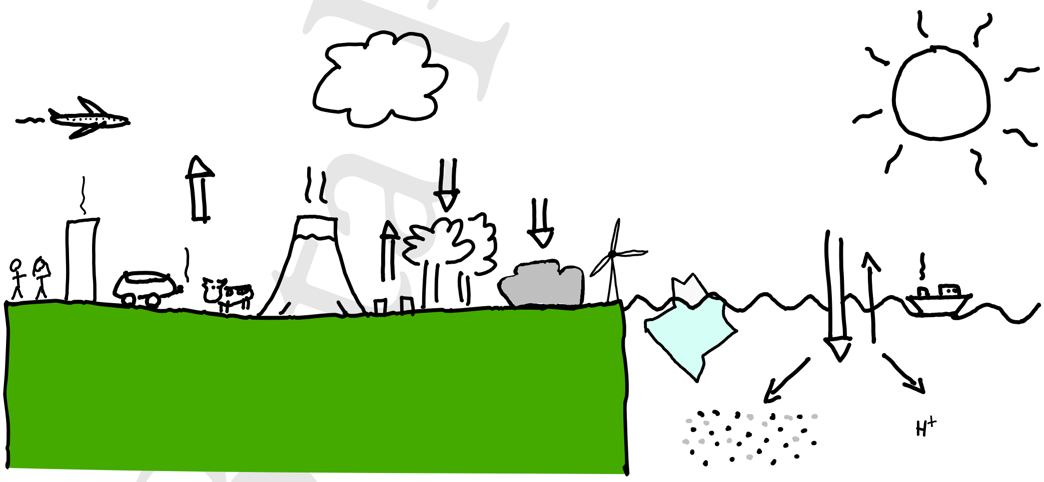
Back-of-the-envelope climate change



Marco Cosentino Lagomarsino

Back-of-the-envelope climate change

Marco Cosentino Lagomarsino

Beta preprint edition, 2023, not for diffusion.

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 lomited 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].

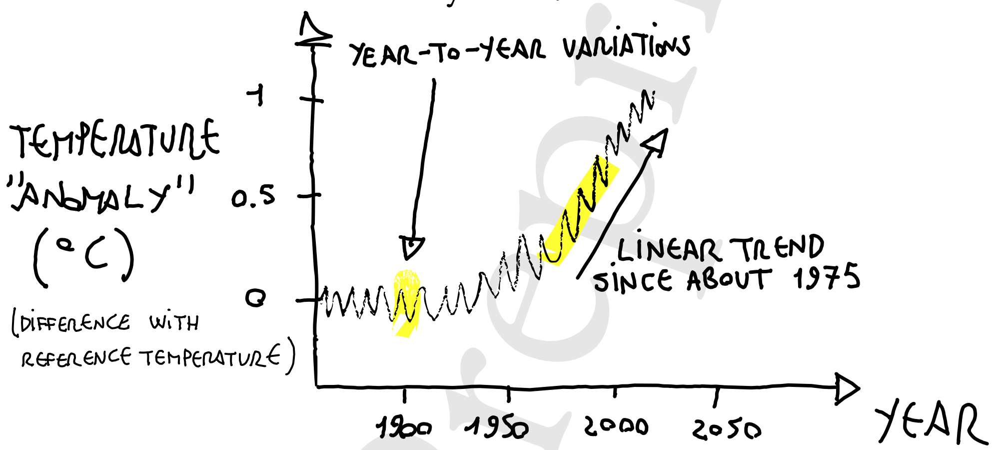
The other problem that we need to face when trying to assess if temperatures are increasing is that temperatures vary a lot. They vary between day and night, from place to place, and they vary across seasons. Many of these variations follow specific trends, mostly due to the illumination from

[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

[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.

the "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 teperature 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 than 5000

[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.

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 latitude "ring" around

[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.

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°C above the 1850-1900 reference.

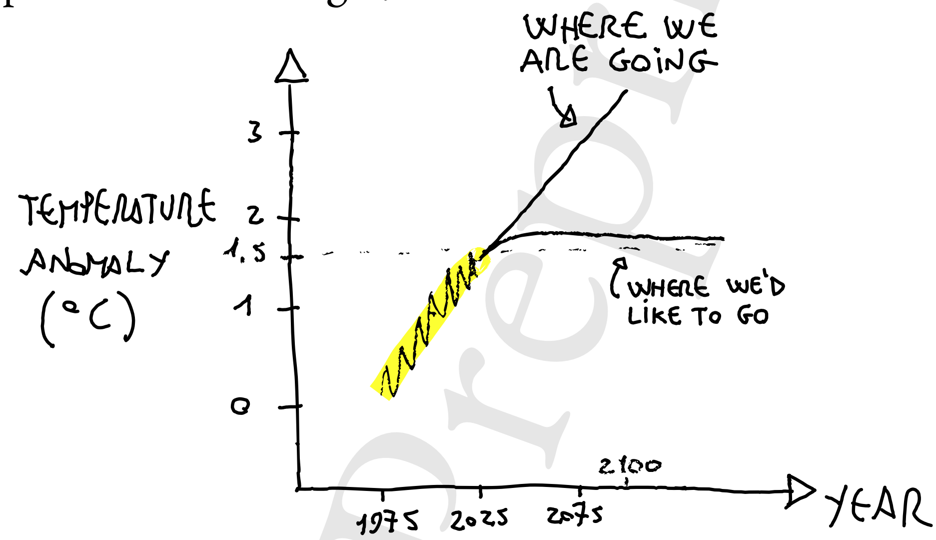
Looking more closely at the plot sketched above, the second thing one could notice is that, despite of the averaging over all Earth locations and all times over the year, there are interesting "noisy" year-to-year variations. Those are the zigzags in the plot; the ones in my drawing are not the real 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^oC change over 25 years, and with a total increase of almost 1^oC 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

