PART ONE

Data, Statistics, and Software

1

HURRICANES, CLIMATE, AND STATISTICS

"Chaos was the law of nature; Order was the dream of man."
—Henry Adams

This book is about hurricanes, climate, and statistics. These topics may not seem related. Hurricanes are violent winds and flooding rains, climate is about weather conditions from the past, and statistics is about numbers. But what if you wanted to estimate the probability of winds exceeding 60 m s^{-1} in Florida next year. The answer involves all three, hurricanes (fastest winds), climate (weather of the past), and statistics (probability). This book teaches you how to answer these questions in a rigorous and scientific way. We begin here with a short description of the topics and a few notes on what this book is about.

1.1 HURRICANES

A hurricane is an area of low air pressure over the warm tropical ocean. The low pressure creates showers and thunderstorms that start the winds rotating. The rotation helps to develop new thunderstorms. A tropical storm forms when the rotating winds exceed 17 m s^{-1} and a hurricane when they exceed 33 m s^{-1}.[1] Once formed, the winds continue to blow despite friction by an in-up-and-out circulation that imports heat at high temperature from the ocean and exports heat at lower temperature in the upper troposphere (near 16 km), which is similar to the way a steam engine converts thermal energy to mechanical motion. In short, a hurricane is powered by moisture and heat.

Strong winds are a hurricane's defining characteristic. Wind is caused by the change in air pressure between two locations. In the center of a hurricane, the air

[1] The winds are estimated over a one-minute duration at 10 m above the ocean surface.

pressure, which is the weight of a column of air from the surface to the top of the atmosphere, is quite low compared with the air pressure outside the hurricane. This difference causes the air to move from the outside inward toward the center. By a combination of friction as the air rubs on the ocean below and the spin of the earth as it rotates on its axis, the air does not move directly inward but rather spirals in a counterclockwise direction toward the region of lowest pressure.

Pressure differences between the cyclone center and the surrounding air determine the speed of the wind. Since the pressure outside most hurricanes is the same, a hurricane's central pressure is a good measure of a hurricane's intensity. The lower the central pressure, the more intense the hurricane. Pressures inside the most intense hurricanes are among the lowest that occur anywhere on the earth's surface at sea level. In the largest and most intense hurricanes, the strongest winds are located in the wall of thunderstorms that surrounds the nearly calm central eye. If the hurricane is stationary (spinning, but with no forward motion), the field of winds is shaped like a bagel, with a calm center and the fastest winds forming a ring around the center.

The distance from the center of the hurricane to the location of the hurricane's strongest winds is called the radius to maximum winds. In well-developed hurricanes, the strongest winds are in the eye-wall and the radius to maximum winds ranges from several kilometers in the smallest hurricanes to several hundred or more kilometers in the largest. While the wind just above the ocean surface spirals counterclockwise toward the center, the air at high altitudes blows outward in a clockwise spiral. This outward flowing air produces thin cirrus clouds that extend distances of thousands of kilometers from the center of circulation. The presence of these clouds may be the first visible sign that a hurricane is approaching.

Landfall occurs when the hurricane center crosses a coastline. Because the fastest winds are located in the eye-wall, it is possible for a hurricane's fastest winds to be over land even if landfall does not occur. Similarly, it is possible for a hurricane to make landfall and have its fastest winds remain out at sea. Fortunately, the winds slacken quickly after the hurricane moves over land. High winds destroy poorly constructed buildings and mobile homes. Flying debris such as signs and roofing material add to the destructive power.

Hurricanes also cause damage as a result of flooding rain and storm surge. Rainfall is the amount of water that falls over an area during a given time interval. Hurricanes derive energy from the ocean by evaporating water into the air, which then gets converted back to liquid through condensation inside the thunderstorm clouds. The liquid falls from the clouds as rain. The stronger the thunderstorms, the more the rain and the potential for flooding.

