1
Introduction

The big question we researchers today are confronted with is whether the ongoing climate changes are associated with changes of climate extremes. Are floods occurring more often, are they getting bigger? The first part denotes the frequency aspect, the second the magnitude aspect of the question. Are we facing a higher probability of summer heatwaves in the future? Are we going to see more devastating storm events such as hurricane Katrina in August 2005? This big question plagues also politicians and stakeholders, who need to decide which measures to take to guard against the risk of those costly extreme events.

You will learn in this book how this big question can be approached using the scientific method. We researchers follow, most of us subconsciously, the philosopher Plato (427 BC–348/347 BC) and assume that there exists an objective but unknown climate reality in space and time. Key to coming closer to examining this reality is through the use of data and their mathematical analysis.

The aim of documenting data and detailing mathematical steps is to permit the reproduction of results as well as an understanding of what has been done. This allows other scientists to criticize our work by obtaining more and better data or to present improved analytical tools. The purpose is not to attack each other but to jointly come closer to the climate reality. It is inevitable that on your way to becoming a climate researcher you will meet mathematical formulas. These are things of beauty! They allow the performance of data-analytical science in the most concise way. They help us to communicate to others as effectively as possible what has been done.

Box 1.1 Personal Reflection: Philosophy of science and irreproducible papers

At the beginning of my PhD studies at the University of Kiel in Germany in the year 1992, I decided to put what had been my private speculations about what can be known on a more solid basis and enrolled in a subsidiary course in the philosophy of science. The influence of the teachers Hermann Schmitz and Hans Joachim Waschkies at Kiel strengthened my existing conviction of the usefulness of the working hypothesis of realism. Plato is one leading thinker on this; others are Immanuel Kant or Karl Popper. Later I found that realism corresponds to the axiomatic approach to probability, which uses data and statistical tools to infer reality with error bars.

How far should a student go beyond the boundaries of a PhD project? Is it a waste of time? My advice is to follow your curiosity, intuition, and aesthetic education. Schopenhauer's recipe is to consume original sources and not "masticated food." The philosophy books whose spines face me while I sit here have to wait until I have finished writing this textbook.

The beginning of my PhD studies brought also frustration. I wished to adapt, advance, and apply methods for analysing proxy climate time series from the seafloor archive. I browsed through papers on geology, climatology, and physics. Many of them presented fanciful tools and colorful plots – but I could not reproduce the results. The major reason for this irreproducibility was the lack of a clear description of the methods used and how the data were treated. I suffered also because I could not find a textbook that explained the ideas behind the methods in an accessible manner, or described the methods at a level that permitted the reproduction of the results. On the other hand, the philosophers I studied in my subsidiary course wrote clearly, which made their works accessible and thus criticizable.

The years subsequent to my PhD were spent writing a book on climate time series analysis (Mudelsee 2014) that aims to be accessible and clear in its descriptions. Websites with the data and the source codes of the methods can therefore make the lives of hungry students easier. The aims of accessibility and reproducibility also guide this book on analysing climate extremes.

Climate firstly can be understood as the state of the atmosphere. The Sun's incoming shortwave radiation is partly reflected by the atmosphere, partly reflected by the Earth's surface, and partly absorbed and re-emitted by the Earth as longwave radiation (Figure 1.1).

This radiative balance corresponds to a certain surface-air temperature. This variable, temperature, is the most important climate variable since it characterizes the Earth's state. Temperature interacts with many other climate variables. Relevant for the context of this book are the variables precipitation and wind speed. The radiative balance can be perturbed by changes in the atmospheric concentration of radiatively important gases, such as carbon dioxide, aerosols, or by changes in the surface properties of the Earth. These factors are the drivers of climate change (Figure 1.1).

Climate can be further understood as the state of other compartments than the atmosphere, such as the marine realms, the cryosphere (ice or snow), or the biosphere. The state in all those compartments, including the atmosphere, is not constant over space, and this spatial variability is the first reason why the climate system is complex. Furthermore, size of spatial variability depends on the climate variable. For example, it is larger for precipitation than for temperature.

The second reason for the complexity of the climate is temporal variability. Everybody experiences that day is followed by night and then another day, or that after spring there comes summer, fall, winter, and then the next spring. If these two periodic variations, the daily and the annual cycle, were the only variations, then predicting weather and climate over long time ranges would be easy. These cycles with periods of one day or one year form a part of the weather. Longer-term variations, above 30 years, form a part of the climate.

However, the climate system comprises many compartments and variables (Figure 1.1) that interact and generate a wide range of timescales of variations. We could in principle consider variations at timescales up to the age of the Earth ($\sim$4.6 Ga), the paleoclimate, but we will restrict ourselves to the climatically comparably stable Holocene (past $\sim$11.5 ka). The data from this interval (Figures 1.2 and 1.3) display a wide range of timescales of climate variations.

