

ABOUT

ENVIRONMENTAL DATA ANALYSIS WITH *MATLAB*

Environmental Data Analysis with MatLab is for students and researchers working to analyze real data sets in the environmental sciences. It teaches the basics of the underlying theory of data analysis, and then reinforces that knowledge with carefully chosen, realistic scenarios. *MatLab*, a commercial data processing environment, is used in these scenarios; significant content is devoted to teaching how it can be effectively used in an environmental data analysis setting. The book, though written in a self-contained way, is supplemented with data sets and *MatLab* scripts that can be used as a data analysis tutorial.

Environmental Data Analysis with MatLab is organized around two principles. The first is that real proficiency in data analysis requires analyzing realistic data on a computer, and not merely working through ultra-simplified examples with pencil and paper. The second is that the skills needed to perform data analysis are best learned in a series of steps that alternate between theory and application and that start simple but rapidly expand as one's toolkit of skills grows. The real world puts many impediments in the way of analyzing data – errors of all sorts, missing information, inconvenient units of measurements, inscrutable data formats, and more. Consequently, real proficiency is as much about confidence and experience as it is about formal knowledge of techniques. This book teaches a core set of techniques that are widely applicable across all of Environmental Science, and it reinforces them by leading the student through a series of case studies on real-world data that has both the richness and the blemishes inherent in real-world things.

Two fundamental themes are used to tie together many different data analysis techniques:

The first is that measurement *error* is a fundamental aspect of observation and experiment. Error has a profound influence on the way that knowledge is distilled from data. We use probability theory to develop the concept of *covariance*, the key tool for quantifying error. We then show how covariance propagates through a chain of calculations leading to result that possesses uncertainty. Dealing with that uncertainty is as an important part of data analysis as arriving at the result, itself. From Chapter 3, where it is introduced, through the book's end, we are always returning to the idea of the propagation of error.

The second is that many problems are special cases of a *linear model* linking the observations to the knowledge that we aspire to derive from them. Measurements of the world around us create data, numbers that describe the results of observations and experiments. But measurements, in and of themselves, are of little utility. The purpose of data analysis is to distill them down to a few significant and insightful *model parameters*. We develop the idea of the linear model in Chapter 4 and in subsequent chapters show that very many, seeming-different, data analysis techniques are special cases of it. These include curve fitting, Fourier analysis, filtering, factor analysis and empirical function analysis and interpolation. While their uses are varied, they all

share a common structure, which when recognized, makes understanding them easier. Most importantly, covariance propagates through them in nearly identical ways.

As the title of this book implies, it relies very heavily on *MatLab* to connect the theory of data analysis to its practice in the real world. *MatLab*, a commercial product of *The MathWorks, Inc.*, is a popular scientific computing environment that fully supports data analysis, data visualization and data file manipulation. It includes a scripting language through which complicated data analysis procedures can be developed, tested, performed and archived. *Environmental Data Analysis with MatLab* makes use of scripts in three ways. First, the text includes many short scripts and excerpts from scripts that illustrate how particular data analysis procedures are actually performed. Second, a set of complete scripts and accompanying datasets is provided as a companion to the book. They implement all of the book's figures and case studies. Third, each chapter includes recommended homework problems that further develop the case studies. They require existing scripts to be modified and new scripts to be written.

Environmental Data Analysis with MatLab is a relatively short book that is appropriate for a one-semester course at the upper-class undergraduate and graduate level. It requires a working knowledge of calculus and linear algebra, as might be taught in a two-semester undergraduate calculus course. It does *not* require any knowledge of differential equations or more advanced forms of applied mathematics. Students with some familiarity with the practice of Environmental Science and with its underlying issues will be better able to put examples in context, but detailed knowledge of the science is not required. The book is self-contained; it can be read straight through, and profitably, even by someone with no access to *MatLab*. But it is meant to be used in a setting where students are actively using *MatLab* both as an aid to studying (that is, by reproducing the case studies described in the book) and as a tool for completing the recommended homework problems.

Environmental Data Analysis with MatLab uses six exemplary Environmental Science datasets:

- Air temperature;
- Chemical composition of sea floor samples;
- Ground level ozone concentration;
- Sea surface temperature;
- Stream flow; and
- Water quality.

Most datasets are used in several different contexts and in several different places in the text. They are used both as a test bed for particular data analysis techniques and to illustrate how data analysis can lead to important insights about the environment.