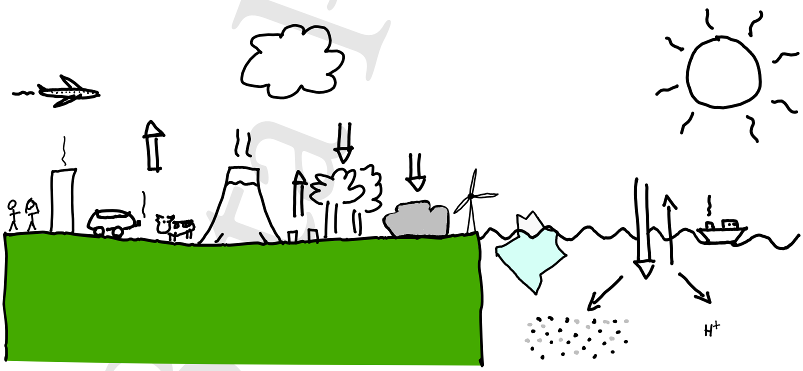
Back-of-the-envelope climate change



Marco Cosentino Lagomarsino

Back-of-the-envelope climate change

Marco Cosentino Lagomarsino

Beta preprint edition, January 2023.

DOI 10.5281/zenodo.8190386

Illustrations by MCL.

Climate-change fact checking and text revision by Maurizio Maugeri.

I wish to thank Barbara Vischioni and Pietro Cicuta for useful feedback and copy-editing help.

To Barbara.

Tu sei rimasta una regina,
L'unica che nel male e nel malissimo mi è rimasta vicina.
(Emis Killa)

In memory of Daniele Colombaroli

Contents

Preface

What is this book about, who is writing it, and who is it for? I, the author, am not a specialist of climate change. So why this book? We live in the era of data, but this is also a time where data are often used instrumentally to impress and to bend reality in order to support *a priori* positions, and not to interrogate reality. Often even genuinely honest attempts to present data end up overwhelming us, because we lack the proper reference points. On one hand, we'd like to be able to "call bullshit" on this body of political and scientific pressure (and there's a very nice book about this, entitled "Calling Bullshit: The Art of Skepticism in a Data-Driven World", by Carl Bergstrom and Jevin West). On the other hand, if we became very good at this, and we spent our lives calling bullshit on each other, we would not achieve much. There is also a constructive, creative part of interpreting the data, and attempting to build a coherent picture from different sources and measurements. Trying to get a grasp on this positive part is what drove me to attempt writing this book. Often, this constructive part is more difficult than the negative one. Even if I am no climate scientist, I am a scientist, that sort of scientist that uses plots and mathematical models to understand data. My background is in theoretical physics, but the reader should not be mislead by this. The job of a physicist is not only to know and apply a certain set of mathematical laws, but also to extract minimal "phenomenological" and operative descriptions and predictions ("models") from data. The same approach is valuable in many contexts ranging from other scientific disciplines to consulting or planning and logistics, and “data science”. So this book is a project on data literacy, trying to convey simple techniques for “reading” the data and constructing positive interpretations, while being aware of possible caveats, biases and limitations. As far as I know there isn’t much work of this kind around, which was my main motivation.

If you are worrying about the climate science side of this book, you should know that so did I at some point. Many climate scientists are physicists like me, but I did not specialize or work in their area, and I did not acquire the same specific knowledge. In order to reassure the reader (and me) about this, my colleague Prof. Maurizio Maugeri at the University of Milan[1] kindly agreed to revise the manuscript regarding the climate-science aspects. In any case, a disclaimer for the reader is that this book should not be seen as a popular science book on climate change: it is more about data literacy than anything else. It does not aim to be exhaustive, and

[1] Maurizio Maugeri is a Full Professor at the Department of Environmental Science and Policy.

it does not aim (or claim) to be exact. Note also that a full understanding of climate change is an enormous challenge even for the specialized scientists who dedicate their lives to this goal. Some of the estimates and considerations you will read could be based on debatable assumptions, simplifications or approximations, or biased by my own limited knowledge, and have to be taken as they are: efforts to extract meaning from data, with the main purpose of stimulating informed critical thinking and making climate change data something we can all begin to grasp by ourselves.

Indeed, my reasons for using climate data were that I thought this could be a very motivating example for a reader and also that I wanted to look into these data myself. Greenhouse emissions and climate change are big in the media, where gigatonnes of greenhouse gases, mass extinctions, impending climate disasters, are thrown at us by unclear numbers, which we need to connect to our everyday experience of life. This is not to say that things are not as bad as they seem. They are very bad, as far as I can judge. But the point is that, regarding climate data as well as other kinds of data, we should be clear-minded about our reactions. My feeling is that there is a huge divide between the narratives about what is happening produced by science, which are difficult to grasp, and typically filtered by the oversimplified translations produced by media, and distorted by politics, and what we experience every day around us. This divide is a big problem. We need a better intuition on the numbers that are involved, and we need to be empowered with the ability to produce rough common-sense estimates just based on the basic facts, as we do with other aspects of our lives (but typically don't indulge and attempt on matters at this scale).

If we had these skills, we would perhaps improve our understanding of the changing world around us. This could also enhance our perception of how we should contain the negative changes and how we should adapt. For example, one key question in the mind of a reader may be whether the climate change we observe is really anthropogenic. Many popular works implicitly say "but the recent change coincides with industrialization and this cannot be a coincidence". However, one can argue that this is not completely convincing and crave for more arguments to support a cause-effect link between atmospheric greenhouse gases of anthropogenic origin and climate. There are tools to argue in support of causality, and there are ways to gather different lines of correlative evidence in support of the idea that this cannot be a coincindence[2].

Hence, this book is an attempt to give some useful tools for reading the data constructively. The book should be readable by anyone who would like to be empowered by a data-driven approach to reflect on world's issues, and in particular on climate change. The mathematical tools used in the book are elementary (middle-school level), but

[2]As we will learn in chapter III, the concerns about climate change don't just come from the observations of warming, but they are above all linked to the scientific knowledge of the radiative balance of our planet and the role that greenhouse gases play in it.

some of the reasoning and analysis techniques parallel those used professionally within quantitative sciences. As a consequence, it is possible that, although the math is elementary, some parts appear unintuitive to some readers who do not have a college or high-school level technical education[3]. In my experience, younger people tend to be the most aware (and arguable the most affected) by our "era of data". Hence, a narrower target for this book could be early college and late high-school students and teachers, particularly in science and engineering. I use some of the material from the book myself as a teaching tool for a 3rd-year college course I teach to physics students, with hands-on projects on the data.

The questions that I aimed to answer concern the simplest things we can establish about climate change *directly from plots or data* that are available surfing the web. The chapters attempt to present the data gradually, and interpret them by straightforward reasoning, and by using the "Fermi approach"[4] of estimates relying on minimal knowledge (or no knowledge at all) of the physics and chemistry that are relevant to climate science. Breaking each question

[3]Most of these parts can be safely skipped.

[4]Enrico Fermi, the famous Italian-American Nobel-prize physicist, was a master of this art. He could always surprise his peers churning out solutions to difficult problems with a set of simple quick-and-approximate techniques. Fermi also had an extraordinary insight into which problems were important and which ones were not . I try to apply some of these techniques professionally (in the interdisciplinary area between statistical physics and biology), and over my career I have been striving to imitate Fermi and others in the insightful art of providing simple descriptions of complex problems, which is what makes physics so beautiful in my eyes.

into smaller sub-questions, and focusing on global trends, I tried to achieve a feeling of the "order of magnitude" of the quantities and processes involved in climate change, and to use this knowledge to draw simple conclusions based on straightforward reasoning.

To improve readability, instead of showing data directly through plots, I decided to use hand-drawn and annotated simplifications, where I could highlight what I thought could be the most important take-home features. The end of each chapter contains a section entitled "Data science take-home messages", which recapitulates the essential tools introduced and used in that chapter. Additionally, all the relevant sources and procedures are described in the "Sources" sections at the end of each chapter. If you are curious about the real plots, the best way is to look at these sources. However, if you are in a hurry, I made a "Sketch2Plot" folder in my GitHub page, where you can take a look at those plots[5]. The sources include scientific publications as well as non-academic sources, such as Wikipedia or data projects (mainly the "Our World in Data" project, by Max Roser and others), where data sets and ready-made plots are more easily accessible to a wide audience[6]). Finally, we have said that the book does not require a prior knowledge of physics. However, if you are a physicist, it should be an easy read, because you will find that it is based on most of the cultural practices of this discipline. In one case, I have also added

[5] Here's a direct link.

[6] In any case, I cite also the original academic sources in the Sources sections.

a short "Physics track" section, because knowing some elementary physics would add precision or substance to some of the arguments.

SOURCES. The book "Enrico Fermi. The last man who knew everything.", by David Schwartz (Basic Books, 2017) is an interesting biography of the famous physicist. The book "Calling Bullshit: The Art of Skepticism in a Data-Driven World", by Carl Bergstrom and Jevin West, is edited by Penguin Random House, and now at its second edition (2021). The web site `https://www.callingbullshit.org/` describes the initiative, which includes a university course and a set of videos and tools.

Chapter I

Global temperature is rising

Is global temperature increasing? And how much? If you read newspapers and you are trying to be a conscious citizen, you are probably willing to accept that global temperature has been visibly increasing for some decades now. But you are equally likely not to have actually looked at the data. And if you were to look at these data, you would have to make sure that this increasing trend is meaningful. Admitted that we can establish that there is a trend, a further question is whether this trend is not a fortuitous coincidence, a "fluctuation" as in, "yes, the temperature has been rising for the past 50 years, but it's equally likely to decrease for the next 50". Therefore, at some point, I decided to download the data and look at this record myself, in order to dive directly into these questions and return to you with some representation of the data that we could look at together, as a useful interpretation exercise.

If we think about temperature, we immediately think of a thermometer. This is the instrument by which we measure temperatures in our daily life, but it was not always available to us. Thermometers were known to the ancient Greeks, but a systematic approach to their calibration was achieved only across the 17th and 18th centuries. Hence, we have a reliable temperature record from measurements with thermometers only starting from the mid 19th century. Starting from the 1950s, the record becomes precise and systematic. The period covered by this record gives us very precise and reliable data (although it starts a bit after the beginning of the industrial revolution). If our goal is to check if temperatures are rising, we want to determine if these data show a significant change over this period [1].

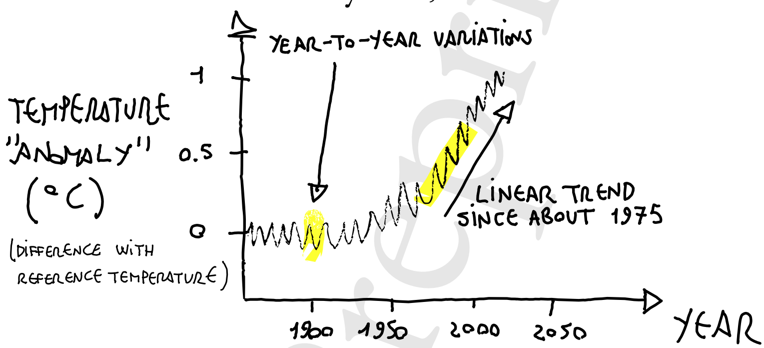
The other problem that we need to face when trying to assess if temperatures are increasing is that temperatures vary a lot. They vary between day and night, from place to place, and they vary across seasons. Many of these variations follow specific trends, mostly due to the illumination from

[1]Specifically, we should be able to see a change corresponding to when humans burned enough coal and other hydrocarbons to manifestly affect the atmospheric pools of grenhouse gas, beyond our uncertainty in the measurement, but this will be the subject of the next two chapters.

the sun. We all know that July and August are hotter (by many degrees) than December and January in Italy, and vice versa in Brazil; and, quite obviously, we also know that days are warmer than nights. Other changes are more subtle and we might perceive them as random. A detailed understanding of all these oscillations and changes is way beyond our scopes. For us, these changes are a disturbance, as we'd like to capture overall temperature changes beyond any trend with latitude, ocean currents, daily and seasonal oscillations and other fluctuations.

A simple approach is to average temperatures globally, and also average temporally over a year. This process should remove the effect of all the daily and seasonal oscillations, and should "iron out" all the effects that we wish to disregard in order to capture our trend. Hence, it's a good idea to consider these global averages (and climate scientists often do this). This book will only consider global trends, also when we will move to the other quantities that are relevant for climate change. Since I am no climate scientist, I started from processed data - trusting the work that has been performed by expert scientists to derive this information from raw data (and, I guess, trying to cross-check information whenever possible). As you can imagine, performing reliable estimates of these global averages may be a lot of work. For example, the spatial locations of data points are spread out, and measurements vary over time and location from 1850 to now, and so do the sparsity of the coverage of the planet by measurements and the precision of measured temperatures.

We would like to use these data to quantify how much global temperature has increased compared to pre-industrial times. Here's my sketch, after looking at these plots, of what the data for global temperature roughly look like in the last almost 200 years[2],



The first thing that hits the eye is perhaps that this plot does not report the actual temperatures, but a temperature anomaly, which is a temperature difference. In other words, it subtracts a reference value (here the global average of the data in the 1850-1900 period, but sometimes this can be the average over anotther reference period, such as 1991-2020, or another convenient reference) to obtain what is called the

[2]Note that all the sketches in this book are rough representations, meant to convey what I judged as the most important aspects, to show how I would represent the data if I had to draw them on the back of an envelope. Obviously, they likely contain some deviation from the original. You are encouraged to look at the real data to get a more precise idea. There are several reasons why I did not use the real data. First, I wanted to integrate the salient trends from different sources. Second, I wanted to try and simplify the essential information, making it more accessible. Third, I wanted to focus on the interpretation part and not on all the tecniques associated to producing plots. Fourth, real plots sometimes look very technical and I did not want to discourage the reader.

"temperature anomaly". This expedient is necessary to average effectively the data across space. Weather stations on land are found at different elevations, they may take temperatures at different times of the day, and stations in different countries may calculate average monthly temperatures using different formulas. To avoid biases that could result from these differences, monthly average temperatures are reduced to anomalies. Empirically, anomalies tend to be constant over hundreds of kilometers, while absolute temperatures can vary over much shorter length scales (this is why we know the absolute temperature of earth with much higher uncertainty than its average temperature anomaly). Averaging the anomaly avoids possible offsets due to calibration and eases the comparison of data from different land and sea areas, as well as time spans[3].

Let's look a bit more closely into how these data are averaged. The procedures differ across projects, and the global average data mostly come from weather stations. For example, HadCRUT5 is a global temperature dataset from the Climatic Research Unit in the UK (see the Sources section at the end of this chapter), and provides gridded temperature anomalies across the world as well as global averages. It starts from monthly-mean temperatures data from over 10000 stations. The number of stations was small during the 1850s, but increases to over 2000 by 1900 and to more

[3]Although it can also be a source of confusion, since if you look at this plot from different sources, you will see different values in the y axis, depending on whether the anomaly has been computed using the average from the period, say, 1991-2020, or 1951-1980, or another period of choice.

than 5000 after 1951. The stations are not evenly distributed across the earth's surface. There are many in North America for example, and much fewer in North Africa. Over the ocean areas, the data mostly consist of sea-surface measurements, which need to be averaged together with the air temperature available from land measurements. Combining these data implies assuming the anomalies of sea surface temperature agree with those of marine air temperature. This assumption is based on previous studies comparing large-area averages of both quantities. Today, the spatial averages are updated every month and performed dividing the earth surface in cells using a 5^o latitude by 5^o longitude grid (different projects use different grids and sometimes the same project has data for different grid choices). They use anomalies referenced to the 1961-1990 period, over which they have the best coverage. For each cell, the averaging procedure weighs the data based on a "statistical model", a mathematical model that treats the data as they were generated randomly with some prescriptions[4]. In the model, nearby locations are expected to vary in a similar way, and distant locations are expected to vary more weakly. This model also keeps into account uncertainties in the measurements. For each hemisphere they consider a weighted average, where the weights are the cosines of the central latitudes of each grid cell (this gives the same weight to each fixed lat-

[4] In other words, a statistical model assumes that the data are random variables and formulates a set of assumptions about the "probability distribution" that generated the observed data. We will define a probability distribution more precisely in chapter VI.

itude "ring" around the planet) taking all the non-missing grid-cell anomalies. Finally, they average the northern and southern hemisphere data.

The goal of such elaborate procedures is to reduce "biases", which are systematic errors in the data due to different sources of error, including time and location of observation, instrumentation type and changes in measurement practice over time, the fact that urban areas are hotter, use of sea surface temperatures (probably the strongest potential source of bias, which also depends on the historical changes in the use of engine-room sensors or buoys) and variations in the number of data points ("sampling") available in a grid cell. So the procedure is quite complex, and there are many choices. How important is the procedure, if different projects use different procedures? On one hand, the experience of many researchers and projects has identified a series of crucial steps to improve the accuracy and avoid important biases, so it is very important. On the other hand, there is theoretical knowledge that tells us that as long as we average enough data (and we manage to avoid systematic biases), we should get close to estimating the underlying global average. The law of large numbers, in probability and statistics, is a theorem that states that as a sample size (in our case, the number of sampled stations) grows, its mean gets closer to the underlying expected value, which we could measure if we had complete information on the data. Another theorem, the central limit theorem, also tells us how big a sample needs to be to represent a population distribution with prescribed accuracy. Therefore, one could also compare methods by verifying how the means change with the number of sampled weather stations (but we do not attempt it here).

Since the plot can be produced with data coming from different projects, one way to be confident that these data are trustworthy is that several independent studies, made by different sets of scientists, doing things differently in terms of the choices described above, and also using different (though overlapping) data, arrive to very similar results for the time series of yearly global temperatures, with discrepancies across data sets that are typically about one tenth of a degree Celsius for a given year. We can take this discrepancy as an indication of the uncertainty in the data. There are several studies that quantify these uncertainties in more direct ways, and give similar figures. Naturally, the error decrease with time of measurement. For example, HadCRUT5 gives "confidence intervals", a very stringent evaluation of the error defined as the interval where we are 95% sure that the true average would fall into (in a scenario where history is repeated many times), decreasing from 0.9 ^{o}C in the 1850s to 0.16-0.2 ^{o}C today. In order to estimate these uncertainties they generate 200 random realizations of the data from all individual weather stations using the statistical model involved in the averaging process, where data are drawn keeping into account the known ranges of uncertainty that emerge from the gridding process.

Back to our global trend of temperature anomaly, since we want to quantify the increase compared to pre-industrial time, an important question is that one should define what exactly is meant by "pre-industrial". By convention, in reports from the IPCC (Intergovernmental Panel on Climate Change), warming is expressed relative to the global mean temperature of the period 1850-1900, taken an approximation of pre-industrial temperatures. So the figures discussed in the news typically refer to this definition (but as I said the plots may not). I tried to make the plot that I sketched comply to this definition. Today (in 2022), global temperatures have risen by about 1.1^{o}C above the 1850-1900 reference.

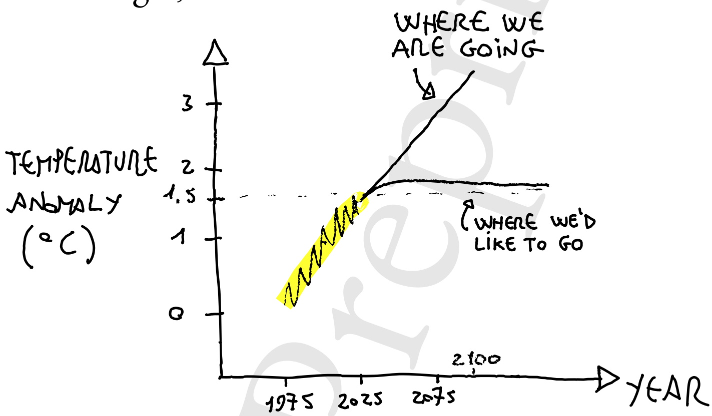
Looking more closely at the plot sketched above, the second thing one could notice is that, despite of the averaging over all Earth locations and all times over the year, there are interesting "noisy" year-to-year variations. Those are the zigzags in the plot; the ones in my drawing are not the real ones (because they were way too difficult to draw precisely for me), but they represent more or less the real wiggle. What is important is to look at how big these zigzags are compared to the trend of the plot. Understanding the source of these changes (whose extent is above the uncertainty that we have defined above) is difficult (and beyond our scopes), but as we will see in a minute, these variations can be our "ruler". Looking at the trend beyond these year-to-year variations, the temperature increase seems to be quite sluggish or absent between 1850 and 1900 (so this is a good period against which to define the anomaly), then it picks up increasingly faster over the next 100 years, and finally it becomes really marked after 1975. After that year, the roughly linear increase hits the eye, with a slope of almost 0.5^{o}C change over 25 years, and with a total increase of almost 1^{o}C between 1975 and today. Surely this change is well beyond our uncertainty in the data, which we defined as the discrepancy of estimated global temperatures across different studies, hence we are quite sure that we are not getting this trend because of measurement errors.

So we have established that there is a rising trend in the global temperature beyond our measurement errors, during the industrial era[5]. But this is not enough. In order to claim that temperature has been changing significantly over this period, we need a meaningful term of comparison for the "natural variation" of our variable, beyond experimental error. For example, we could ask whether the global average yearly temperature has varied more over the years than local temperatures vary from day to day, or, perhaps more fairly, than the global value for, say, the northern hemisphere varies across seasons. Since we are familiar with daily and seasonal temperature variations, we know that the answer to both of these questions would clearly be no, but it is also easy to argue that this comparison is not fair, because these local or faster changes cancel out when we consider global averages. Then how can we conclude that the temperature increase that we see in the past 45 years or so is considerable?

[5] Note that in this chapter we are just asking whether there is a trend during the industrial era. This is consistent with the interpretation that this trend is caused by human activity, but it is not a proof. Firmly establishing causality is very difficult, and in this case relies on our direct experimental knowledge of the physics of the greenhouse effect, as we will discuss in chapter III.

One way is to compare them with the year-to-year variations, (the "fluctuations") of the same global average. As we noticed before, these are about one tenth of a degree, and the trend of the increase we are seeing in the data since 1975 is way above this amount. Hence, just looking at global temperature data we conclude that we cannot dismiss the 1 ^{o}C increase between 1975 and now as an irrelevant or purely coincidental change, based on what we know on temperature variations from the past two-hundred years or so (as I could have guessed, all those newspaper headlines had a meaning!). So the increase in global temperatures is small compared to the temperature changes we experience between day and night, but very big compared to its natural year-to-year variation, which is a better ruler, because it considers that local and fast variations cancel out in our global averages. We'll get a confirmation of this point looking at more ancient (indirectly) recorded temperatures.

BEFORE WE DO THAT, let's take a look at what we would predict if we extend this trend to future temperatures. Making meaningful predictions for the future is difficult, but the pressure is much less if we clarify our assumptions. There is a very simple assumption that we can make, which is that things in the near future will continue to behave as they did in the past years (which is a good approximation of the scenario where we will do nothing about climate change)[6]. In this case, since the trend has been increasing as a straight line for the past forty-something years, it seems reasonable to assume that it will keep doing the same. We can then keep drawing the almost straight line we see from 1975 to now, extending the trend to future years. If I assume this "linear extrapolation", the sketch below roughly depicts trend that I get,



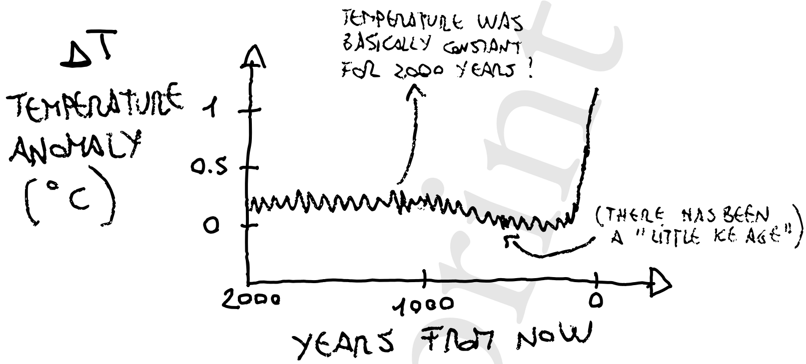
The highlighted part is basically the current trend. To show how radical a change of trend we need, compared to the current state of things, in order to reach our internationally set climate change targets, I have also drawn an imaginary trend line that reaches this target. Clearly, where we want to go[7] is radically different to where we are currently going. Looking at the literature, it seems that the actual climate models are more pessimistic than my linear extrapo-

[6]There are, of course, better approximations, and there are connections between these global averages and extreme events. Please hold on until subsequent chapters for more information on these points.

