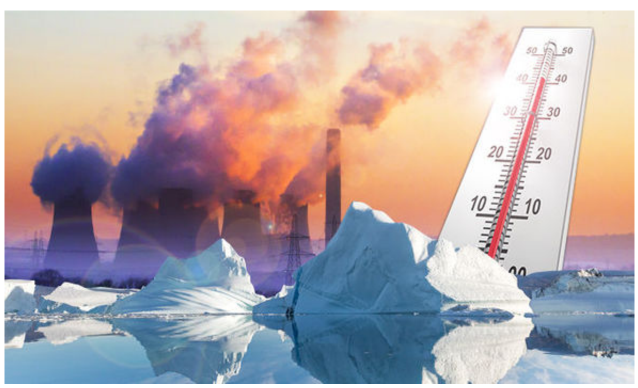
University of Liège

PROJ0016 - Big data project

Is global warming for real ?



Pierre-Loup Nicolas, Lionel Mathy, Evgeny Ivanov

Academic year 2018 — 2019

Contents

Part I

Introduction

1 Problem statement

Global warming is a well-known[1] and globally acknowledged[2] issue linked to the broader problem of climate change. It is characterized by an increase in the global temperature around the globe, with consequences ranging from more frequent natural disasters, ice cap melting (and the subsequent sea level rise), and multiple other consequences, both environmental and economical.
It directly affects vegetation patterns, migration of animals, and puts in danger of extinction some of the most vulnerable species. Overall, global warming is a threat for both nature as whole and humanity.

But in spite of receiving constant coverage in the mass media, the issue of global warming is still underestimated by a large margin of the world's population. The recent withdrawal of the United States from the Paris Agreement has shown yet another time that a lot of skepticism remains in some countries.
Under the Paris Agreement, each country must determine, plan, and regularly report on the contribution that it undertakes to mitigate humanity's impact global warming. Every member of the United Nations signed this agreement.
However, immensely populous developing countries such as China and India, despite signing and ratifying the convention, have no plan to reduce their environmental impact, as it would harm their fast-growing industries. Other major countries like the United States and Russia also prioritize their economical interests. As of today, Europe is the only region where environmental policies are taken seriously and a transition from "dirty" energy to renewable energies is somewhat underway. Indeed, this is not sufficient, as shown by many studies worldwide.

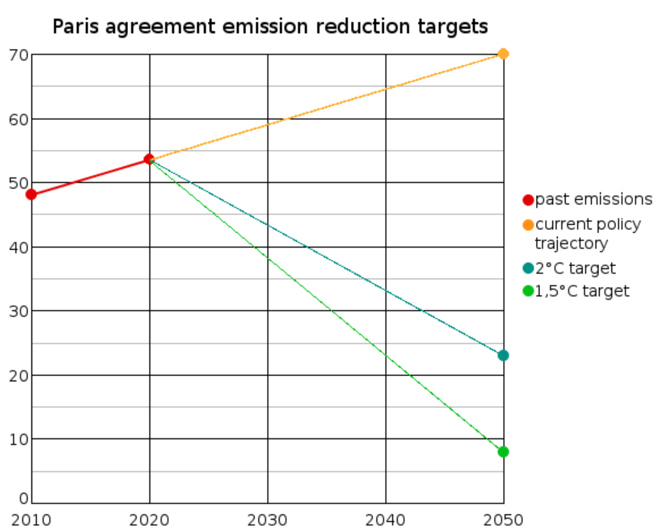


Figure (1) Paris climate accord emission reduction targets and current real-life reductions offered.

[1] https://en.wikipedia.org/wiki/Climate_change_opinion_by_country
[2] https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=XXVII-7-d&chapter=27&clang=_en

The current situation shows that most people either do not believe in global warming, or do not consider it as a relevant preoccupation[3]. Despite being a hot topic in the news, the human origin of global warming is surprisingly a topic of debate, even within the scientific community. Some people will agree on the fact that global warming is a reality, but deny humanity's responsibility.
The so-called "Climategate" scandal[4] further contributed to the skepticism of the general public towards climate change.

2 Literature review and purpose of the project

The present project aims at producing a general investigation using statistical analysis and Big Data tools in order to demonstrate not only that global warming is a reality, but also to show the degree of human involvement into it.
Data science is particularly well suited for this purpose (Guo et al., 2015 [1]) But data science faces a lot of technical challenges with respect to climate science. Due to the relative abundance of data related to climate science, it is sometimes difficult to proceed them wisely. These technical difficulties (Faghmous and Kumar, 2014 [2]) are detailed thereafter.

The data collection in itself is of utter importance. There are several ways of gathering meteorological data, with their pros and cons. The first and main method is *in situ* observations. The big advantage of this method is the availability of fairly old data records in some stations, particularly in Europe, which allow to conduct a long-term climatic analysis. This may prove useful to evaluate humanity's impact on global warming, particularly if it is possible to go as far back as the Industrial Revolution.
However, it is difficult to obtain a good spatial coverage by meteorological stations, particularly in the remote lands of Russia and Canada.
The second method is remote sensing. The satellite technology has the advantage of providing good spatial coverage of some meteorological parameters, for instance soil temperature and cloudiness. However, the scientific community does not fully rely on satellites, due to data contamination by the atmosphere and the general precision of the data, thus this technique is mostly used to recover spatial information on some meteorological parameters, especially for the empty spaces within the stationary network. It is also worth to mention that the technique exists since 1960 and is still constantly improved, thus a limited number of climatic records is available.
The third method is a collection of paleoclimate techniques, with which it is possible to reconstruct some meteorological variables, particularly the gas composition of the atmosphere, with a certain precision far back into the past, much earlier than the industrial revolution or any human impact on the climate.

After the collection phase arises the issue of the heterogeneity of the datasets. Even though the scientific community all over the world exchanges data and methods of processing on a regular basis, data is being collected using different instruments, sometimes too old, sometimes poorly calibrated. Verifying the trustworthiness of the datasets used is crucial.
On top of that, there exists a phenomenon of auto-correlation of climatic data in time and and space, due to the potential location proximity of meteorological stations. For instance, the climatic conditions in Belgium and northern France are closely related to each other, and collecting data in these two regions might lead to similar results most of the time. The existence of this auto-correlations phenomenon implies an additional difficulty in inferring statistical models that emphasize the different variables implied in specific climate phenomena. Since the data is not independent and identically distributed, the complexity of our statistical models increases.

[3] Indeed, most people in underdeveloped countries have most likely other concerns.
[4] https://en.wikipedia.org/wiki/Climatic_Research_Unit_email_controversy

Lastly, there are some issues related to the application of machine learning in climate change analysis. It is difficult to use supervised machine learning algorithms, mainly due to the fact that it is not possible to separate the existing datasets into two parts: those who represent global warming and those who are not "contaminated" by it. Under the hypothesis that global warming appeared and is evolving since the industrial revolution, almost no data that has not been contaminated by the effects of global warming is available. Since supervised learning methods rely on trustworthy learning samples, their application to climate change studying can be difficult.
Moreover, many machine learning algorithms try to be good at predicting the output of a new feature vector given to the model, without explanation of the processes who led to the current state. In other words, most machine learning techniques are good at imitating reality, but not so much at simulating it. Since climate change is a physical phenomenon, the whole point of its investigation lies towards the explanations and insights we can get from the models rather than blindly predicting an output. This implies to reject all machine learning algorithms working as a black box where results are difficult to interpret and confront to the well established theories in climate science and global warming.

For instance, we could consider the evolution of the temperature in different regions. Interesting questions are for example to know if the mean temperature increases or decreases in a given region and if yes, since when ? What can explain such results ? Many parameters have to be taken into account: the CO2 concentration, the topology of the region, the history of the region,... can be assessed in order to see if this evolution is the result of a global warming phenomenon or if it is the consequence of other phenomena.
Obviously, many techniques can be used to answer these questions. The most obvious one is the extensive use of plots to see the trends and the abnormal variations. Machine learning algorithms can also be of great utility. Linear regression models for instance can be good evaluation models for the temperature. These regression models can be done globally or for specific regions, taking into account the time and temperature, or other parameters. However, we have to assess the statistical significance of such methods.

We face the same challenge when trying to evaluate humanity's impact on the climiate. The evolution of temperature around the globe can be correlated with some variables like the concentration of greenhouse gas or the evolution of sea level. Determining the correlation between temperature and the CO2 concentration in the atmosphere could allow us to see if the observed variations of temperature are noticeably caused by human activities or if other factors can explain this evolution. It is well-known that the earth's climate is cyclic and goes trough periods of glaciation and periods of global warming. Therefore, we have to be cautious about our conclusions.

This project's objective is not to forecast the future evolution of worldwide climate, but instead to focus on the diagnosis of major climatic variables to answer the question "Is global warming real ?". In the frame of the project, not only are the temporal dynamics studied, but also, whenever possible, the spatial dynamics as well, even though due to atmospheric circulation the contribution of a tracer from one particular point in time to another will spread across the globe.

Part II

Study & results

3 Initial data exploration

To properly begin the project, our objective was to perform a first analysis of some datasets we found in the pre-analysis part, on three variables of central importance in the context of climatology and global warming: temperature in itself, of course (air and soils), but also ice extent and the atmospheric CO2 concentration.

3.1 Air and soil temperatures

For the global air and soil temperature analysis, the Era-Interim reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) was used (D. P. Dee et al, 2011 [6]). Since the goal here was to get familiar with the data and to perform a primary analysis, a spatial resolution of 3x3 degrees was chosen as a trade-off between the speed of processing and geographical generalization. The time period of 1979-2017 was chosen as the longest available from the current reanalysis dataset. Soil temperatures were taken on the deepest available level (the 4th, corresponding to a depth of 1 to 2.89 meters) to avoid instant correlations with the air temperature dynamics.

In order to reflect both the global and the local potential effects of global warming, both global and local temporal trends were calculated. The global trends for soil and air temperatures were calculated by using the area of each individual grid cell as weights. Hence the huge size differences between the grid cells located near the poles and the ones closer to the equator were taken into account.
Furthermore, a trend for the air temperature was calculated for each individual grid cell and plotted on a world map for easier visualization. In addition to that, trends for each month were calculated for big cities in different parts of the globe, in order to show the internal annual dynamics of global warming. Those cities are Brasilia, Brussels, Johannesburg, Karachi, Kinshasa, Los Angeles, Mc Murdo, Melbourne, Moscow, New York, Singapour, Tokio, Valletta, Yakutsk and Yellowknife.
In the present report, not all results are presented, only the most relevant; the others are available on the project's GitHub.
Figure 2 shows the global trends for the air and soil temperatures. Both curves demonstrate a stable increase of 0.5 Celsius degrees over the last 38 years, and the same behaviour, due to a direct dependence of the soil temperature on the air, even on deep horizons. A relatively flat part of both curves, between the years 2000 and 2010, is compensated by a significant temperature jump in the last five years.