Climate is a complex system that varies on a wide range of scales in time and space. This means, we do not know everything about the climate, we are uncertain. Therefore, a purely mathematical apparatus is insufficient to quantitatively describe the climate. We need the statistical language. We assume that a probability (a real number between 0 and 1) can be assigned to an uncertain event, such as "it rains tomorrow" or "during the Holocene the rate of occurrence of Indian Ocean monsoon droughts more than doubled." Statistics then infers events and probabilities from the data.

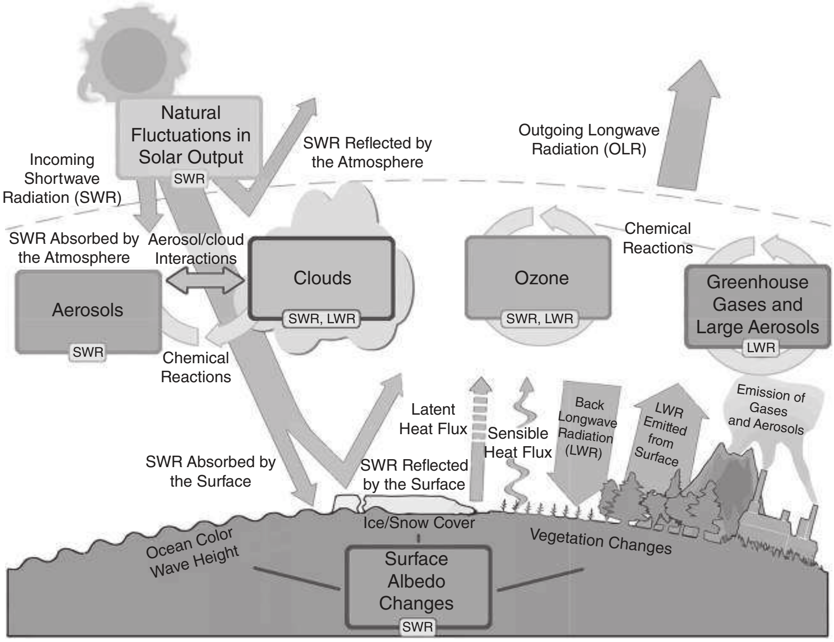

Figure 1.1 Main drivers of climate change. The radiative balance between incoming solar shortwave radiation (SWR) and outgoing longwave radiation (LWR) is influenced by global climate "drivers." Natural fluctuations in solar output (solar cycles) can cause changes in the energy balance (through fluctuations in the amount of incoming SWR). Human activity changes the emissions of gases and aerosols, which are involved in atmospheric chemical reactions, resulting in modified O_3 (ozone) and aerosol amounts. O_3 and aerosol particles absorb, scatter, and reflect SWR, changing the energy balance. Some aerosols act as cloud condensation nuclei, modifying the properties of cloud droplets and possibly affecting precipitation. Because cloud interactions with SWR and LWR are large, small changes in the properties of clouds have important implications for the radiative budget. Anthropogenic changes in greenhouse gases (GHGs, e.g., CO_2, CH_4, N_2O, O_3, CFCs) and large aerosols (>2.5 μm in size) modify the amount of outgoing LWR by absorbing outgoing LWR and re-emitting less energy at a lower temperature. Surface albedo is changed by variations in vegetation or land surface properties, snow or ice cover, and ocean color. These changes are driven by natural seasonal and diurnal changes (e.g., snow cover), as well as human influence (e.g., changes in vegetation types) (Forster et al. 2007). Original source of figure and legend: reproduced with permission from Stocker et al. (2013: figure 1.1 therein)

Climate evolves over time, and stochastic processes (that is, time-dependent random variables representing climate variables with not exactly known values) and time series (that is, the observed or sampled process) are central to