At a given location, the amount of rain that falls depends on factors including hurricane intensity, forward speed, and the underlying topography. Forward speed refers to how fast the center moves. Rainfall can be enhanced by winds blowing up a mountain. Antecedent moisture conditions also play a role. Freshwater flooding can be a serious danger even hundreds of kilometers from the point of landfall. Bands of showers and thunderstorms that spiral inward toward the center are the first weather experienced as a hurricane approaches. High wind gusts and heavy downpours occur in the individual rain bands, with relatively calm weather occurring between the bands. Brief tornadoes sometimes form in the rain bands especially as the cyclone moves inland.

The force of the winds moving around the storm creates the storm surge by pushing the ocean water toward the shore. Over the open ocean, the water can flow in all directions away from the storm. Strong winds blowing across the ocean surface create a stress that forces the water levels to increase downwind and decrease upwind. This wind setup is inversely proportional to ocean depth, so over the deep ocean away from land the water level rises are minimal. As the hurricane approaches shallow water, there is no room for the water to flow underneath so it rises and gets pushed by the wind as surge. Slope of the continental shelf and low atmospheric pressure also play a role. A shallow slope allows for greater surge. The ocean level rise due to the low air pressure adds to the surge, but its magnitude (about 1 cm per hectopascal drop in pressure) is less than that caused by the wind. The advancing surge may increase the water level 5 m or more above the sea level. In addition, wind-driven waves are superimposed on the storm surge. The total water level can cause severe surge impacts in coastal areas, particularly when the storm surge coincides with astronomical tides caused by the moon and sun.

1.2 CLIMATE

Climate is a set of weather patterns. One description of hurricane climate is the number of hurricanes over a given region and time. Other descriptions include the average hurricane intensity, the percentage of hurricanes that make it to land, or metrics that combine several attributes into one. Elsner and Kara (1999) present a wealth of climatological information on North Atlantic hurricanes.

On average, 50 hurricanes occur worldwide each year. Hurricanes develop during the time of the year when the ocean temperatures are the hottest. Over the North Atlantic, this includes the months of June through November, with a sharp peak from late August through the middle of September when the direct rays of the summer sun have had the largest impact on sea temperature. Worldwide, May is the least active month while September is the most active. During the twentieth century, hurricanes made landfall in the United States at an average rate of five every 3 years.

Hurricanes vary widely in intensity as measured by their fastest moving winds. Hurricane intensities are grouped into five categories (Saffir–Simpson scale) with the weakest category-one winds blowing at most 42 mps and the strongest category-five winds exceeding speeds of 69 mps. Three category-five hurricanes hit the United States during the twentieth century, including the Florida Keys Hurricane in 1935, Hurricane Camille in 1969, and Hurricane Andrew in 1992. Hurricanes also vary considerably in size (spatial extent), with the smallest hurricanes measuring only a few hundred kilometers in radius (measured from the eye center to the outermost closed line of constant surface pressure) and the largest exceeding a thousand kilometers or more.

Hurricanes are steered by large-scale wind patterns in the atmosphere above the surface and by the earth's spin. In the deep tropics, these forces push a hurricane slightly north of due west (in the Northern Hemisphere). Once north of about 23 degrees of latitude, a hurricane tends to take a more northwestward track and then eventually a northeastward at still higher latitudes. Local fluctuations in the magnitude and direction of steering can result in tracks that deviate significantly from this pattern.

Hurricane climate is linked to what is happening to the earth's global climate. Greater ocean heat, for example, enhances the potential for stronger hurricanes. This can occur on time scales of weeks to years. El Niño is a good example. It is the name given to climate changes over the tropical Pacific region caused by warming of the ocean surface by a few degrees. This leads to a shift in the tropical thunderstorms normally over the western Pacific near Indonesia eastward toward the central and eastern Pacific, closer to the Atlantic. The thunderstorms create upper air winds that inhibit hurricane formation. Another example is the position and strength of the subtropical high-pressure ridge, which on average is situated at a latitude of 30° N. A hurricane forming on the equatorward side of the ridge gets steered toward the northwest. During some years, the subtropical ridge is farther south and west, forcing the hurricanes westward through the Caribbean Sea and toward Mexico and the United States.