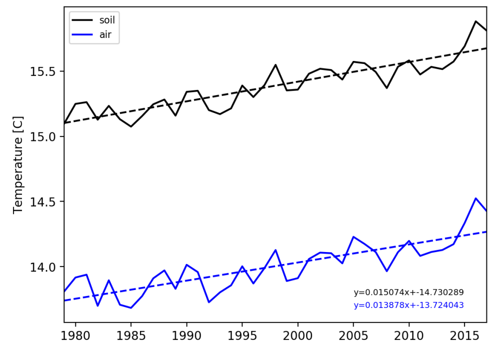


Figure (2) Inter-annual trends for globally averaged soil and air temperatures for the years 1979-2017.

Figure 3 shows air temperature differences according to the local trends. The Arctic ocean and all continents show a clear warming trend over the last 38 years. The oceans and Antarctica are more resistant to the phenomenon and, in some places, show an opposite trend.

But as will be seen in the next section, the Arctic warming and the resulting reduction of the ice sheet unfortunately constitute a self-accelerating system: snow and ice melt due to the warmer temperatures, and this leads to a decrease of the Earth's surface albedo (i.e. its ability to reflect solar radiations), as snow and ice reflect almost all solar radiations. This in turn accelerates the warming process because the solar radiations are less reflected by the Earth's surface.
Generally speaking, land is less thermally inertial than water; that is why continents absorb incoming radiations faster and thus are more exposed to global warming. The Antarctic continent is an exception because it is almost entirely covered with ice and it is the coldest place on Earth; therefore, a huge amount of energy is required to melt its surface, and most of it is not yet affected by temperature increases.

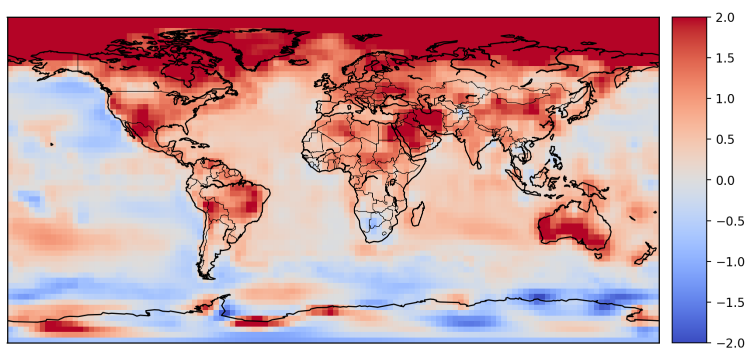


Figure (3) Spatially varying trends for the air temperature (the colorbar correspond to warming (red) and cooling (blue) trends).

Figure 4 shows the local internal annual trends for two locations: Brussels and the Antarctic station of Mc Murdo (the others are available on the GitHub). Brussels demonstrates a warming trend for all seasons, especially in winter, which is probably because of the activity of the Gulf Stream. Mc Murdo, an Antarctic station situated near the ocean shore, shows a significant warming trend for several months, exceeding two degrees. Both locations follow the global trend of the others cities studied. Some cities are less exposed (as Singapour, which has got warmer of 0.5 degrees). Cities in the northern Hemisphere, such as Yellowknife, Yakutsk and New York show important warming trends. The equatorial cities of Kinshasa and Brasilia also became significantly warmer. However, it is worth to mention that in some places, some particular seasons became visibly colder, such as summer in Los Angeles and in Johannesburg (in the Southern Hemisphere, summer starts in December). Among all chosen cities, only Karachi does not show a clear warming trend.

It is worth mentioning that all of those cities have a “warm dome”, which makes them warmer than the surroundings. However, the results for the cities were not obtained from the local meteorological stations, but from the global product, and thus are not contaminated by the microclimate uncertainties. As a result, all analyzes (the global, the local and internal annual) show, up some extent, a clear trend of global warming.

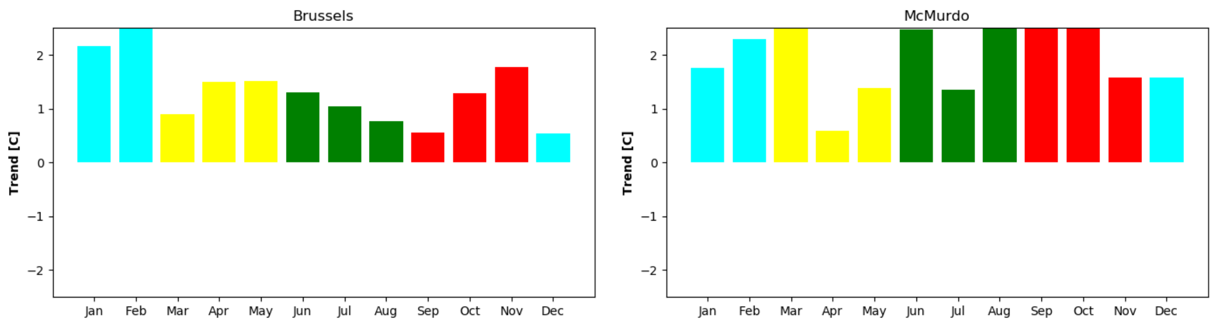


Figure (4) Monthly air temperature trends for Brussels and the Antarctic station of Mc Murdo.

3.2 Ice Analysis

The initial idea for ice analysis was to consider ice thickness in the Arctic sea, to determine whether a thinning trend could be observed or not. A large dataset grouping ice thickness measurements from stations all around the Arctic was explored. In particular, the data of Russian stations spread throughout northern Siberia were used because of their important temporal range (1936 to 2000), although only one measurement was available for each year. However, it appeared that the mean variation in ice thickness through the years did not show a particular trend. Therefore this parameter is insufficient to observe the potential effects of global warming.

The second obvious variable related to the ice cap was the variation of the sea ice extent and area.
It might be useful to first define the notions of extent and area. Those are closely related but different measurements of the extension of sea ice and they give slightly different information.
Extent simply defines a region as “ice-covered” or “not ice-covered”. In the literature, a satellite "data cell" is generally considered as "ice-covered" if the ice concentration is greater or equal to 15%.
On the other hand, area takes the percentages of sea ice within data cells and adds them up to report how much of the region is covered by ice; meaning that it can be seen as a more accurate measurement of the actual amount of ice. But summertime measurements of ice concentration can be inaccurate and thus, extent is generally preferred by scientists, as it is more reliable.
Using a dataset (Fetterer, F. et al., 2017 [5]) listing the global sea ice extent and area for both the Arctic sea and Antarctica from 1979 to 2018, we analyzed the evolution of these two values over the years in January (winter season in the Arctic, summer in Antarctica) and June.

Let us first consider the Arctic sea, as it is often considered in the literature and media to be more vulnerable than Antarctica.

The ice extent shows a mean annual loss of approximately 60 000 squared kilometers ($\approx$ 2 times the size of Belgium) in January (46 500 in June), with a total difference of roughly 2.5 million squared kilometers between 1979 and 2018 in January, and 2 million in June.
That loss of sea ice is not linear as several years show an increase in ice extent compared to the previous year; but the cumulative variations show a distinct tendency of the ice extent to decrease over the years (Figure 5).

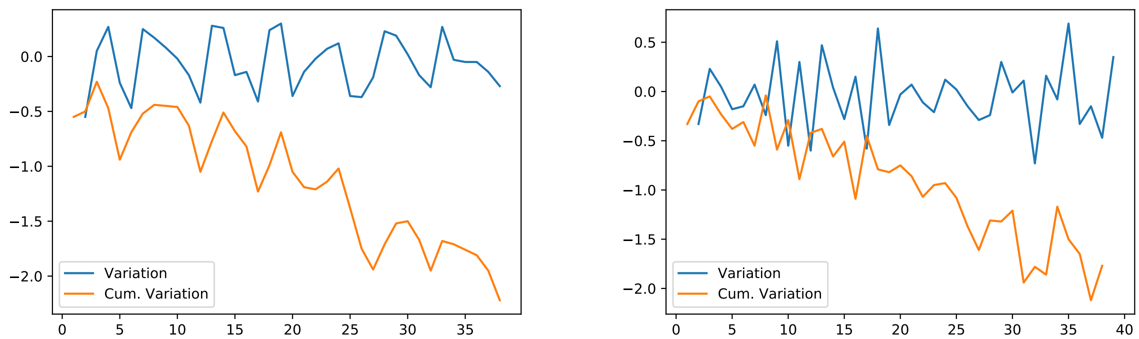


Figure (5) Annual variation of the ice extent in the Arctic from 1979 (year 0) to 2018, in January and June. The cumulative variations are indicated in orange to help visualize the global trend. The variations are expressed in millions of squared kilometers.

But as mentioned before, ice area is a better measurement of the actual amount of sea ice in the Arctic. However, a similar trend is observed, though not as spectacular.
In the Arctic, ice area shows a mean annual loss of approximately 28 000 squared kilometer in January (18 000 in June), with a total difference of roughly 800 000 squared kilometers between 1979 and 2018 in January, and 600 000 in June.
The decrease in sea ice area is thus slower than for ice extent. It is also a more recent phenomenon: the first obvious signs of decrease started in the nineties. However, the situation is worsening (Figure 6).

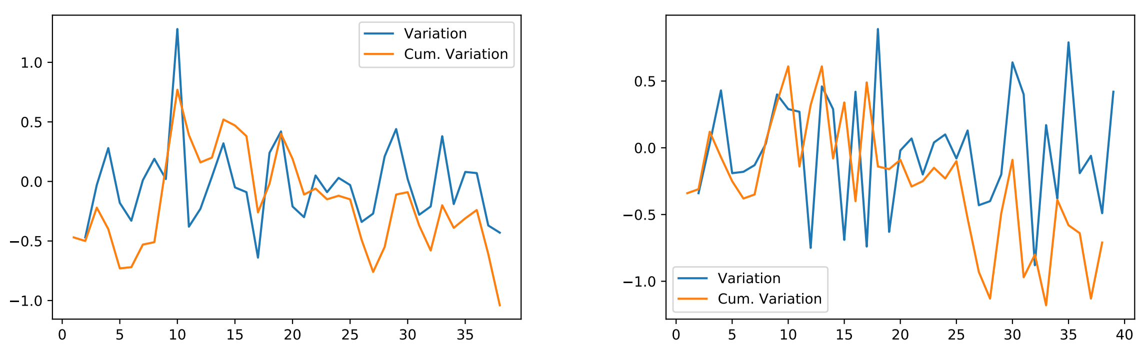


Figure (6) Annual variation of the ice area in the Arctic from 1979 (year 0) to 2018, in January and June. The cumulative variations are indicated in orange to help visualize the global trend. The variations are expressed in millions of squared kilometers.

