021	[0.01, 0.3]	0.025607	0.014277	0.034399	0.87	0.564423

Table III: Correlation between average land temperature and sum CO2 emissions **per country** (top row) and increase of CO2 emission **per country** (bottom row).

The results are mixed. We can see that there is no correlation between the total emissions of a country and its temperature increase, however we do find a small correlation which could be significant ($p = 0.034$) between the increases of a country's CO2 emissions and its average temperature.

V. Conclusion

We have conducted a cross dataset analysis on temperature, temperature anomalies and fossil fuel CO2 emission from 1750 to 2015. Our thorough analysis using time series decomposition has verified the increasing trend of global warming and has found extremely strong correlations between the total CO2 Fossil Fuel emissions and the global average temperature. The global warming effect also seems to be mostly global and not local since we found no significant correlations between the sum of CO-2 emissions per country over the there and its average land temperature, and small indication on the correlation between its increase of CO2 emissions and its temperature.

References

[1] M. Allen, P. Antwi-Agyei, F. Aragon-Durand, M. Babiker, P. Bertoldi, M. Bind, S. Brown, M. Buckeridge, I. Camilloni, A. Cartwright *et al.*, "Technical summary: Global warming of 1.5° c. an ipcc special report on the impacts of global warming of 1.5° c above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty," 2019.

[2] R. Djalante, "Key assessments from the ipcc special report on global warming of 1.5 c and the implications for the sendai framework for disaster risk reduction," *Progress in Disaster Science*, vol. 1, p. 100001, 2019.

[3] A. P. Møller, W. Fiedler, and P. Berthold, *Effects of climate change on birds*, 2010.

[4] M. Mudelsee, *Climate time series analysis*. Springer, 2013.

[5] M. A. R. Sarker, K. Alam, and J. Gow, "Exploring the relationship between climate change and rice yield in bangladesh: An analysis of time series data," *Agricultural Systems*, vol. 112, pp. 11–16, 2012.

[6] A. Raha, S. Das, K. Banerjee, and A. Mitra, "Climate change impacts on indian sunderbans: a time series analysis (1924–2008)," *Biodiversity and Conservation*, vol. 21, no. 5, pp. 1289–1307, 2012.

[7] "https://cloud.google.com/public-datasets," 2021.

[8] "https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data," 2021.

[9] "https://data.world/adamhelsinger/fossil-fuel-co-2-emissions," 2021.

[10] "https://data.world/kcmillersean/global-temperature-anomalies," 2021.

[11] C. Robert, C. William, and T. Irma, "Stl: A seasonal-trend decomposition procedure based on loess," *Journal of official statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[12] C. Chatfield, "The holt-winters forecasting procedure," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 27, no. 3, pp. 264–279, 1978.

[13] T. Hastie and R. Tibshirani, "Generalized additive models: some applications," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 371–386, 1987.

[14] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on scientific computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

[15] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[16] "https://en.wikipedia.org/wiki/Kuwaiti_oil_fires#Environmental_impact," 2021.

[17] "https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation," 2021.