[7]At least referred to those people who agree that increasing global temperatures are a concern, the rest of this book will provide more arguments for why they should be.

lation, as they tend to predict that with no intervention, global temperature would increase even faster (the current estimates say that the peak increase, reached before 2100, could easily be almost 3 degrees Celsius above pre-industrial levels). But what these models really want to establish is how much we need to reduce our greenhouse gas emissions (see chapter II and III) in order to stay within the targets of 1.5 or 2 degrees Celsius above pre-industrial levels that we are trying to set ourselves. We will come back to this problem later, and I'll try to address it with some back-of-the-envelope arguments. Why we set these targets is also something that we would like to understand more directly, as we go through the data. For now we can say that the current consensus among the scientists that study these things is that an increase of 1.5 to 2 degrees Celsius above pre-industrial levels could already be dangerous, but given the situation we are in we cannot really hope to do any better.

BEFORE THERMOMETERS WERE INVENTED, temperature changes left a set of different "footprints" that we can measure. These footprints give rise to many possible indirect measurements of temperatures that evaluate diverse sources of information, including the chemical composition of ice cores, the width of tree rings, ocean sediments, fossil pollen and corals. These data give a very instructive perspective on climate change, so at some point I reached the conclusion that I should look myself into these data as well. Here I try to hand-draw global temperature data from the past 2000 years that integrate multiple sources, and that are calibrated with high-precision contemporary measurements.



The plot sketched above is very definite in telling us that the temperature changes we are seeing today are exceptional, and other than that global temperature stayed remarkably constant since people like Julius Cesar and Cleopatra strolled around the Mediterranean[8]. What is curious and remarkable is that when the industrial revolution started, and humans began pumping tonnes of CO_2 in the atmosphere by the millions, temperature was actually slightly decreasing. This "Little Ice Age" began in the 14th century and the coldest period ends around 1850 (possibly thanks to our intervention). Glaciers were expanding all around the planet, and Londoners would skate on the Thames during Winter. The reasons for this climate change are not known precisely, but probably lay in volcanic erup-

[8] There is a "Medieval Warm Period" between 900 and 1300 AD which I have not emphasized in my sketch, and is not clearly visible in the original plot. The warming is well documented across the North Atlantic region, and the current agreement is that peak warmth occurred at different times for different regions, hence the event was not uniform across the globe.

tions and/or reduced activity of the Sun. The Little Ice Age is another important ruler. Interestingly, we know that it had severe consequences in different areas of the planet. There are records of bad harvests and famines in northern and central Europe, Japan and the Mississippi valley, and of a reduction in North Atlantic cod fishing[9]. Yet, the temperature change observed during the Little Ice Age is quite small compared to what we witnessed in the last 50 years. These considerations, grounded on data, confirmed again my growing opinion that it is a good idea today to pay attention to climate changes.

ICE CORES are a precise way to measure temperatures. They are part of the data set I just discussed, but they can go much further back in time, so by looking at these data we can venture into the more distant past of our global temperature. The way scientists can obtain these temperature estimates relies on the quantification of the ratio between heavier and lighter hydrogen and oxygen isotopes that form the water molecules in the ice. An isotope is a version of an atom that is identical in terms of chemical reactions, but has a nucleus with a different weight. Intuitively, the principle behind this measurements is that heavier isotopes precipitate faster and need more energy to evaporate. Hence, the relative amounts of heavier and lighter isotopes. This justifies a correlation[10] of isotope ratios with temperature. The

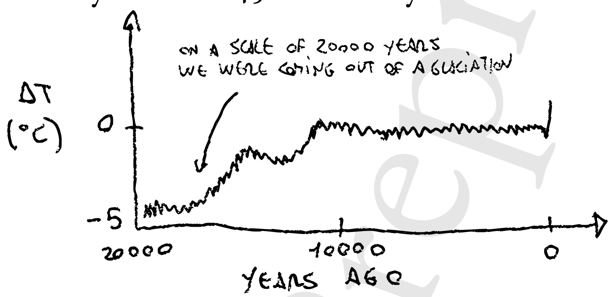
[9] I did not check myself the significance of these findings. Let us take them as "anecdotal evidence" that the global temperature change during the Little Ice Age had an impact. The question, however, may be subject to debate.

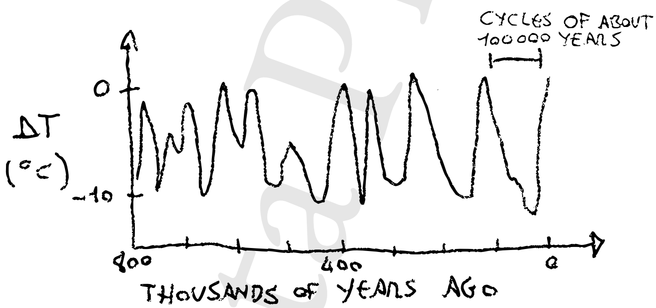
[10] We will define more rigorously the concept of correlation in chapter III.

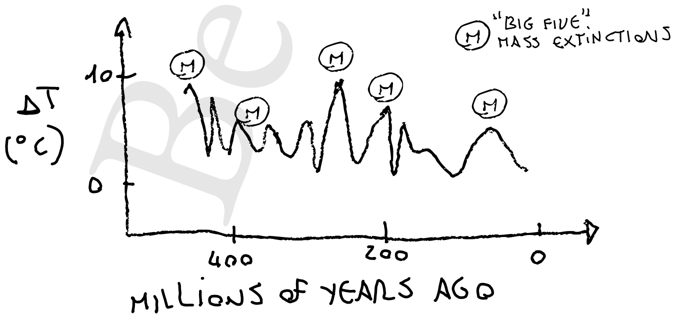
question arises of whether this method is precise. As a newbie, I initially thought that this method could be sloppy, but it is not. In order to check, I looked a study where they calibrate the measurement using yearly seasonal temperature variation (the exact 1997 study by van Ommen and Morgan where I found these data is cited in the "Sources" section at the end of this chapter). This study contains a "calibration" plot, where the x axis reports a range of measured temperature (at a given time in the year), and the y axis reports the corresponding isotope ratios. The data cluster very tightly on a straight line, meaning that the isotope ratio variation for a given temperature is very small. A 1 percent increase of oxygen isotope ratio corresponds to a change of about 1.8 degrees Celsius. From the plot, the uncertainty appears to be between 1 and 2 degrees. However, the instrumental uncertainty (evaluating the isotope ratio) is very small, and most of the errors are due to relating the ice core layer to a specific period. Hence, measurements could be much more precise on time scales of many years. As I will discuss, ice cores also give us information on the greenhouse gas content of the Earth's atmosphere in the past, just because we can find those gases dissolved in the frozen water.

Amazingly, we can obtain ice core information on temperatures from up to 800000 years ago. Beyond that date it is still possible to have access to temperatures from geological sources, for example looking at the chemical composition of sediments. These records are obviously less precise in terms of dating, but they can still swiftly measure temperatures back to hundreds of millions of years ago, which is tremendously impressive to me. A big warning is that, as we will see, in general data look different on different time scales, just because the time resolution is different. In practice, we can define the time resolution of a time series (or time scale) as the time interval between two subsequent data points, for example two consecutive layers in an ice core. On some time scale, a time series (for example, temperature versus time) follows some trend or fluctuation, or varies with some periodicity, but all this may cancel out if we go to a larger time scale. For example, if we sample the mean temperature in Europe every month, we see that it oscillates because of seasons, but if we take a time point every January each year, or we average the temperature of all the months each year, we will lose this oscillation. Equally, if every first Tuesday of each month the temperature would halve, but just for one day, we would not see it in the monthly records if we average or we measure on any other day. We can think of a change in time scale as an average, where each time we go to a larger time scale the "fast" changes cancel out because of the averaging, and a clear trend might emerge (but obviously we will have fewer data points). In fact, ice core layers contain averages of isotope levels from snow deposits of many years, and it's very reasonable to think of the measurements as averages, or totals, from all those years. The same will happen if we compare recent data on a scale of one year to old data on scales of thousands to millions of years. At each change of scale, we are looking at averages, and the only truly fair comparisons can be made between data taken at the same time resolution.

WARPING BACK TO THESE DISTANT TIMES, with these warnings in mind, we can discover surprisingly dramatic changes in our planet's temperature. The following sketched plots take us on a breath-taking ride to three different scales of the Earth's past, 20000 years, 800000 years and 450 million years.







The authors of these plots[11] made efforts to put all temperatures on the same scale. Clearly this task is difficult and the results are subject to uncertainty, but we will take their calibration for good and assume that the temperatures that we see on the different plots are comparable. The first plot sketched above shows that before the little ice age and the very recent radical increase connected in the industrial age, temperature has been rather stable since about twelve thousands years ago[12]. Remarkably, about twelve thousands years ago is also when we humans began transitioning to farming, after living as hunter gatherers for almost two hundred thousand years. This transition was the key to a set of dramatic changes in our social structure and living standards that made it possible for human populations to start a phase of exponential growth, and eventually build the technological society we have now. Are you thinking that such a remarkably stable climate might have played a role in the dramatic turn in human history during this period? That's also what I thought when I saw this plot for the first time. And, of course, we are not the first, a scientific article by Feynman and Ruzmaikin formulated this hypothesis in 2007 (see Sources section).

[11] See the "Sources" section for details.

[12] Note that this statement simplifies the extent of natural climate swings in this period. In view of our goal of extracting the most important information from the data, and in view of the recent global temperature changes, we can consider it to be valid, but it should be seen as a strong approximation and of course we can very well imagine that there are a lot of relevant things that happened climate-wise during twelve thousands years.

The second plot shows that on a scale of hundreds of thousands of years global temperatures underwent dramatic oscillatory cycles, with changes that reached an amplitude of up to 10^{o}C, toward colder temperatures than today. These cycles and (which also change their periodicity going further in the past) were proposed to depend on changes in the Earth's orbit around the Sun, affecting the irradiated heat. They are called Milanković cycles, after the astronomer that first hypothesized their existence in the 1920s. While the link between Milanković cycles and ice ages is not debated today, there are some aspects that are still open.

Milanković believed that variation of the summer exposure in northern high latitudes had the strongest effect, and based on this he deduced a 41 ky (kilo years, 41000 years) cycle. This is observed in older data, from 1 to 3 million years ago (though not visible in my sketches). However, as we see from my sketched plots, the last 800 ky show a clear period of about 100 ky (which matches the eccentricity cycle). Various explanations for this discrepancy have been proposed, but this and other discrepancies indicate that the links between global climate and orbital variations of our planet are complex and not controlled directly by exposure (in the sense that many other factors may play a role). We will return to this question in chapter VI.

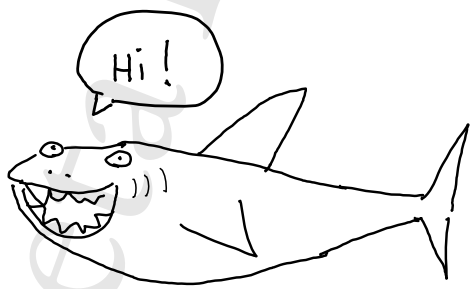
Humans (homo sapiens) have been around for two to three hundred thousand years, so, as a species, we have experienced at least two, maybe three cycles of these dramatic climate changes. This proves that our species can survive radical climate changes. But of course we achieved that feat as small sparse population of ape-like hunter gatherers. Sustaining a large, complex and costly technological civilization such as the one we have today through such changes appears to me like a daunting task. What happened to humans in the other "peak" periods when temperatures where reasonably warm? The most recent of these peaks corresponds more or less to the period when we moved out of Africa to migrate to Europe and Asia. Specifically, there is evidence that 120000 to 90000 years ago, when the climate in Africa may have become too hot, humans started to migrate to the Fertile Crescent and Arabia, although the primary human migration is dated around fifty thousand years ago, in a cold and dry period (suggesting that other local factors played a role in the migration).

The other important consideration that we can make looking at this second plot is that the temperatures at which we thrived as a species are hot for the Earth's standards, within this time range spanning hundreds of thousands of years. However, looking at the third plot, which spans millions of years, we get a completely different impression. Today's global climate appears much colder considering the averages on a scale of hundreds of millions of years in the past. Strikingly, five of the temperature peaks correspond to huge mass extinction events known as "the big five", which were first detected from dramatic reductions in the diversity of the fossil record of key sets of species. Dinosaurs lived between the last two mass extinctions, and a widely accepted theory[13] is that the last mass extinction, which wiped them away, was due to an asteroid impact that took place sixty-six million years ago in Yucatan, Mexico. Scientists estimated what happened to temperature "shortly" after this event by looking at oxygen isotope composition of fish debris. Clearly, for times so far away in the past, the time resolution that can be achieved is limited, but what they found is quite interesting. After a short (a few decades) artificial winter caused by dust blocking sunlight, it seems that temperatures quickly rose by about 5^{o}C, and this situation lasted for about a hundred thousand years. This is believed to have occurred due to increased CO_2 emissions because the asteroid hit rocks that are rich in carbon, directly releasing it in the atmosphere, as well as triggering vast wildfires. This situation of fast CO_2 increase and fast temperature increase is probably the closest "natural experiment" that we can use as a term of comparison to the situation of the last two hundred years, where CO_2 emissions and temperatures are rising very rapidly. In most other occasions in the historical record these changes occurred very slowly. The following two chapters will deal with CO_2 emissions and their link with temperature changes.

Going back to the wider picture of million-year temperature changes, the visible correlation between temperature

[13] This hypothesis encountered criticism. A leading critic is Gerta Keller, who proposes that volcanism is a likely cause of a gradual extinction. We know about the asteroid impact and about many of its consequences, but establishing causality with a very long-term process of mass extinction is very difficult. More on causality in chapter III.

peaks and extinction events (which does not only involve the big five mass extinctions, but also smaller detectable mass extinction events) at these time scales triggers the idea that global warming is not good for biodiversity, which we will discuss a bit more later on. Curiously, sharks survived all five of these mass extinctions events. This diverse and adaptable species can live from deep oceans to shallow seas and even rivers, and they eat a wide variety of food, from plankton to fish, to mammals like seals. Consequently (if that is of any comfort), we can expect that sharks as a group are likely to survive future oceans changes, including the ones that are triggered by humans. Unfortunately, all this Darwinian fitness did not make sharks a more intelligent or sympathetic species (at least from our standpoint). Sharks remain relatively not fun to hang out with.



DATA SCIENCE TAKE-HOME MESSAGES. This chapter is centered on the concept of "time scale". In order to compare two quantities we need a "scale". You probably remember from high school that it is only possible to compare things that have the same unit of measure (we cannot compare 200 years with 400 meters), but even when we correctly compare things with the same units (in our case, temperature changes), we need a "scale". The scale can be different depending on the question we ask. We have used different terms of comparison: we used a comparison based on the discrepancy between data sets coming from different projects in order to get an estimate of uncertainty in the data, but all the rest was based on comparisons between typical changes on different time scales. For example, looking at our data we have used (symmetric) year-to-year variations to define a time scale and an amplitude for *fluctuations*, which we have distinguished from a *trend* followed by the data in the past 200 years. In data science, the analysis of quantities that vary in time is called time-series analysis. A time series is a sequence of data points collected over time intervals. We have seen that time-series data can report changes over different time scales, milliseconds, days, years, millions of years. And we also noticed that we can understand a change (a reduction) of time resolution as an average. We can say that time series data represent a mixture of variation patterns at different time scales, and we have analyzed the variation of the global temperature time series across different time scales. We found that, for global temperature, different patterns manifest at distinct time scales, as it is typical in many complex systems. For this reason, we cannot compare too stringently time series that "live" on different time scales. If we say that 200 million years ago the temperature was 5 degrees Celsius higher than in 2020, this statement is inaccurate, because we are comparing an average over many thousands of years with an average over a single year. Equally, if we want to compare the temperature change of the past 100 years with the ancient record, we can only do it with past steps of a 100 years (and in data sets that have sufficient resolution to sample data avery 100 years). Incidentally, doing this with a data set that has data every 100 years for the past million years, shows that the increasing trend of about 0.01 ^{o}C/y, is *only* found for the past 100 years. Additionally, we have learned that an average can be seen as an approximation of an underlying typical behavior, and the more data we have the better the approximation (in probability theory these statements translate into the law of large numbers, and the central limit theorem). this holds unless our data have systematic errors: in such case the systematic errors will add up and skew our averages with respect to the underlying value. Finally, we have learned the concept of linear extrapolation, if we isolate a region where our data look like a straight line, we can formulate a bare-bone simple testable prediction by assuming that this trend will continue for some time.

SOURCES. The global average temperature datasets 1850-2020 from NASA, NOAA, Berkeley Earth, and meteorological offices of the U.K. and Japan, as well as the temperature record of the last 2,000 years, which comes from the PAGES 2k consortium (Nat. Geosci. 12, 643-649, 2019) were downloaded from Wikipedia at the page "Global temperature record" (link), or directly from the sources: NASA GISS HadCRUT5, from the Met Office Hadley Centre (the web page at this link also provides useful information on how the data are collected and averaged, and how their uncertainty is evaluated), NCDC NOAA National Centers for Environmental Information, National Oceanic and

Atmospheric Administration, Japan Meteorological Office, Berkeley Earth Database. Detailed scientific information on the precise estimates of global temperature uncertainties and the establishment of a trend beyond incertainties can be found on the scientific article by Folland et al. Geophys Res Lett 28, 13, 2001. This study used climate models to correct for biases due to the use of sea-surface temperature, and concluded that the change in global temperature between 1861 and 2000 was 0.61 degrees Celsius, with an uncertainty of 0.16 degrees. The uncertainties quoted in the text come from the most recent HadCRUT5 study, Morice et al. JGR Atmospheres 126, 3, 2021. A wide-audence article by Alan Buis describing the correction procedures and why they are necessary can be found on the ASK NASA website at this link. Information on climate models can be found on the article by Jeff Tollefson on Nature 580, 443-445 (2020) The 800 ky Ice core data was downloaded from the NOAA database (`https://www.ncdc.noaa.gov`). It is the EPICA Dome C Ice Core 800 kYr Deuterium Data and Temperature Estimates, and the original publication is Jouzel and coworkers, Science, 317, 5839, 793-797 (2007). The calibration plot for the isotope ratio discussed in the text comes from Figure 3 in the study van Ommen and Morgan, J. Geophys. Res., 102(D8), 9351-9357 (1997). The million-year time scale ice core plot is hand drawn from Song and coworkers Nat Commun 12, 4694 (2021). The hypothesis on a role of climate stability for the development of agricultural societies originally comes from J. Feynman and A. Ruzmaikin, Climatic Change 84, 295-311 (2007). The climate context of the out-of-Africa human migration is investigated by Jessica Tierney and coworkers Geology 45 (11): 1023-1026 (2017). The information on the consequences of the Yucatan meteorite impact comes from MacLeod and coworkers, Science 360 (6396) 1467-1469 (2018), and the cover article by Cristophe Lécuyer on the same issue. On the debate around the impact hypothesis for the Cretaceous-Paleogene mass exctinction, see Schulte and coworkers Science 327, 5970, 1214-1218 (2010) and Keller and Coworkers Science 328, 5981, 974-975 (see also Chapter 3 on the intricacies of establishing causal links from data.

Chapter II

Atmospheric greenhouse gases are rising

A greenhouse gas is a gas that is able to absorb heat from the Earth's surface and reflect it back to the Earth's surface. The news tell us every day that release of carbon in the atmosphere (our "carbon emissions") is the main contributor to temperature change, and naturally I needed to look at the data to obtain some back-of-the envelope estimate of how much gas we need to generate a certain amount of "atmospheric heat". However, before entering that problem, it was necessary for me to get an idea of which amounts we are talking about when we deal with global greenhouse gases; for carbon dioxide, but also for the other relevant greenhouse gases. The present chapter is an account of what I found.

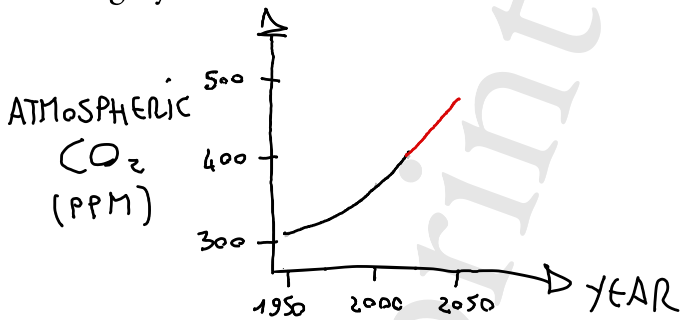
Water vapor is a major greenhouse gas, but it does not stay around very long, because increasing amounts of water vapor lead to condensation to form liquid water (clouds, rain). Because of condensation, temperature change from water vapor is mitigated, and, as it turns out, we do not need to worry too much about water vapor for climate change. Besides, water vapor emissions are largely unrelated to human activity. However, in presence of global warming, water vapor plays an important indirect role, since warmer air holds more vapor and condenses less, hence, it absorbs more heat, triggering a self-reinforcing phenomenon, often called "positive feedback loop", towards increasing temperature. Positive feedback can greatly enhance the effect of a small perturbation and, in brief, makes water vapor enhance the greenhouse effect[1]. Incidentally, this kind of positive feedback can be very dangerous because it also threatens the stability of a system. I will try to address it further in a later stage (Chapter VI).

In any case, the greenhouse gases that we need to worry about the most are the ones that do not condense (and hence can stay around very long), and that are linked to industrial-age human activities. Carbon dioxide (CO_2 in chemical djargon, we will use both terms hereon) is the first of the list. It is, at the concentrations we are considering, a

[1]More precisely, this holds until more vapor leads to more clouds, which can lead to more reflected sun light.

non-poisonous inert gas, formed by respiration of all living systems. For example when we breathe we combine oxygen and glucose and emit CO_2, and the equivalent process exists in the metabolism of microbes. A different microbial process, fermentation, breaks down glucose in absence of oxygen, and can also produce CO_2, for example CO_2 is produced when yeast populations ferment to make wine, beer or bread. Conversely, photosynthesis from plants or plankton takes up CO_2 and water, using light to make glucose. Thus, there is a sort of "natural balance" between emissions and uptake from photosynthetic and non-photosynthetic life forms. Human industrial-age CO_2 emissions perturb this balance, and come mainly from two activities: the burning of fossil fuels and the production of cement. Natural phenomena such as volcanic eruptions may also release a lot of CO_2.

Similarly to what we did for temperature, we can look at the evolution of global CO_2 levels in recent years (for example since the 1950s), where we have direct and precise measurements. We should also bear in mind that how much CO_2 we find in the atmosphere differs from how much we emit. Instead, atmospheric CO_2 is the result of the complex balance between all the "sources" and "sinks" of global atmospheric carbon. Considering all these processes is very complicated (see below), but if we imagine for the time being that all that matters for the greenhouse effect is how much carbon is in the atmosphere, we can consider directly atmospheric CO_2 levels, and worry later about where they come from (and how to reduce them). Here's my sketch of what the data for atmospheric CO_2 since 1950 until recent times roughly look like,



The y axis units of this plot are in "parts per million" (PPM). This quantity tells us how many particles of carbon dioxide there are in one million particles of air (which is made of different substances, including CO_2, nitrogen and oxygen). The sketched data can be considered as global averages by location and yearly fluctuations (as we did for temperatures, and for the same reasons), in order to visualize a gross trend. The plot, whose slope increases steadily since the 1950s, looks more markedly bent or "banana shaped" than we found for temperature in the previous chapter, but more or less follows a similar trend. In the 1950s and 1960s the increase still looked mild (for temperatures it was almost flat). In the late 1960s, the increase of atmospheric carbon dioxide was around 1 PPM/year. Over the next decades, the yearly increment has increased almost invariably, reaching 2.5 PPM/year (and more) in the 2010s. The red line in my sketch is a linear extrapolation of the trend from 2008 to 2018 (about 2.3 PPM/year) to future years, up to 2050. According to this prediction, we would reach about 480 PPM of global atmospheric carbon in 2050. The prediction is conservative, because it assumes that every year we will increase atmospheric CO_2 by the same amount, but so far we've done increasingly worse, adding more PPM than the previous year to our atmospheric CO_2 almost every year since the 1950s. As for temperature, the trend is much beyond the uncertainty in the data, and in this case the seasonal variation of global CO_2 levels is actually small (a few PPM). We will use the historical record to "calibrate" these values below, but before we do that, I need to discuss the other greenhouse gases that we should not ignore.