Now let us consider the case of the Antarctic continent, in the southern hemisphere. The geographical situation is completely different, as Antarctica is an actual landmass (contrary to the Arctic), covered in ice.

However, the dataset only considers sea ice (hence the lower values than for the Arctic), for a fair comparison with the Arctic.
As a reminder, it was observed in section 3.4 that Antarctica mostly resists to the worldwide warming trend. We can thus expect a different situation than for the Arctic sea.

The ice extent shows a mean annual loss of approximately 44 000 squared kilometers in January (46 800 in June), with a total difference of roughly 1.6 million squared kilometers between 1979 and 2018 in January, and 1.8 million in June.[5]
It can be observed that the loss of sea ice is much more chaotic than in the case of the Arctic sea; and the cumulative variations even show an increase over the years (Figure 7). This corroborates the observations made in section 3.4. However, it should be noted that the recent years saw a dramatic decrease in ice extent, meaning that the situation could change abruptly in the near future.

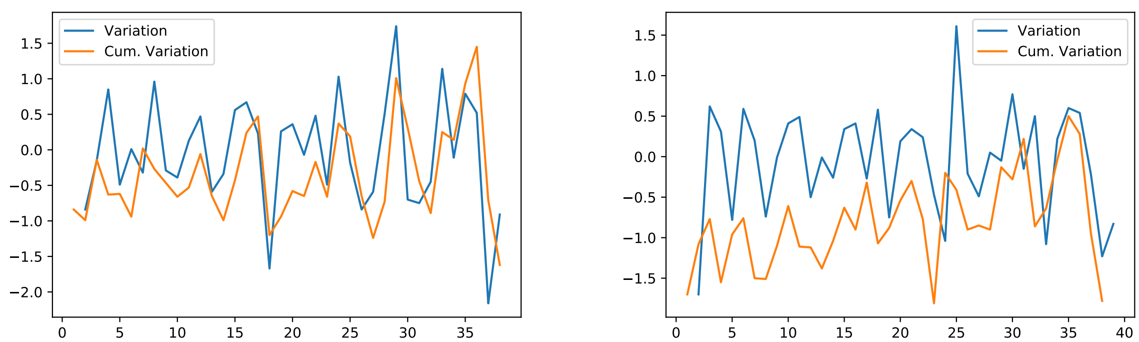


Figure (7) Annual variation of the ice extent in Antarctica from 1979 (year 0) to 2018, in January and June. The cumulative variations are indicated in orange to help visualize the global trend. The variations are expressed in millions of squared kilometers.

As for ice area, Antarctica shows a mean annual loss of approximately 24 500 squared kilometer in January (37 000 in June), with a total difference of roughly 900 000 squared kilometers between 1979 and 2018 in January, and 1.4 million in June.
The same observations as for ice extent can be made here. Ice area had a tendency to increase over the years, but an important decrease can be seen in recent years (Figure 8).

Judging from these observations, we can suppose that some natural phenomenon protected the Antarctic continent from the effects of global warming until recent years, where the situation started to change. It is plausible that the next few years will lead to a warming trend similar to the one observed in the Arctic.

[5]As Antarctica is in the southern hemisphere, January is in summer, and June near the beginning of winter.

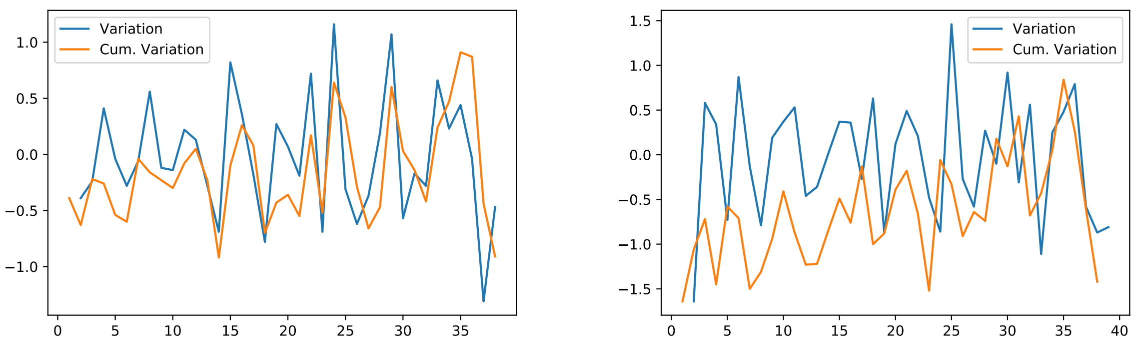


Figure (8) Annual variation of the ice area in Antarctica from 1979 (year 0) to 2018, in January and June. The cumulative variations are indicated in orange to help visualize the global trend. The variations are expressed in millions of squared kilometers.

3.3 CO2 Analysis

The purpose here is to analyze CO2 concentration measurements made in many places in order to see the evolution of this variable. In order to do that, very interesting datasets are provided by the Scripps CO2 program (C. D. Keeling et al., 2001-2005 [7] [8]). This program was initiated in 1956, at a time where it was suspected that the emission of fuel burning gas increased the atmosphere's CO2 concentration. In situ observations were made before this program. However, the techniques used before the Scripps CO2 program are nowadays considered as unreliable. This program provides many measurements since 1957 and is still continuing nowadays. The number of observation locations has increased over the years.

For each observatory, the records were taken by the mean of flasks and then analyzed in an external laboratory, except in the Mauna Loa Observatory and in the South Pole. For the Mauna Loa obsevatory, both flask and direct in situ observations were made. The South pole is a particular case. Most of the time, observations were based on in situ observations except sporadically from 1960 to 1963.

The seasonally adjusted CO2 concentrations for every observatory in the Scripps program are shown in Figure 9.
"Seasonally adjusted" means that the influence of the seasonal cycle has been removed. Indeed, the photosynthesis cycles of vegetation have an impact on the atmosphere's CO2 concentration; during the day and especially in summer, plants are stocking some carbon dioxide in order to produce glucose and oxygen. During the night and especially in winter, plants reject this CO2. To get accurate results, the influence of this seasonal cycle is removed.

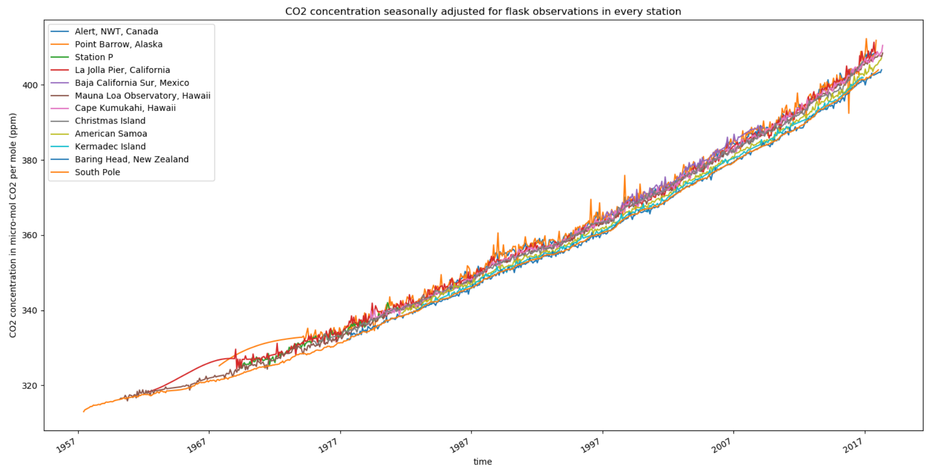


Figure (9) Seasonally adjusted CO2 concentrations for flask observations in every station of the Scripps CO2 program.

The key result is that the increase in CO2 concentration with time is not only a local phenomenon, but rather a global trend. The most surprising observation is that the concentrations observed in each location at a given moment are practically the same. The CO2 concentration's evolution follows almost the same law in every location.
One could think that the increases in CO2 concentration around the observation sites are due to nearby human activities. However, the measurements done in the South Pole contradict this assumption.

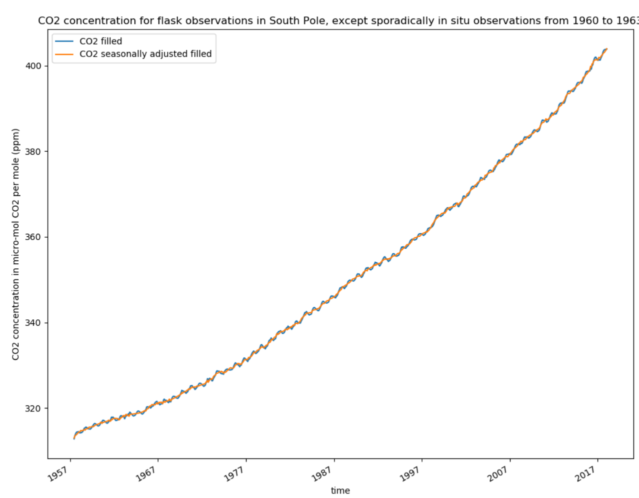


Figure (10) CO2 concentration in the South Pole.

It is clear that even in locations far away from human activities like the South Pole, there is an increase in CO2 concentration and from Figure 9, we can notice that the speed of this evolution is practically the same as for the other stations. In Figure 10, the seasonal adjustment is practically the same as the initial measurement due to the fact that there is no real seasons in the South Pole and due to the lack of vegetation, which is the main origin for seasonal cycles.

The problem with all these measurements is the lack of observations before 1958. With this data, it is impossible to know if the CO2 concentration has always increased or, if it is not the case, when it started increasing. It is quite difficult to give an explanation of this phenomenon and, in particular, to test the hypothesis that this evolution is due to fuel burning gases. One way to quantify the CO2 concentration far in the past is to use ice cores. The Scripps CO2 program organized expeditions in order to analyze ice cores and deduce from them the CO2 concentration in the past. The plot of the result is shown in Figure 11:

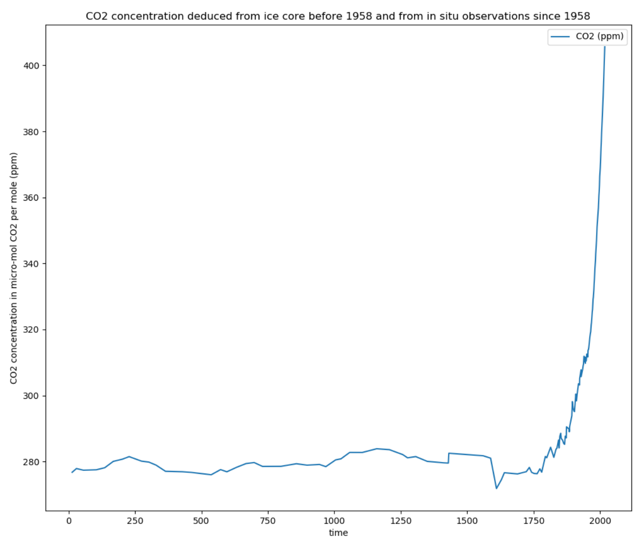


Figure (11) Long-term evolution of the CO2 concentration. Results for the years prior to 1958 where obtained using ice-core observations.

