Preface

A Third Need in Engineering Education

At its inception, engineering education was predominantly *process oriented*, while engineering practice tended to be predominantly *system oriented*[1]. While it was invaluable to have a strong fundamental knowledge of the processes, educators realized the need to have courses where this knowledge translated into an ability to design systems; therefore, most universities, starting in the 1970s, mandated that seniors take at least one design/capstone course. However, a third aspect is acquiring increasing importance: *the need to analyze, interpret and model data*. Such a skill set is proving to be crucial in all scientific activities, none so as much as in engineering and the physical sciences. How can data collected from a piece of equipment be used to assess the claims of the manufacturers? How can performance data either from a natural system or a man-made system be respectively used to maintain it more sustainably or to operate it more efficiently? Such needs are driven by the fact that system performance data is easily available in our present-day digital age where sensor and data acquisition systems have become reliable, cheap and part of the system design itself. This applies both to experimental data (gathered from experiments performed according to some predetermined strategy) and to observational data (where one can neither intrude on system functioning nor have the ability to control the experiment, such as in astronomy). Techniques for data analysis also differ depending on the size of the data; smaller data sets may require the use of "prior" knowledge of how the system is expected to behave or how similar systems have been known to behave in the past.

Let us consider a specific instance of observational data: once a system is designed and built, how to evaluate its condition in terms of design intent and, if possible, operate it in an "optimal" manner under variable operating conditions (say, based on cost, or on minimal environmental impact such as carbon footprint, or any appropriate pre-specified objective). Thus, data analysis and data driven modeling methods as applied to this instance can be meant to achieve certain practical ends—for example:

(a) verifying stated claims of manufacturer;
(b) product improvement or product characterization from performance data of prototype;
(c) health monitoring of a system, i.e., how does one use quantitative approaches to reach sound decisions on the state or "health" of the system based on its monitored data?
(d) controlling a system, i.e., how best to operate and control it on a day-to-day basis?
(e) identifying measures to improve system performance, and assess impact of these measures;
(f) verification of the performance of implemented measures, i.e., are the remedial measures implemented impacting system performance as intended?

[1] Stoecker, W.F., 1989. *Design of Thermal Systems*, 3rd Edition, McGraw-Hill, New York

Intent

Data analysis and modeling is not an end in itself; it is a well-proven and often indispensable aid for subsequent decision-making such as allowing realistic assessment and predictions to be made concerning verifying expected behavior, the current operational state of the system and/or the impact of any intended structural or operational changes. It has its roots in statistics, probability, regression, mathematics (linear algebra, differential equations, numerical methods,…), modeling and decision making. Engineering and science graduates are somewhat comfortable with mathematics while they do not usually get any exposure to decision analysis at all. Statistics, probability and regression analysis are usually squeezed into a sophomore term resulting in them remaining "a shadowy mathematical nightmare, and … a weakness forever"[2] even to academically good graduates. Further, many of these concepts, tools and procedures are taught as disparate courses not only in physical sciences and engineering but in life sciences, statistics and econometric departments. This has led to many in the physical sciences and engineering communities having a pervasive "mental block" or apprehensiveness or lack of appreciation of this discipline altogether. Though these analysis skills can be learnt over several years by some (while some never learn it well enough to be comfortable even after several years of practice), what is needed is a textbook which provides:

1. A review of classical statistics and probability concepts,
2. A basic and unified perspective of the various techniques of data based mathematical modeling and analysis,
3. an understanding of the "process" along with the tools,
4. a proper combination of classical methods with the more recent machine learning and automated tools which the wide spread use of computers has spawned, and
5. well-conceived examples and problems involving real-world data that would illustrate these concepts within the purview of specific areas of application.

Such a text is likely to dispel the current sense of unease and provide readers with the necessary measure of practical understanding and confidence in being able to interpret their numbers rather than merely generating them. This would also have the added benefit of advancing the current state of knowledge and practice in that the professional and research community would better appreciate, absorb and even contribute to the numerous research publications in this area.

Approach and Scope

Forward models needed for system simulation and design have been addressed in numerous textbooks and have been well-inculcated into the undergraduate engineering and science curriculum for several decades. It is the issue of data-driven methods, which I feel is inadequately reinforced in undergraduate and first-year graduate curricula, and hence the basic rationale for this book. Further, this book is not meant to be a monograph or a compilation of information on papers i.e., not a literature review. It is meant to serve as a textbook for senior undergraduate or first-year graduate students or for continuing education professional courses, as well as a self-study reference book for working professionals with adequate background.

---

[2] Keller, D.K., 2006. *The Tao of Statistics*, Saga Publications, London, U.K

Applied statistics and data based analysis methods find applications in various engineering, business, medical, and physical, natural and social sciences. Though the basic concepts are the same, the diversity in these disciplines results in rather different focus and differing emphasis of the analysis methods. This diversity may be in the process itself, in the type and quantity of data, and in the intended purpose of the analysis. For example, many engineering systems have low "epistemic" uncertainty or uncertainty associated with the process itself, and, also allow easy gathering of adequate performance data. Such models are typically characterized by strong relationships between variables which can be formulated in mechanistic terms and accurate models consequently identified. This is in stark contrast to such fields as economics and social sciences where even qualitative causal behavior is often speculative, and the quantity and uncertainty in data rather poor. In fact, even different types of engineered and natural systems require widely different analysis tools. For example, electrical and specific mechanical engineering disciplines (ex. involving rotary equipment) largely rely on frequency domain analysis methods, while time-domain methods are more suitable for most thermal and environmental systems. This consideration has led me to limit the scope of the analysis techniques described in this book to thermal, energy-related, environmental and industrial systems.

