Preface

The environmental movement of the 1960s and 1970s resulted in the creation of several laws aimed at protecting the environment, and in the creation of federal, state, and local government agencies charged with enforcing these laws. Most of these laws mandate monitoring or assessment of the physical environment, which means someone has to collect, analyze, and explain environmental data. Numerous excellent books, guidance documents, and journal articles have been published to explain various aspects of applying statistical methods to these kinds of environmental data analyses. Also, several specialty software packages for specific niches in environmental statistics exist, such as for ecology, forestry, and climate modeling. Not very many software packages provide a comprehensive treatment of environmental statistics in the context of monitoring the physical environment as mandated by current environmental law.

EnvStats is an R package for environmental statistics. It is the open-source successor to the commercial module for S-Plus© called EnvironmentalStats for S-Plus, which was first released in April 1997. The EnvStats package, along with the R software environment, provides comprehensive and powerful software for environmental data analysis. EnvStats brings the major environmental statistical methods found in the literature and regulatory guidance documents into one statistical package, along with an extensive hypertext help system that explains *what* these methods do, *how* to use these methods, and *where* to find them in the environmental statistics literature. Also included are numerous built-in data sets from regulatory guidance documents and the environmental statistics literature. EnvStats combined with other R packages (e.g., for spatial analysis) provides the environmental scientist, statistician, researcher, and technician with tools to "get the job done!"

EnvStats and this user's manual are intended for anyone who has to make sense of environmental data, including statisticians, hydrologists, soil scientists, atmospheric scientists, geochemists, environmental engineers and consultants, hazardous and solid waste site managers, and regulatory agency analysts and enforcement officers. Some parts of EnvStats incorporate statistical methods that have appeared in the environmental literature but are not commonly found in any statistical software package. Some parts are specifically aimed at users who are required to collect and analyze environmental monitoring data in order to comply with federal and state Superfund, RCRA, CERCLA, and Subtitle D regulations for environmental monitoring at hazardous and solid waste sites. All of the functions in EnvStats, however, are useful to anyone who needs to analyze environmental data. In fact, all of these functions are useful to anyone who needs to analyze data.

This manual is divided into 9 chapters. Chapter 1 is an introduction to environmental statistics in general and the EnvStats package in particular, and includes information on system and user requirements, installing the software, loading and using the package, and getting technical support. The last section of the chapter includes a tutorial.

Chapters 2, 3, 4, 5, 6, 7, and 8 contain information about how to use the functions in EnvStats to design sampling programs and perform graphical and statistical analyses of environmental data. Chapter 9 shows you how to use EnvStats to perform Monte Carlo simulation and probabilistic risk assessment.

At the back of the book is an extensive list of references for environmental statistics as well as an index of key words and terms. In addition to using the index, you are encouraged to use the online hypertext help system as well.

Companion Scripts

Companion R scripts to reproduce the examples in this user's manual, as well as scripts for reproducing examples in US EPA guidance documents, are located in the **scripts** subdirectory of the directory where the package was installed. See Chap. 1 for more information.

Companion Textbook and Help Files

This user's manual provides ***brief*** explanations of various topics in environmental statistics. A companion textbook, currently in preparation and titled *Environmental Statistics with R* (Millard et al. 2014), provides more details and can be used as a textbook for a course in environmental statistics. (The predecessor to this textbook is Millard and Neerchal 2001.) Space constraints dictate that the examples in this user's manual convey a general sense of how various EnvStats functions can be used. The companion help files list all of the arguments associated with these functions and give more examples.

Technical Support

Technical support for R is available through the R-help mailing list (see the URL www.r-project.org for more information). Technical support for questions or problems specifically related to the functioning of the EnvStats package is available by e-mailing the author at EnvStats@ProbStatInfo.com.

Typographic Conventions

Throughout this user's manual, the following typographic conventions are used:

- The **bold font** is used for chapter and section headings, as well as operating system commands and file names. Sometimes it is also used for emphasis. R menu selections are shown in an abbreviated form using this font and the arrow symbol (>) to indicate a selection within a menu, as in **Packages>Load package…**.
- The *italic courier font* is used to display what you type within an R Command or Script Window.
- The `courier font` is used to display output from R and the names of R objects.
- The *italic font* is used for chapter and help file titles within the text, emphasis, and user-supplied variables within R commands.
- The ***bold italic font*** is used for emphasis.

R commands are preceded with the "greater than" sign, i.e., >, which is the default R prompt. For commands that require more than one line of input, the line or lines following the first line are indented, whereas within R they are preceded with the "plus" sign, i.e., +, which is the default R continuation prompt. Note that page size and formatting for this book determine how command lines are split. As a user, you can choose to split lines differently or not at all.