The key result of ice-core analysis is the fact that CO2 increase is a relatively recent phenomenon that is still continuing today. In the past, there were variations in the atmospheric CO2 concentration, but it mostly remained stable until 1800. Since circa 1800, the CO2 concentration is increasing at an abnormal rate. It should be noticed that the beginning of the XIXth century corresponds to the beginning of the first Industrial Revolution and the advent of steam machinery and factories. No obvious natural cause might explain this sharp evolution in CO2 concentration. The hypothesis that the abrupt increase in atmospheric CO2 concentration is the consequence of human activities is thus probably true, given these scientific measurements.

3.4 First conclusions

This first round of data exploration already gave us some insight into the global trends around the world and a first impression on the current situation. It also outlined several topics that would be discussed in the next step of the project: a correlation analysis between air and soil temperatures and the melting of the ice cap, or a discussion of the potential origins of the worldwide increase in CO2 concentration, for instance.

4 Towards a model

4.1 Regression analysis

At this point, we wanted to try a simple modelization of the dynamics of temperature, using the three variables we studied in the previous section.
And in order to achieve a first modelization of global warming, we decided to build incrementally a model, starting from simple assumptions and dependencies between variables. In particular, we considered the possible dependencies between ice extent and CO2, and between atmospheric temperatures and CO2, in order to build linear models to predict the evolution of these variables in the next few years. Of course, we were aiming at building a very simplified model (an already daunting task, as we quickly found out), as climatic models are extremely complex and depend on a myriad of factors.

4.1.1 Ice extent

Here, we considered the link between CO2 concentration in the atmosphere and ice extension (both in the Arctic and in Antarctica), and performed a linear regression over the two variables and their respective evolution between 1979 and 2018 (using the same datasets as in the previous step of our project [5] [8]) to obtain a linear model giving us the extension of ice as a function of CO2 concentration.

Not surprisingly, the observations made in section 3.2 concerning the differences between the situation in the Arctic Sea and in Antarctica are confirmed by the obtained models: while in the Arctic ice extension is a decreasing function of the atmospheric CO2 concentration, in Antarctica it is a (slightly) increasing function of it (Figure 12). This corroborates the negative dynamics of air temperature along a significant part of the Antarctic shoreline seen in the previous chapter.

Numerically, we found the following relations (ice extent is expressed in millions of squared kilometers, CO2 concentration in ppm):
For the north pole: $Ice\_extent = -0.02717175 * CO2 + 24.24842255$.
For the south pole: $Ice\_extent = 0.00908372 * CO2 + 10.06414641$.

If we perform the regression on the global ice extent (Arctic sea + Antarctica) however, the resulting function is indeed a decreasing one (Figure 13): $Ice\_extent = -0.01825151 * CO2 + 34.37638671$

The fact that, in Antarctica, ice extent is an increasing function of CO2 is *a priori* non-sensical; on the other hand, the global relationship seems more plausible, but does not give much information. This indicates that CO2 is clearly not sufficient to explain the variations in ice extent (which is quite logical given the complexity of this phenomenon); we will have to build a more complex model.

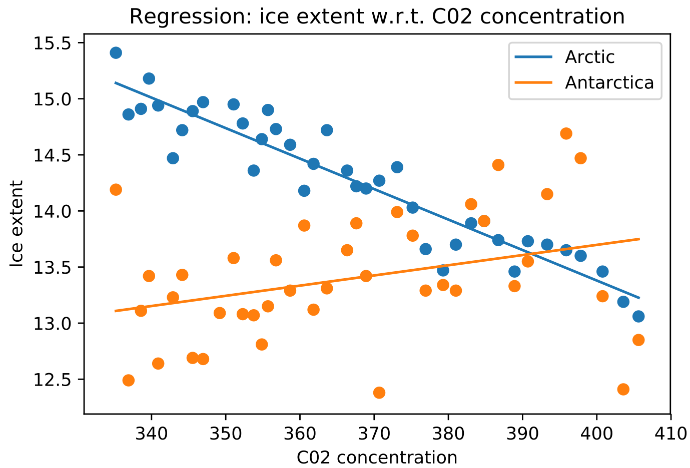


Figure (12) Regression over the ice extension and CO2 concentration, for both poles.

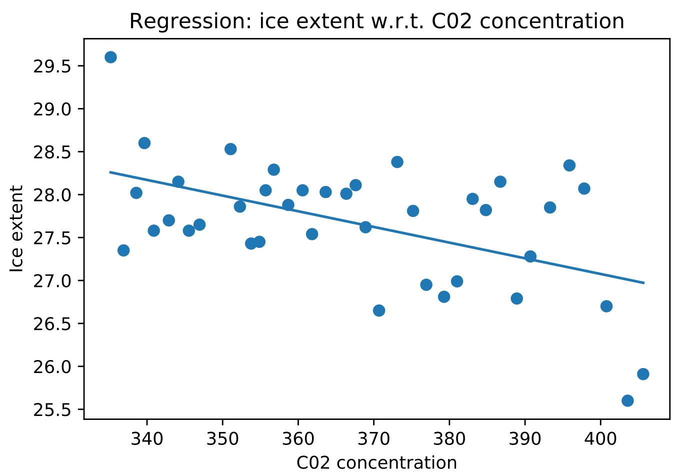


Figure (13) Regression over the ice extension and CO2 concentration, when taking the global ice extent (both poles together).

4.1.2 Temperatures

In this part, we propose to analyze correlation between the time, the CO2 concentration and temperature records from the Mauna Loa Observatory. Due to its location far away from major continents, it provides an opportunity to deal with data without influence of any local pollutant. Thanks to that, we can see the direct influence of CO2 concentration on the temperature at that place. We also try to model the potential correlation between the different variables. In order to have relevant results showing a tendency, we decided to keep the average values of the variables for each year. Thanks to that, we erase the influence of seasons in the different measurements.

The equation for the CO2 concentration obtained via regression is the following: $CO2\_concentration = 1.4967 * year - 2623.70246$.

Figure 14 shows the relationship between the CO2 concentration and the air temperature. No clear dependence can be observed. This contradicts the common belief that the temperature rises because of the increase in atmospheric CO2. In Mauna Loa, we observe an increase in CO2 concentration but the temperature does not follow the same tendency. As mentioned in section 3.4, air temperature is not increasing everywhere in the world at the same rate. On the other hand, CO2 concentration follows practically the same law in every measurement location, thus it is possible to make conclusions on the global CO2 state based on local evidence. That leads to the conclusion that the air temperature dynamics and their causes are a really complicated phenomenon that cannot be explained only by the increase in CO2 concentration, as many seem to believe.
The linear model built for these variables is given by $Temperature = 0.0023 * CO2\_concentration + 15.8341$.

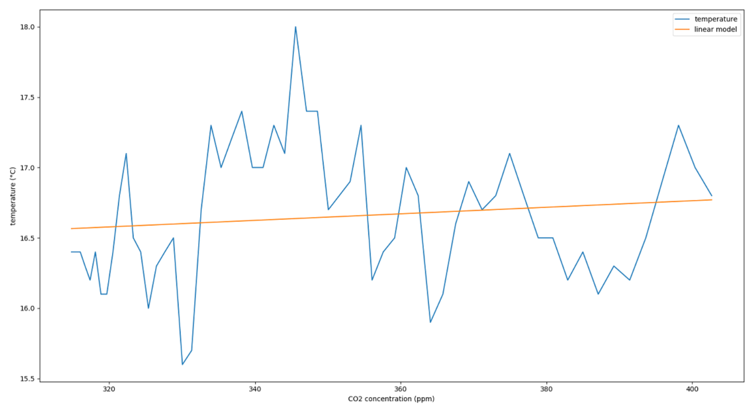


Figure (14) Air temperature as a function of the CO2 concentration, with the corresponding linear trend. Data from Manua Loa observatory.

A linear regression was also performed to evaluate temperature with regards to CO2 and time. We found the following relation:

$$Temperature = 0.059 * Year - 0.0364 * CO2 - 87.8058$$

One interesting result is that, when considering the 3 variables to build a linear model, the year variable influences more the increase in temperature than when considering a linear model using only the time and temperature. On the other hand, the CO2 concentration is a little bit inversely correlated with temperature. This means that time has more influence on temperature than CO2 concentration. This shows that global warming is not a simple effect of atmosphere CO2 concentration increasing.

4.2 A simple zero-dimensional model

The increase in atmospheric CO2 and the ice melting that we quantified and discussed in the previous sections are widely recognized as variables linked to global warming. But are they causes or consequences of it ? Which role do they play in climate change ?

In climate science, there is almost no definite causes. The only exception is the Sun, which is the main engine of all Earth processes. There is no definite consequences neither: through the so-called butterfly-effect, even

a small perturbation in the system can cause its change.
The melting of sea ice appears to be a consequence of global warming. Ice melts due to the rise of air temperature. However, when ice turns into water, its ability to reflect the incoming solar short-wave radiations changes. While the middle-aged ice of the Arctic has an albedo of 0.6, for water in the polar latitudes it is equal to 0.35; which means that the less ice there is, the less solar radiations will be reflected. Thus, the system self-accelerates.
Greenhouse gas (CO2), in another hand, appears to be a cause. It absorbs the outgoing long-wave radiations from the Earth, and thus decreases its effective emissivity, which leads to an increase in temperature. Although many natural organisms and surfaces emit CO2, the main responsibility on its rise in the atmosphere lies on humans, as discussed in section 3.3.