Our plot shows us that we currently have slightly more than 400 PPM of CO_2 in our atmosphere. But how much is that? Can we put our hands on this quantity? Suppose we had the technology to take up all the carbon from the atmosphere and compress it into a diamond. How big would it be (if we know, we can also plan to place it somewhere nice)? Diamond is one of the densest possible forms of carbon; its density is around 3500 kg/m^3, about half that of steel. We can go on the IPCC web site and find out that 1 PPM of CO_2 corresponds to about 2.12 Gt of carbon (Gt means gigatonnes, or billion tonnes, which is 10^{12} kg, see below for a more direct estimate)

So, we are dealing with 400 PPM times 2.12 Gt, so about 850 Gt carbon, which is 8.5 10^{14} kg. Divided by 3500 kg/m^3 and then by 10^9 m^3/km^3, it gives us a block of diamond of almost 250 km^3. This is a cube whose side is a bit larger than 6 km, or a sphere whose diameter is almost 8 km. This purediamond ornament would not fit in the Tower of London. On the other hand, it would be large enough to embellish our highest mountain ranges, such as the Himalaya, the Andes, or the Alps. Imagine watching the sun set behind a diamond mountain, with reflections of a million colors... how romantic![2]

This was our first example of estimate in the style of Enrico Fermi. As I mentioned in the preface, the physicist Enrico Fermi was famous for his amazing ability in approximate problem-solving, using (seemingly) little or no actual data (a proverbial example where Fermi operated in this way is his estimate of the power of the atomic bomb that exploded in the Trinity test, based on the movement of a few pieces of paper that he dropped from his hand as the shock wave due to the explosion passed through). This has given the name "Fermi problem" or "Fermi question" to situations of this kind. Fermi himself used exercises of this type for teaching purposes in his courses. In a Fermi problem, the goal is not to arrive at an "exact" answer, but to obtain a "good enough" answer, as well as (more importantly) some insight on the structure of the question under exam.

In our Fermi estimate, we used the value of 2.12 Gt per PPM of carbon by getting it from the internet. This could be considered cheating by some (or "standing on the shoulders of giants" by others). Instead, we can try to get it by another Fermi estimate. If we want to do that estimate by

[2] But don't stay out too late, nights can get chilly on this planet without any carbon in our atmosphere! (see the next chapter.)

ourselves we need to know the volume of the atmosphere, and how many particles of air per unit volume there are in the atmosphere (the number density of air), then by knowing the mass of carbon we can get to our goal. To determine the volume of the atmosphere, we can use the fact that the Earth has a radius of approximately 6300 kilometers and the atmosphere extends up to an altitude of about 40 kilometers above the surface. This gives a volume of the atmosphere of about $20\,10^{18}$ m^3. The number density of air is not simple to estimate without some physics. It can be estimated roughly by the "ideal gas" law, which relates it to temperature and pressure. At room temperature and pressure (25 ^{o}C and a pressure of 10^5 Pa), this law gives about $2.5\,10^{25}$ molecules/m^3. In practice, air density changes with height (as pressure and temperature do) and the atmosphere does not have a very well defined height as it does in our estimate. In order to compensate for this effect we will divide the volume by a factor of 5[3]. So there are about $2.5\,10^{25}$ times $4\,10^{18}$, which gives 10^{44}, particles of air, and 1 PPM of that is obtained by dividing by one million, or 10^6. This is 10^{38} carbon atoms. The mass of carbon in molecular mass

[3]This is adjusted, because getting the number of air molecules in the atmosphere right is a bit the weak point of this not-so-simple estimate. There is another way to get to the same number. The total mass of the Earth's atmosphere is about $5\,10^{21}$ grams. If we take air to be a mixture of about four molecules of nitrogen to one of oxygen, based on their molecular weight, the mass of 1 mole of air will be about 29 grams. One mole of any substance contains by definition about $6\,10^{23}$, an Avogadro number, of molecules. So there are about 10^{44} molecules in the Earth's atmosphere. Based on this estimate we can also claim that every time we take a breath of air, we inhale some of the atoms breathed by Leonardo da Vinci or Cleopatra

units (u) is 12, because it's made of six protons and six neutrons (and its electrons are very light). A molecular mass unit is the mass of a proton, about $1.7\,10^{-27}$ kg. If we multiply this mass by 12 to estimate the mass of a carbon atom, and then by 10^{38} carbon atoms in one PPM, we get about 2 Gt, which is the quantity that we were looking for. If we hadn't adjusted the total number of air molecules we'd be a factor of 5 off. Not too bad, but not sufficiently precise for our scopes. So, getting this directly was a bit laborious (sometimes Googling is easier!) but see the "Physics track" section at the end of this chapter to justify this point with a bit of physics insight.

We said that CO_2 is the main contributor to greenhouse gases. However, other greenhouse gases produced by farming and industry significantly affect the Earth's atmospheric global temperatures. We can break it down into three main contributors: nitrous oxide, methane, and a whole class of fluorinated gases. All these gases are quantitatively much less than CO_2, but unfortunately their potential for causing global warming is much higher, and for this reason we cannot neglect their contribution to climate change. Indeed, different greenhouse gases radiate heat differently. To compare the global warming caused by different gases, scientists have introduced the notion of "global warming potential" (GWP). The GWP is defined as the heat absorbed by any greenhouse gas over a certain period of time, divided by the heat absorbed by the same mass of carbon dioxide over the same period. By definition, GWP is 1 for CO_2. Additionally the GWP of another substance depends on the number of years over which the ratio is calculated. Often the time span is indicated by a subscript, as in GWP_{100} for the GWP for 100 years. This hundred-year period is considered frequently by the practitioners, as this is the time frame over which we need to contain the consequences of industrial-age emissions, so sometimes when the subscript is absent it is implied that we mean a hundred years.

Using the GWP we can compute a "carbon dioxide equivalent" (CO_2e) as the amount of CO_2 which would cause the same greenhouse effect as a given combination of greenhouse gases (over a certain time span), multiplying the amount of each substance in the atmosphere by its GWP and summing over all the substances. Over a hundred years, the GWP for nitrous oxide is about 250-300[4], that of methane is about 25-30[5], and fluorinated gases range from one thousand to tens of thousands. Fortunately they are quite rare.

Fluorinated greenhouse gases are used today as replacements to so-called ozone depleting substances, which were major contributors to stratospheric ozone depletion (as well as to the greenhouse effect), hence since the 1980s they have been first banned and gradually replaced by alternatives (in refrigerants, as aerosol propellers and as foam blowing

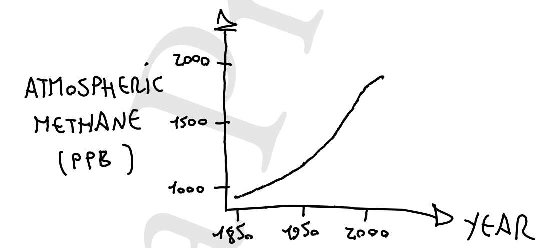
[4]The estimated value in the IPCC sixth assessment report is 273, we will use 300 in the following.

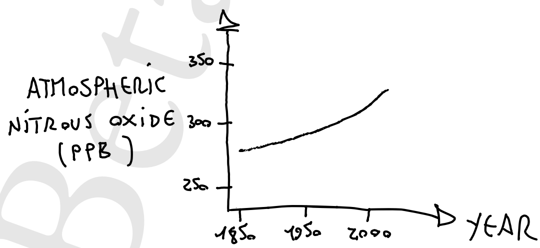
[5]We will use 30 in the following.

agents). Fluorinated greenhouse gases do not damage the atmospheric ozone layer. However, most of them are powerful greenhouse gases, with a very high warming potential (which varies from gas to gas). Fortunately, their concentration is low in the atmosphere. Hence, despite of the huge GWP of fluorinated gases, for the moment nitrous oxide and methane appear to be more important today as contributors to the greenhouse effect. However, we should note that the cumulative warming effect of fluorinated gases and ozone-depleting substances (which are different in terms of effect on the ozone layer, but can be considered as the same family in terms of their behavior as greenhouse gases) currently still surpasses that of nitrous oxide.

Nitrous oxide is mainly (about 70%) linked to the growing use of nitrogen fertilizers (and is also a depletant of the stratospheric ozone layer). Excess fertilisers based on nitrogen end up into emissions of nitrous oxide essentially because crops cannot use them all. Careful use of fertiliser when crops need it would reduce these emissions. Turning to methane, the first contributor (about 40%) to its emissions is animal agriculture, through livestock, mostly cows and sheep, which produce methane in their digestive processes. There are roughly 1.5 billion cows and 1 billion sheep around the world, and reducing this number would reduce overall methane emissions. Rice production is also an important contributor to agricultural methane emissions, because in flooded fields the water blocks oxygen from penetrating the soil, creating a habitat for bacteria that emit methane. Farming methods that reduce or eliminate flooding could reduce methane. Unintentional leaks of methane from fracking and oil and gas extraction and transportation are the second largest contributor (about 30%) to methane emissions. The third largest contributor (about 15%) is waste landfills, from decomposition of organic materials. Chapter V will attempt a more systematic approach to the numbers of global greenhous gas emissions, for the moment let's try to understand the global trend of the main greenhouse gases in our atmosphere.

The hand drawn plots below give us the global amounts of atmospheric nitrous oxide and methane since 1850.



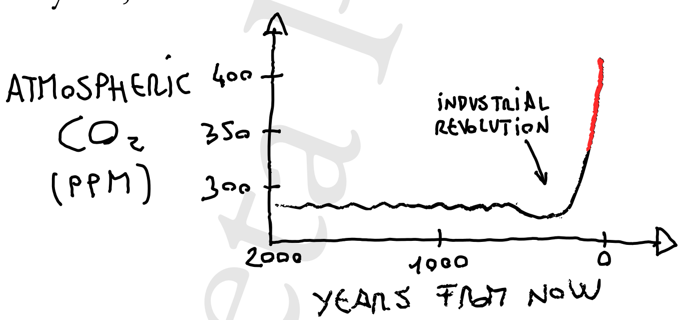


The y axis units are now in parts per billion (PPB). As for CO_2, the plots show clearly that both gases have been increasing steadily over the period we are considering, with methane roughly doubling its amount. There are roughly

2000 PPB of atmospheric methane today, which multiplied by its GWP_{100} give roughly 60 PPM of CO_2e from methane, not negligible compared to the 370 PPM of CO_2. The 325 PPB of nitrous oxide give an ever larger contribution of about 100 PPM CO_2e over a hundred years. Perhaps a better way to compare is by using the concentration differences from pre-industrial levels and the current ones. By this metric, we get a difference of about 150 PPM for CO_2, about 1 PPM for methane and about 0.05 PPM for nitrous oxide. If we now multiply by the GWP of each gas, methane gives about 1/5 of the contribution of CO_2, and nitrous oxide gives about 1/10 of the contribution of CO_2.

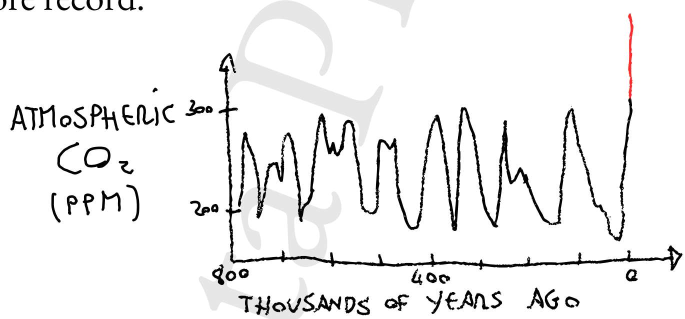
Going back to CO_2, we can look, as we did for global temperatures, at the record of emissions in the more distant and very distant past, to gain some reference that we can use to frame the recent record. As for global temperature, we can use this information to put a scale on the proportion of the changes that we see today. As for temperature, ice-core data give us a fairly precise idea of CO_2 amounts, and in this case the measurement is direct, as it amounts to measuring how much CO_2 is found in a layer of ice corresponding to some time in the past. To be precise, ice cores measure the CO_2 dissolved in ocean waters, since carbon dioxide dissolves in water. Assuming the global amount of dissolved carbon is in chemical equilibrium, scientists can deduce the atmospheric amounts. Roughly, there is about fifty times as much CO_2 dissolved in the oceans as is found in the atmosphere. Dissolved carbon is present in all water basins. The reaction involves three main constituents, free CO_2, the bicarbonate ion and the carbonate ion. Most of the dissolved CO_2 persists as CO_2 molecules, but the dissolved ions are sufficient to induce a decrease in the pH of the oceans, the so-called "ocean acidification". At lower pH, carbonate becomes undersaturated, which means that the equilibrium point between ions dissolved in seawater and attached to crystal (precipitated) is shifted in a way that there are more ions around. In this condition, living beings using shells made of calcium carbonate become vulnerable because their shells dissolve more easily in this acidic environment.

The plot that I sketched below describes the estimated trend in atmospheric CO_2 from ice core data for the past 2000 years,



The y axis is in PPM as above, and the last part of the plot is highlighted, as it corresponds to the roughly two-hundred years period just discussed. As we found for global temperatures, the plot tells us that CO_2 levels have been remarkably constant until the sixteenth century or so, and after that only slightly decreased. Instead, during the industrial revolution atmospheric CO_2 started to increase quickly and radically, and it never stopped since. This plot helps us addressing the important question of where we should aim for atmospheric carbon levels, by an educated guess. Arguably, our target is the pre-industrial level, slightly lower than 300 PPM (around 280 PPM), since we know from thousands of years of human history that it creates temperature ranges that are comfortable enough for the human body to allow (more primitive forms of) our civilization to thrive. Possibly we can get by with slightly more. This reasoning holds as long as we do not drive the climate system irreversibly too far out of its current state (see Chapter VI).

Looking at a more remote past, the following sketch shows a plot of 800000 years of CO_2 levels from the ice-core record.



Once again, the recent record is the last highlighted part of the drawing. This part would barely be visible, but I highlighted it in the drawing, in order to show that the amounts of atmospheric CO_2 that we see today are completely out of range compared to what happened for an enormously long time. This 800000-year record shows multiple variations in CO_2 levels where the atmospheric values of this gas changed by more than 100 PPM, but they stayed in all cases lower than 300 PPM. We were already at a peak when the industrial revolution started. This is at least what we see at the time resolution that is accessible with these data. The time resolution of ice core data depends on the thickness of the layers that can be sampled reliably. For these data, the scientific literature estimates it to vary with time between 200 and 1000 years [6], meaning that we can regard each data point as an average over such a period. In other words, if at some time in the distant past there had been a peak of atmospheric CO_2 that lasted only say 400 years (we can hope that the current peak will not last longer than this), it might not show up in this plot, or be only slightly noticeable. If we remember the equivalent plot for temperature in the previous chapter, we can observe that the oscillations of temperature and CO_2 are remarkably similar. We can conclude that on a time scale of about 10 ky if CO_2 goes up, so does global temperature, and conversely, if CO_2 goes down, global temperature decreases. In brief, these 800000 year-old data seem to be telling us the same story as the latest news: there is a strict correspondence between temperature and atmospheric CO_2. This subject will be the center of the next chapter.

Beyond ice core data, and further in the past, it is possible to estimate (more roughly) CO_2 levels from geochemical proxies in the fossil record, to get information from hundreds of millions years ago. For example, scientists have

[6] See for example Masson-Delmotte and coworkers, Quaternary Science Reviews 29 (2010) 113-128.

looked at stomata (pores) in leaves of fossil plants, and related their amounts to carbon levels. I collected some of these plots, and took a closer look. These more ancient data suggest that the closest time when CO_2 in the atmosphere may have reached the abundance it has today was 20 million years ago. Once again we should bear in mind that the comparison is not completely fair due to the (much poorer) time resolution of these data. Possible transient increases that lasted even thousands of years, would be completely erased out by the averaging in these data. But the geological data still give us a good idea of how exceptional today's atmosphere is in terms of CO_2 content (even if it is compared to these mean values). They also tell us that multiple times tens to hundreds of millions years back, the Earth has seen average CO_2 levels that compare to the levels that we see today, and these increases were correlated with increases in the global temperature.

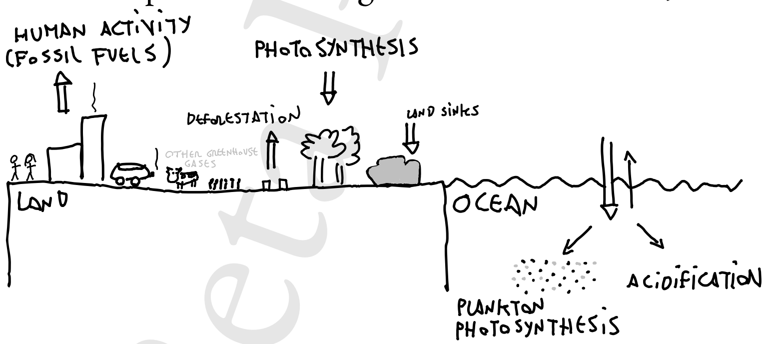
It is mind-blowing to think how dramatically different the climate of our planet has been over its history, but on a more pragmatic level we found out how the question of time resolution is crucial in these data. Let's go over it again. Today's record where CO_2 levels are measured every day with high precision contains day-to-day and seasonal variation. If we plot these data, we can see sinusoidal oscillations during the year, with an amplitude of a few PPM, due to the collective photosynthesis of all plants and plankton (which is seasonal). If we average the yearly data or take one point every year we get a somewhat smoother trend line. Taking data with lower time resolution is equivalent to performing these averages on longer and longer time intervals. Ice core data have varying time resolution, depending on how far in the past they go. Recent data can be compared to the year-to-year averages, but going back by hundreds of thousands years ago each data point (in the current data sets) corresponds to a time step of 4-500 years. In geological records the effective time resolution can become a million years or more. Importantly, the effective time resolution of a data set in the form of time series can also be limited by the precision of the measurements. For example, suppose we can take geological CO_2 data reliably over "slices" of 100000 years and over this period the quantity we measure varies by say 1 PPM. If the resolution of our measurement technique were 30 PPM we'd have to pile up many slices before we are sure we see a change that is above our experimental uncertainty.

To make the issue of time resolution more concrete, we can try to address a question that borders with science fiction. Suppose another complex life form lived on Earth many millions of years ago, and started a civilization similar to ours, developing fossil fuel engines and heating systems, building cities and cars, and computers and/or other fancy gadgets. Let's assume that this civilization lasted a few hundreds to a few thousands years. Would we be able to tell? This crazy question occurred to me looking at these data, and curiously it turns out that somebody had actually considered it seriously in the recent past. The study by Gavin

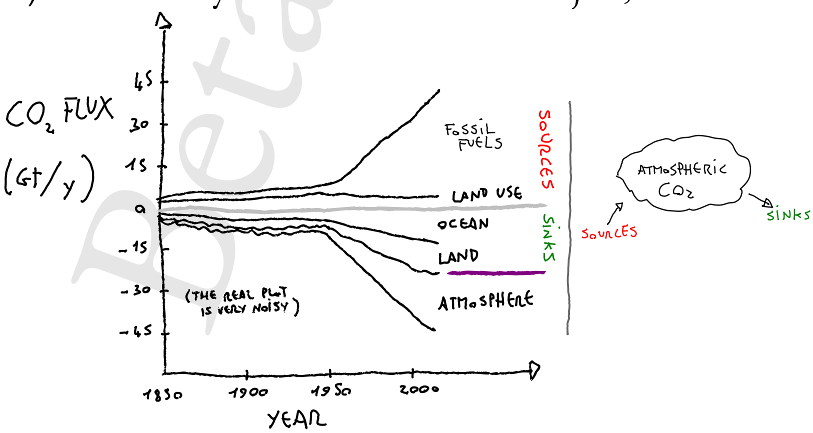
Schmidt Adam Frank is cited in the Sources section at the end of this chapter. The answer they found is intriguing. As could be expected, buildings and cities would be long gone, mainly because of tectonic movements, and it would be extremely unlikely to find artifacts and fossils as, on average, we find one single fossil for every 10000 years of Earth time. One could hope to be able to see some chemical trace, such as residuals in sediments (plastics, for example), and it seems that CO_2 levels from fossil fuels could probably be our best hope to gather such evidence. But the current time resolution would not really allow us to detect short-lived peaks of 300 years, such as the one we are currently producing (in the best of hypotheses). As it turns out, it would be easy to miss carbon emissions from an industrial civilization that lasted 100000 years, 500 times longer than our industrial age so far. Hence, it seems we cannot really exclude that we are not the first technological society on this planet that has faced (and caused) the climate problems we experience today.

THERE IS A "CARBON BALANCE", similarly to what happens in a financial budget, and it is a crucial point that we need to discuss in this chapter. Understanding why there is a budget is necessary to get a simplified idea of how much of the CO_2 we emit ends up staying in the atmosphere (affecting climate), and where the rest ends up. We have already mentioned that atmospheric CO_2 is the result of the complex balance between different "sources" and "sinks" of carbon, that plants and some microorganisms carry out photosynthesis, and that CO_2

dissolves in oceans. In this sources/sinks metaphor we can think of carbon fluxes as a plumber thinks of water loss in our house's hydraulic system. In the budget metaphor we can think of it as an accountant collection and tracking of a firm's incomes and expenditures. If we join the two metaphors we probably get the idea that plumbers and accountants are more similar than they actually are in real life, but I guess you get the message. A carbon sink is any identifiable process that absorbs more carbon from the atmosphere than it releases, and vice versa a carbon source has a (net) positive emission towards atmospheric CO_2. The sum of all the fluxes of carbon in and out of the atmosphere (and other earth compartments) is called the carbon cycle. The simplified sketch below illustrates some of the main processes acting as sources and sinks,



An important additional piece of information (which we already touched upon talking about GWP) is that atmospheric carbon, left alone, does not decay or degrade. Methane, by contrast, is removed from the atmosphere by chemical reactions on time scales of tens of years. Thus, although methane is a potent greenhouse gas, its effect is short-lived. Equally, nitrous oxide is destroyed in the stratosphere and removed from the atmosphere, persisting for around a hundred years. Without these spontaneous decay processes, atmospheric carbon levels mainly depend on their emission and absorption by the sources and sinks. Roughly, 70% of atmospheric CO_2 may dissolve into the ocean over a period of 200 years. The other removal processes (for example rock formation) require several hundreds of thousands of years. This means that once in the atmosphere, carbon dioxide can stay there practically indefinitely unless it is removed by sinks. The balance of positive sources and sinks is called the carbon budget. Naturally, if we want to normalize our current situation we can act on both sources and sinks: emit less and help the land and ocean to sequester more carbon. To get a rough impression of the role of sources and sinks, here's a sketched plot of carbon "flux" (in billion tonnes, or gigatonnes, Gt, of CO_2 year) estimated by the Global Carbon Project,



Reasoning on this plot, when I first saw it, was very useful to me. The plot shows two main sources of CO_2. The greater one, rising steadily since the 1950s, is the carbon released from the burning of fossil fuels. The other, which stayed more constant over the years (about 5 Gt/y), is the release of carbon from land use, such as through deforestation (degraded forest emits CO_2, from all the carbon that was stored in the plants), land clearing for agriculture, and degradation of soils. The negative flux part of the plot shows that there are two main carbon sinks, ocean and land. The land sequesters carbon mainly through forests. Plants capture carbon dioxide from the atmosphere for photosynthesis, and some of this carbon is eventually transferred to soil as plants decompose. The ocean captures CO_2 because CO_2 dissolves in it. Phytoplankton in the ocean can absorb some by photosynthesis, and this amount can be stored in oceanic sediments because these organisms tend to sink when they die. It is important to note that the sinks are dynamic. The plot shows that since the 1950s the ocean and land sink have increased their negative fluxes adapting to the increased emissions, but that at the same time atmospheric carbon has also increased steadily. From these data, ocean and land seem to have been able to accommodate (roughly) about one half of the annual budget so far.

However, we also know that this happened at the cost of an increased ocean acidification, which has heavy consequences on ocean ecosystems. At this rate, the estimates tell us that oceanic carbon may saturate in about 100 years, after which the ocean will loose its capacity to dissolve more carbon. Sometimes if you have a clogged drain or sewage system in your home it is not easy to notice it at the beginning; water encounters a partial obstruction, and redirects itself through your plumbing system, which is a complex branched network of pipes, causing initially small reactions. You may experience a slow drain in one or multiple areas of the home, or a funky smell from your sink, and underestimate the situation. But after some point the inner workings of your piping system become saturated, and they are not able anymore to tolerate the perturbation. You flush the toilet and brackish water backs up in your sink, or your yard floods with water from your sewage system. Ideally you want to act before the only solution is hiring a plumber with a honey wagon to pump out your discharge.