There are those students for whom a mathematical treatment and justification helps in better comprehension of the underlying concepts. However, my personal experience has been that the great majority of engineers do not fall in this category, and hence a more pragmatic approach is adopted. I am not particularly concerned with proofs, deductions and statistical rigor which tend to overwhelm the average engineering student. The intent is, rather, to impart a broad conceptual and theoretical understanding as well as a solid working familiarity (by means of case studies) of the various facets of data-driven modeling and analysis as applied to thermal and environmental systems. On the other hand, this is not a cookbook nor meant to be a reference book listing various models of the numerous equipment and systems which comprise thermal systems, but rather stresses underlying scientific, engineering, statistical and analysis concepts. It should not be considered as a substitute for specialized books nor should their importance be trivialized. A good general professional needs to be familiar, if not proficient, with a number of different analysis tools and how they "map" with each other, so that he can select the most appropriate tools for the occasion. Though nothing can replace hands-on experience in design and data analysis, being familiar with the appropriate theoretical concepts would not only shorten modeling and analysis time but also enable better engineering analysis to be performed. Further, those who have gone through this book will gain the required basic understanding to tackle the more advanced topics dealt with in the literature at large, and hence, elevate the profession as a whole. This book has been written with a certain amount of zeal in the hope that this will give this field some impetus and lead to its gradual emergence as an identifiable and important discipline (just as that enjoyed by a course on modeling, simulation and design of systems) and would ultimately be a required senior-level course or first-year graduate course in most engineering and science curricula.

This book has been intentionally structured so that the same topics (namely, statistics, parameter estimation and data collection) are treated first from a "basic" level, primarily by reviewing the essentials, and then from an "intermediate" level. This would allow the book to have broader appeal, and allow a gentler absorption of the needed material by certain students and practicing professionals. As pointed out by Asimov[3], the Greeks demonstrated that abstraction

[3] Asimov, I., 1966. *Understanding Physics: Light Magnetism and Electricity*, Walker Publications.

(or simplification) in physics allowed a simple and generalized mathematical structure to be formulated which led to greater understanding than would otherwise, along with the ability to subsequently restore some of the real-world complicating factors which were ignored earlier. Most textbooks implicitly follow this premise by presenting “simplistic” illustrative examples and problems. I strongly believe that a book on data analysis should also expose the student to the “messiness” present in real-world data. To that end, examples and problems which deal with case studies involving actual (either raw or marginally cleaned) data have been included. The hope is that this would provide the student with the necessary training and confidence to tackle real-world analysis situations.

Assumed Background of Reader

This is a book written for two sets of audiences: a basic treatment meant for the general engineering and science senior as well as the general practicing engineer on one hand, and the general graduate student and the more advanced professional entering the fields of thermal and environmental sciences. The exponential expansion of scientific and engineering knowledge as well as its cross-fertilization with allied emerging fields such as computer science, nanotechnology and bio-engineering have created the need for a major reevaluation of the thermal science undergraduate and graduate engineering curricula. The relatively few professional and free electives academic slots available to students requires that traditional subject matter be combined into fewer classes whereby the associated loss in depth and rigor is compensated for by a better understanding of the connections among different topics within a given discipline as well as between traditional and newer ones.

It is presumed that the reader has the necessary academic background (at the undergraduate level) of traditional topics such as physics, mathematics (linear algebra and calculus), fluids, thermodynamics and heat transfer, as well as some exposure to experimental methods, probability, statistics and regression analysis (taught in lab courses at the freshman or sophomore level). Further, it is assumed that the reader has some basic familiarity with important energy and environmental issues facing society today. However, special effort has been made to provide pertinent review of such material so as to make this into a sufficiently self-contained book.

Most students and professionals are familiar with the uses and capabilities of the ubiquitous spreadsheet program. Though many of the problems can be solved with the existing (or add-ons) capabilities of such spreadsheet programs, it is urged that the instructor or reader select an appropriate statistical program to do the statistical computing work because of the added sophistication which it provides. This book does not delve into how to use these programs, rather, the focus of this book is *education-based* intended to provide knowledge and skill sets necessary for value, judgment and confidence on how to use them, as against training-based whose focus would be to teach facts and specialized software.

Acknowledgements

Numerous talented and dedicated colleagues contributed in various ways over the several years of my professional career; some by direct association, others indirectly through their textbooks and papers-both of which were immensely edifying and stimulating to me personally. The list of acknowledgements of such meritorious individuals would be very long indeed, and so I have limited myself to those who have either provided direct valuable suggestions on the overview and scope of this book, or have generously given their time in reviewing certain chapters of this book. In the former category, I would like to gratefully mention Drs. David Claridge, Jeff Gordon, Gregor Henze John Mitchell and Robert Sonderegger, while in the latter, Drs. James Braun, Patrick Gurian, John House, Ari Rabl and Balaji Rajagopalan. I am also appreciative of interactions with several exceptional graduate students, and would like to especially thank the following whose work has been adopted in case study examples in this book: Klaus Andersen, Song Deng, Jason Fierko, Wei Jiang, Itzhak Maor, Steven Snyder and Jian Sun. Writing a book is a tedious and long process; the encouragement and understanding of my wife, Shobha, and our children, Agaja and Satyajit, were sources of strength and motivation.

Tempe, AZ, December 2010 T. Agami Reddy

Applied Data Analysis and Modeling for Energy Engineers and Scientists
Reddy, T.A.
2011, XXI, 430 p., Hardcover
ISBN: 978-1-4419-9612-1