Using the linear trends for ice extent and CO2 calculated in section 4.1, it is possible to estimate the dynamics of these variables until 2030, neglecting the non-linearity of their temporal dynamics due to a short prediction time.
Instead of just correlating them with temperature, which is possible statistically but has little sense scientifically, we can use a simple 0-dimensional physical-statistical model to calculate their impact on temperature in the near future. In its basis, the model uses the law of the radiative equilibrium. The model is defined as follows:

$$T = \sqrt[4]{\frac{(1-\alpha)*S}{4\epsilon\sigma}}$$

where T is the global temperature in Kelvin, α is the surface albedo, S is the solar constant (equal to 1367 $Watt * m^{-2}$), ϵ is the effective emissivity of the Earth and σ is the Stephan-Boltzmann constant (equal to $5.67 * 10^{-8}\ J * K^{-4} * m^{-2} * s^{-1}$.

In this model, the state variable is the air temperature, and the arguments are the albedo and the effective emissivity of the Earth. According to the approximation commonly found in the literature, the Earth's albedo is currently 0.3, thus the effective emissivity is believed to be 0.612. However, it is taken here as 0.617 to match the initial state of the system, where the global temperature is around 14.4 C by the year 2017. Then, for each predictive year, the global albedo was calculated, taking into account the predicted extent of the ice by the common "mixtures formula":

$$ice\_surface*ice\_albedo+non\_ice\_surf.*non\_ice\_Earth\_albedo = total\_Earth\_area*Earth\_albedo$$

where the albedo of the surface not covered by ice was estimated based on the data from 2017 and then further used in calculation of the Earth's albedo.

The same "mixture approach" was taken into account for the calculation of the effect of CO2 on the global emissivity. According to meteorological research (`http://www.biocab.org/ECO2.pdf`), CO2 at concentrations of 300 ppm in the atmosphere has an emissivity of 1.69963e-03, while raised up to 600 ppm it decreases to 1.6988e-03. Based on this data, a linear function was built to calculate the CO2 contribution to the global emissivity for all predicted concentrations. The non-CO2 contribution on emissivity was estimated based on the approach explained above and then, the global emissivity was calculated for each year until 2030.

Figure 15 shows the difference in prediction of the global temperature based on the the second order polynomial approximation and based on the 0-D model. While the pure statistics show an increase in temperature of 0.4 degrees C, which generally corresponds to the IPCC scenarios (Figure 16), the 0-D model gave an increase of only 0.035 degrees C.
This difference can be explained by the fact that CO2 is not the main greenhouse gas. Water vapor (i.e. clouds) absorbs much more long-wave radiation, and its concentration in the atmosphere is dominant. Methane contributes to global warming as well, and it has a greater potential to heat the atmosphere

than the CO2, due to the huge quantity of it stored in the permafrost. The thawing of the permafrost and the associated release of methane is a self-accelerating system. And in order to correctly predict the temperature rise with the statistical-physical model, these greenhouse gases should be taken into account. Finally, the albedo change due to ice melting was found to have little impact on the air temperature. According to NASA, the global albedo does not have any particular trend, which proves the fact that the ice melting phenomenon does not significantly influence the climate.

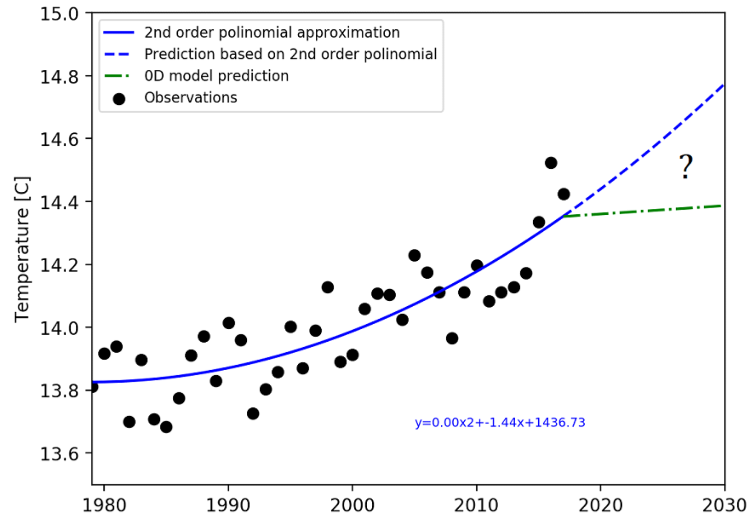


Figure (15) Temperature prediction based on polynomial approximation (its equation is written in blue) and the zero-dimensional physical climatic model, which uses the trends for CO2 and ice discussed above. The question mark shows the huge difference between the predictions and opens a discussion about the other processes contributing to climate change.

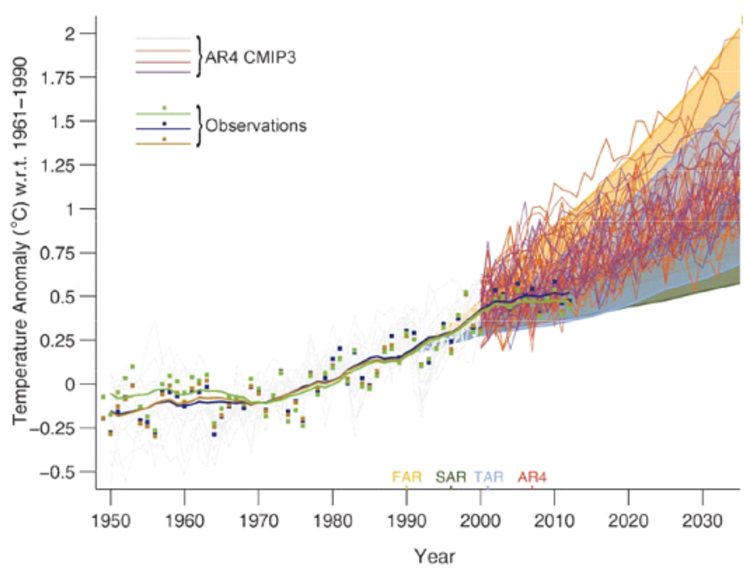


Figure (16) IPCC global warming projection.[9]

4.3 More in-depth greenhouse gas study

In section 4.1.2, we tried to make models of the temperature in function of the CO2 concentration. We took the CO2 concentrations and temperatures at Mauna Loa observatory in Hawaii in order to make such models. One limitation of these models was that we only considered CO2, a well-known but not unique greenhouse gas. This logically led us to interest ourselves in other greenhouse gases, in particular methane (CH_4) and nitrous oxide (N_2O), whose effects are much worse than CO2. We found in the scientific literature that releasing 1 kg of CH_4 in the atmosphere is equivalent to releasing 25 kg of CO2. It is also told that releasing 1 kg of N_2O in the atmosphere is equivalent to releasing 298 kg of CO2.

In order to analyze the impact of these two gases, we decided to use paleoclimatic data derived from ice-core in Law Dome, in Antarctica (Etheridge et al., 2011 [10]). This paleoclimatic data allows us to go far in the past. This dataset provides CO2, CH4, and N2O concentration from the year 1 AD to 2004 AD.
We decided to compare this greenhouse gases dataset with a dataset giving the global annual mean temperature on Earth. This temperature dataset starting in 1880 AD, we can model the evolution of temperature for the period 1880-2004.

The concentration in CO2 largely dominates the two other gases. However, by plotting the evolution of the three gases separately, we can observe that the two other gases follow the same increasing trend as the CO2.

We first computed the linear coefficient of correlation between each greenhouse gas and the temperature. The results are the following:
The correlation between CO2 and temperature is 0.9000304158363414
The correlation between CH4 and temperature is 0.8815914696665186
The correlation between N2O and temperature is 0.8947778500950238
We can see that the linear correlation between each greenhouse gas and the temperature is high. Given this observation, it seems appropriate to consider linear models to explain the evolution of temperature (we will give a statistical justification for that in the next section). In order to take noise into account, we used Bayesian ridge models.

The Bayesian ridge model of the temperature as a function of all three greenhouse gases performs really well (mean squared error of ≈ 0.012052), though it is practically the same for the Bayesian ridge model of the temperature as a function of the CO2 only (MSE of ≈ 0.012036).

In addition to linear Bayesian ridge regression models, we also made two non-linear models, based on random forests: one with a constrained depth of 10 for the trees and the other one with no constraints on the depth of the trees. For the two models, the number of estimators used is 100000.
The mean squared error of the model with constrained depth is ≈ 0.000969, and the returned feature importance are as follows: CH4 concentration importance : 0.3559104213405303
CO2 concentration importance : 0.3475801119480182
N2O concentration importance : 0.2965094667114437
The model with unconstrained depth gives almost the same values.

One interesting observation here is that, for the random forest models, the atmospheric concentration in methane (CH_4) is the most important factor contributing to global warming. However, the importance is not much higher than the two other types of greenhouse gases. Consequently, each of the three greenhouse gases has a real impact on the observed increase in temperature. This conclusion is also valid for Bayesian ridge models because adding or removing variables does not affect a lot the accuracy of the model, meaning that each greenhouse gas type has an impact on global warming. This impact must not be neglected.

In conclusion, we see that a priori, the three types of greenhouse gases analyzed (CO_2, CH_4, N_2O) have an impact on global warming (once again, we will verify this claim from a statistical point of view in the next section). The impact of each greenhouse gas on global warming seems similar to the two others. CO2 is, according to climatologists, less harmful than the two others, but the disproportionate concentration of CO2 in the atmosphere compared to the two other greenhouse gases makes CO2 as impacting as the two others on global warming.

5 Evaluating our models

In the course of our project, we analyzed several climatic variables of interest in order to detect global trends and hints of a warming of the climate. These experiments yielded interesting results but are not sufficient, on their own, to give a conclusion on the reality of global warming.
In particular, one could argue that the trends we observed for various variables and that might give signs of global warming, could be a statistical anomaly and not a real, long-term phenomenon. Therefore, it is of interest to perform some forms of statistical test on the basic models we considered throughough this project.

We tried to evaluate the plausibility of two models in particular: our bayesian ridge model for temperature in function of greenhouse gases concentrations, and the slightly more complex zero-dimensional model.

First off, we performed a test for significance of regression (TSR) on the linear model we built with the three greenhouse gases CO_2, NH_4 and N_2O. The test is used to check if a linear statistical relationship exists between the predicted variable (here temperature) and at least one of the predictor variables (the greenhouse gases). In other words, we try to see if a linear model linking temperature with greenhouse gases is better than a constant model or if we just try to force a linear model on a phenomenon that cannot be explained this way.

The test is based on an f-test of equality of variance. We define the regression sum of squares as

$$SS_R = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$$

and the associated variance (regression mean squares) is given by

$$MS_R = \frac{SS_R}{dof(SS_R)}$$

where $dof(SS_R)$ corresponds to its degree of freedom, i.e. the number of explaining variables. In the case of the greenhouse gases model, there are 3 variables, thus 3 degrees of freedom.