WITH THE DATA COLLECTED SO FAR, we can try a back-of-the envelope estimate of how our emissions are linked to the increase of atmospheric carbon. The last plot tells us that today (in 2019-2021), we emit about 35 Gt of CO_2 from fossil fuels, which becomes about 40 Gt including land use. We need to convert this in PPM, which is not completely trivial, but we already said that 1 PPM of CO_2 corresponds to about 2.12 Gt of carbon. Now, since CO_2 is not made only of carbon (there are also two oxygens), we need to further convert the mass of carbon into the mass of CO_2. Chemistry tells us that 1 Gt of carbon is about 3.67 Gt of CO_2. The way to work this out is simple. CO_2 has a carbon, molecular weight 12 u (six protons and six neutrons, as we said above), and two oxygens, molecular weight 16 u each. So the mass ratio of CO_2 to carbon is just

$(16+16+12)/12 \simeq 3.67$. Hence, 1 PPM of CO_2 amounts in about 7.7, say 8 Gt of CO_2. Or vice versa, 1 Gt of emitted CO_2 will make about 1/8 PPM (one eighth of a PPM). However, we know that only a percentage of what we emit stays in the atmosphere. From the plot, this looks like about 40%, probably a bit more, let's say 45%. Therefore we can say that

$$1 \text{ Gt } CO_2 \text{ emission} \simeq \frac{1}{8} \text{ PPM } 45\% \text{ in atmosphere}) \,,$$

and

$$40\,\text{Gt/y } \; CO_2 \text{ emission} \simeq \frac{40}{8} \text{PPM}\, 45\% \text{ in atmosphere}) \,,$$

which is 40% of 5 PPM/year, about 2.25 PPM/year which fits reasonably well with the trend we found above, and with the average trend of 2.5 PPM/year found between 2010 and 2020.

To sum up, we have our own rough way to estimate how much of how our emissions end up in the atmosphere. Let's see if our model works for the past. In the period 1950-1955, the global annual CO_2 emission was about 7 Gt, which according to our calculation amounts into 0.35 PPM/year. In reality, it was about 0.5 PPM/year, so we are a bit off but we are not doing too bad. This model is quite naive but tells us something not far from the truth, that if we want to quickly reduce atmopheric greenhouse gases we need to cut our emissions to zero, or nearly so. This should ring a bell, because we just said that until now we have basically just been increasing! At least until very recently; to add an optimistic note, now in 2022 the recent trend seems more sluggish, or weakly decreasing. More precisely, we can only emit as much as plants and microorganisms can absorb, in order to keep the balance. In practice, as our plot shows, these sinks are adaptable (to some extent), and will hopefully work more as long as there is more CO_2 around, helping us a bit.

Our estimates point to two big questions. The first is how fast we need reduce our emissions to nearly zero, in order not to increase too much global temperatures, the second is if we also need to get rid of the excess carbon with additional technologies. Regarding the first question, this is what climate models are all about. Today I think they tell us to cut emissions by 7% every year to be able to stay 2^{o}C above pre-industrial temperatures.

REGARDING the second question, all scenarios where the excess temperature will stay within 1.5^{o}C above pre-industrial levels depend on "carbon capture" technologies that sequester carbon as they are emitted (typically from a power plant or a factory) or from the atmosphere[7], making yearly emissions negative, by the second half of the 21st century. In other words, we would have to remove carbon from the atmosphere using different kinds of large-scale infrastructure. Some technologies, and small scale applications of these technologies exist, but the cost (in terms of

[7]Currently the systems that capture CO_2 directly from air (apart from plants) appear to be quite far from providing reasonable solutions.

design, industrial conversion, workpower, etc. and also in terms of further emissions!) of deploying them to the required scale (Gt per year) may be a formidable challenge. For example, equipping a large-scale coal power plant could cost a billion euros over its lifetime, about 40 years, and result in a 40% use of the energy it produces (about a half). In other words, the equivalent factory with carbon sequestration technology will have 600 MW of usable power. A 1000 MW (1 GW, gigawatts) factory could release about 5 Mt (million tonnes) of CO_2 per year, so a lifetime total of 200 Mt. MW (megawatt, 10^6 watts) is power, or energy per unit time, whereas kWh (kilowatts times hours) is energy. There are 8760 hours in a year, so ideally a 1 MW factory can release about 8.76 10^9 kWh, about 10 GWh, per year. In practice, because of limited efficiency, we can say that, for coal, this is a fifth of the theoretical value, about 2 GWh per year. If the running cost per kWh is 0.01 euros, and we assume that due to the discount rate (economic growth) only one tenth of the 40 running years (the first 4 years) matter (which might be reasonable for a 10% discount rate), then an "overall cost" of the carbon capture technology can be quantified in

$$10^9 \text{ euro} + (400\;2\;10^9 \text{ kWh/year})\times$$
$$(40 \text{ years } 0.1)(0.01 \text{ euro/kWh}) ,$$

which divided by 200 Mt gives about 150 euros per tonne of CO_2. This reduces to about 45 euros per tonne if the cost of carbon sequestration is only 10% of the produced energy.

This very rough reasoning seems compatible with the available estimates (based on very complex reasoning involving economic models and carbon emissions), which say that a carbon price of 100 euros or more per tonne CO_2 would be needed to make industrial carbon sequestration viable. In this estimate, the total cost is nine billion euros, the impact of the one-billion euros construction cost is small (a bit more than 10%), and the energy cost has the most impact. We neglected the carbon emissions related to building the carbon sequestration technology (which might also play a role). I also note that the discount rate (which mirrors how optimistic we are about future economic growth) plays a very important role in this estimate. Roughly, we can say that in absence of the 10% projected economic growth the cost could be 10 times larger.

If many power plants were forced to capture all or most of the CO_2 they emit and store it permanently underground (a technology that is currently not well established), and the conversion cost in terms of emitted CO_2 were not too heavy, we could sensibly cut our emissions. The estimated global coal power capacity is currently about 2000 GW (responsible for about 20% of our global emissions); this is 2000 power plants such as the one considered in our example. Hence, the total conversion cost over the lifetime of the plants would be about $2\ 10^{13}$ euros; these are about twenty trillion euros, each of which is a thousand billions. By comparison, the gross natural product (defined as the total value of all goods and services produced by all citizens in a given financial year) of he US and china is around 22-23 trillion euros, and the world gross national product is about 130 trillion ($1.3\ 10^{14}$) euros. An alternative (perhaps more conservative) way to quantify the cost is to look at the investment needed to obtain the same power. We can say that if the carbon sequestration takes 40% of the produced energy in any given time, a 1000 MW factory would have a reduced effective power of 600 MW once the emissions are sequestered. As a consequence of that, instead of 2000 coal power plants we would need about 3300, hence a sheer investment of "only" 3300 billion or 3.3 trillion euros (the current GNP of France). One could repeat similar estimates for oil and gas power plants, but the bottom line is that if these figures are reasonable, global-scale carbon-capture technology would take a significant fraction of all the money we make.

This does not mean that we should not try to deploy carbon capture technologies to a feasible scale in order to improve the global carbon budget, and also it does not mean that we should not try to improve the figures to make them more feasible. Precisely to what scale we should aim to deploy carbon capture technologies seems to me like a very big question, as it involves evaluating carefully delicate trade-offs between advantages and costs from multiple origins, and also a concerted global planning of course, given the budget. What should be apparent however, is that carbon capture technology (today and in the coming years) alone does not seem to be a feasible global solution to human-induced climate change, and that massive and quick "upstream" cutting of emissions remains a top priority.

As I mentioned above, there are technologies that capture carbon directly from the atmosphere, potentially providing negative emissions, but this is difficult, because CO_2 is very rarefied (a part per million means one gram for each tonne of atmosphere). Essentially, these are huge arrays of fans that filter air and put it through chemical reactions that remove CO_2, after which they let the air go back in the atmosphere (as long as we are emitting CO_2, it makes sense to deploy these technologies as close as possible to where we emit, but this is not simple in the case of cars or planes for example). Similarly to the case of carbon capture from the source, there are challenges for scaling up this technology to the required thoughput, in terms of investments and energy requirements, which lead to considerations that parallel those we attempted above. As a consequence, the current consensus on direct air capture is also similar. Alone, it will not provide a solution, but it might help, especially in the long run. Many are afraid of the risk of planning the near future relying too much on the assumption that carbon capture technologies can be deployed at a global scale, since if they are stopped or delayed by unexpected later hurdles the consequences could be catastrophic.

A simpler (and more ecological) "technology" that removes CO_2 from the air is photosynthesis by trees (and plankton). A fully grown tree can absorb at least 10 Kg of CO_2 per year. How much forest do we need to capture a

Gt of CO_2 every year? One Gt is 10^{12} Kg, meaning that we need to plant 10^{11} trees (a hundred billion) to offset one fortieth of our current total emissions. If we suppose that each fully grown tree needs an area of 10 m^2 (roughly a box of 3x3 meters), it means we need 10^{12} square meters of trees, or a million square kilometers for each Gt/year. It's a lot, but on a global scale it does not seem impossible. It is roughly twice the area of France, but only one seventh of the area of the Amazon forest. There is an ongoing project to conserve, restore and grow a trillion trees by 2030 (`https://www.1t.org/`). Importantly, in the long run, a lot of carbon would accumulate in this huge set of trees (mostly in the wood of mature trees), and one would have to make sure that the stored carbon does not go back in the atmosphere too fast (for example by burning the wood), or the effort would be undermined. Once again, my numbers maybe be a bit off, but they give us an idea of the magnitude of the problem, and, to me, they show that planting trees once again can help, but will most probably not work without cutting fossil fuel emissions as well.

What certainly we should *not* do is cut or let die too many trees and let the carbon that makes them up get to the atmosphere, because in those trees there is (roughly speaking) basically all of the carbon absorbed cumulatively over their lifespan, so that they can in principle emit in a very short time much more CO_2 than they can sequester every year. How much CO_2 actually gets released is quite complex, as it depends to what happens to the dead plant (for example it could be degraded by microorganisms, burnt, or the wood could persist for a long time, with different outcomes). On a positive note, wildfires appear to be relatively inefficient in releasing the CO_2 stored in a forest, because they are sparse. Harvesting, instead, causes a higher mortality than wildfire, and consequently (and counterintuitively) it seems that increasing harvest of mature trees to save them from fire may actually increase emissions, rather than reducing them.

To conclude, my back-of-the-envelope arguments suggest that while we should do our best to enhance the carbon sinks, we should definitely cut our emissions. The "where do we start" part is the subject of chapter V; before I get there, the next chapter tries to grasp how atmospheric greenhouse gases translate into temperature changes, and the following one how temperature changes affect weather and extreme events.

DATA SCIENCE TAKE-HOME MESSAGES. In this chapter, we performed our first exercises in the area of quantitative estimates and theoretical modeling. First, we warmed up with an estimate of the size of a cube mate of the atmospheric carbon. This estimate uses the classic expedient of translating the quantification of something into something that we can relate to (a solid cube). Second, we constructed a more ambitious model to relate CO_2 emissions to atmospheric carbon (in PPM), formulating a testable prediction, which we used to validate the model. Third, we used typical methods from Fermi estimates to reason on the carbon capture costs for a big coal power plant, and for all coal power. During the first estimate, we learned that guessing precisely the conversion from PPM CO_2 to Gt was not very easy. A lesson to learn is that this kind of estimate is very effective to guess the order of magnitude of something, but it may be difficult to refine its precision (or, more precisely, it may require very specific information, scientific insight, and detailed modeling). In terms of methods, here's a few suggestions to construct a good Fermi estimate: (i) break down the problem into sub-problems; (ii) think about bounds and take central values; (iii) think about "sanity checks", to verify the validity of the estimates (iv) dare to approximate, because sometimes it is important to start with an estimate of some kind start getting a grasp on the problem, and sometimes a not-so-good solution opens the way for a better solution; (v) "Google it", if there is no possibility of estimating some value that makes an essential part of our estimate with our knowledge, it's not a crime to rely on the knowledge of others. Other than this, this chapter reiterated the analysis of time series across time scales found in chapter I, this time for atmospheric CO_2 and other greenhouse gases. For example, we learned that because time scales correspond to averages, the carbon or temperature footprint of a hypothetical ancient advanced civilization would be likely be "averaged out", hence not visible in our data.

PYSICS TRACK. We can use hydrostatic equilibrium ("Stevin's law") in order to estimate the height of a constant-density atmosphere as follows. The law states that $\mathrm{d}P = -\rho g \mathrm{d}z$, where P is pressure, ρ is the atmospheric density g is gravitational acceleration at the earth's surface, and z stands for the vertical coordinate. By integrating this relation up to a height h ("the height of the atmosphere"), where we assume that $P = 0$, we get

$$\int_{P_0}^{0} \mathrm{d}P = -\int_{0}^{h} \rho g \mathrm{d}z \,,$$

where P_0 is atmospheric pressure at sea level. From this relationship we get

$$h = \frac{P_0}{\rho g} \,.$$

Using $P_0 \simeq 10^5$ Pa, $\rho \simeq 1$ Kg/m^3 and $g \simeq 10$ m/s^2, we get $h \approx 10$ Km, justifying the assumption we used within the chapter. This estimate is itself limited by several aspects, perhaps the principal of which is that temperature, density and pressure are related, and they vary with height. Their relation can be approximated by the ideal gas law $PV = Nk_BT$, or $P = \rho(R/M)T$, where k_B is Boltzmann's constant, T is temperature, R is the universal gas constant and M is the molar mass. The last equation can be solved for density as

$$\rho = \frac{PM}{RT} \,.$$

Assuming constant temperature, substituing ρ into Stevin's law and integrating gives

$$P(z) = P_0 e^{-\frac{gM}{RT} z} \,.$$

One can further refine the estimate assuming that T is not constant.

SOURCES. The plots on recent greenhouse gas emissions were found on Wikipedia, at the following pages
`https://en.wikipedia.org/wiki/Carbon_dioxide_in_Earth's_atmosphere`,
`https://en.wikipedia.org/wiki/Atmospheric_methane`,
`https://en.wikipedia.org/wiki/Nitrous_oxide`,
and on the curated plots by Ed Dlugokencky and Pieter Tans, NOAA/GML (`gml.noaa.gov/ccgg/trends/`). To make my own plots of atmospheric CO_2 levels, I downloaded two data sets from the NOAA database (`https://www.ncdc.noaa.gov`). The recent record (1751-2018) comes from the Oak Ridge National Laboratory, Carbon Dioxide Information Analysis Center, Scripps Institute of Oceanography CO_2 program, and the U.S. Energy Information Administration, International Energy Statistics, accessed December 7, 2020. The 800 ky Ice core data set is called "Antarctic Ice Cores Revised 800KYr CO2 Data" and collects data from several studies. The collection was published as Bereiter and coworkers Geophysical Research Letters, 42(2), 542-549 (2015) (link). The CO_2 data on My time scales come from Retallack Nature 415, 38 (2002), and can be visualized at the web page earth.org. The time series of the global carbon budget was taken from Our World in Data, Ritchie and coworkers (2020), Published online at OurWorldInData.org. Retrieved from this link. It is based on data issued in 2019 by the Global Carbon Project (`https://www.globalcarbonproject.org/`). For my own analyses, I downloaded the 2020 edition of the Global Carbon Budget. Andrew GCP, 2020. Global Carbon Budget 2020, link. A quantification of Nitrous Oxide sources and sinks can be found in Tian *et al.* Nature 586, 248-256 (2020). This article by Andrew

Moseman from the MIT climate portal argues that CO_2 levels between 280 and 350 parts per million created the climate where humanity can thrive (link). A NOAA document by Stephen A. Montzka (link) provides a non-technical comparison of the climate-warming influence of different greenhouse gases, including fluorinated gases (grouped as "HFCs") and ozone depleting substances (grouped as "CFCs" and "HCFCs"). The same information can be found in Chapter 7 of the IPCC-AR6 (Cissé *et al.* Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press 2022, pp. 1041-1170, doi:10.1017/9781009325844.009, see this link.) The question of the time resolution in ice-core data is discussed in Masson-Delmotte and coworkers Quaternary Science Reviews 29 (2010) 113-128 (2009). The study on the traces left by previous civilizations is discussed by Steven Ashley in Scientific American, Apr 23, 2018, and by Adam Frank in The Atlantic, Apr 13 2018. The original publication is Schmidt & Frank International Journal of Astrobiology, 18(2), 142-150 (2019). The article by Jorge Sarmiento and Nicolas Gruber, Physics Today Volume 55, Issue 8 (2002) discusses very clearly the carbon budget and the the oceanic sink. More recent (and technical) information can be found on Gruber *et al.* Science 363, 6432 1193-1199 (2019). The estimates on coal power plants are derived from figures on the global energy wiki, at the link gem.wiki. They come from the 2021 IEA report "World Energy Outlook" (link), from the 2007 MIT report "The Future of Coal" (link) and from the 2009 report "New Coal-fired Power Plant Performance And Cost Estimates" published by Sargent & Lundy (link). The data on carbon capture can be found on Wikipedia (link). In particular, the data on costs of carbon capture come from Thorbjornsson et al. Energy Strategy Reviews. 7: 18-28 (2015). The article by Realmonte and coworkers, Nature Communications 10, 3277 (2019) discusses feasibility estimates for direct air carbon capture. A brief and clear document on the impact of forests on climate change can be found on the 2003 FAO Forests and Climate Change Working Paper (link). The arguments on wildfires come from Bartowitz and coworkers, Front. For. Glob. Change, 5, 2022.

Chapter III

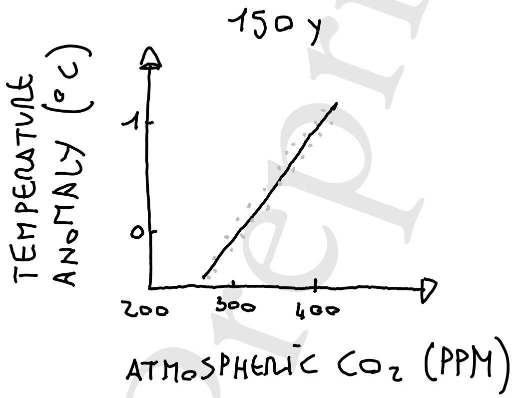
Atmospheric greenhouse gases are tied to temperature

Can we explore the link between rising global temperatures and rising atmospheric greenhouse gas, from the data? This was my next question, but I knew it would be difficult to go very far. In fact, in order to gain a little more insight, it is necessary to cover some ground on the physics of global warming. In terms of bare data, it is difficult to address this question, but what is possible is to cross the data sets presented in the previous chapters and try to learn something. Consequently, this chapter is a mix between a brief historical account of how a mix of laboratory experiments and global climate observations lead scientists to discover and understand global warming, and what I could gather from the simple data we have already seen.

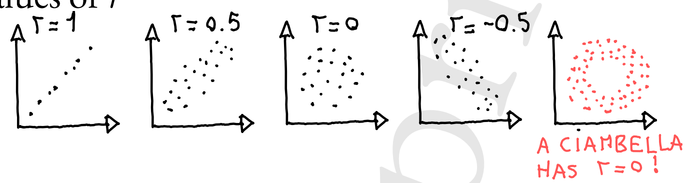
Our knowledge of the greenhouse effect comes from fundamental physics, and chemistry. The Earth absorbs radiation from the sun, and warms up, also emitting infrared radiation. Greenhouse gases absorb this radiation and re-emit it, preventing from escaping the atmosphere. Since more energy is absorbed than is radiated back into space, the result is a warming effect, the greenhouse effect. A French scientist, Joseph Fourier, was the first to realize (around 1824) that the Earth's atmosphere retains heat radiation. He did not have the full theoretical tools to calculate how the energy balance (involving heat radiated from the sun, reflected away, and absorbed by the atmosphere and the ground) places the Earth at a steady temperature. However, he realized that the planet would be significantly colder if it lacked an atmosphere. Currently, scientists believe that without the greenhouse effect the Earth would have probably been too cold to sustain complex life forms such as ourselves. This is the principle, but clearly a whole planet is not a system where controlled experiments can be performed, and we have already discussed that there are a lot of complex processes and feedback mechanisms. Earth's climate is driven by many factors, including solar activity, variations in the Earth's orbit and rotation, and changes in ocean and wind currents. Today, we can perform direct (to some extent) experiments in the atmosphere, to establish for example that carbon dioxide and water vapour are among the gases that absorb heat. However, it is only since recently that we can take sophisticated atmospheric measurements, for example using satellites. What happened originally is due to a careful combination of "field observations" and generalization of concepts from designed laboratory-scale experiments.

John Tyndall performed a crucial set of these controlled experiments in the 1850s. In a series of carefully designed observations, he sent "radiant heat" (heat transmitted by infrared radiation) through a tube containing various gases and vapors. He discovered that water vapor absorbs much more radiant heat than any other atmospheric gas. He also discovered that absorption due to carbon dioxide was very large. Extrapolating these results to the atmosphere, he argued the that variations in the atmospheric concentrations of CO_2 and water vapour could be very important in moderating the Earth's climate, essentially discovering the greenhouse effect.

Carrying out laboratory experiments is definitely beyond the scopes of our back-of-the-envelope approach, but, provided we believe the theory, we can still ask how much heat corresponds to how much greenhouse gases. In other words, if all these considerations hold, the record of global temperature and global CO_2 during the last 150 years should be tightly linked. We already know that these data are easy to obtain on the internet. So far we have been looking at these two data sets separately, but it is not hard to join them together. For every year we can obtain information on both data. We just need to "scatter plot" one versus the other, we use the information on the time period. Here's a sketch of how the scatter plot of temperature anomaly vs CO_2 looks like, each point corresponds to a year



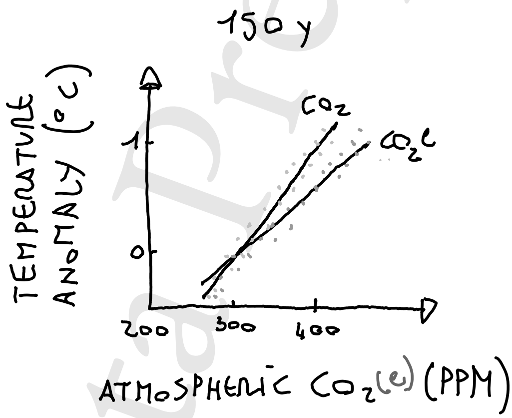
The points clearly cluster along a straight line, meaning that when CO_2 levels change, temperature also changes, with a tight interdependency between the two variables. The interdependency is tight because all the points are very close to the trend and do not form a scattered cloud. A way to quantify the strength of a relationship is the correlation, r, a numerical measure that expresses the extent to which two variables are linearly related (without giving an indication about cause and effect). There are many ways to measure correlation, but ours the simplest correlation coefficient r (which was first defined by Auguste Bravais and then developed by the pioneer statistician Karl Pearson in the 19th century), works well with scatter plots that look like a single cloud with an elongated or circular shape (one needs better scores for multiple clouds or clouds that look like donuts or bananas). The maximum value of r is 1 (perfect correlation, and the minimum is -1. When $r = 0$ there is no correlation. Here's a sketch n illustrating how different trends in a scatter plot translate to their corresponding values of r



The sketch also tells us that the Pearson correlation is not a good way to quantify correlation if the points in our scatter plot are correlated but not linearly, or in other words they cannot be pictured as clustering around a straight line (in the example I have drawn they form a circular pattern, or a "ciambella", which is the name of a donut, in Italian).

Coming back to our case, we found $r = 0.72$ between temperature anomaly and CO_2 (and we also found that in our case the shape of the scatter plot justifies using linear correlation). Even more interestingly for our current scopes, the slope of our trend (which is not the same as the correlation, because our data points can be more or less scattered around the same slope) tells us roughly how much CO_2 corresponds to a change of one degree Celsius in the global temperature, about 100 PPM, or more precisely I find 1.1 ^{o}C/ 100 PPM. At an increase of 2.5 PPM / year, which is 100 PPM in forty years, this translates into an increase of slightly more than one degree every forty years. As we discussed previously, these linear predictions are conservative, because the trends of the 2010s are increasing.

Now, there's not only CO_2. We can also wonder about about other greenhouse gases, since we have data for those as well. The easiest thing to do is to lump them up in the carbon dioxide equivalent CO_2e. We can download data and compute CO_2e as the sum of the contributions of CO_2, methane and nitrous oxide (see the previous chapter). The following rough sketch represents how the plot comparing CO_2e with temperature looks like, compared to the the one with CO_2.



We can see that the plot has a shallower slope. As a consequence, our back-of-the-envelope estimates do not change much if we use CO_2e instead of CO_2. The emissions increase, but the temperature increase per emissions is a bit less. The slope from this plot is about 0.7 ^{o}C /100 PPM. I get for the recent years an increase of 2.7 PPM/ year, which means an increase of 0.95 degrees in 50 years. In part, we can expect this to be a natural result of our definition of CO_2e, since CO_2e is CO_2 plus something, while each year the temperature is the same in the two scatter plots. Hence, each of the CO_2e points can only move to the right in the plot compared with its CO_2 "twin" from the same year. The slope changes because in the recent years the emissions of other greenhouse gases increased more compared to earlier. The strength of the correlation is about the same as in the previous plot. We can also check whether the contribution from methane and N_2O only is coupled to temperature anomaly. This is not shown in our last sketch, but the answer, readily obtained from a scatter plot of the temperature anomaly versus these quantities, is yes. So, as expected, the data tell us that the other two greenhouse gases also contribute to warming, although our estimates indicate.