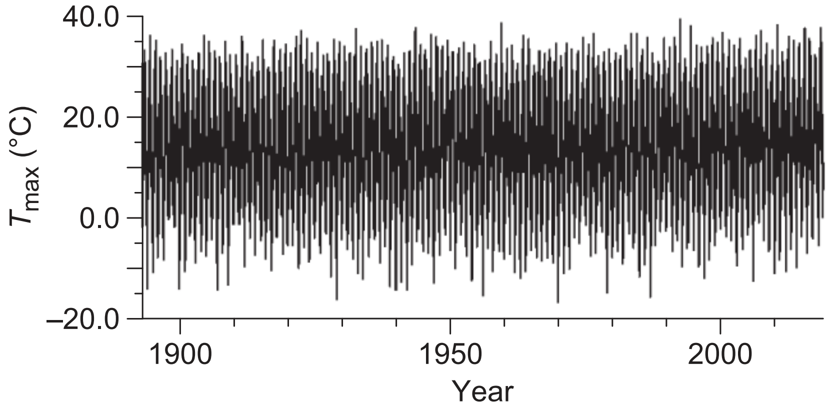

Figure 1.2 Maximum daily surface-air temperature at Potsdam, Germany, during the instrumental period. The interval is from 1 January 1893 to 31 December 2018; the data size is 46,020. The maximum is taken from the three daily temperature readings at 7:00 a.m., 2:00 p.m., and 9:00 p.m. Data from Friedrich-Wilhelm Gerstengarbe (Potsdam Institute for Climate Impact Research, Germany, personal communication, 2014) and www.pik-potsdam.de/services/climate-weather-potsdam/climate-diagrams (6 February 2019).

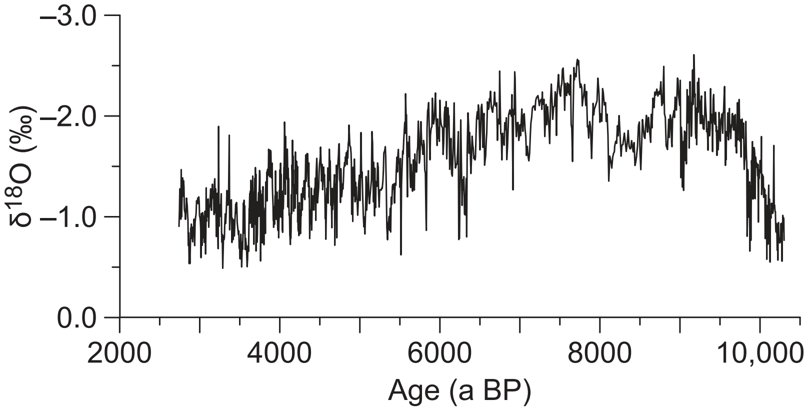

Figure 1.3 Indian Ocean monsoon rainfall during the Holocene. The interval is from 2741 to 10,300 years before the present (BP); the data size is 1345. The climate archive is stalagmite Q5 from Oman, which grew during the shown interval. The proxy variable oxygen isotopic composition (δ^{18}O) indicates the amount of rainfall, with low δ^{18}O reflecting strong monsoon. Time runs from right to left. The vertical scale is inverted in a paleoclimatic manner so that the transition from the last glacial to the present Holocene interglacial at around 10,000 a BP is "upward." Data from Fleitmann et al. (2003)

statistical climate analysis. We use a wide definition of trend and decompose a stochastic process, X, as follows:

$$X(T) = X_{\text{trend}}(T) + X_{\text{ext}}(T) + S(T) \cdot X_{\text{noise}}(T), \tag{1.1}$$

where T is continuous time, $X_{\text{trend}}(T)$ is the trend component, $X_{\text{ext}}(T)$ is the extreme component, $S(T)$ is a variability function scaling $X_{\text{noise}}(T)$, the noise component. The trend is seen to include all systematic or deterministic, long-term processes, such as a linear increase, a step change, or a seasonal signal. The trend is described by parameters, for example, the rate of an increase. Extremes are events with a large absolute value and are usually rare. The noise process is assumed to be weakly stationary with zero mean and autocorrelation. Giving $X_{\text{noise}}(T)$ standard deviation unity enables the introduction of $S(T)$ to honor climate's definition as not only the mean but also the variability of the state of the atmosphere and other compartments (Brückner 1890). A version of Eq. (1.1) is written for discrete time, $T(i)$, as

$$X(i) = X_{\text{trend}}(i) + X_{\text{ext}}(i) + S(i) \cdot X_{\text{noise}}(i), \tag{1.2}$$

using the abbreviation $X(i) \equiv X(T(i))$, etc. The observed, discrete time series from process $X(i)$ is the set of size n of paired values $t(i)$ and $x(i)$, compactly written as $\{t(i), x(i)\}_{i=1}^{n}$.

The big question concerns $X_{\text{ext}}(T)$. There are parameters that describe the extreme component. Statistical methods help to estimate the parameters using the time series data.

Box 1.2 Personal Reflection: Statistical notation and the climate equation

I was puzzled, as perhaps you now, when I first saw in statistics books the distinction made between X and x or between T and t. It indeed makes sense to keep the numerical value of a sample (e.g., $x = 20\ ^\circ$C) apart from the probabilistic concept of a random variable (e.g., $X =$ temperature at Potsdam). We do not know the exact value of the current temperature at Potsdam before it is measured. Before the measurement, however, we may be able to draw a curve of the chance of observing a certain temperature, x, against x on the basis of past measurements. This curve is called the probability density function (PDF) of the random variable X. (Strictly speaking, we have to speak of observing a value between x and $x + \delta$, with δ being arbitrarily small.)