We also define the error sum of squares and the associated variance

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$MS_E = \frac{SS_E}{dof(SS_E)}$$

where $dof(SS_E)$ = number of observations – (number of explaining parameters + 1) = number of observations – 4.

Now we can do the f-test. The statements for the hypotheses are as follows:

- H_0: No real relationship between the predicted and predictor variables; all parameters of the model are equal to 0.
- H_1: At least one predictor variable is significant (at least one parameter different from 0).

We define $F_0 = \frac{MS_R}{MS_E}$. From theory, we know that if the null hypothesis H_0 is true, then the statistic F_0 follows the F distribution. We can thus derive a p-value, which denotes the probability of the data given the null hypothesis. Using this test, we obtain a p-value $p \approx 1.711980278725769 \times 10^{-43}$ for the whole model. We can thus safely reject H_0 and consider the regression model as significant.

However, if we look at the significance of individual regression coefficients using a *t-test*, we get the following results:

- p-value for CH4: 0.9926910951843249
- p-value for CO2: 0.0467134306567861
- p-value for N2O: 0.8149190162681534

Therefore, only the CO_2 can be considered as statistically significant in the evolution of temperatures, which justifies the fact that it is often considered as *the* greenhouse gas linked to global warming.

We also tried to consider a likelihood ratio test to evaluate the plausibility of our linear 0D model.
The ECMWF data averaged over the globe (see section 3) was used in the test. The variance of the data was calculated for the null hypothesis over the untouched dataset, while for the linear model its slope was taken into account and subtracted from the observations before calculating the variance. Therefore, two Gaussian functions were constructed (see GitHub for the details).
Then, the logarithms of the probability density functions were calculated for both distributions.

By computing the maximum likelihood ratio, we found a p-value of 0.000000000041 for the linear model, meaning that we can safely reject the null hypothesis. Even when using the same variance for both distributions, we obtain a p-value of 0.0000003.

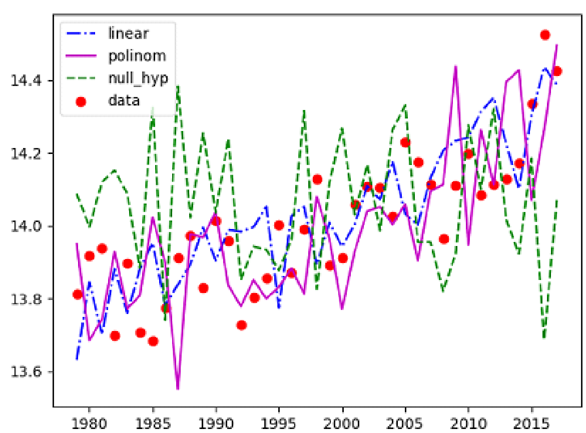


Figure (17) Tested probabilistic models (blue – linear + variance; magenta – second-order polynomial + variance; dashed green – null hypothesis (bias is fixed); red dots – data)

Then, the "superiority" of a second-order polynomial function over the linear model was tested, with the same procedure. With a p-value of 0.01437, we can consider that a polynomial model is a more realistic approximation of the temperature trend.

We can thus reasonably conclude that the increasing trend in temperature that we observed and tried to model (simplistically) in this project is not a mere statistical fluke, but a real phenomenon.
However, we have to keep in mind that the simple tools used in the context of this project are not quite sufficient to assess the exact magnitude and consequences of such a complex phenomenon; they are not sufficient to conclude on the real impact of humanity either. But the reality of global warming, be it an entirely natural process or not, is quite clear.

6 Other phenomena of interest

On top of the simple models presented and evaluated in the previous sections, we performed additional experiments on other climatic phenomena that could be of interest. These phenomena mostly illustrate side-effects of global warming, and can also help us identify the magnitude of the changes our climate is undergoing.

6.1 Study of the climatic seasonal shift using self-organizing maps

Indeed, because of climate change, one could expect a shift of boundaries in the four climatological seasons (spring, summer, autumn and winter). To prove or discard this hypothesis, we used self-organizing maps, an unsupervised learning algorithm that is often used in climatological studies (Liu and Weisberg, 2011 [11]).
To conduct our analysis, climatological data giving monthly mean air temperatures for 1979-2017, covering the globe with a spatial resolution of 3x3 degrees, was used. The Neupy Python library (`http://neupy.com/pages/home.html`, retrieved 21/02/2019) was used to construct the SOM. The period of 1979-1994 was used as a training set (thereafter referred to as the reference period) to define our reference climatological seasons using the SOM. The 3D matrix of temperatures was flattened along the spatial dimension and fed to the algorithm. We chose to use four clusters, according to the number of seasons. Initially, four random samples were assigned as the centers of the clusters. Two hundred iterations were used to correct positions of the clusters.

To represent classified neurons and the positions of clusters over the reference period, the mean temperatures averaged over the southern hemisphere was plotted against those over the northern hemisphere (Figure 18). Interestingly, the groups of neurons are quite distinct, with the exception of the few yellow and magenta dots referring to 0 degrees for the Southern Hemisphere and 10 degrees for the Northern Hemisphere. The picture shows a clear distinction between the seasons for the reference period.
We have to keep in mind that summer in the Northern Hemisphere corresponds to winter in the Southern Hemisphere, due to the inclination of the Earth's rotation axe and Earth's exposure to the incoming solar radiations. On top of that, climatological seasons do not exactly match the general season definition: for instance, due to inertial warmth, September lies to the definition of a summer season more than June does.

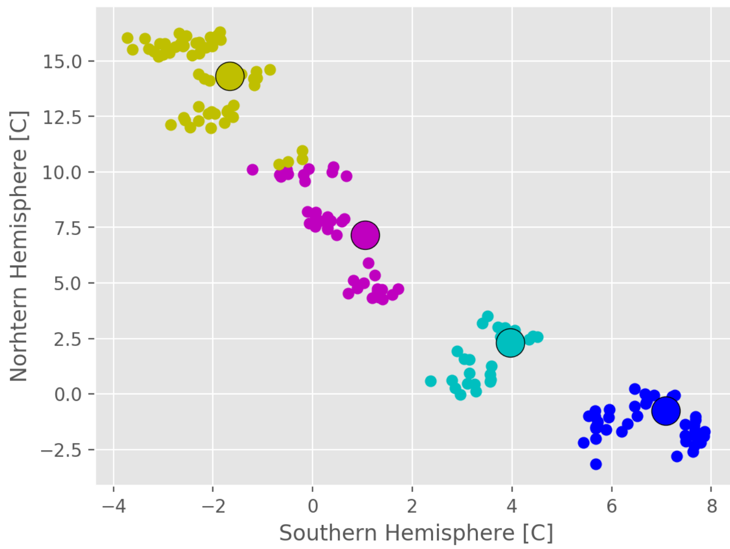


Figure (18) The SOM. Each individual neuron (small point) represents an average temperature over the corresponding hemisphere for a certain month within the reference period. Positions of the clusters, indicated by the big circles with black edges, indicate the SOM execution and define seasons (yellow – summer, magenta – early autumn and late spring, cyan – early spring and late autumn, blue – winter) for the current study.

Then, the definitions of seasons obtained as a result of the SOM execution were applied for the testing period of 1994–2017, thereafter referred to as the application period. The results (Figure 19) show a visible shift of seasons. Visible differences can be seen in the classification of February and May. For the reference period, February was a typical winter month, but for the application period, it already shares features of a transitional cold month 18% of time. May was a summer month 25% of time in the reference period, but for the application period this number increases up to 62%; which redefines it as a summer month. In general, summer has become significantly longer due to its early start in May; and spring has become shorter. Winter has become slightly shorter and the length of autumn remains pretty much unchanged. Therefore, the hypothesis of a seasonal shift is proven.

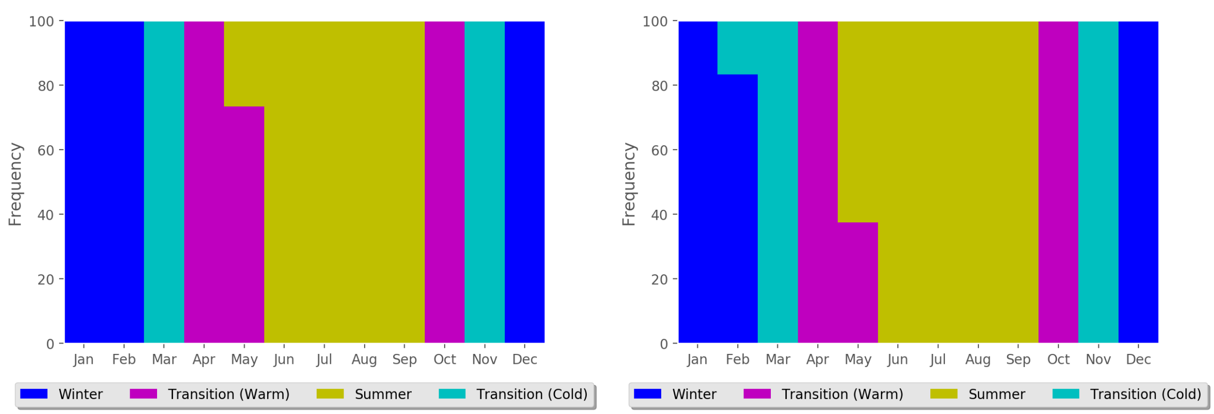


Figure (19) Frequency [%] of classification of a given month in a given season: for the reference period (1979-1994, shown left), and for the application period (1994–2017, shown right).

Our study, however, has some flaws. The result strongly depends on our training and testing sets. Since the rule of respecting a 70 to 30 proportion (training set to testing set) cannot be applied here due to the non-random character of our data, it is recommended to find a period for which global temperature does not change over the years. However, we showed previously that for the period 1979-2017, the global air temperature demonstrates a stable increasing trend.
Thus the present analysis gives conclusions based on data which is already impacted by global warming. However, this analysis was also performed to show the capabilities of self-organized maps to study climate change and our claim on the existence of a seasonal shift still remains valid.

6.2 Snowfall as an indicator of climate change

Snowfall data is not a straightforward indicator of climate change. Casual snowfall depends on many factors, such as the availability of water (in liquid state it is more exposed to evaporation), air temperature (the hotter it is, the faster evaporation goes), wind speed (removes water molecules, leading to further evaporation). However, strong snowfalls occur only when different air masses with different temperatures (for instance, when the warm Gulf Stream air meets the cold Labradorean air) get in contact. Taking into account the fact that climate change supposedly changed all individual climatological constituents, we could expect a change in the snowfall as well. The fact that excessive snowfall can be harmful for people and the economy makes it even more interesting for us to investigate.