Despite the limitations of my rough modeling attempts in capturing the quantitative details of the data, we see that they work in getting us the rough numbers, converting our emissions to atmospheric greenhouse gases, and then converting the amount of greenhouse gases into expected global temperature changes. This has the important conceptual consequence of indicating that the temperature changes during the industrial age can fully (or almost fully) be explained by the changes in greenhouse gases. We do not really need to search for other concurrent explanations.

We can ask whether it is possible to produce a quantitative estimate relating greenhouse gas emissions to temperature based on science, and not just on empirical observations as we did. This was done by Svante Arrhenius (1859-

1927), a Swedish scientist with exceptionally broad views. Arrhenius was a chemist, but he explored the interdisciplinary territory between chemistry, physics and mathematics. It was this combination of talents that led him to a great achievement: the construction of a quantitative mathematical analysis of the influence of CO_2 on the Earth's energy budget, culminating in the publication of his famous paper, "On the influence of carbonic acid [CO_2] in the air upon the temperature of the ground". He came to these considerations because he was interested in implicating variations in atmospheric CO_2 as a cause of the ice ages. The calculations involved the balance between the radiated heat (solar radiation arriving at the Earth's surface) and the subsequent absorption of re-emitted infrared radiation by the atmosphere. Arrhenius used measurements of absorption by water vapour by the American physicist Samuel P. Langley (which were more precise than Tyndall's). Based on these, he proposed a relation between atmospheric carbon dioxide concentrations and temperature. He found that the average surface temperature of the earth is a result of the infrared absorption capacity of water vapor and CO_2. Arrhenius suggested a doubling of the CO_2 concentration would lead to a 5^{o}C temperature rise (in fact, Arrhenius's estimate of climate sensitivity appears too high compared to today's figures). Arrhenius was the first (with Thomas Chamberlin), to claim in the 1890s that human activities, through fossil fuel combustion, would result in enhanced global warming, by adding CO_2 to the atmosphere. In a 1896 lecture, Arrhenius observed that a doubling of CO_2 levels would occur three-thousand years in the future, based on the then-current rates of burning fossil fuel (which were about one-thousand times smaller than today). Clearly, it didn't look like an urgency at that time.

REMOVING CO_2 and other greenhouse gases from the atmosphere seems the most sensible way to mitigate the greenhouse effect. However, the work we just discussed by Arrehnius and other scientists suggests another way, which is to interfere with the flux of heat from the sun that is absorbed in the first place, by trying to reflect back more sunlight. These methods are studied since the 1970s and go under the name of solar geoengineering, or solar radiation modification. They are different from the carbon capture technologies discussed in the previous chapter, because they do not act to reduce atmospheric CO_2 and other greenhouse gases from the atmosphere. Crucially, they are a *temporary measure*, because the moment you stop reflecting sunlight back, the atmospheric greenhouse gases will continue doing their job as usual. Equally, solar geoengineering would not stop ocean acidification and other consequences of the high concentration of greenhouse gases. As such, it is the Earth-climate equivalent to taking anti-histamines for allergic rhinitis (hay fever). Everyone that has hay fever or other allergies (like me) knows that anti-histamines might relieve the symptoms but they do not cure the cause. Allergens trigger an immune reaction, which leads to the release of histamine; this produces the vasodilation leading to a runny nose. With anti-histamines we block histamine receptors, but the immune reaction still goes on; we may feel it by other symptoms, and the runny nose will surely come back the minute we stop taking drugs. Thus, geoengineering can at most be seen as a complement to reducing our greenhouse gas emissions, but it cannot be regarded as an alternative.

One can reflect sunlight in simple ways, for example by painting the roofs of buildings in reflecting colors (white is the most reflecting). Can this reflect enough light to really be of help? It is difficult to give precise estimates, but we can try to keep things simple and fix our ideas on the orders of magnitude. The estimate involves the energy balance between absorbed and reflected light from the Sun. The Earth reflects on average about 30% of the light it receives (in technical jargon the Earth albedo is 0.3). However, this coefficient includes reflection from clouds (which are very reflective) and surface (where reflectivity varies a lot). The mean balance of irradiated energy is known. The global annually averaged amount of solar irradiance per unit surface is about 340 W/m^2 (Watts, which are energy per unit time, which is how much energy in Joules arrives in a given time interval, in seconds, all per square meter), and about 200 W/m^2 of this radiation gets to the Earth's surface. At the atmosphere level, about 50% of the radiation that does not get to the earth's surface is reflected, and the the other 50% is absorbed. Finally, about 15% of the radiation that reaches the earth's surface (so an average of 30 W/m^2) is reflected there.

For the surface of urban areas we could take, roughly, 1% of the planet (it is about 3% but not all of it is available for our white roofs operation). So if we put it at the average surface reflectivity it can contribute for 15% of 2 Watts per earth's square meter (1% of the 200 W/m^2 quoted above), 0.3 W/m^2, and urban areas could reflect 2 W/m^2 if they were perfect mirrors, with a net gain of 1.7 W/m^2. Note however that white roofs also mean that in winter you need more heating. In colder areas, if you heat buildings by fossil fuels the total contribution of white roofs may also become detrimental. In hot areas, you can also gain a bit more by using less air conditioning.

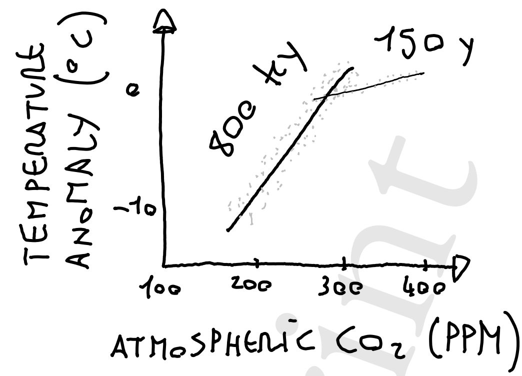
So we can estimate that maybe only on a third or so of the effective surface of our planet it would be actually convenient to repaint all roofs, and we get about 0.6 W/m^2. This works for perfect mirrors, so maybe we could also reduce it by an extra factor. We can compare this quantity to the total radiative forcing due to greenhouse gases, which is about 3 W/m^2. It is not totally negligible but clearly far from the total (and probably our estimate is still on the optimistic side). I think the current IPCC viewpoint is that cool roofs may help by giving a small contribution. Green roofs are also proposed, which reflect less, but have other benefits. In this very simplistic estimate, we have also disregarded any feedback effects from the Earth-atmosphere system (which are beyound our reach to evaluate, and arguably not sufficiently understood by the scientific community, see below). Space-based mirrors that reflect sunlight were also proposed, but this method appears (at least with current technology) to be costly and not easily scalable.

Other propositions involve playing with the sunlight that gets reflected at the atmospheric level. They are called “marine cloud brightening” (whereby, for example, clouds are sprayed with seawater) and “stratospheric aerosol injections” (whereby, for example, sulfate molecules added in the stratosphere would help reflecting incoming light from the Sun). As we have seen, clouds reflect a lot of sunlight. The idea of marine cloud brightening is to add a spray of salt from seawater (though other aerosols are considered) to forming clouds. This would favor cloud formation, incrementing the reflected light. This process would be more effective over the ocean, where normally there are fewer forming clouds. The general role of aerosols in cloud formation is well understood. However, aerosol-cloud interactions are quantitatively uncertain and complex to account for in climate modeling.

The effect of sulfur stratospheric injections parallels the cooling that happens after volcanic eruptions. The catastrophic Yucatan asteroid impact produced the same effect, a winter lasting tens of years, by launching enormous quantities of sulfur into the stratosphere, creating an artificial winter (before the massive long-lasting global warming), which might have played an important role in the global mass extinction. Alumina, calcite and salt are also considered as aerosol alternatives, and the main proposed method to deliver the aerosols is using modified airplanes or balloons. Economically, there are arguments that a large-scale deployment may be feasible. Climate models indicate that these methods may be capable of affecting temperatures at both the local and global temperatures, and that they might do it quickly. However, models are not experiments, and a major problem seems to be that our knowledge on existing natural aerosols in the stratosphere is limited, so that very little is known empirically and experimentally about the possible global effects of these methods. As the net solar radiation decreases, the drop in global temperatures might result in erratic weather patterns, such as increase in the intensity of droughts and floods. The induced temperature and climate changes may also impact adversely the cultivation of crops, affecting food sustainability.

Hence, global geoengineering projects could alter the Earth-climate system in unintended and unforseeable ways, and as such, they encounter fierce opposition within the scientific community. There is a strong consensus that solar-geoengineering research brings risks and that other ways to address global warming appear more promising. Going back to the parallel with anti-histamines, if I take these drugs, I know that they have been extensively tested in stringent clinical trials, and I know that the side effects are limited: I can get headache, sleepiness, reduced clariy of mind, but I can count on not getting hallucinations, heart-failure, or seizures. Today, we cannot set clear boundaries in a similar way with the "side effects" of solar geoengineering techniques. However, while mainstreaming these techniques appears very risky, this should not discourage further research. On the contrary, it can be argued that we need more research, if only to provide solid evidence that geoengineering is harmful, and to which extent, or alternatively to provide clear bounds on its applicability.

LET US NOW LOOK BACK on a scale of 800 ky. We can wonder whether and how the coupling of CO_2 to temperature acts on these much longer time scales, witnessed by ice core data. The previous chapters told us that there have been huge oscillations in both quantities, which looked very similar. We can now join the data and try get a feeling of how much, compared to recent years. Once again, for each range of years corresponding to a ice core layer, we have an estimate of global temperature and values of greenhouse gases, and we can draw a scatter plot using all these points. CO_2e does not make too much sense on this larger time scales, because it is defined based on the warming effects on shorter time scales. Hence we consider only the plots for CO_2. Here's a sketch of how it looks like, compared to the other one.



You have probably already noticed earlier that the range of temperature anomalies is much larger on the 800 ky time scale, but the CO_2 range is (slightly shifted but) similar. As a result, the slope of the plot is much steeper on the longer time scale, almost 10 degrees per 100 PPM of CO_2. We conclude that the relationship between CO_2 and temperature anomaly is very different on the two different time scales. In particular, on the scale of hundred thousands years, lower amounts of CO_2 corresponded to much larger changes in global temperatures. In brief, our rough back-of-the envelope approach will not extend to the distant past. And, on a positive note, we are lucky that the slope connecting CO_2 and temperature changes is so shallow in the last 150 years!

Can we rationalize why this is possible? First, the scatter plots which we have used only assess correlations. They cannot look at the presence or effect of other variables outside of the two being explored. We can conclude that other variables are probably important: we already know that oscillations in temperature can at least in part be attributed to an external cause, the different amounts of heat from the

Sun due to changes in the Earth's orbit. Our plot tells us that these also corresponded to changes in CO_2 (probably with a complex cause-effect relationship). More generally, the data show us that the tangled climate processes acting on different time scales might be quite different, so that, as we already mentioned, observation time scales are very important in comparing old and recent data. For example, while the Yucatan asteroid impact probably created a similar situation, we have no data from such distant past with a 1-year resolution, on a time scale of 200 years (during which there was a major injection of CO_2 in the atmosphere), which we can compare with the recent past. Our best data has a data point every 4.4 ky, and probably an effective resolution of 20 ky. In practice, for scientists it may be easier to use climate models "trained" with the recent past to understand the end-Cretaceous mass extinction than vice versa. Additionally, correlation does not tell us about cause and effect. Causality in climate is not so straightforward on very long time scales: CO_2 can influence global temperature, which in turn changes the planet weather, affecting the emissions. These scary feedback loops can lead to different stable long-term scenarios for the climate on our planet, that are very different from the current one (but may be similar to the climate of tens of millions years ago) as we will discuss a bit more in detail in chapter VI.

So what is the cause of climate change? A question that most of us have is how are we so sure that it is really anthropogenic, so whether we are really the *cause*. However, we have mentioned before how establishing causality is very difficult. Correlation is not causation. Many times this doesn't matter, but in our case it matters a lot. If climate change were not anthropogenic, then maybe all our attempts to fight it would be futile, and we should just devote all our efforts to living with it. So we care greatly about causality. If we mistake correlation for true causation, we could end up wasting a lot of resources and efforts.

So how can we measure causation? In the past chapters, we have worked around this question, but never faced it directly. First, we established that there is a significant change in global temperature, beyond its natural variation, beyond its trend in the past thousands of years. We can also argue that such a sudden jump is not compatible with the 200-year steps of global temperature that we can observe in the past record. Then we have followed the same line of argument for atmospheric greenhouse gases. Finally, we have seen how atmospheric greenhouse gases and temperatures are strongly correlated (both in the industrial era and on longer time scales, but with very different trends). All these findings are consistent with an interpretation that sees global temperature changes as a consequence, but they are not a proof of causality.

Generally speaking, to establish causality you must have the following three things, jointly: (i) the cause must come before the effect in time, (ii) the relationship between cause an effect cannot occur by chance, and (iii) there are no other unaccounted variables that explain the relationship. In a laboratory setting we'd be able to state that a variable is causally related to another variable, because we can perturb the system, and make different realizations of it. For example, we could use a randomized controlled experiment, where you have two groups, and only one gets the treatment, while the other group does not. If the two groups can be considered equivalent, and the group that gets the treatment reacts positively, then we know there is causation between the treatment and the positive effect that we observe. Unfortunately we don't have two copies of our planet to try this. However, we can argue that the earth 1000 years ago was our replicate. Everything was very similar to now, for sure on the astonomical level, except for our emissions, and the trend in temperature. We can go back to the plots shown in chapters I and II and work out our comparison. This argument relies on an assumption, the fact that the earth in year 1000 was to a sufficient extent "equivalent" to the earth at the beginning of the industrial revolution, but supports causality. We still have to check that an increase of global temperature did not cause all the greenhouse gas we now find in the atmosphere (this is easy because we know we have done it), and that there isn't another "upstream" factor that caused both temperature and atmospheric greenhouse gases to increase (but currently we have no candidate explanation).

But there is something else, which I find even more convincing (and is in general a powerful tool of science), which lets us argue strongly towards causality. This is the fact that we know the basic physical principles underlying the links between from experiments performed in the laboratory, where perturbations can be applied, and causal determinants can be tested carefully, one by one. We have spent some words on the experiments of Arrhenius and his predecessors in this chapter, and of course there are a lot more laboratory-scale experiments that have been carried out since Arrhenius, by which we know to a great level of detail and causal hierarchy the greenhouse effect. The assumption here is that a laboratory-scale experiment can teach us about planetary-scale behavior. This is of course another assumption, but physics has shown us that this can be the case (remember the old story of Isaac Newton, the apple, and the planets?) So yes, we can be convinced that global warming is caused by anthropogenic greenhouse emissions in the same way that we are convinced that an apple falls from the tree following the same law by which our planet orbits around the sun.

DATA SCIENCE TAKE-HOME MESSAGES. This chapter is centered on the concepts of correlation and causation. First, we learned that a scatter plot is a way to investigate the interdependency between two variables. Correlation is a measure of a mutual relationship between the two variables. We have used Pearson correlation, which is one of the most used measures of correlation, but also hinted at its limitations: it works well when data points form a single cloud without holes around a linear relationship, but it fails for nonlinear relationships such as a ciambella. We can imagine that a Pearson correlation is the result of attempting to draw a straight line best representing the data in our scatter plot, and the Pearson correlation coefficient, r, reflects how distant are the data points from the line. We also noted that the Pearson correlation reflects the strength of a linear relationship but not the slope of that relationship (in our case $r = 0.72$, and has no units, but the slope is 1.1 ^{o}C/ 100 PPM). Subsequenlty, we discussed how measuring correlation does not imply any cause-consequence relationship between the measured variables, we understood that it is much more difficult to detect causal relaitonships in data, and we addressed the arguments in favor of causation in the case the an anthropogenic effect on climate change after the industrial revolution. Finally, we used Fermi-like estimation techniques on white roofs, and discussed Geoengineering, which we can see as a way to alleviate the symptoms without removing the cause.

SOURCES. The sources for the CO_2 vs temperature plots are Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, Earth Syst. Sci. Data, 12, 3469-3479 (2020) and Oak Ridge National Laboratory, Carbon Dioxide Information Analysis Center, Scripps Institute of Oceanography CO_2 program, and the U.S. Energy Information Administration, International Energy Statistics, accessed December 7, 2020. The article by Westerhold and coworkers, Science 369(6509):1383-1387, (2020) provides a high time-resolution data set for climate data up to 66 million years ago. The article by Chiarenza and coworkers PNAS, 117 (29) 17084-17093 (2020) analyzes the climate effects on dinosaur habitats caused by an impact winter triggered by the Chicxulub (Yucatan) asteroid. The article by Thomas Anderson and coworkers, Endeavour 40,3 178-187 (2016) contains a detailed account of the understanding of global warming since Arrhenius. IPCC has issued another interesting and detailed history of climate-change science (Le Treut, H., R. Somerville and coworkers 2007: Historical Overview of Climate Change. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change), which can be downloaded at ipcc.ch. To learn more about geoengineering, Wikipedia provides a good starting point, `https://en.wikipedia.org/wiki/Solar_geoengineering`. The unsigned editorial on Nature 593, 167 (2021) argues that research into solar geoengineering should be preserved, and points to further literature. The estimates on the avearage global Earth's radiation energy balance come from Kiehl, J.T. and Trenberth, K.E. Bulletin of the American Meteorological Society, 78, 197-208 (1997).

Chapter IV

Temperature increase is tied to weather changes

What are the consequences of global temperature change? In the previous chapters, you saw me trying to work my way (on the back of an envelope) through how global temperatures are rising, how greenhouse gases are rising in our atmosphere, and how the two things are connected. Next, I wanted to set my hands on on how the global temperature increase affects more local weather patterns and other processes that are closer to our daily lives. If global temperatures could increase without any consequences, we'd be free to burn all the coal and oil we wanted, and enjoy the warm weather. Instead, our everyday experience increasingly suggests us that things are different, and that global warming has deep consequences on our weather.

We are now at the point where these consequences are very noticeable. Summer 2022 in Italy was the driest in 40 years, and even before the summer most regions hardly saw any rain in months. The drought has led to the loss of a substantial part of the crops, and many farms face the possibility of bankruptcy. We'd like to check, using evidence from data, whether events like these are not a coincidence, but they are significant consequences of climate change.

Especially in recent years, we've seen news of gradual and extreme changes in our weather. Intuitively, and based on elementary science, we know that some link must exist. Warming leads to evaporation, which may lead to droughts and heavy rains or snows. Additionally, extreme heat events should have a straightforward link with (global) mean temperature. But a devil's advocate could tell us that this need not be the case. In practice, "attribution" of specific events or classes of events to climate change may be quite difficult.

For example, to address how climate change has affected the likelihood of an extreme storm, or the magnitude and probability of having extreme storms in northern Italy in the summer, one would have to take into account many factors. First, the natural variability of such an extreme event is hard to evaluate. Extreme events are rare (almost by definition) and statistically difficult to address. Second, the links between global temperature and storms may depend on local phenomena and climate details, like circulation, or El Niño. Finally, there may other factors that may make us label a storm as extreme, for example, poor construction or poor land or river maintenance may increase the damage caused by a storm of a given intensity. Additionally, there are kinds of events that may have more indirect links with climate change, and many different concurrent causes. For example if 70% of wildfires in Northern Italy are caused by humans (intentionally or not), this factor should not be correlated to climate change. If some of these people become more environmentally aware (or less dishonest) in the next ten years, the trend of wildfires may decrease, while global temperatures will likely continue to increase, as we have seen.

Most importantly, it is difficult to set fair terms of comparison and agree on what is extreme and what is a proper "negative control". Let's define a scientific control. Controls are designed to rule out alternative causes and alternative explanations of a given results, and to minimize the effects of confounding variables. Typically they are a parallel experiment carried out in different conditions, but as we will see the concept does not really require being able to design and perform experiments, and is valid in data analysis, and even for theory. If we apply a proper set of controls, our results appear more reliable. In experimental science (particularly in life sciences), properly defining controls is considered essential.

In particular, a negative control is a condition which we can reasonably expect to produce a null result or no result at all. For example, we may expect that a specific ingredient could be essential for the result of one experiment. Hence, to validate our interpretation, we can perform the experiment leaving out the ingredient, and use it to verify that it does indeed produce a null result, confirming our hypothesis. In the design of an experiment, there are several ways to define a negative control. For example, we can leave out an essential ingredient, inactivate or remove a hypothesized causal factor, or test for an effect that would be impossible by our hypothesized mechanism. A simple example is administering a placebo to a sub-group of patients in an experiment on the effects of some drug. In theoretical or computational studies, we can still apply the concept of scientific control, and in particular of negative control, by performing calculations without certain ingredients and comparing them with the original ones. In observational studies, such as when we look at natural data, we cannot repeat the experiment, and we do not have the ingredients under control, but we can only observe and measure what happens around us. Consequently, finding a proper control may be more tricky, because the design is not accessible to us, but we can only observe natural phenomena. However, in some cases, we can define proper controls by comparing different phenomena that are similar but differ by some ingredients and/or by some effect.

LET'S FOCUS ON THE CONSEQUENCES of extreme events. During the recent years, we have all experienced, directly or through the news, an increase of reported extreme events such as wildfires, storms[I], extreme temperatures and droughts. The frequency, the duration and the intensity of these events appear to be changing, and we would like to firmly establish these trends using the data. However, we have also discussed why the science of weather attribution is difficult. We might make our life easier if we focus on events that are expected to be more related to temperature (extreme heat and cold, drought, etc.), and possibly even easier if the physical principles linking this kind of event and global temperature changes are direct. But in any case, we have to establish a plan of action for our analysis.

Usually attribution proceeds by two main methods. The first is using climate models. Extreme events that are sufficiently well known from the physics, and chemistry viewpoint can be simulated by computer models, especially those more directly connected to temperature (e.g. heat waves). Scientists can run identical climate models under

[I]As a matter of fact, storms can be be very complex. For example, an increase in the frequency of storms in Italy does not emerge significantly from the data. In other parts of the world, however, this signal is seen much more clearly. The reason why the signal does not emerge appears to be linked to the greater probability of having so called "anticyclonic circulation", circulation of winds around a central region of high atmospheric pressure, which in turn is related to general circulation rules in the atmosphere and global warming.

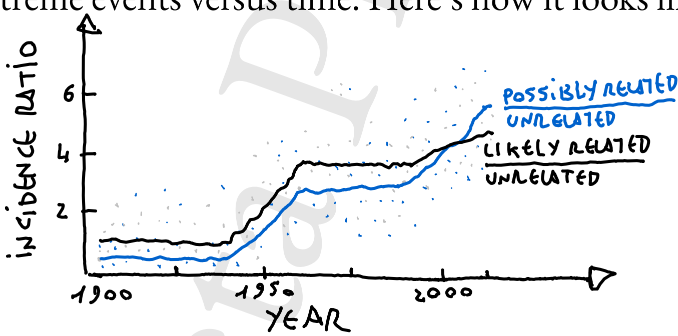
two scenarios, a positive and a negative control. In the first simulation, the negative control, greenhouse gas concentrations are kept constant at some level from the past, for example before humans started burning fossil fuels, and the climate model is run over some period. In the second positive scenario, the climate model includes the data from the actual historical record of greenhouse gas levels, and the ingredients to link it to our process or event of interest. By comparing the results from the two scenarios, scientists can estimate (and quantify statistically) how much human emissions have determined a change. This method is very effective, but clearly hard to access for the non scientist, and even for the non specialist scientist. It is clearly beyond our reach here, we would like something more accessible.

The second method relies on using observational records, and defining proper observational controls. For example we could make statistical comparisons between a carefully defined "before" and "after" used as positive and negative control. This seems compatible with our back-of-the-envelope approach, and I tried something of the kind myself. What I did does not use a "before versus after" comparison, but attempts to define another kind of negative control. Here is the recipe that I followed. First, I got myself a record of extreme events, from the Our World in Data website. This record comprises the numbers of extreme events of different kinds reported globally, by year, from 1900 to 2019, divided in classes that are more or less likely related to global temperature changes.

I divided these classes into three groups, that (to my arbitrary taste) looked like they could be likely related, possibly related or clearly unrelated to climate change. The "likely related" class included wildfires, extreme weather, extreme temperature, drought. The "possibly related" class included: mass movements, landslide, floods. Finally, the "clearly unrelated" class was made of volcanic activity and earthquakes. I figured that volcanic eruptions and earthquakes cannot be the consequences of climate change due to greenhouse gas emissions or at least this seems very unlikely to me. However, the record of these volcanic activity and earthquakes increases in time in my data, especially from the 1960s. I figured that this was probably because of increased awareness, organization in collecting the data, media and scientific tools. All factors that I needed to control for in my analysis. The total amounts of my clearly related, and possibly related classes of extreme events also increase in time, and this could come at least in part for the same reasons. Hence, I can imagine that there is a natural "observation bias" in these reports that makes them all increase in time. To attempt to compensate for this bias, I used the clearly unrelated group as a negative control. In other words, it seemed like a good idea to use the summed counts of volcanic eruptions and earthquakes as an "offset" for the events that I want to score against global temperature increase.