1.3 STATISTICS

Statistics is the collection, organization, analysis, and interpretation of data. Here you will use it to describe, model, and predict hurricane climate. Like people, all hurricanes are unique. A person's behavior might be difficult to anticipate, but the average behavior of a group of people is quite predictable. A particular hurricane may move eastward through the Caribbean Sea, but hurricanes typically travel westward in this part of the world. Statistical models quantify this typicalness. Furthermore, statistical models provide a syntax for describing uncertainty.

Statistics is not the same as arithmetic. A strong hurricane might be described as a 1-in-100-year event. If you do the arithmetic, you might conclude that after the event there is plenty of time ($100 - 1 = 99$ years) before the next one. This is wrong. Statistics tells you that you can expect a hurricane of this magnitude or larger to occur once in 100 years *on average*, but it could occur next year with a probability of 1 percent. Moreover, hurricanes a bit weaker only will be more frequent and perhaps a lot more so.

The adage is that you can use statistics to anything. But this assumes deceit. Without statistics anything you say is mere opinion. Recent upward trends in hurricane activity have spurred a debate as to the cause. Statistics has helped to clarify the issues. A good reference textbook for statistical methods in weather and climate is available in Wilks (2006). Statistics can be divided into descriptive and inferential. Descriptive statistics is used to summarize data. Data are hurricane records. Inferential statistics is used to draw conclusions about the processes that regulate hurricane climate. They help you understand how climate operates.

Inference is used to test competing scientific hypotheses. Inference from data requires attention to randomness, uncertainty, and data biases. It also requires a model. A model is a concise (as few parameters as possible) description of your data. Set of counts representing the number of hurricanes occurring in a region each year, can be described using a Poisson distribution (model) and a rate (parameter). Given the rate, the modeled counts should match the observed counts. This allows you to understand hurricane frequency and how it varies with climate through variations in the rate.

Statistical models are needed because proof is impossible. The primary goal is to determine what is *likely* to be true. When you toss a frisbee, you know it will come down and where. This is physics. By contrast, if you smoke you might get lung cancer. If the seas get hotter, hurricanes might get stronger. You, know that your odds of getting cancer are higher if you smoke. And you know that the odds of stronger hurricanes are higher if the seas are warmer. You can work out these odds with a statistical model, but you cannot know with certainty whether you as an individual smoker will get cancer, or whether next hot year will bring stronger hurricanes.

The relationship between hurricanes and climate can be subtle. Differences in the spatial and temporal scales are large. Hurricanes form in many different ways, are impacted by various environmental conditions, and dissipate under all sorts of influences. Environmental factors, genesis mechanisms, and feedbacks interact in complex and multifaceted ways, so establishing proof is impossible. So instead, you observe, compare, and weigh the evidence for or against various hypotheses.

Despite limitations to predicting individual behavior of hurricanes beyond a few days imposed by thus complexity, it is possible to predict average behavior at lead times of weeks to decades. Hurricane climate predictability arises from the slow evolution of ocean heating and due to processes associated with monthly to seasonal dynamics of the atmosphere–ocean system. Progress has been made, much of it from using statistical models. Most of what we know about hurricanes derives from observation. Statistical models are built from observational data. Observations become more precise over time leading to heterogeneous data sets, but a good modeling strategy is to use as much information as possible.

Inferential statistics come in two flavors: frequentist and Bayesian. Frequentist (classical) inference is what you learn in school. It involves methods of hypothesis testing and confidence intervals.

Bayesian inference is a way to calculate a probability that your hypothesis is true by combining data with prior belief. The term "Bayesian" comes from Bayes theorem, which provides the formula for the calculation. For example, what is the probability that you are correct in claiming that the strongest hurricanes are getting stronger? With Bayesian inference, probability is your degree of belief rather than an intrinsic characteristic of the world.

The underlying Bayesian principle is the accumulation of evidence. Evidence comes from historical and geologic data that, by their very nature, are incomplete and fragmentary. Much has been written about the differences in philosophy between classical and Bayesian approaches. Here we are pragmatic and use both. Frequentist methods are used except in cases where we think Bayesian methods provide an advantage.