The interesting thing with time series is that we now concatenate the values $x(i)$ according to the observation times $t(i)$; the symbol i is the counter. This allows the climate's memory to be taken into account: hot today, likely also hot tomorrow. This memory is mathematically referred to as autocorrelation, persistence, or serial dependence. The concatenated random variables, $X(i)$, are called a stochastic process. Our modeling of the climate process has to take into consideration also trend, extremes, and variability – and then we have the climate equation (Eq. 1.2). Our task is statistical inference: to guess the properties of the climate process, mainly its extreme component, using the sample.

A note on the climate equation: This is a conceptualization that has proven itself to be descriptive of what we know about the course of climate over time. It is not an equation derived from the first principles of physics – this is impossible for the complex, "dirty" climate system (Figure 1.1).

Since the climate variable, $X(T)$, includes the trend, $X_{\text{trend}}(T)$, we have either to assume absent trends or to estimate the trend component and remove it. On top of this, we have to separate $X_{\text{ext}}(T)$ from the noise component, $S(T) \cdot X_{\text{noise}}(T)$, which is also contained in the climate equation. Detection of extremes, their sizes, and when they occurred is the theme of Chapter 2. This chapter also details the different types of extreme data. Appendix A informs about the various climate variables being measured (observations and proxy) and Appendix B informs about climate archives. A special case is the time series that is produced by a climate model (Appendix C).

The statistical language, the methodological core, and the various statistical tools used to analyze the extreme component data are detailed in Chapter 3. These first three chapters put us in a position to apply the tools to real data.

The application chapters focus on the climate variables of strongest influence on society and economy: precipitation, temperature, and wind speed. We will study both positive and negative extremes. From precipitation extremes (Chapter 4), floods refer to positive high values and droughts to small but still positive values. Regarding temperature (Chapter 5), there are heatwaves and cold spells. In the case of wind speed (Chapter 6), we look at hurricanes and other storms but ignore doldrums.

For each of the three variables, precipitation, temperature, and wind speed, many new articles appear each week in the scientific literature, and it is difficult to keep pace. Therefore, we select a few, topical examples of climate extremes, which we analyze as case studies. Each chapter has at the end a section entitled "Reading Material," which contains a selection of major papers or books that have advanced the scientific field.

We hope that the book's structure will provide a solid basis for understanding data and methods. You will be able to apply the described methods to new data and obtain new results, which can then be interpreted. You can test your understanding in the exercises at the end of Chapters 2 and 3. Background information, which may be helpful for doing the exercises, is given on statistical estimation (Appendix D), numerical techniques (Appendix E), and data and software (Appendix F). For course instructors: The solutions to the exercises are given on the book's URL (secured part).

Finally, distributed throughout the book, boxes with "Personal Reflections" appear: stories, personal assessments, or explanations from the author, written in hopefully clear and understandable language. The intention is to make this major theme accessible to the student. The statistical analysis of climate extremes is, as you will experience, a cutting-edge scientific area of high socioeconomic relevance.

Reading Material

Bradley (1999) is a textbook on paleoclimatology, which is structured according to climate archives, dating tools, and proxy variables. Brönnimann et al. (2008) is an edited book on climate during the instrumental period, with a focus on temperature. Cronin (2010) is a textbook on paleoclimatology, which is structured according to the geologic timescale. Peixoto and Oort (1992) is a textbook on climate from a physics viewpoint. Pierrehumbert (2010) is a textbook on climate from a geophysics viewpoint, also on other planets than Earth; see www.cambridge.org/pierrehumbert (6 April 2019). The Fifth Assessment Report of the IPCC (Stocker et al. 2013) gives a comprehensive assessment of climate change; it can be downloaded from www.ipcc.ch (6 April 2019). The Sixth Assessment Report is due in 2021. Brückner (1890) is a long, early paper contributing to climate's definition not only in terms of the mean but also the variability. Fleitmann et al. (2003) is an early paper utilizing the stalagmite climate archive and the δ^{18}O proxy variable to study past monsoonal rainfall at high temporal resolution.

Bryant (1991) is a textbook on geologic and climate extremes, excluding temperature. Kropp and Schellnhuber (2011) is an edited book on climate extremes and their statistical analysis, with a focus on precipitation and floods. Mudelsee (2014) is a book on the statistical analysis of climate time series; it contains a chapter on extremes; see www.manfredmudelsee.com/book (6 April 2019).

The book by Bell (1986) can be described as mathematics narrated via biographies. Popper (2004) is a talk about realism and accessibility. Russell (1996) is a classic, an original, and it is no masticated food.