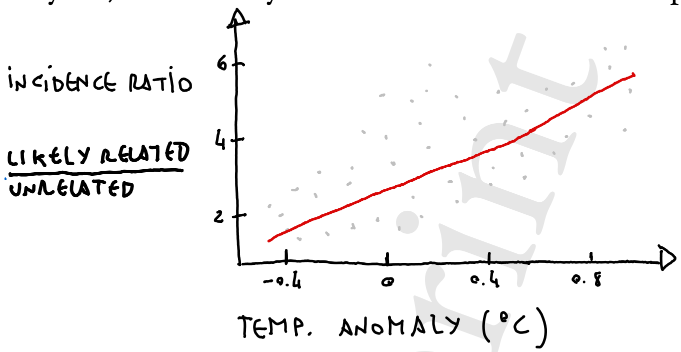
With my positive data and negative control, I could ask the following questions. First, do the extreme events in the likely related class really increase when compensated against the awareness bias (quantified by my negative control), and how much? Second, does the possibly related class behave like the negative control (hence I would be drawn to conclude that the events that make up this class are not related to global warming, or at least they cannot associated with it based on my data), or does it deviate (hence I would be drawn to change its label)? Third, does the frequency of events, once compensated, correlate with temperature anomaly and how strongly? The last question has to do with quantifying how much "disaster per temperature change" we can expect to find.

To answer the first two questions, I plotted the ratio of likely to totally unrelated and possibly to totally unrelated extreme events versus time. Here's how it looks like.



The points are very sparse, and the solid lines show the trend, which is obtained by averaging the data over a 25-year time span (hence their time resolution is 25 years). The black line refers to the the ratio of likely related events to the clearly unrelated ones, and it unquestionably shows that it increases after 1950, by a factor of about six. Thus, the awareness bias that (likely) makes the counts of earthquakes and volcanic eruptions increase in time is insufficient to explain the increase observed in reported wildfires, extreme weather, extreme temperatures, droughts, and we are led to believe that these have been really increasing. Even more interestingly, if we look at the blue line, we can examine the events that I previously classified as possibly related. These follow a very similar trend (and in recent years become comparable and even surpass the likely related events in their total offset counts). This means that, if we trust our assumptions, we would conclude that the increases in the total counts of extreme mass movements, landslides, and floods (after the 1950s) are also a likely consequence of global climate change, or at least they are independent. If we were to perform a "before versus after" analysis, we could compare the y-axis coordinates of the data points between 1900 and 1950 with those after 1980, and notice that they have a very different range. On average their ratio is about four, but the range of each of the two subsets of points is only about two. Finally, an interesting feature of the data is that, on average, they plateau between 1960 and 1987, which means that, in this observation range, an increase of these events would have been difficult to detect. The reason for these plateaus is unclear. Possibly the plateaus are due to a change in reporting customs (maybe due to lower awareness or different priorities), or to natural fluctuations (a simple coincidence).

To answer the third question, I used the time series of the temperature anomaly from chapter I and produced a scatter plot of the ratio of likely related to clearly unrelated extreme events (of a given year) against temperature anomaly (of the same year). Here's my hand-drawn sketch of the plot,



This plot shows a clear correlation (r=0.58, which becomes 0.72 if I join the classes of likely related and possibly related extreme events, as the previous analysis would suggest I should do). The slope is such that an increase in global temperature by one degree would lead to a four- to five-fold increase in the overall chances of an extreme event in any of the classes considered. If we want to extrapolate the trend (there are obviously a lot of caveats, but let's do it just for the sake of argument), our naive model would predict that the frequency of extreme events could be eight to ten fold higher compared to pre-industrial levels, if we hit a two-degree increase in global temperatures (which we will likely do). This does not look like something we can underestimate.

Obviously, one has to be careful with negative controls, and some things might have gone wrong in my simple analysis. Possibly my control set of earthquakes and volcanic eruptions was actually decreasing in time because of unknown and unreported geological reasons. In this case, I would be underestimating the awareness bias and overestimating the offset counts of my likely related and possibly related events. Additionally, awareness could have grown at different rates for different phenomena. Volcanic eruptions and earthquakes were probably reported in a fairly thorough way already at the beginning of the 1900s, while a flood or a landslide may have raised a much lower level of attention at that time. Another possibility is that expanding settlements are related to the incidence of these extreme events. If increasingly more people went to live in the floodplains, then flooding events that would not have affected anyone earlier on could be perceived as disasters (more precisely for our negative control to be ineffective we would need that more people went to live in the floodplains than closer to volcanos). One could react to this criticism by finding additional (non-geological) negative controls, for example to account for expanding settlements, and proceed with a more constrained analysis. We can end it here and move on, as I guess I have more or less reached my goal of getting a feeling of what it means to perform such an analysis.

If we turn our attention to "smooth" events, the more gradual changes are easier to notice. For example, ice melting at the poles[2] and in glaciers leads to oceans rising. There is NOAA data for global ocean levels, they have been increasing steadily for the last 40 years or so, by a total

[2] Note that melting of floating ice does not contribute directly to sea-level change, because it is already in the ocean water. However, Greenland ice makes a strong contribution.

amount of 125mm. This is about 3 mm / year. We can look at the global temperature changes of the same forty years period and translate it into a sea-level change of a bit more than 10 cm per degree Celsius. Note however that this naive estimate does not take into account that the response of sea-level change to temperature change occurs with a time delay, which can actually be quite long, because when the temperature chenges, it takes time for the glaciers to melt (and also for oceans to show thermal expansions).

We still would have to translate this into some local consequences, such as floods, by the use of attribution methods of extreme events. Ocean acidification is another easily measured smooth global change, with a lot of consequences, as we have already discussed. There is data for the global mean surface sea water pH, which has been decreasing at remarkably constant speed over the period 1985-2020, by about 0.002 units per year. Currently it is about 8.055, and at this rate it will be 7.9 in 2100. Data indicate that it was around 8.2 in the pre-industrial era, and since pH is based on a logarithmic scale, base ten, a 0.3 change in pH means a factor of $10^{0.3}$, which is about a factor of 2 in hydrogen ion concentration or "acidity".

Finally, as in the case of ocean acidification (and ocean temperatures), often weather consequences carry other consequences on ecosystems, such as extinctions, migrations, changes in the dominant species, or even loss of ecosystem functioning, which we also should put under the account of global climate change.

DATA SCIENCE TAKE-HOME MESSAGES. Besides using tools developed in the previous chapters (correlation, linear extrapolation, etc.) this chapter addressed the question of establishing a trend using a negative control or a null scenario. In our case, we used the incidence ratio of extreme events that (according to our knowledge) should be mostly unrelated to climate change, as a negative control for those that are likely related. We have seen how this simple strategy can be productive in terms of addressing and answering different questions, constructively, while being very careful of its limitations. One key aspect is that the choice of the control is always - to some extent - arbitrary. We also reviewed the typical data-science strategies used by attribution studies.

SOURCES. The problem of attribution is discussed by Mark Buchanan in Nat. Phys. 17, 978 (2021). The data on natural disasters was retrieved from Our World in Data at the URL `https://ourworldindata.org/natural-disasters`. The article appeared in 2014 and was updated in 2020 (Hannah Ritchie and Max Roser 2014 - "Natural Disasters". Published online at OurWorldInData.org). The data on sea level changes come from the NOAA climate.gov web site, and data on ocean pH can be found at the copernicus.eu web site.

Chapter V

Breakdown of emissions

This chapter is a summary (that should fit on the back of an envelope) of the sources of our greenhouse emissions, in order to try to begin understanding where we can start, if we want to reduce our emissions.

Let's start by an example. I mentioned in chapter II that yeasts emit CO_2 when fermenting in wine, beer, and bread. But, as it turns out, we should not worry too much about that for global emissions. The most CO_2 emitted in bread production may come from nitrogen fertilizers. For beer, ingredients and fuel/electricity used in the production should dominate. For wine, the bottle usually has a bigger footprint than the wine, and the road miles (but not the shipping) could have a strong impact. We immediately see how complex and unexpected things can turn if we try to reason on how to contain our emissions! I will try to report some rule of thumb principles below, but it is clear that computing precise carbon footprints is quite a challenge.

One systematic way to approach the question, but also to understand its complexity is called "input-output analysis". This analysis was developed (building on the work of classic economist Karl Marx) by the Soviet-American economist Wassily Wassilyevich Leontief (1906-1999), who was awarded the Nobel prize in connection with this achievement. The Italian economist Piero Sraffa (1898-1983), also developed the same theory independently. The basic idea of input-output analysis is the "production of commodities by means of commodities" (as put by Sraffa). We can imagine a very large input-output "table", which reports the goods that are necessary in order to build or produce other goods. Goods can be all sorts of things, for example we can imagine that if we feed ten hens with one hundred kilos of wheat, and we have one rooster, we will end up with thirty hens. This is not unlike a chemical reaction, so that we can represent the table also as a set of relationships like

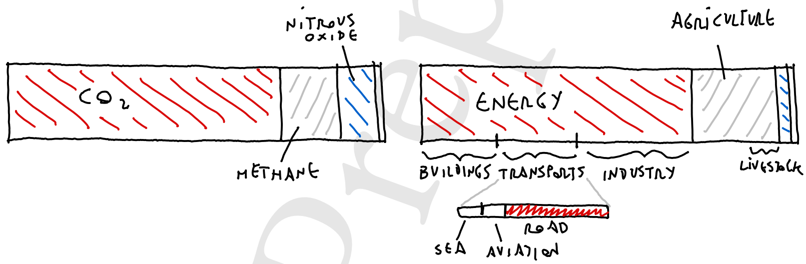
$$10 \text{ hens} + 100 \text{ Kg wheat} + 100 \text{ L water} + 1 \text{ rooster} \rightarrow 30 \text{ hens} .$$

The table contains the list of all possible production "recipes" and also their stoichiometry, which is how much of each ingredient you need to get the output.

In a second stage we can enrich the table with information regarding the energy and the carbon emissions of each production step. The carbon footprint of a good is then given as the sum of direct and indirect emissions, where indirect emissions are computed by telescopic sums involving the input-output matrix. This sum also relates carbon footprints to production activities and production sectors. For example, for our hens, we will have a footprint for the CO_2 produced by the hens, and a footprint from the wheat and water. In turn, the wheat will need water and energy and water will need energy, each with its associated carbon emissions, and so on. This reasoning also gives us a glance of how carbon emissions are intimately coupled to macroeconomic cycles

The idea of applying input-output analysis to estimating carbon costs dates back to the late 1960s and early 1970s (including works by Leontief himself), and seems to be the major "top-down" approach today. As you can imagine, the results you get can only be as good as the data you have in your input-output matrix. The problem then becomes estimating accurately production and transport constraints and habits related to each specific good, which can be very hard. The quantification of carbon footprints has been presented in several books, by people who are much more competent than me. In particular, for me it has been very instructive to read the book "How Bad are Bananas" by Mike Berners-Lee (his more recent books go even deeper, and other authors also address these questions). This information is clearly essential to determine our behavior as individuals, to drive policy making and globalize our efforts towards mitigating climate change. It is probably close to the core of the social question of climate change today. It is also beyond the reach of this book, which is less ambitious, and mostly aims at reading the data and raising some "back-of-the-envelope" awareness.

LET'S BREAK DOWN CARBON EMISSIONS using different subdivisions, on a global level. Here's a sketch with the main approximate figures



By compound, the figures are simple. During these years (2016-2021), have been emitting (roughly) about 50 gigatonnes (Gt) of CO_2e per year. Of these, CO_2 makes up about 75% of our greenhouse emissions (in terms of CO_2e). Most of it (about 65% of the 50 Gt) is from fossil fuels, but about 10% of our total CO_2 emissions is due to forestry and land use. Methane makes up 16-17% of our total CO_2e and nitrous oxide about 6-7%. The rest (about 1-2%) is fluorinated gases. This makes it clear that the main problem to solve is atmospheric CO_2 from fossil fuels. Deforestation, meat and rice are still an issue, but addressing these would only solve one-third of the problem at most.

By sector, and by processes, we can understand more. Energy takes up 75% of the total emissions: about 25% of the total 50 Gt is energy for industrial use, while buildings and transport take up about 16-17% each. It is instructive to look at the breakdown of transport. Most of the emissions (about 12 % of the total 50 Gt) are due to road transport (divided 60% -40% between private transport, such as cars, and freight, such as trucks), while aviation and shipping are around 2% of the total. Aviation emits just under one billion tonnes of CO_2e each year. Note that for a single person trip, aviation is by far the most costly in terms of emissions. Taking into account the extra effects of emissions at high altitudes could be 10 times more costly per kilometer than an efficient train, and twice as much as a car with one passenger. Thus, as individuals, we should still prefer the train. However, if we can trust these figures, we should conclude that on a global level we should concentrate on reducing *road* transport. Finally, rails look good, as they only take up less than 1% of the total emissions. There are also fairly large contributions from unallocated fuel combustion (about 8% of the total emissions), energy-related emissions from the production of energy from other fuels, and fugitive emissions (about 6%), which occur mainly during the extraction, transport, storage and processing of fossil fuels.

Agriculture is responsible for 18-19% of the total 50 Gt emissions. As a sub-category, livestock and manure makes up 6% of the total emissions, and rice cultivation 1.3%. The rest of this category is made of deforestation and cropland

(3.5%), crop burning (3.5%) and agricultural soils (4.5%). Finally, waste (landfills and wastewater) make up about 3% of our total emissions and cement and other chemicals add up to roughly 5% of our emissions. Note that there is a considerable "fragmentation" of the contributions into many "small" components, which means that policymakers will need to address many different problems one by one to contain emissions. For example, the agricultural and waste subcategories compare to, and surpass, global aviation and shipping. Nevertheless, one should probably start from the largest slices of the pie, which appear to be the ones related to energy (especially its use in industry and buildings, and for transport).

By country, China is well known to be the country with highest emissions (27% of the roughly 50 Gt CO_2e), but there are also important caveats. First, the *per capita* emissions (emissions per individual) are much less in China (about 7 tonnes per individual per year, comparable to European countries) than in the United States (roughly 15 tonnes per individual per year, surpassed only by Canada, Saudi Arabia and Australia). Second, China (and Asia in general) is the outsourcing venue for the production of many goods that are then shipped in the whole western world (for example, electronic equipment, clothes, sneakers etc.) As a consequence, a lot of the Chinese CO_2 production should not be regarded only as a Chinese responsibility. Adding the emissions of India (about 7% of the total), Japan (about 3.5%) and the other countries, we find that Asia adds up to producing 53% of the global emissions. Europe (as a continent, including Turkey, Russia and Ukraine) and North America produce about the same, 17-18% of the total each, while Africa and South America place themselves around 3.5% each.

Considering our personal carbon footprints, as individuals (if we live in developed countries, or in economically booming countries such as India), it is clear that we should adapt our behavior to contain emissions related to what we do and consume. However, it is not always easy to gain a clear intuition of what exactly we should do, and to what extent what we will do actually matters (which can be frustrating). What follows is a back-of-the envelope list of what I understood so far.

Road and air transports clearly matter, and limiting our car usage and airborne travel, preferring trains whenever possible will contain our personal footprint. Another considerable part of our personal footprint is probably the energy consumption from our home (heat and electricity). Reducing meat consumption can also be important. There are many nontrivial hidden aspects related to food consumption, due to the input-output relations: for example the CO_2 overheads from transport associated to the shipping of our food. Bottled water can have a heavy footprint if it is transported on long distances by road transport, because water is so heavy. Fruit from the other side of the world can be reasonable in terms of carbon footprint, if shipped by sea. However, they are not ok if shipped by air. Tomatoes produced in the Netherlands in January thanks to electrically heated greenhouses, then shipped to France are not ok due to the carbon footprint overhead from the greenhouse heating.

The tortuous sides of personal carbon footprints are not limited to food. For example, one might try to go paperless to reduce her/his carbon footprint, but there are comparable carbon costs related to internet and computing (storage, cloud, streaming). Some estimates put the carbon footprint of the internet and the systems supporting them to about 3-4% of global greenhouse emissions. Most of it should be due to streaming videos and cloud storage/computing. Electric cars are also connected to CO_2 emissions at the moment. One has to account for the CO_2 emission related to building the car, and also the CO_2 emissions of producing electricity, which depends on the country. In countries where electricity generation is mostly based on fossil fuels (coal, oil, gas), an electric car can have similar lifetime emissions to an efficient conventional vehicle, such as cars with hybrid-electric engines. This said, there are things that are incontrovertibly not emission friendly of course, like buying SUVs and using them for a long commute all the year long.

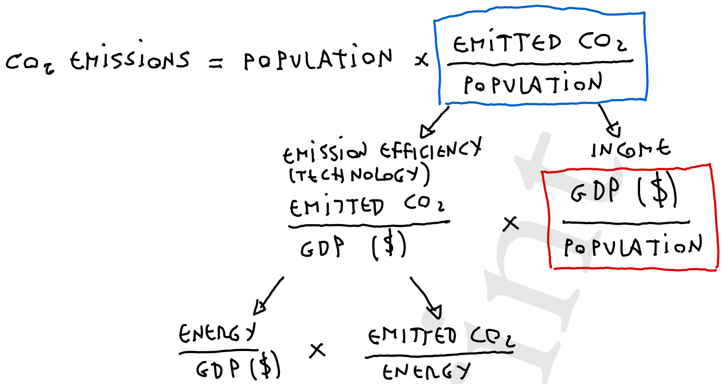
In this complex scenario, my personal feeling is that currently it may be (at least in many cases) difficult to "behave well" in terms of the emissions that we determine, even if we mean well, and that we should think carefully about our footprints. My best advice is to try to "place it in your head".

That's what I am trying to do at least. If you google things like "carbon footprint calculator" you could try to compute your footprint based on your behavior and identify where you can act. In other google searches you will also find many lists of possible actions that you can take, which are typically different from each other, and will likely confuse you (as it happened and happens to me), making it difficult to set clear priorities. A recent study by Ivanova and coworkers tried to evaluate precisely and rank the mitigation potential of several individual behavioral choices. Here's their shortlist, ranked by median[1] effect (1) live car free, (2) shift to EV (this has a large average effect but also a huge variability and some potential for backfire - for the reasons discussed above), (3) take one less long-haul flight per year, (4) shift to renewable electricity, (5) shift to public transport, (6) refurbish and renovate your house in order to make it more energy efficient, (7) switch to a vegan diet (a vegetarian or partly vegetarian diet also has some effect), (8) use a heat pump for heating, (9) improve your cooking equipment (10) use renewable energy-based heating. The authors claim that these "ten commandments" for consumption options together (accounting for the overlaps) yield an average annual mitigation potential of 9.2 tonnes CO_2e per capita per year, which, multiplied by the population of richer countries (where these options are possible and meaningful) could lead to a relevant effect.

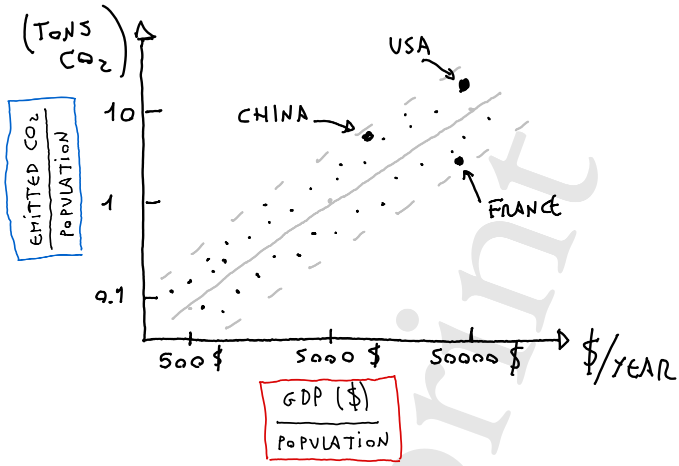
[1] median means, roughly, that 50% of the data lie above this value, and the other 50% lie below.

So individual behavior matters, but in my opinion the top priority is also to get this thing into the agenda of your politicians. It would be very helpful for example to be guided by our governments, starting from simple things, for example by forcing whoever sells us a house, a car, a tomato, or anything else to give us clear information on the direct and indirect CO_2 emissions associated to building, shipping, using whatever good we are buying, or (equally useful) to force them to stay within certain boundaries. I am writing in the middle of an electoral campaign in my country, and, sadly, nothing of the sort seems to be on the table. Key problems to address (focusing on democratic systems, as non-democratic systems seem hopelessly in the hands of few capricious individuals) are that (i) governments are "spatially" local, in terms of circumscription and also because they are engineered to represent local needs (or, worse, needs of the niche of our society that manage to control them), and (ii) they are also "temporally" local, typically with a time scale of 5 years or so. Hence, politicians have a clear incentive to make life easier for their voters on these short time scales, in order to be reelected. However, to solve such global problems, we need bodies with vision on a wider time range and a much wider scale than a single country, because in the end we are all in the same boat. There are bodies for that, as we all know, but in the current equilibria their steering power seems to have some room for improvement. Re-thinking global democracy in a way that it has a global-scale vision and a long time range sounds like a good thing to do (also because it could pressure more effectively non-democratic countries), but unfortunately this seems hard to do on the back of an envelope.

GOING BACK TO GLOBAL EMISSIONS, there's a famous way of trying to break down the total emissions into different contributions. This is called the "Kaya identity" and was proposed in 1990 by Japanese energy economist Yoichi Kaya. The Kaya identity can be seen as another example of a Fermi problem. We have resorted to this technique throughout the previous chapters, so it should be familiar by now. But now we will bring this to another level. An essential step of working our way through a Fermi problem (or any problem, the point really is that this is a training ground for any modeling exercise) is decomposing the problem into simpler sub-problems on which the answer may depend. In some cases it is possible to further decompose each sub-problem into lower-level questions that are easier to address, and so on. In some other cases there are different ways to decompose a problem, each with its own instructive value. By comparing these expectations one usually achieves a better understanding of the constraints, characters, assumptions, and ingredients defining the question under investigation. Following this strategy and line of thought, the Kaya identity breaks down the total CO_2 emissions in the product of four factors, total population, GDP per capita, used energy per GDP, and emitted CO_2 per unit energy, as sketched below.



The clever part is that this time we are not focusing only on the final number as usual, but we also want to know the roles of the different factors in establishing the answer. In other words, by evaluating the different factors we can analyze the human impact on climate to a deeper level. By definition, the factors in the Kaya identity cancel out, but making them explicit is useful to evaluate the contribution of different factors to global emissions, in terms of available data, population, GDP per capita, energy per GDP, and emissions per unit energy. Equally clearly, there are other useful possible ways to decompose the problem, depending on the available data. Reasoning on these factors one can highlight the elements of the global economy on which one could act to reduce emissions. The plot below (which I have taken from trends observed in real data) makes an attempt at this kind of analysis,



When compared to the scheme of the Kaya identity, the plot adds an important contribution to the analysis of our decomposed factors. Each point is a country, and the data show that, across countries, CO_2 per capita is basically directly proportional to GDP/population. This means that poorer countries emit much less. However, the data points do not really cluster tightly on a straight line, but they look more like they lie within a "stripe" around a trend line. This tells us that the other contribution to CO_2 per capita in the Kaya estimate, CO_2 per GDP, is responsible for the scatter around the direct proportionality. In other words, being below or above the trend line is an index of the CO_2 emissions per income. France, for example, is found below the trend line. It has a higher GDP per capita than China, but it is more efficient in its emissions than China, which is less rich but has a comparable amount of CO_2 per capita. Conversely, the United States are rich but also above the trend line, highlighting the fact that they are less "emission efficient" as a country than others. Actually - since the plot is in a logarithmic scale[2], it also says that emission efficiency (the product result of using energy efficiently and using forms of energy that emit less CO_2) has a much greater impact for richer countries (and with greater impact come greater responsibilities)[3].

Another important consideration is that if a country is very poor, its emissions are much lower, regardless. For example, it is sometimes argued that population growth is a root of rising CO_2 emissions, but our analysis tells us that things are more complex than this. Note that the GDP in the plot is per capita (per person), so for example if we want to compare the total emissions of Africa and Europe, total populations also matter. However, if we look closely in the data (not shown in my sketch), the average per capita emission of in Europe is more than ten times larger than in Africa, but the population of Africa is less than twice as large as the population of Europe. In fact, on a global level, countries with lower incomes (where population growth is highest) make up a small share of total emissions. Instead,

[2] on a linear scale, the trend is visibly linear, but one sees well that the spread of the data around this trend increases with increasing x-axis values. You can imagine that the dashed lines for a diverging "V" shape around the trend. I chose to represent it as a log-log scale because it looks more tidy.