To conduct this study, we proposed two hypotheses for verification. First, that the snowfall rate should decrease over time, signifying a warming of climate. Second, that the severity of extreme snowfalls, on the other hand, should increase, since every cold wave is subject to a massive media coverage, relating each event to climate change.
The climatological reanalysis data Era-Interim was downloaded from the European Centre for Medium-Range Weather Forecasts (ECMWF). In this dataset, snowfall had a temporal resolution of 12 hours for the period $1979 - 2018$, and represented the amount of solid precipitation collected during a half day. Since the definition of "solid precipitation" includes not only snowfall, but hail as well, even Equatorial Africa has non-zero data records. The period was divided in two semi-equal parts: period A ($1979 - 1997$) and period B ($1997 - 2018$). Their inequality was always taken into account in the calculations. The total amount of data had a size of roughly 420 megabytes.

Figure 20 represents the change in the average daily snowfall between the periods A and B. It shows that the amount of snow increased over the past two decades around Antarctica and in the Aleutian Low, meaning the intensification of cyclonic dynamics. The amount of snow decreased over the whole Gulf Stream, signifying that the contact surface between the Gulf Stream hot air and the Labradorean cold air shifted northward.
The decrease of snowfall in Western Europe, visible in Belgium, is due to the shift of the winter air temperatures from negative to positive (in Liège, January is around $+1$ C, although half a century ago it was below zero). This zero-temperature boundary and related snowfalls shifted eastward, leading to an increase in snowfall in Eastern Europe. As a result, the hypothesis was partially approved for the Northern Hemisphere, but rejected for the Southern Hemisphere.

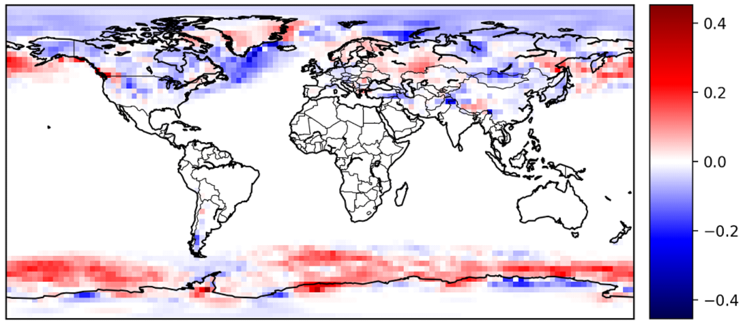


Figure (20) Change in the snowfall rate between 1979 – 1997 and 1997 – 2018 (the difference in the length of periods is taken in account). The colorbar is in millimeters per day.

For the second hypothesis, an "extreme event" was defined as an event where the daily amount of snow exceeds 99% of its records for the given period (such an event, according to climate science, happens once a year).

Figure 21 represents change of the number of extreme events over the Globe between the periods A and B. It shows its visible reduction for the whole Europe (except the South of France) and for the USA (except New England). It means, that despite the fact, that their actual frequency has decreased for the Western World, the severity of each individual extreme event has increased. However, the global picture shows no clear trend and requires more detailed investigation. Thus, the second hypothesis was partially approved for the Northern Hemisphere and rejected for the Southern Hemisphere.

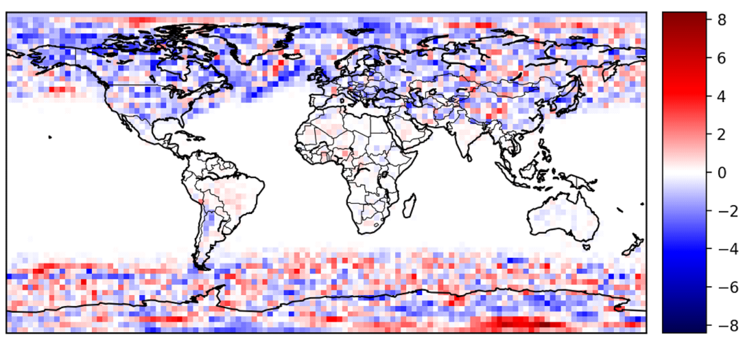


Figure (21) Change in the number of extreme events between 1979 – 1997 and 1997 – 2018 (the difference in the length of periods is taken in account). The colorbar is in numbers per period.

6.3 Sea level analysis

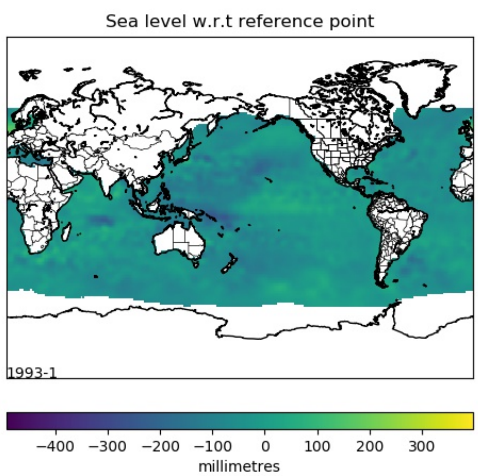
A more commonly studied indicator related to global warming is the evolution of the sea level. This natural phenomenon, a direct consequence of interglacial periods where massive amounts of ice melt, is directly impacted by climate warming in two ways:

- Thermal dilatation of water: as the temperatures get warmer, thermal expansion dilates water, which takes more space;
- Ice melting, a phenomenon already studied in this project.

Sea level rise since the start of the 20th century has been dominated by retreat of glaciers and thermal expansion of the ocean, but the contributions of the two large ice sheets (Greenland and Antarctica) are expected to increase in the 21st century, as, as previously studied, ice melting is accelerating. The ice sheets store most of the land ice (more or less 99,5%), with a sea-level equivalent of 7,4 meters for Greenland and 58.3 meters for Antarctica; this could dramatically increase the rise of sea level in the next century.

In any case, studying the evolution of the sea level would be of interest. We used a dataset combining data from TOPEX/Poseidon, Jason-1, Jason-2 and Jason-3, four major satellites for oceanographic measurements. The date is near-global (65S to 65N), with a 1 x 1 degree spatial resolution.
The sea level is expressed in millimeters, with regards to a reference level. We know from the literature (Ablain et al., 2019 [12]) that these measurements are highly reliable and will hopefully get even more reliable in the future, as new data is collected.

Perhaps counter-intuitively, the level of the sea is not the same everywhere; in fact, there is sometimes a difference of up to one meter between, for instance, the Caribbeans and the coast of Greenland. This is due to several factors, oceanic streams (most notably the Gulf Stream) being one of them.
In order to visualize the spatial trends in sea level variations, we took advantage of the spatial information (latitude, longitude) contained in the dataset to plot the data on a map and produce a video animation (which can be consulted on the github) showing the evolution of the sea level around the globe.
When looking at the animation (or, equivalently, in the dataset), it appears that, overall, an increase in sea level is the norm in most places, though some regions (most notably the Eastern Pacific) show a slight decrease.
Figure 22 shows two snapshots from the animation; one taken in january 1993 (first month of the dataset), and the other in november 2018 (last month). Although not blatant, the general increase in see level described above is indeed showing all around the globe (brighter yellow).



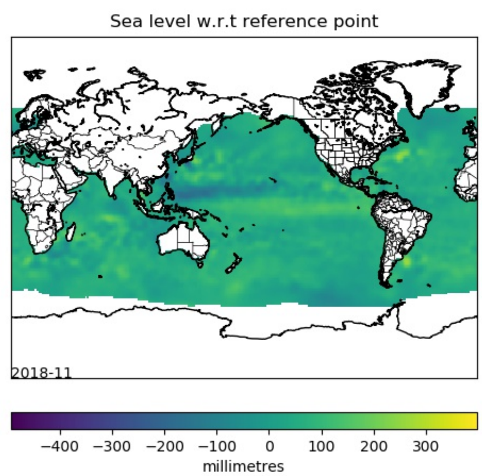


Figure (22) Compared sea levels in 1993 (left) and 2018 (right), in millimeters with regards to the reference level.

We can also visualize the global trend by considering the evolution of the global average sea level over time. Figure 23 shows the plot of the monthly global average sea level over time. Indeed, we can clearly see that overall, the sea level is increasing steadily. Yearly, we observe an increase of 3,2 millimeters in the sea level; this can seem rather modest, but it already has important consequences in some parts of the globe.
Furthermore, we have to keep in mind that the increase we visualize is mostly caused by thermal expansion

and the retreat of glaciers; we have yet to witness the effects of the rapid increase of ice melting in the two major ice sheets, Greenland and Antarctica, and their soon-to-come increased contribution to sea level rise. Climatologist predict an important increase, that could lead to an increase of several meters by the end of the 21st century.

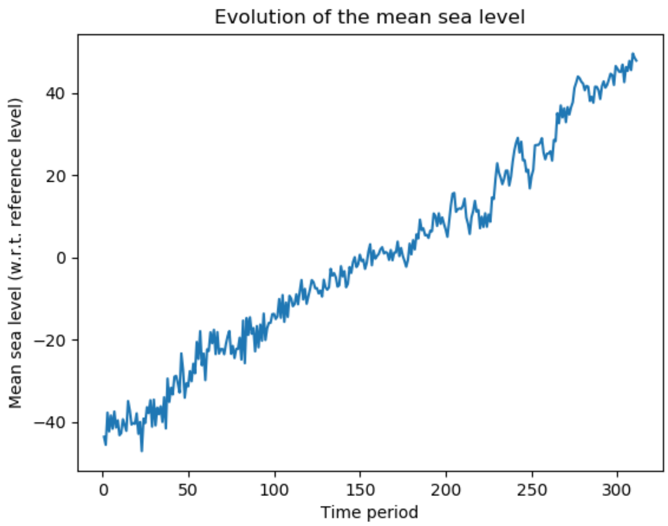


Figure (23) Evolution of the mean sea level (global average) between january 1993 and november 2018, with regards to the reference point.

6.4 Precipitation

The amount of precipitation is one of the climatic variables for which data is collected since a long time, and is reliable. It is also an interesting variable to study if we want to detect changes in meteorological events. The dataset we used had precipitation data recorded from 1891 to 2016. The precipitations are available globally, from all around the globe, with a 1 x 1 degree spatial resolution. In order to make temporal analysis of the data, videos of the evolution of precipitation on earth have been created as well, and are available on the github.