A Note About Figure Titles

To conserve space, very few of the plots shown in this user's manual have titles above them since all plots are labeled at the bottom with a figure number and title. However, several plotting functions in EnvStats produce figures with titles at the top by default if the argument `main` is not supplied. Thus, there are several examples in this manual where the figure shown does not include a title at the top, but if you type in the commands as shown or run the companion script to produce the figure, you will produce a figure with a title at the top. To create a different title from the one shown, use the `main` argument or, to suppress the title, set the argument `main=""`.

Acknowledgments

In the early 1980s, while pursuing a degree in biostatistics, I became aware of a knowledge gap in the field of environmental statistics. There was lots of research going on in the academic field of environmental statistics, but there were lots of poor designs and analyses being carried out in the real-world "field." There were

even federal laws mandating incorrect statistical analyses of groundwater samples at hazardous waste sites.

One of the first steps to improving the quality of environmental statistics is to improve the quality of communication between statisticians and experts in various environmental fields. Several researchers and practitioners have done this by publishing excellent books and journal articles dealing with general and specific problems in environmental statistics (see the references listed at the back of this book).

The next logical step is to provide the necessary tools to carry out all the great methods and ideas in the literature. In the mid-1980s, I sat down with two friends of mine, Dennis Lettenmaier and Jim Hughes, to talk about building a software package for environmental statistics. Dennis and Jim gave me some good ideas, but I never acted on these ideas because at the time I didn't feel like I had the tools to make the kind of software package I wanted to.

The emergence of S-Plus© and help-authoring tools in the early 1990s changed all that. S-Plus at that time was one of the few statistical software packages with great graphics that allowed users to write their own functions and create pull-down menus. ForeHelp© was one of several help-authoring tools that allowed anyone who can figure out a word processor to write a hypertext Windows, HTML, or JAVA help system. The availability of these tools resulted in the creation of EnvironmentalStats for S-Plus, which was first released in April 1997. The emergence of open-source software, including R, has led to a revolution in the availability of specialty tools for data analysis. Porting EnvironmentalStats for S-Plus to the R environment in the form of the EnvStats package, where it can be integrated with hundreds of other R packages, is the next logical step in its evolution.

There are several people who helped and encouraged me over the past several years as I developed and extended first EnvironmentalStats for S-Plus and then the R package EnvStats. Most recently, thanks to Charles Davis (Envirostat & Statistics Ltd) and Kirk Cameron (MacStat Consulting Ltd) for help with updating the functions for simultaneous nonparametric prediction intervals, and to Phil Dixon (Iowa State University) for early feedback on EnvStats.

Thanks to Insightful Corporation (now part of TIBCO) for their past marketing, distribution, and support of EnvironmentalStats for S-Plus. Several people at Insightful contributed to the success of EnvironmentalStats for S-Plus, including Kim Leader, Scott Blachowicz, Christopher Disdero, Cheryl Mauer, Patrick Aboyoun, Tim Wegner, Rich Calaway, Charlie Roosen, Stephen Kaluzny, and the Insightful technical support team. Special thanks to the members of the S-news group for constantly giving me insight into the workings of S-Plus. I would also like to thank Gilbert FitzGerald, former Director of Product Development at MathSoft, for his encouragement and advice. Thanks to Jim Hughes (University of Washington) for being a statistical consultant's statistical consultant. Also thanks to Tim Cohn and Dennis Helsel (US Geological Survey), Chris Fraley (Insightful and University of Washington), Jon Hosking (IBM Research Division),

Ross Prentice (Fred Hutchinson Cancer Research Center), and Terry Therneau (Mayo Clinic) for their help in answering technical questions. Dick Gilbert was a constant source of encouragement and advice – thanks! Thanks to the Beta testers of EnvironmentalStats for S-Plus for their feedback on early versions.

There would be no EnvStats package for R unless there was R, so I am indebted to Ross Ihaka and Robert Gentleman, the creators of R, and the R development core team. There would have been no EnvironmentalStats for S-Plus if there had been no S-Plus, so I am grateful to Doug Martin for leading the creation of S-Plus. There would be no S-Plus and no R if there had not first been S. I am grateful to the researchers of the Bell Laboratories S team at AT&T who first created S, including Richard A. Becker, John M. Chambers, Alan R. Wilks, William S. Cleveland, and others.

I am grateful to Marc Strauss and Hannah Bracken at Springer-Verlag for their help in transforming this user's manual into a book. Finally, and most gratefully, thanks to my wife Stacy and my son Chris for their moral support. I couldn't have done it without you!

EnvStats is the culmination of a dream I had over 25 years ago. I hope it provides you with the tools you need to "get the job done," and I hope you enjoy it as much as I have enjoyed creating it!

Seattle, WA
March 2013   *Steven P. Millard*