1.4 R

Science demands openness and reproducibility. It is essential that you explain exactly what you did and how you did it. All computations should be reproducible. The R language makes this easy to do. Redoing all fieldwork is asking too much. But with climate research, reproducibility is perfectly possible.

R contains a number of built-in mechanisms for organizing, graphing, and modeling your data. Directions for obtaining R, accompanying packages, and other sources of documentation are provided at `http://www.r-project.org/`. R is an open-source project, which means that it depends on a community of active developers to grow and evolve. No person or company owns it. R is maintained and supported by thousands of individuals worldwide who use it and who contribute to its ongoing development. R is mostly written in C (with R and some Fortran); R packages are mostly written in R and C.

The book is interlaced with R code, so you can read it as a workbook. It was written in the Sweave format of Leisch (2003). Sweave is an implementation of the literate programming style advocated by Knuth (1992) which permits an interplay between code written in R, the output of that code, and commentary on the code. Sweave documents are preprocessed by R to produce a LaTeX document. The book also shows you how to use R to construct graphs and maps for presenting your data including modern tools for data inspection and visualization. The text uses computer modern font, with italics for data set names outside of R. A bold face is used for package names and a typewriter font for the R code.

1.5 ORGANIZATION

The book has two parts. The first part (Chapters 1–6) concerns the background material focusing on software, statistics, and data. The second part (Chapters 7–13) concerns methods and models used in hurricane climate research. Chapter 2 is a tutorial on using R. If you have never used R before, start here. The material is presented using examples from hurricane climatology. Chapter 3 provides a review of introductory statistics. Topics include distributions, two-sample tests, correlation, and regression. Chapter 4 is an introduction to Bayesian statistics. Chapter 5 is how to make graphs and maps from your data. Chapter 6 gives information about the various data sets used in the later chapters.

Chapter 7 is on models for hurricane occurrence. Emphasis is on the Poisson distribution for count data. Chapter 8 is on models for hurricane intensity including quantile regression and models from extreme value theory. Chapter 9 is on spatial models. We examine models for areal units and field data. Chapter 10 looks at various types of cluster models including temporal, spatial, and feature clusters. Chapter 11 examines time series models including a look at the novel application of networks. Chapter 12 provides examples of Bayesian models including those that make use of Markov chain Monte Carlo methods. Chapter 13 considers a few impact models.

Chapters 2–6 provide introductory material and a tutorial in using R. The material can be presented as part of a course on statistical methods in the environmental sciences at the undergraduate level. Chapter 6 provides details on the data sets that are used in the later chapters and can be skipped on first reading. Chapters 7, 8, and 9 provide the basic building blocks of models for hurricane climate research. This material is appropriate for advanced undergraduate- and graduate-level courses in climatology. Chapters 10–13 show examples from more recent research that can be used in a graduate seminar on methods and models for hurricane analysis and prediction.

The examples come from our research on hurricanes of the North Atlantic. All the data sets and code used in the book are available at www.hurricaneclimate.com, and readers are encouraged to reproduce the figures and graphs. We provide references to the relevant literature on the methods, but do not focus on this. We use standard statistical notation and a voice that makes the reader the subject.

This book does not cover global climate models (GCMs). Interest in GCMs is predicated on the assumption that further progress in forecasting will come only through improvements in dynamical models. If the hurricane response to variations in climate is highly nonlinear or involves many intervening variables then dynamical model skill will outpace skill from statistical models. On the other hand, if the response is near linear and associated with only a handful of variables, improvements to dynamical models will result in forecasts that largely duplicate the skill available using statistical models but at substantially greater expense.

Consider that a dynamical model of all the cells in your body and the associated chemistry and physics is not likely to provide a more accurate prediction of when you will die (from "natural" causes) than a statistical model based on your cohorts (age, sex, etc.) that takes into account a few important variables related to diet and exercise. The analogy is not perfect as there is not a cohort of earth climates from which to build a statistical model. But under the assumption of homogeneity and stationarity, you have pseudoreplicates that can be used as cohorts. If we are right, then greater attention needs to be paid to statistics.

And there is no better way to do statistics than with R. Chapter 2 provides a tutorial, on how to get started.