First, a video of the monthly evolution of precipitation have been done. We observed that the distribution of precipitation is fluctuating. It is fluctuating nowadays, but also in the first years available in the dataset. However, some regions are more stable than others in term of precipitation quantity. For example, Europe is a relatively stable region in terms of rainfall (and other meteorological phenomenons). On the other hand, the regions surrounding the equator are much more unstable. We can clearly see that these regions can be split in two seasons: the dry and wet season.
Obviously, even if the seasons are not disappearing, it does not mean that there is no change in precipitation. In order to detect potential changes over time, it is easier to consider global annual precipitation. Another video has been created. In that video, we can see that the distribution of precipitation is not stable with time. However, the pattern of the evolution of precipitation is not as clear as when we were considering the trends in the evolution of temperatures, for instance. Further investigation are needed to analyze the variations in precipitation.

To quantify the evolution in precipitation, one can first make some plots giving the global tendency of the evolution of precipitation. Figure 24 gives the yearly average precipitation (in millimeters) per month. There is no increasing tendency as it was the case when we studied temperatures. However, the plot clearly shows that precipitation vary more than in the past. This also opens the possibility to consider that "extreme" (i.e., out of the ordinary) climatic periods are more and more frequent.

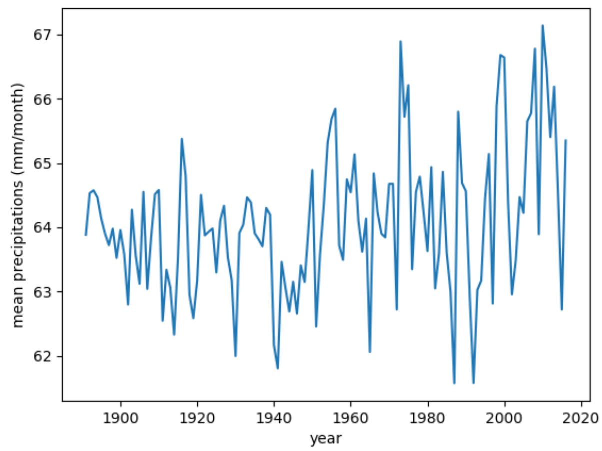


Figure (24) Yearly average month precipitation quantity in millimeters per month.

In order to quantify these increasing fluctuations, one interesting statistical parameter is the standard deviation of the yearly average quantity of monthly precipitation. Figure 25 shows the standard deviations of the yearly average precipitation per month. We can clearly see that the standard deviations is globally increasing with time. Furthermore, the variations are also becoming sharper with time. The dynamics of the water cycle seem to change. This could mean that some regions of the globe are receiving more precipitation than before while others suffers from droughts.

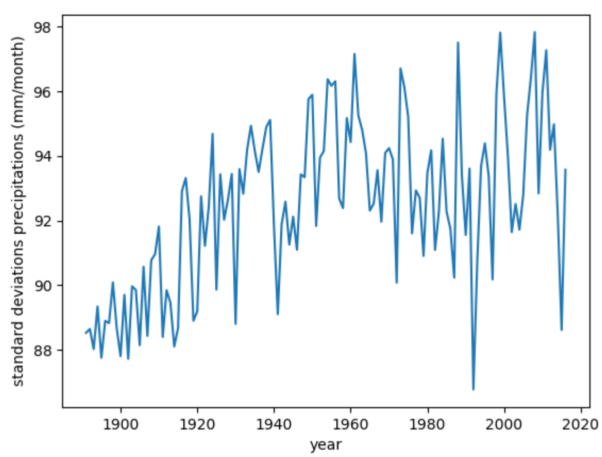


Figure (25) Standard deviations of the yearly average precipitation per month.

In order to visualize the spatial and temporal changes in precipitation and to see which regions on the globe are impacted and how they are impacted, the idea is to compare the yearly mean precipitation of each regions of the globe with a reference. A first approach would be to take as reference the first year available in the dataset (i.e. 1891). However, this is not very reliable because exceptional weather conditions and exceptional years have always existed. Thus, to have reliable averages, we need to consider periods of many years.

In this analysis, we chose to cut the dataset in two identical part. The years from 1891 to 1953 included are used as reference years to compute the annual precipitation for each region on the globe. The years from 1954 to 2016 are used to check if there is an evolution in annual precipitation. The results are shown in figure 26. We can see that, indeed, there are some variations in annual precipitation. The results show that some regions become drier and drier and, conversely, others become wetter and wetter. The global tendency is that tropical areas are loosing precipitation and that temperate regions are gaining precipitation. This shows that, even if the global annual average precipitation are not evolving a lot, the distribution of rainfall on Earth is changing.

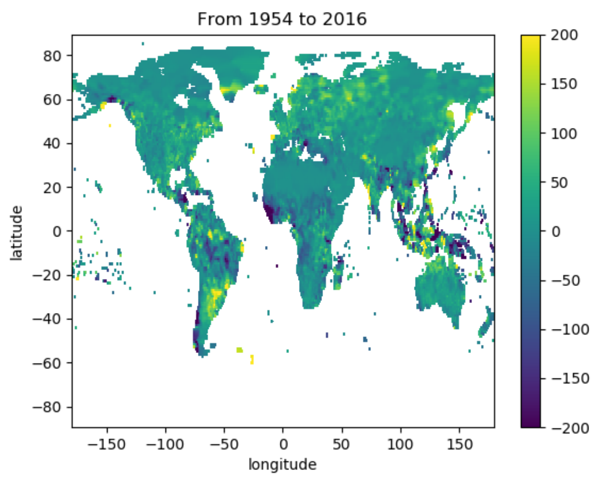


Figure (26) Differences in average annual precipitation (in millimeters per year) between the period 1954-2016 and the period 1891-1953.

The changes are quite significant. It cannot be imputed to small variations as we can see between each year. We can see variations of more than 200 millimeters of precipitation. In comparison, in Belgium, the annual precipitation quantity is about 800 millimeters. The climate of some regions in the world is thus really impacted. However, in many areas, there is no significant changes in precipitation. Thus, contrary to temperature where the evolution and changes could be measured everywhere in the world, precipitation changes are a more local phenomenon. It is difficult to determine the cause of these changes. One good hypothesis is to consider that changes in temperatures have modified water movement dynamics, resulting in variations in precipitation compared to the past.

Part III

Conclusion

Throughout the course of this project, we studied several climatic variables in relation to the issue of global warming and climate change in general, and found some interesting results that tend to confirm the reality of global warming and its side-effects.

We tried to build simple models of temperature as a function of a few other parameters, ranging from simple regressions to a zero-dimensional model, and evaluated the relevance of these models and their conclusions using statistical tools. It appeared that although very simplistic, these models explain global warming much more satisfactorily than a mere statistical fluke, confirming that something is indeed happening to our climate.
In particular the likelihood ratio analysis showed that among the statistical approximations we considered, the best one is a polynomial of the second order, which proves non-linearity, i.e. acceleration, of the warming process.

Quantifying precisely the phenomenon is clearly out of the reach of such a simple study, however. But we can observe that the predictions of our zero-dimensional model are somewhat similar to those of the IPCC annual reports.

On top of these models, we analyzed several climatic phenomena that are often associated to the issue of global warming and climate change: seasonal shift, sea level rise, snowfall and precipitation. This gave us some insight on the consequences of global warming throughout the globe, though these phenomena cannot help us conclude on the reality of global warming by themselves. These results can thus be seen as an addendum to the main results.

Among the difficulties we encountered in the context of this project were the issues of data correctness and statistical significance.
Applying mathematical models and (un)supervised learning methods to a dataset is quite easy from a programming point of view, but as we already mentioned in the first part of this report, climatological data can suffer from several problems, such as dataset heterogeneity, auto-correlation and so on.
Moreover, we must always be cautious about the statistical significance of the results, as the dynamics of the climate are extremely complex and unpredictable in essence.

The problematic could have been addressed in more details with the extensive use of paleoclimatic data, and by considering other climatic variables that are more directly influenced by human activity (like deforestation and water consumption) to quantify the degree of human responsibility in the phenomenon.
But all in all, we can quite safely conclude that global warming is indeed a reality, with palpable consequences on several climatic phenomena.

References

[1] Guo H.-D., Zhang L. & Zhu L.-W., *Earth observation big data for climate change research.* Elsevier B.V., September 2015.

[2] Faghmous, James H. & Kumar, Vipin, *A Big Data guide to understanding climate change: The case for theory-guided Data Science.* Mary Ann Liebert, September 2014.

[3] Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie, *An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3.* Geophys. Res., 116, October 2011.

[4] Russell S. Vose et al., *NOOA's merged land-ocean surface temperature analysis.* American Meteorological Society, November 2012.

[5] Fetterer, F., K. Knowles, W. Meier, M. Savoie, and A. K. Windnagel, *Sea Ice Index, Version 3.* Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center, 2017. [Accessed November 18, 2018]

[6] D. P. Dee, S. M. Uppala et al, *The ERA-Interim reanalysis: configuration and performance of the data assimilation system.* Quarterly Journal of the Royal Meteorological Society, 2011. [Accessed November, 2018]

[7] C. D. Keeling, S. C. Piper, R. B. Bacastow, M. Wahlen, T. P. Whorf, M. Heimann, and H. A. Meijer, *Exchanges of atmospheric CO2 and 13CO2 with the terrestrial biosphere and oceans from 1978 to 2000. I. Global aspects.* SIO Reference Series, No. 01-06, Scripps Institution of Oceanography, San Diego, 2001.

[8] C. D. Keeling, S. C. Piper, R. B. Bacastow, M. Wahlen, T. P. Whorf, M. Heimann, and H. A. Meijer, *Atmospheric CO2 and 13CO2 exchange with the terrestrial biosphere and oceans from 1978 to 2000: observations and carbon cycle implications, pages 83-113.* in "A History of Atmospheric CO2 and its effects on Plants, Animals, and Ecosystems", editors, Ehleringer, J.R., T. E. Cerling, M. D. Dearing,Springer Verlag, New York, 2005.

[9] *IPCC global surface warming projections have been accurate*, Skeptical Science website. `https://www.skepticalscience.com/ipcc-global-warming-projections.htm?fbclid=IwAR3EMIb5UATC_pXWffflntckA9koaVMaI7WsXdgG8F8vjU6ep08QGYlGQVU` Consulted December 13th, 2018.

[10] Etheridge, D.M., et al. *Law Dome Ice Core 2000-Year CO2, CH4, and N2O Data.*, GBP PAGES/World Data Center for Paleoclimatology Data Contribution Series 2010-070. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA, 2011.

[11] Liu, Yonggang, and Robert H. Weisberg. *A review of self-organizing map applications in meteorology and oceanography*, Self Organizing Maps-Applications and Novel Algorithm Design. IntechOpen, 2011.

[12] Ablain, Michaël, et al. *Uncertainty in Satellite estimate of Global Mean Sea Level changes, trend and acceleration*, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-10 Manuscript under review for journal Earth Syst. Sci. Data, 2019.