[3] In the past years, many western countries are showing a trend towards achieving economic growth while reducing emissions, thanks to replacing fossil fuels with low-emission energy, as well as increasing of GDP constant total energy use. This reduction in emissions in rich countries was not only achieved by offshoring production in (and hence transferring emissions) to countries like China and India. These statements can be verified by plotting the trend of per capita CO_2 emissions and GDP as a function of time for the past 20 years, and using "consumption-based" CO_2 emissions (as computed by the Global Carbon Project), which are adjust for exported production

high-income countries (with low population growth) take a big share of our global emissions. In particular, we can regroup the data, and find out that roughly fifty percent of the world's population can be connected to almost ninety percent of our global emissions[4].

Finally, we can observe that being energy efficient may not be sufficient for a country to radically reduce its emissions, it might also matter, in some sense, to be "poorer". An analysis of the past sixty years shows meaningful correlations between global GDP (the summed GDP of all countries) and global emissions, confirming the suggestion that there might be important trade-offs to understand and optimize between economic growth and greenhouse gas emissions. Clearly it's not just a matter of poverty, and this is of course the case in the current conditions (we already mentioned that some wetern countries are successfully reducing this caoupling). Development initiatives also play a role. It may be that in the future value and added value can be associated with new processes that are carbon neutral or with low energy consumption. Moreover, if we achieve carbon-neutral energy production (also in developing countries), this correlation will break.

HOW MUCH RENEWABLE energy can we produce? This question is kind of specific compared to what we have discussed so far, but I thought it could be useful

[4] These analyses are not evident directly from the plot because I have not labelled the points corresponding to countries and theor populations, but I have performed the analysis for you.

to get an order-of magnitude idea of what can be achieved. As we have seen, a central issue is cutting fossil fuels, and carbon-free energy production would in principle solve many problems. In 2019, around 11% of global primary energy came from renewable technologies. The International Energy Agency (IEA) issued a report in 2021, on past, needed and projected rates of increase for renewable energy. They estimate that we need to double the rate of transition to renewable energy, in order to have any chance of achieving "net zero" (a balance between the carbon sources and sinks) carbon by 2050. The total current power in renewable energy production is estimated in about 3000 GW, and it is forecast to reach almost 4800 GW by 2026. We can compare it with the 2000 GW of coal power that we mentioned in chapter II. Most of the margin of improvement appears to be in wind turbines and solar power plants, so we will focus on these two. Hydroelectric power is the largest component currently, but its capacity to grow seems lower. The other renewable alternatives are small currently.

Let's start from wind power, and we can run a comparative estimate with coal power, along the lines of chapter II. How much electricity can a wind turbine create? It depends on the model and whether they are on land vs off shore, with a range that is 1-10 MW for the maximum capacity; we can take 3 MW as a value in the middle. But wind comes and goes, and every wind turbine has a range of wind speeds in which it will produce at maximum capacity. At slower wind speeds, the production falls off dramatically. If the wind speed decreases by half, power production may decrease by a factor of eight. On average, therefore, wind turbines generate annual outputs that could be 15-30% of their capacity. We can boil down our estimate to an effective 1 MW per turbine corresponding to 2 GWh of electric energy per year, more or less the same figures that we had for coal. In this case we can compare directly the power (in MW). A 1000 MW coal power plant (the one we considered in chapter II) then corresponds to about 300 turbines. The total power emitted by coal is 2000 of those. Suppose we want to replace the global coal production by wind turbines, we then need about 600000 turbines. In 2021, there was a 743 GW wind-power capacity worldwide corresponding to 250000 of our model turbines. We would need 350000 more to reach the the coal power. To see whether this is feasible, we could try to estimate how much surface is needed. To maximize output, turbines should be spaced in such a way that they capture the most wind while not blocking the flow of air from each other (in absence of other obstructions). The optimum spacing is usually between 8 and 12 times the rotor diameter in the direction of the wind, and between 2 and 4 times in the direction perpendicular to the wind. We can assume that the turbines need to stay in 10×3 diameters array, so that a square $600\times600 = 3600$ turbines occupies a surface of $600\text{ km} \times 200\text{ km} = 120000\text{ km}^2$. This is comparable to the surface of England, or of the state of New York, or North Korea. Would you sacrifice the entire surface England for wind power? It doesn't have to be England itself of course. The surface equivalent of England is a lot, because this needs to be unoccupied surface, where there is sufficient wind for turbines to work effectively, etc., but it does not look impossible on a global scale. It's not larger than the land surface of the planet, for example, and is not as large as a continent either.

If we think about photovoltaic solar power, recent data from the US National Renewable Energy Laboratories (NREL) shows that the national average output is around 70 kWh/m^2 for one year. With this figure, our estimate is easy, we just have to match the total coal energy for one year. In chapter II we estimated this that 2 TW would give 4000 TWh in one year with coal power plant efficiency. This would be equivalent to 60000 km^2 of solar panels. That's half of the surface needed for replacing the coal with wind turbines, and equivalent to the surface of Croatia or West Virginia. So again, unless the estimate is flawed somewhere, it does not look impossible.

Now, solar panels have a considerable energy cost (so that in some cases might take even a few years to break even with the energy needed to manufacture them) and a carbon footprint, which are both related to the manufacturing process, which costs energy; and if this energy is produced with fossil fuels, it comes with CO_2 emissions. The energy cost should be subtracted from our total, hence we would need to take an extra fraction of land surface of solar panels in order to produce the same energy as our coal power plants, say, over a life time of 40 years. This extra surface would also have an energy cost, which we would need to offset and so on. Fortunately, this kind of infinite sum is known to converge, but we need to be aware that the estimate above is likely a lower bound. NREL also analyzes in detail manufacturing costs associated with different photovoltaic technologies, considering their economic and energetic sustainability, so that many interesting details can be found on their web pages. Concentrated solar power (CSP) is another promising low-emission technology using the sun's energy. It simply uses mirrors or lenses to concentrate sunlight onto a receiver; electricity is generated from heat, for example by a steam turbine. CSP had a global maximum power of 6800 MW in 2021 and this figure is forecast to grow fast (but also to remain a small fraction of the total).

Regarding greenhouse emissions related to manufacturing, those are relevant also for windmills. The best way to compare technologies is to consider the full life cycle and average the emission over this period. NREL also issues these figures, from detailed comparisons of life-cycle assessments of the carbon production of different technologies for generating electric power. According to these figures, solar panels emit on average around 50 g of CO_2e per kWh, and windmills around 10 g CO_2e per kWh. Nuclear power plants are also estimated to emit on average around 10 g CO_2e per kWh. By comparison, gas electricity generation emits on average 400 g CO_2e per kWh and coal power plants emit on average 1000 g CO_2e per kWh, so we are talking about a difference of a factor of 10 to a 100.

Nuclear power (currently about 10% of global electric power), probably deserves a separate discussion, as it is the subject of a hot political debate. The figures above clearly show that the subject of the debate is not its greenhouse gas emissions. Rather, the criticism concerns other problems related to nuclear energy. What hits most people is the possibility of (rare but very large and dramatic) accidents, such as Chernobyl or Fukushima. Estimates show that the overall death toll of fossil-fuel energy generation is actually much higher; my feeling is that the rare but very big events related to nuclear power-plant failure may be very difficult to compare in a fair way with the more steady death toll of other technologies. Another issue is that the long-term storage of nuclear debris is (to say the least) problematic. However, possibly the most important criticism claims that nuclear power has limited scalability. For example, using current figures, it does not seem unrealistic to say the current (known) reserves of uranium would last us about 200 years at the current rate of consumption [5]. If we were to use ten times as much nuclear power (hence converting all our production in this direction), our fuel would last us only 20 years. By contrast, we can count on the fact that wind and solar power will not run out, hence they might be a better investments for a country.

[5]This is also debated, since the unknown reserves may be much more. Also, there are fast-breeder reactors that can use uranium 60 times more efficiently, but this technology is not fully ready today and it is not very clear if it can ever work on a large scale. Finally, there is a lot of (known) uranium dissolved in the oceans, but currently we have no efficient ways to extract it.

While this subject is highly debated, it seems fairly clear that nuclear power cannot provide by itself a global solution to climate change (also *because* of the hot political debate around it). On the other hand, while nuclear (fission [6]) power may not the horse to bet on in the long run, it is seems almost certain that it is a necessary interim solution for the transition to low-emission energy production. For sure it makes sense to leverage on the existing infrastructure (possibly investing on renewal, maintenance and safety).

To conclude, we should note that a big problem with renewable energy from wind or the sun is intermittency. The sun gives a roughly 12h downtime every day (more in winter, especially at high latitudes) and is not effective when the weather is bad. Equally, wind can be intermittent as well as unpredictable (this depends on the exact site). One possible solution is keeping a certain amount of fossil fuel, hydropower, and/or nuclear power generation in order to maintain the continuity. Another complementary solution is to improve the storage and transport infrastructure, in order to be able to transport the excess energy to far away places or to store it for future times when needed. Electric energy storage comes in many forms, but is never very easy. At the moment it seems that batteries are the most common solution to the intermittency problem. Lithium ion batteries can be made very big for example. Australia is adopting a large-scale lithium ion battery technology combined with

[6] Nuclear fusion, when available, could be a very different story, but it seems likely that this will require at least tens of years.

renewable generation, including a famous 100 MW battery built by Elon Musk's Tesla in South Australia, whose main purpose is to prevent blackouts due to the discontinuity of the network.

Several arguments have been proposed to support the idea that the problem of intermittency is not insurmountable, but it certainly puts on the table a new set of complex problems compared to the current technology. For example, the effect of geographic diversity may cause an effective cancellation of positive and negative fluctuations. In other words if a grid is sufficiently large and well connected, it will include at any given time regions where it's day and night, where the wind is strong or still, where it's sunny or cloudy, each of which could send energy to the other regions when it experiences a positive fluctuation. Provided this is the case, it can still be quite difficult to predict the needed routing for the next, say, few hours, in a way not to cause transient blackouts. This is too much for the back of an envelope; as a matter of fact, this looks to me like an interesting problem where the statistical physics of complex systems (my area) might have something to say.

To elaborate on all these energy-related topics, the 2008 book by David MacKay "Sustainable Energy - Without the Hot Air" (free on line at `https://www.withouthotair.com/`) is arguably the golden reference. It contains many quantitative estimates and careful evaluations of sustainable energy scenarios. It is mostly based on possible plans for Britain, but it also contains interesting estimates considerations, and suggestions for plans “that add up” for Europe, North America, and the whole planet.

DATA SCIENCE TAKE-HOME MESSAGES. This chapter is mostly about how data can be "sliced" in different ways to reason on different questions. The data structure that we used throughout the chapter is a "partitoning", the allocation of different fractions or percentages that make a whole, based on various classification strategies. Mixing this strategy of describing the data with Fermi estimation, we brought the use of Fermi estimates to a new level with the Kaya equation. In discussing this estimate, we started up with the usual plan of action of breaking down the problem into sub-problems. However, instead of using the sub-problems to get to the final answer, this time we indulged in studying the role of the different sub-factors. Specifically, we used scatter plots to learn about the correlative trends from each sub-problem. This fully formalized Fermi estimate was instrumental to analyze the determinants of some trends and also to reason on the "outliers", the points that do not follow the trends (and on why they do not). Additionally, we performed some standard Fermi estimates on the deployment scales of wind turbines, using the tricks established in chapter II. Finally, we have briefly mentioned input-output models, which are important in climate science as well as in economics.

SOURCES. The 2010 book by Mike Berners-Lee "How Bad Are Bananas? The Carbon Footprint of Everything" (London, Profile Books) now at its third, updated edition, gives the approximate carbon footprint many products and activities from sending a text message to a football world cup. The figures quoted for transport emissions come from the UK Department for Business, Energy and Industrial Strategy (link). The International Civil Aviation Organization (ICAO) offers an on line tool that calculates the CO_2 emissions from air travel (link.) See also Caserini *et al.* Ingegneria dell'Ambiente 6 2019 (DOI 10.32024/ida.v6i1.207). I have already mentioned the 2008 book on sustainable energy by the physicist David MacKay (`https://www.withouthotair.com/`). Physics World defined it "a book every budding physicist should read and perhaps also the one every working physicist would like to have written": I confess I would like to have written it myself. The article by Ivanova and coworkers Environ. Res. Lett. 15 093001 (2020) attempts to quantify the overall effects of different consumer behavioral changes on overall emissions. The "macroscopic" breakdown of current emissions presented here comes from Our World in Data, Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "CO_2 and Greenhouse Gas Emissions". Published online at OurWorldInData.org. The article can be found at this link The Kaya identity is presented on Wikipedia, at the page Kaya identity, and the empirical evaluation of the different contributions comes from the following articles on Our World in Data (links): "emission drivers" and "world energy". The link between economic growth and greenhouse gas emissions is also discussed in a 2021 Time magazine on-line article by Ciara Nugent and Emily Barone (time.com link) The IEA report on renewable energy can be found at the URL `https://www.iea.org/reports/renewables-2021`. The figures used for the estimates on wind turbines were taken from `https://energyfollower.com/` and the figures on solar power come from NREL (URL `https://www.nrel.gov`). The NREL web site also gives life-cycle assessments of the typical CO_2e emissions of different technologies for electricity generation (link), as well as manufacturing costs associated with different photovoltaic technologies (link). The paper by Pehl and coworkers Nature Energy 2, 939–945 (2017) uses life cycle assessment and modelling to show that indirect greenhouse gas emissions induced by wind, solar and nuclear power are very small compared with other emissions sources. Contrasting views on nuclear energy can be compared looking at the article by Hannah Ritchie and coworkers (2020) - "Energy". Published online at OurWorldInData.org (link), the article by D. Abbott, Bulletin of the Atomic Scientists, 68(5), 23–32 (2012), the article by T.W. Brown, and coworkers, Renewable and Sustainable Energy Reviews, 92, 834-847, (2018), and the article by M. Lehtveer and F. Hedenus, Journal of Risk Research, 18:3, 273-290 (2015). The book by David MacKay also discusses this issue in detail. A 2022 opinion article by M. Jacobson (Stanford University) from the web magazine "Physics" of the American Physical Society argues how and why intermittency can be bypassed (URL `https://physics.aps.org/articles/v15/54`).

Chapter VI

Homeostasis

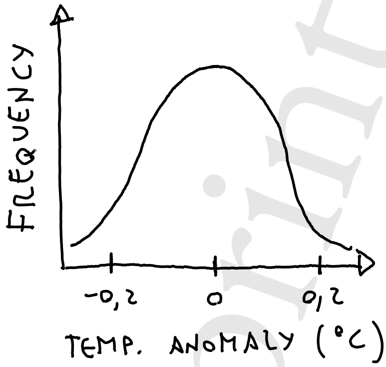
In biology, a system is called homeostatic (from Greek, meaning “staying equal”) if it tends to maintain a relatively stable equilibrium against fluctuations or perturbations. For example, body temperature is homeostatic. While we are in normal health, our physiology maintains our body temperature within a range of 36-37°C. The body temperature is set by a dedicated area of the hypothalamus, which works as a thermostat. When body temperature rises, receptors in the skin and the hypothalamus sense a change, triggering a response such as increased sweating to cool the body. In other words, there are processes in place that respond to perturbations in order to dampen their effect. When we have a fever, we do not loose homeostasis; our body temperature remains strictly controlled but the hypothalamus changes the setpoint of our thermostat, because an increased body temperature could make our body a more hostile habitat for pathogens. This change of the setpoint is a response to substances ("pyrogens") that "drive" the system, which may come directly from microorganisms in our body or as a signal from our own cells. A different kind of hyperthermia may occur when heat-dissipating mechanisms fail (for example during a heat stroke), and there is an imbalance between heat production and loss. In this case, temperature increase gets out of control, as the stabilizing response fails.

The path I took of the previous chapters suggests that we humans have been perturbing (probably quite brutally) our planet-climate system for the past two-hundred years or so. Suppose we want to start behaving; it is pretty useful to know if we've been perturbing a system that will go back to normal once we release the external perturbation, or whether it will relax to a different setpoint due to our disturbance, or whether our perturbation will or may trigger an escalation that, once ignited, becomes completely beyond our control. So my question is, is the earth-climate system homeostatic, and can I see it from simple data? At some level, the answer to this question is very difficult. We need to identify the natural (physical, chemical) processes that implement a possible the thermostat-like response. There will be multiple ones and understanding requires physics, chemistry and such. At a lower level, the answer may be within our reach. There are simple observational criteria to check whether a "dampening" response to perturbations is in place. First, if a variable is under homeostatic control its range should be limited. Second, if our variable finds itself due to some spontaneous "fluctuations" above its setpoint, the event should be followed (possibly with some delay) by a spontaneous decrease towards the setpoint, and vice versa the variable should make a positive increase if it is found below its setpoint.

Having looked at time series data for a while now, we know that the observation time scale is very important (see chapters 1 and 2), and we can expect that the answer will strongly depend on the time scale by which we look at the data. For example, at the scale of 100 ky we can expect that orbital forcing or other unknown drives that determine the oscillations that we have seen in chapter I will affect our setpoint. But let's say that we are interested in the next few-to-a-few-hundred years, the time span where our behavior can hopefully have a useful effect. On these time scales, we already know that nitrous oxide and methane spontaneously decay in the atmosphere (in some time, we said tens of years for methane and roughly a hundred for nitrous oxide). Since the decay rate is constant per unit particle of our gases, we can expect that the overall decay rate will be proportional to their total amount or to their concentration. This simple fact plays the role of a natural negative feedback (putting their setpoint to zero in absence of any emission). Hence, we can optimistically expect a dampening effect of fluctuations for the abundance of these two gases in the atmosphere, and consequently for their effects on temperature. For the case of atmospheric CO_2 the question is much trickier. As we have seen, there is no spontaneous decay for this greenhouse gas (at least "in a foreseeable future" for the time scales we are interested in), and its atmospheric levels depend on a complex balance of sources and sinks (whose rates however may respond to the system state).

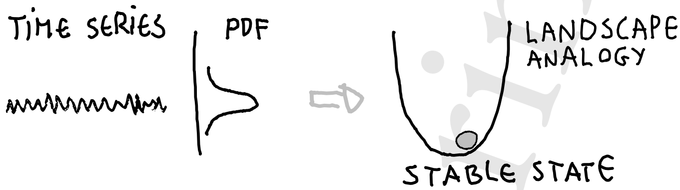
So here's what I did. First, I decided to look at global temperature directly, and pulled out of the drawer the data sets used in chapter I, ready to reprocess them with new questions. Because of the time scale story, the time series on 100 ky time scales or more are pretty useless to our scopes, since there are a lot of external drives and changes. The recent years (industrial age) time series is very precise, but it looks a bit challenging since everything is increasing due to our massive and increasing emissions. I could recycle it only if I find an effective way to remove the trend due to human "external" perturbations. But in any case I might get into some trouble because I'd be looking at data only in presence of a strong forcing on the earth-climate system. Rather than doing that, I looked for a period where there was no trend. I figured that maybe it would be possible to ask the question on a 1000 y time scale using ice core data. You will remember that the period between 1000-2000 years ago showed a remarkably flat trend of temperature anomaly. We are lucky that this 1000-year period was quite boring weather-wise (other human matters were less boring, especially around Europe, there was an Empire in decline, religious persecutions, "barbarian" invasions, etc.) So I extracted this stretch of the data set and performed a few analysis. The plot that

I sketch below already contains some answers to my first question.



The plot is my way to show that the range of the global temperature anomaly is small (roughly in the period between the birth of Jesus Christ and the first crusade). It is a histogram of all the values of temperature anomaly within my time series, the y axis is the relative counts. I put no quantities in the y axis because the counts can be normalized arbitrarily by any constant (what matters is the relative counts). Usually they are normalized in a way that the sum of all counts is one, and in this case my plot is called a "probability distribution function" or PDF. In this case the PDF of my data looks like a very common PDF, the Gaussian. Many instances of empirical data with some random variability within a finite range follow this probability distribution, for example the probability that man or women have a certain height is Gaussian, and so is their birth weight. In our case, the peak of the Gaussian, zero, is the most probable anomaly, and the spread (over our late Roman empire-early Middle Ages period) is symmetric, spanning about 0.2 degrees on both sides. This is a more quantitative assessment of what I already used in chapter I to set an eyeballed value for the baseline fluctuations of global temperatures.

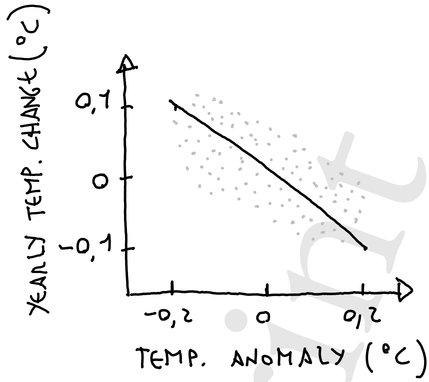
The sketch below illustrates the concept of a PDF extracted from a time series that has no trend.



The same sketch tells us that we can use a "landscape metaphor" to describe our temperature anomaly fluctuations, and see temperature as the coordinate of a one-dimensional rolling ball living in a U-shaped valley (subject to gravity), and subject to some random kicks. This landscape metaphor is used very frequently in dynamical systems, with many applications across science[1] and it will be useful in a short while to understand tipping points.

So, all okay so far, but we also wanted to know whether and how much if temperature anomaly finds itself above (or below) its setpoint it tends to go back. The plot sketched below attempts to answer this question.

[1]For example, in chemistry the ball corresponds to a reaction coordinate, and in evolutionary biology to a phenptype.



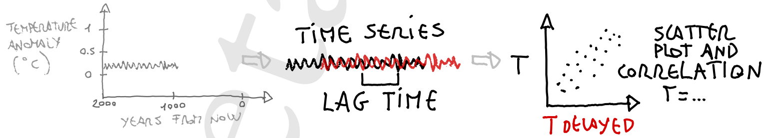
It is a scatter plot comparing the yearly temperature step to the temperature anomaly of a year. For example, if the temperature anomaly was -0.1931 ^{o}C in AD 33 and -0.1117 ^{o}C in AD 34, then I would take -0.1931 ^{o}C in the x axis and 0.0814 (the difference between the two years) in the y axis. The negative trend passing from zero shows that if the temperature was above the setpoint (x axis point above zero), on average the yearly temperature step was negative, and the temperature tended to decrease (y axis smaller than zero)[2]. Vice versa, if the temperature anomaly some year was below zero (below the setpoint) then on average the yearly temperature step was positive, and the temperature tended to increase. Additionally, the size of the homeostatic steps is linearly proportional to the deviation from the setpoint. So (for a change) some good news! Our analysis suggests that we are dealing with a system that (at least under moderate perturbations), is homeostatic. It wants to go back to its setpoint temperature, and this is also what we want it to do.

[2]Note that this plot is one of those cases where the slope is more important than the correlation.

We have no information on how it does it but this, as we said, requires complex physics and chemistry.

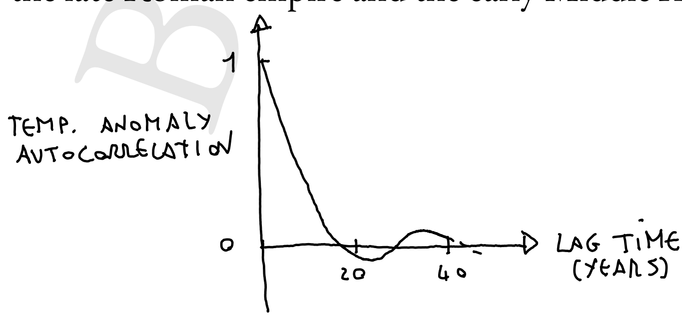
Finally, we would like to know how many years it typically takes for the temperature to go back to normal after a fluctuation (or, more precisely, how long it took in the early Middle Ages). If we can measure this relaxation time, and we extrapolate it it to the present time, it will give us enough information to formulate an informed prediction on how long we will have to wait to see the effect of our reduced emissions (provided we reduce our greenhouse emissions in the coming years). It could be one, ten, or a hundred years, and it makes a big difference.

One technique to do this is called "delayed autocorrelation". Let us define this concept. We already introduced the Pearson correlation r in chapter 3, and we know it comes from a scatter plot. The idea of the delayed autocorrelation is to ask how much a time series is correlated with itself in the past or in the future. See the sketch below:



If we correlate the time series with itself in the present (no delay, and the lag time is zero), we get trivially $r = 1$, because the scatter plot just has every time the same point in the x and y axis. For any time delay, we can look at the scatter plot of the value of our variable at some time and its value after the delay. For example, we take a delay of 10 years, and we scatter plot between year 0 and 990 AD and the value of the temperature anomaly that year and ten years after. We get a scatter plot, from which we get a value of r (presumably now less than one). If the time series remains correlated with itself for long time delays, it means that it will tend to remain above the setpoint if it were above initially, and vice versa it will tend to remain below the setpoint if it were initially below. Hence, the effect of a perturbation would not have fully be absorbed (yet). If the correlation is lost for some time delay, the variable has lost the memory of its earlier value, achieving full relaxation. In other words, by looking at the autocorrelation as a function of the time delay, we quantify the relaxation time that we are interested in (technical note: this works if the time series has a flat trend and its PDF does not change over time).