[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.

considerable? 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,



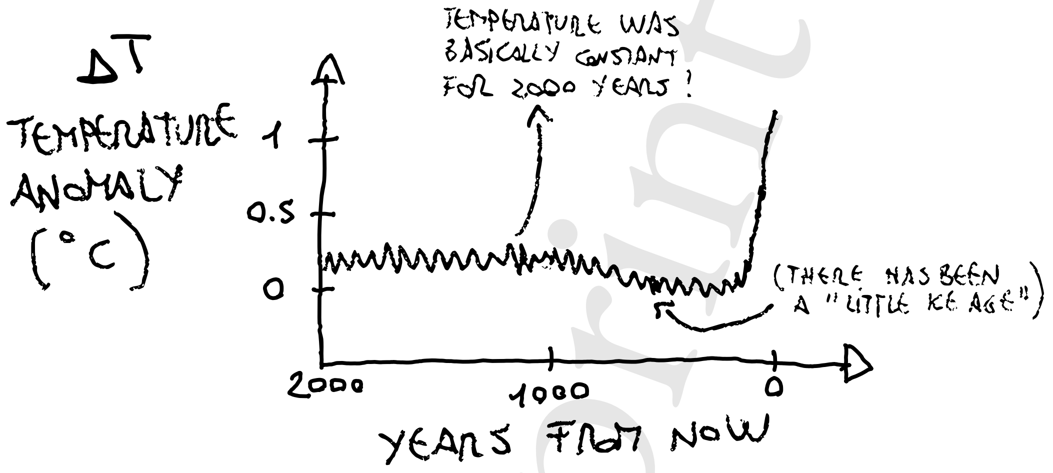
The highlighted part is basically the current trend. To show how radical a change of trend we need, compared to the current state of things, in order to reach our internationally set climate change targets, I have also drawn an imaginary trend line that reaches this target. Clearly, where we want to 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 eruptions and/or re-

[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.

duced 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

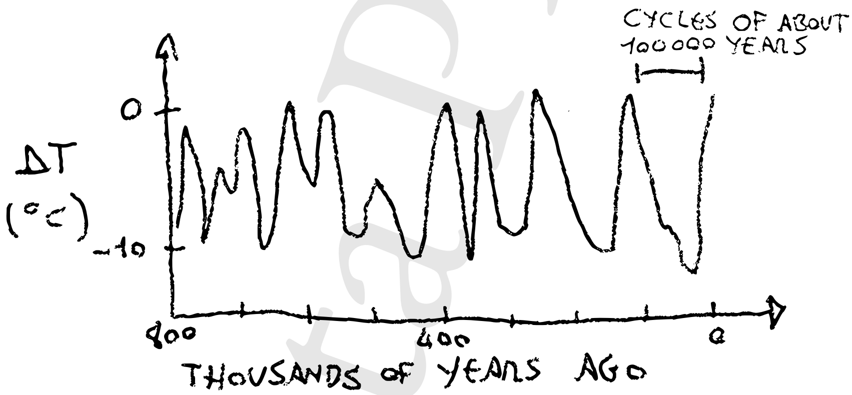
[9] I did not check myself the significance of these findings. Let us take them as "anecdotal evidence" that the global temperature change during the Little Ice Age had an impact. The question, however, may be subject to debate.

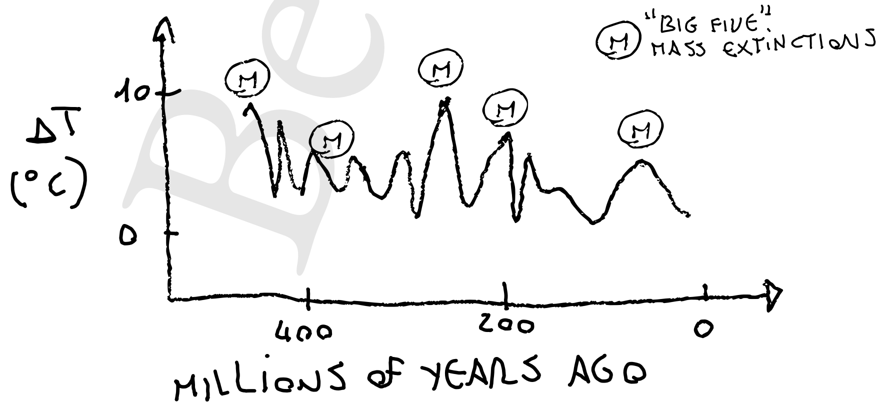
[10] We will define more rigorously the concept of correlation in chapter III.

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.

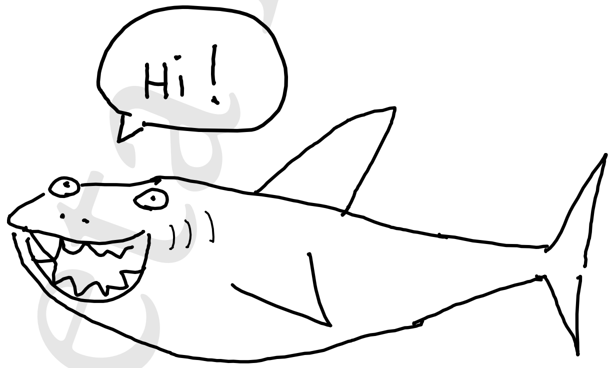
Humans (homo sapiens) have been around for two to three hundred thousand years, so, as a species, we have experienced at least two, maybe three cycles of these dramatic climate changes. This proves that our species can survive radical climate changes. But of course we achieved that feat as small sparse 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^oC, 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 bettwer the approximation (in probability thoery these statements translate into the law of large numbers, and the central limit theorem). 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