This behavior is quantified by a simple plot. For each delay, we can get a scatter plot and a value of r, and plot r as a function of the delay (in our case in years). This is called the delayed autocorrelation function. By looking at how this plot decays we get an idea of the relaxation time. Here's a sketch of what I got for our global temperature anomaly time series for the 1000 "boring" (temperature wise) years of the late Roman empire and the early Middle Ages,



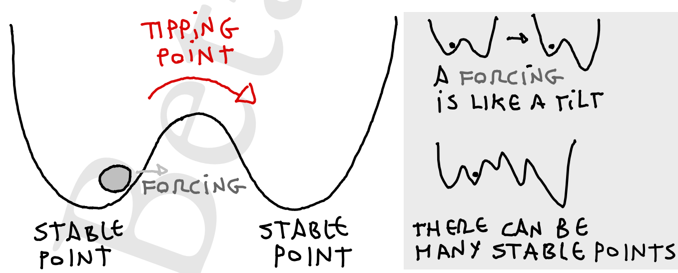
The autocorrelation is completely gone in 20 years (and has a half-life of about 5-6 years[3]). If we can extrapolate this to present times, this is another piece of good news. If we could instantly restore the levels atmospheric greenhouse gases we would expect to see the effect quickly. Our informed prediction is that it will not be immediate, but it will not be centuries either. In five years or so we can expect to see a noticeable effect. Of course we cannot reduce atmospheric greenhouse gases instantly, by reducing emissions we reduce a source, and the atmospheric levels have to slowly go down (depending on the balance between sources and sinks). Additionally, our prediction could be wrong if the situation of the early Middle Ages is not transportable to current times. There is no absolute guarantee that the same rules that applied in the Middle Ages will apply now, because there are some variables in the system that are different now. For example, we know that the perturbation that we are causing (which now puts us almost 1 degree Celsius above the setpoint) is much larger than the natural fluctuations against which we are testing our relaxation time (which we saw are 0.2 degrees Celsius at most). Another point is that we have been applying a forcing on the system (by releasing greenhouse gases) for some time now, and this might have changed its setpoint temperature (see below).

[3]Note that for small delays the measurement uncertainty, by adding some "noise" to the true values, reduces the correlation: a detailed analysis of the effective time resolution of these data is not simple, and we should regard the part of this plot relative to larger delays as more significant.

As you will have noticed, the delayed autocorrelation also makes a couple of small but visible oscillatory "whirls", which is interesting. On the other hand, the whirls are very small, and not really relevant for our scopes so we can probably disregard them for now.

Something perhaps interesting is that the plots that I have shown you are almost completely in line with the behavior of a well known stochastic model from physics, originally describing the motion of a colloidal (micrometer sized) particle in water subjected to a spring. The equation of motion for this model was solved in 1930 by Leonard Ornstein and George E. Uhlenbeck, and had been proposed a few years earlier by Paul Langevin. Curiously, by tracking the motion of a micrometer-sized bead under the microscope inside a laser trap (called "optical tweezers"), I would produce plots that look very similar to the ones that I've drawn above from ice-core derived temperature data. This is not a coincidence, because the Ornstein-Uhlenbeck process captures the behavior of many systems, from biology to the stock market. There are somewhat "deep" reasons that make many systems that are under homeostasis and "close to equilibrium" (in some sense not far from their setpoint) follow this behavior. The reasons are that (i) small perturbations around equilibrium points generate spring-like response, and (ii) noisy systems where the noise variables have a bounded range and are uncorrelated generally pop out Gaussian distributions. Both points, mathematically, are related to the "Taylor expansion" for approximating a function close to a minimum or a maximum, named after the mathematician Brook Taylor, who first used it in 1715). They are also one example of what is called "universal behavior" in physics (generally we physicists, right or wrong, are very fond of this). But I can see that I am drifting dangerously away from the back of my envelope, which is becoming a small-scale historical treatise of physics and mathematics. Apologies.

However, there is one last thing that I'd like to address. As we mentioned it, there is a big elephant our room. We made our predictions based on the assumption that our system was fluctuating "freely" close to its "equilibrium" state, or setpoint, but in reality in recent times we have been applying a strong perturbation for a fairly long time (compared to its natural relaxation time, which we found to be 20 years at the most). In such situation, we might be creating the conditions for the system to cross a "tipping point". Here's a sketch by which we can represent a tipping point.



In the landscape metaphor we can picture the system (our variable, global temperature) as a ball in a valley. Naturally, the ball would like to fall towards the bottom of the landscape, but random noise makes it wiggle a little bit. If we apply a forcing (we can think of it as a "wind" in the landscape), we naturally distort the landscape and bias the average position of our variable. The tricky thing is that the landscape may have multiple minima (two in my drawing). When our system is unperturbed, we can picture it fluctuating close to one minimum. In our case it is the current global temperature setpoint. When a perturbation is applied (a wind to the right, in my drawing) it might increase the chances that our system "tips" to another minimum by a random jump in the wrong direction. In our case the fear is that we could tip our climate to a minimum with higher global temperature. As you can imagine pulling it back would require an opposite drive, hence it could be very hard. Also, due to the structure of the landscape, we expect that after it tips it will fall very fast for a while, until it reaches a new minimum. So we can easily imagine that how the tipping dynamics could be abrupt, as well as largely irreversible. The right panel of my sketch above illustrates how a forcing can be understood as a tilting applied to our landscape, and how a landscape can have multiple minima corresponding to stable points.

A tipping point is another general behavior that can emerge from a complex dynamical system, a critical threshold point beyond which a system deeply reorganizes. Indeed, the term "tipping point" was first used in sociology in the 1950s, to describe the phenomenon (which happened in the Bronx in New York City for example) by which a certain fraction of nonwhite residents in a neighborhood would trigger the move of most white residents to other neighborhoods, switching the ethnic dominance. The dramatic aspect is that if we bring a system close enough to a tipping point, the tipping event can be brought about by an even small random spontaneous disturbance. Hence (i) the specific tipping event will be very difficult to predict because it is the result of some random coincidence (made more likely by our external forcing) (ii) a small but unfortunate "fluctuation" will determine disproportionately large changes in the system. Not good news.

We have mentioned in chapter I the complex link of Milanković cycles with the periodicity of global temperatures, the 41k years cycles seems easy to explain, because its period corresponds to the preiod of the forcing, but the 100 ky cycle observed in the last million years is less easy. We can return to this point with our landscape view of global temperature. There are multiple stable minima, and there are tipping points, then there are external forcings. The radiation from the sun is a periodic forcing. The forcing is small, an oscillation whose amplitude is about 0.1% in the energy flux that hits the earth from the sun. We also observed in chapter I that the the changes are large, and global temperature over the ice-age cycles can change by more than 10 degrees Celsius. If one puts the numbers in (it needs a small mathematical model), it turns out that it is very difficult to explain a 10 degrees variation with this forcing, if we assume that the temperature only has one stable state. However, this becomes much easier if we think that the 10 degree difference is encoded in different stable states (minima of our landscape), and the forcing just varies the probability of random transitions between these states. Additionally, since the transitions are stochastic, the forcing makes them just more or less probable in time, and their period can be different from the period of the forcing. This phenomenon is called "stochastic resonance" and it was introduced in the 1980s to explain the 100 ky temperature cycles. Since it is very general (it just requires a system with multiple stable states and a small periodic forcing), it has applications in many fields, from turbulence to biology.

Mechanistically speaking, tipping points typically require positive feedback mechanisms that counteract the negative feedback mechanisms enforcing the setpoint, and fuel the tipping events. Can the earth-climate system tip, or in other words does its landscape have multiple minima, and are we driving it towards other minima? This is definitely not a back-of-the-envelope question, but leaving it up to the experts, both paleoclimate data and global climate models suggest that (unfortunately) the answer is yes. Even more dangerously, the impression I got is that we do not know very precisely where tipping points may lay (and whether we have already crossed some). The professionals are thinking about this though, and providing their best estimates. A very recent analysis (2022), published in the journal Science, evaluated more than two-hundred previous studies on tipping points, using both past-climate observations and mathematical models, and drew a shortlist of of a handful of global and local tipping elements by confidence, impact, response time scale, and their global temperature thresholds. They conclude that current global warming already lies within the range of confidence where some well-understood tipping point could be found. Several tipping events are likely to occur even if we stay within the range of the 1.5 to 2^{o}C range of global temepraturе increase set by the Paris agreement.

First, geological evidence indicates that the earth-climate system went through different "equilibrium" points during its history. For more than 70% of its history Earth appears to have been in a so-called "greenhouse" (or "hothouse") state, largely ice free, even at the poles, with lots of CO_2 in the atmosphere and much hotter climate than today (this is actually likely not one state, but a set of states, each with its shade of hotness). These greenhouse periods can be rich in life and biodiversity (if we think of dinosaurs living in tropical jungles, for example, this scene takes place during a greenhouse). For at least 2.5 billion years (since the beginning of the Proterozoic), the climate has switched (five times so far) between greenhouse and so-called "icehouse" conditions. Because of the 100 ky time scales cycles we have mentioned before, within icehouse periods there are glacial and interglacial sub-periods. We live in a mild interglacial of a long-term icehouse (known as the quaternary Ice Age, which began approximately 2.6 million years ago), and historical climate records suggest that we are heading towards a new glacial within a few thousand years. However, the antarctic ice sheet has most probably existed for about 34 million years.

Second, there are indications as to how specific tipping points could come about in the current earth-climate system. Losing enough of the ice sheets in West Antarctic and Greenland, of warm-water coral reefs, and the Amazon rainforest may determine tipping points, and some of these some tipping points may be close to being crossed (or may have already been crossed). In the case of the Amazon forest, the positive feedback is due to deforestation leading to more droughts and heatwaves, leading in turn to more tree deaths and more fires. As a consequence of this, since some times the forest is no longer a carbon sink. Additionally theory, simulations, and reconstruction of past events suggest that the gulf stream system, the well-known system of ocean currents, also likely has a tipping point. Currently it sends warm surface water north form the tropics and carries cold fresh water back south. If freshwater input from melting glaciers reaches a certain threshold, it could collapse into a state of reduced flow, and not return to its current state even after melting stops. The new stable state could last for thousands of years. Thawing of frozen permafrost in the northern hemisphere may also provide a relevant positive feedback because it contains massive amounts of carbon that could be released in the atmosphere.

So be careful.

DATA SCIENCE TAKE-HOME MESSAGES. This chapter uses fancier techniques than the other chapters, so if you got lost, we can entirely blame it on me. We have used the histogram as a way to estimate the probability distribution (and therefore the range of variability) of a fluctuating variable. Subsequently, we have used a scatter plot of the "step" (which in time series analysis goes under the name of "discrete derivative") with the value of the variable in order to test for homeostatic behavior (and we found evidence for it). An important catch, which stems directly form the considerations made in chapter I, is that this analysis depends on the time scale of the step. Another important aspect is that these analysis rely on the assumption that the time series is "stationary", or in other words the underlying PDF is not changing, and as a consequence (in particular) there is no increasing or decreasing trend. Our time series had no trend, because we chose the period between 1000 and 2000 years ago, where both the mean global temperature and its fluctuations remain remarkably constant. If a time series has a trend, the analysis is more difficult: there are techniques to subtract the trends in the average and in other features of the PDF. Once the time series is "detrended", one can apply similar analyses. Subsequently, we have used autocorrelation analysis (we have correlated a time series with itself, with a time delay) to quantify a "decay time" of global temperature fluctuations. Finally, to introduce tipping points, we have entered the advanced topic of a "landscape" description of a dynamical system with fluctuations, also mentioning the hypothesis of stochastic resonance.

SOURCES. All the data used in this chapter was already introduced in chapter I. A nice definition of homeostasis is provided by this Scientific American article by Kevin

Rodolfo https://www.scientificamerican.com/article/what-is-homeostasis/. A lot of the information provided here (and more) on earth-climate stable states and tipping points can be found on Wikipedia at the pages Greenhouse and icehouse Earth and Tipping points in the climate system. The original paper proposing stochastic resonance to explain the 100 ky cycles is Benzi et al. Tellus, 34, 10-16 (1982). One of the authors of this study is the 2021 Nobel prize winner Giorgio Parisi.

Chapter VII

Save the planet?

This is a book about "reading" data. So we can forget about climate for a moment and focus on our real goals. I have asked you to trust my data-science skills in turning real data and real plots into a few cartoon-like sketches, so that we could focus together on the interpretation and modeling part. We have learned that any quantification from a plot needs a scale, a unit of measure, that time scales are important, that scatter plots measure correlations, that correlation is not causation, we have learned how to define and use negative controls and "null" scenarios to score against our data, how define and measure fluctuations and how to test for homeostasis by comparing time variations with instantaneous values of our variables. Additionally, we have learned how to build a quantitative estimate by breaking down our problem into sub-problems, setting prcise terms of comparisons, using central values and orders of magnitude, and converting units. Finally, we have learned how to connect plotted data and theoretical estimates by formulating predictions and testing them, and by using the decomposition of a problem into sub-problems.

We have covered informally all these useful tools about reading a plot, considering carefully questions about trends, fluctuations and time scales, formulating a reasonable estimate, venturing in a positive interpretation, and daring a prediction. But at the same time I hope I have shown how important it is to any data-driven analysis to be conscious of the limitations, of the assumptions, and to explore the limits of our data, of our analyses, and of our conclusions. Perhaps the hardest part is gaining consciousness that data analyses are never objective because they rely on assumptions, approximations, limited knowledge, as well as intrinsic limitations of the data. However, this does not mean that *anything* is possible. Data provide the essential information that constrains the possible interpretations of reality. If we chose to follow these constraints we reduce by a great amount the space of possibilities, and we give solid grounds to knowledge. Hence, there is space for a debate, for opinions, for arguments, for different solutions, but only an informed debate can be meaningful. A debate that is not constrained as much as possible easily results into meaningless noise. I believe these tools and concepts are applicable widely today, in the "era of data". They are always applicable within science, of course. What perhaps is more interesting and "new" today is that they are becoming applicable to global and societal problems that intersect with "data science", because data is extensively available. Climate change is only one example, though perhaps the most urgent today. Think of all the emergent (interconnected) global sustainability and well-being problems related to inequality, health, energy, population, food. Perhaps we can be optimistic and think that if we learn about reading, discussing, and interpreting data we might do something good in the long run.

AND WHAT ABOUT CLIMATE CHANGE? Here's what I think after having taken my personal path through these data. It is not a matter of saving the planet, it is a matter of preserving ourselves. This is just my opinion and you are entitled to your own of course. But my impression is that we live on a pretty sturdy planet, which looks quite robust in terms of amazingly adapted, and adaptable, life forms. It had its periods of extremely cold and extremely hot weather, of volcanic eruptions. It had its meteorite hits, it mass extinctions always followed by biodiversity explosions. The planet is fine, and to me it looks like it could survive many other cycles of all these events. Sure, maybe we can drive the environment to lead ourselves to extinction or near-extinction. Likely we can bring down a good number of other species with us, including cute mammals like the panda and the dugongo, but also we should not overestimate our relevance in the biosphere. There are sharks, who have already survived several mass extinctions. There are entire ecosystems thriving on hydrothermal vents at the bottom of the ocean, and my guess is that they would hardly be affected by any of the processes described in this book. Hence (other opinion), there is nothing altruistic in keeping a close eye on the consequences of our existence on the environment (as many of both the ecological and anti-ecological narratives seem to imply). It is mainly about looking after ourselves and the conditions in which we thrive as a species. By something that I would hardly call a coincidence, these conditions include aesthetically appealing features such as blue oceans full of colored fish, white snowy mountains, and green woods (I am sure that the sulphur-loving bacteria that live in hydrothermal vents would have different criteria if they could express their opinion on the most beautiful environments).

So, end of opinions, back to estimates. For fun, or for the sake of argument, we could try to end the book by estimating upper bounds on the mess we can make with the carbon available to us; let's play the role of supervillains and burn all hydrocarbons reserves. Looking at the literature, those are at least 5000 Gt (gigatonnes) of *carbon*, which as we saw is about 18000 Gt CO_2 (because of the oxygens). However, not all of it is accessible, so we can look at intermediate values, guesstimating that, say, 5000 Gt CO_2 could be available in the next century or so (see below for actual figures).

Our estimate from chapter III was 1.1 degree Celsius per 100 PPM CO_2 in the atmosphere. We also said that 1 Gt emitted corresponded to one eighth of a PPM, and that due to the carbon sink, roughly only 45% of that, 0.05 PPM, would stay in atmosphere. Putting the numbers together we get that 20 Gt emitted CO_2 is 1 ppm, so that 2000 Gt CO_2 is 1.1 degree and 5000 Gt CO_2 is 2.75 degrees, the whole 18000 Gt CO_2 would be 9.9 degrees. Let's go for our "central" value of 5000 Gt CO_2. Since we are already plus 1.1 degrees, we would end up at plus 3.8 degrees or so. Of course this estimate is very crude, but if we ask the professionals that's a lot according to all climate models. Consider that plus 2 degrees Celsius is already seen as an extreme scenario by most. For our habitat, of course, I guess this could be considered chilly weather in the Cretaceous. The trend we estimated for the incidence of extreme events was a factor of five for about one degree, so for 3.6 degrees we would extrapolate about an 18-fold increase of extreme weather events. To get an upper bound for the temperature increase for 5000 Gt of emitted CO_2, we can also forget about the 45% factor, because at some point the carbon sinks from ocean and land could (will likely) saturate. In this case 8 Gt is 1 ppm so 800 Gt is a 1.1 degree increase, and we could end up at plus 8 ^{o}C from pre-industrial levels. Big exclamation mark (but still standard Cretaceous weather). As we mentioned before, both estimates can be seen as pretty conservative because they assume that everything is linear, while we said that positive feedback mechanisms exist and are already visible in the data. Most likely the Earth would end up in a greenhouse stable state that has seen before. A more common estimate asks how much carbon we can emit in order to stay within the boundary of 1.5 degrees Celsius above pre-industrial levels. That is about 0.4 degrees from now (2022). So using my figures, this would be 36 more PPM, hence I get about 720 Gt in the optimistic scenario, and removing the 45% (assuming ocean and land sink saturation) I get 36 more PPM times 8 Gt/ppm, which is 288 Gt. Again this is pessimistic on the saturation, but most probably still optimistic because it relies on linear extrapolations.

I may know my way through data, but of course we know that I am not a professional of climate science. Granted. Let's see how good (or bad) we did with our estimates, by looking at figures derived by some professionals of this sector. Carbon tracker is an independent organization concerned with the energy transition. It releases reports on these things. At the start of 2022, they say that only 320 billion tonnes of greenhouse gases (320 Gt CO_2) can be emitted to stay within a 66% chance of limiting warming to 1.5^{o}C above pre-industrial levels. So it seems we were a bit too optimistic (as expected).

They also quantify the total emissions of all fossil fuels in reserves of companies listed on global stock exchanges. This is more than 10 times that level, around 3700 Gt CO_2, so about ten times more than we can emit to stay within the 1.5 degree bound. This is lower than the 5000 Gt CO_2 figure we guessed above. Looking around, different estimates may vary the figures for both reserves and burnable carbon (frankly it becomes a bit confusing), but there is a consensus that the level of potential carbon emissions exceeds the reserves (by a factor of at least two in the figures I have seen). Even from oil companies. Indeed, it seems that BP and Shell have confirmed that according to their figures burning all known fossil fuels would result in more than 2°C of warming.

Side note, the figures tell us that the years of emittable carbon fossil fuels left are not too many. Using the figures from carbon tracker, at the current amounts of yearly emissions, the "burnable" part will be exhausted in just eight years, by 2030 (and the total around 2100).

Additionally, I was not the only one to come up with the supervillain question, so we can compare our results on that side too. A fairly recent study (2016), used whole-Earth climate models to explore how the planet would respond to 5000 gigatonnes of carbon (a likely lower bound of the grand total amount of fossil fuel reserves). Compared to my naive models, their simulations contain a precise description of the saturation process of the land and ocean carbon sinks, as well as other positive sources of feedback. According to their model the land sink will become saturated around 2100. The oceans will not stop absorbing our CO_2 emissions between now and 2300 (when they chose to end their simulations) but their rate of absorption will decrease noticeably after 2100. In terms of predictions of global temperature, the models they run suggest the range between 6.4 and 9.5 degrees Celsius. They also plot temperature anomaly versus Gt of emitted carbon, and different models show a fairly linear relationship, with slopes in the range from 1.1 to 2 degrees Celsius per 1000 emitted Gt of carbon, which is 0.3-0.55 per 1000 Gt of emitted CO_2. Our naive estimates above lead to 0.55 degrees Celsius per 1000 emitted Gt of CO_2 for this figure (and 1.4 degrees Celsius per 1000 emitted Gt of CO_2 for the upper bound assuming full saturation of the sinks), so we see that we did not do extremely bad.

For sure the naive estimates are good in terms of effort, if we think what it means to run and analyze a simulation of these models. Apart from the scientific knowledge that these models contain, running them is a challenge in terms of both computation and storage capacity. The models compute the solution to differential equations of fluid mechanics, chemistry, and thermodynamics across time and space for the whole planet, approximated as a three-dimensional grid with some spatial resolution, and calculate variables such as temperature, winds and ocean currents. Different components of a climate model simulate the atmosphere, the ocean currents and its biogeochemistry, the land and its vegetation, etc. Accounting for all these aspects together in a simulation is a heavy task computationally. Depending on the required resolution, simulations may require a supercomputer and considerable computational times (days, for example), and typically many simulations are required to calibrate a model and explore the parameters.

There is an additional aspect, which is that even these mastodons models are models, they rely on different assumptions, use different approximations, and incorporate different processes. In some sense, all models are wrong, because they represent arbitrary simplifications of complicated systems. Hence, the heavy simulations need validation against a large set of empirical and experimental data. Data can be used as a testing ground to identify what the model may be simulating incorrectly or ignoring. They could range from local measurements, to short-time earth scale data from satellite observations, to paleoclimate data. Since many earth-scale climate models by now have been published and tested, they have already passed several preliminary validation steps, and are available as separate packages. In this "ecosystem" of different packages, another approach is using different earth-scale models to make sure that the differences between the models do not affect the main results of a specific study.

For sure climate models are a "gold standard" to follow. However, once models reach such a scale and complexity they inevitably loose transparency. They become to some extent "black boxes" and require simpler explanatory principles in order to remain interpretable. Nobody likes a black-box prediction, because it sounds like a prophecy; we want to understand the answers given by the big simulations in terms of where they come from, which ingredients are important, and what key explanatory principles lead to the conclusions. Simpler models provide interfaces to make the big models more transparent and connect them to data. The estimates and techniques such as the ones presented here probably represent the first layer of a complex hierarchy of mathematical models that connects data to complex representations, such as planetary-scale simulations. You can do much better of course, but the higher you go in the hierarchy, the higher price you pay in terms of complexity and interpretability of your model.

DATA SCIENCE TAKE-HOME MESSAGES. The first part of the chapter contains the main philosophical statement of this book: we should venture into analyzing and interpreting data, while being conscious about the constraints. If our approach to data is only negative, we can never construct any interpretation. So we should go ahead, look at the data, formulate our questions and hypotheses, try and interpret. But when we do that, there are many constraints, caveats, unknown areas, areas that we don't know very well, and assumptions that we should be aware of. Acquiring that awareness is probably more important than our conclusions. We should be very careful in formulating and circumscribing our conclusions, in a way that they reflect our awareness, and ready to be wrong, and to change our mind. The second part of this chapter concludes the book with the now familiar estimation techniques applied to a supervillain question on climate change, and a reflection on the role of simple estimates in big simulations (which is very topical in contemporary data science).

SOURCES. The figures on the reserves of fossil fuels and unburnable carbon were derived from the 2022 report by carbon tracker (`https://carbontracker.org`). The study using climate model simulations to predict the response to five Gt of emitted CO_2 is Tokarska, K. and coworkers, Nature Clim Change 6, 851-855 (2016). As you will have noticed, I have made use of the "Our World in Data" project (`https://ourworldindata.org`) several times across this book. It is an enormously valid data reference, designed to "make the knowledge on the big problems accessible and understandable." The more recent database "Anthroponumbers" (`https://www.anthroponumbers.org`), also contains a collection of useful numbers for quantifying the impacts of the human presence on Earth.

It also lists other databases with useful and interesting data. You can use these resources to start your own quantitative estimates and considerations concerning global sustainability problems. Happy journey!

We're on a road to nowhere
Come on inside
Takin' that ride to nowhere
We'll take that ride
I'm feelin' okay this mornin'
And you know
We're on a road to paradise
Here we go, here we go

—Talking Heads