The new chapter on surface conditions will be of interest to both novices and veterans. The issue is to clarify the notion of local equilibrium at a phase boundary, i.e., When can one assert that the concentration at a phase boundary is pinned to its equilibrium value, even as an adjacent bulk phase undergoes nonequilibrium processes? Such questions arise, for example, in the catalytic properties of metal surfaces and in the change in doping character of semiconductors during annealing. Ghez discusses how Gibbs's theory of equilibrium can be extended to diffusion processes, which are nonequilibrium. The analysis leads to a generalization of the Stefan condition. There is also a section on diffusion in a grain boundary that separates two different bulk phases.

In summary, this is a highly readable, excellent introduction to diffusion processes and the diffusion equation. It would be useful as a text for an upper division undergraduate course or graduate course in modeling, or in a special topics course on diffusion itself. Several challenging exercises complement the exposition, and notes at the end of the chapters steer the reader to further explorations. Even veteran practitioners of diffusion can get a different perspective.

REFERENCE

[1] R. Ghez, *A Primer of Diffusion Problems*, Wiley, New York, 1988.

J. David Logan
University of Nebraska at Lincoln

Principles of Data Mining. *By David Hand, Heikki Mannila, and Padhraic Smyth.* MIT Press, Cambridge, MA, 2001. $50.00. xxxii+546 pp., hardcover. ISBN 0-262-08290-X.

Is data mining the same as statistics? The distinguished authors of *Principles of Data Mining* struggle to make a distinction between the two subjects. In the end, what they have written is a fine applied statistics text. They state, "the most fundamental difference between classical statistical applications and data mining is the size of the data set." Perhaps not surprisingly then, their "principles of data mining" turn out to be much the same as the "principles of applied statistics." The sad fact is that statistics education, both graduate and undergraduate, has largely failed to respond to the challenges and opportunities presented by larger scale data sets. Breiman lamented the longstanding divorce of statistics from data (see Olshen [1]), and it is certainly true that statistics researchers continue to be more enamored of mathematics than of computing. This text redresses the balance somewhat.

Principles of Data Mining revolves around the notion of a "data mining algorithm." The authors decompose standard statistical/data mining activities in terms of the five basic components of a data mining algorithm:

1. the task the algorithm is used to address (e.g., classification, visualization, etc.);
2. the structure of the model or pattern being fitted to the data (e.g., linear regression, Gaussian mixture, etc.);
3. the score function used to assess the quality of the fit to the observed data (e.g., AIC, squared error, etc.);
4. the search or optimization method used to search over parameters and structure (e.g., gradient descent);
5. the data management technique used for storing, indexing, and retrieving of data.

This systematic view of data mining works well in my graduate-level data mining class and is a valuable contribution. Chapters 5–8 deal with these components in a fairly abstract way. Chapters 9–11 then deal more conventionally with descriptive and predictive models and cover a large range of topics, albeit in a superficial manner. Chapters 12–14 are somewhat detached from the rest of the text and deal with elementary database concepts, rule mining, and retrieval by content.

The extensive scope of the text requires that most topics receive curt treatment. For example, the 30-page chapter "Predictive Modeling for Regression" includes linear models (least squares, maximum likelihood, transformations, variable selection), generalized linear models, artificial neural networks, generalized additive models, and projection pursuit regression. Any statistics textbook targeting readers with undergraduate-level mathematics and statistics yet addressing topics such as projection pursuit regression faces a tough challenge. Competing texts deal with this in different ways. Han and Kamber [2] focused for the most part on detailed descriptions of specific algorithms for rule mining, prediction, and clustering. Hastie, Tibshirani, and Friedman [3] focused on a narrower range of topics (mostly chapters 10 and 11 of *Principles of Data Mining*) but provide much more in-depth treatment. Venables and Ripley [4] assumed their readers have greater prior knowledge and also focus on a particular software system for data analysis. *Principles of Data Mining* succeeds admirably with its approach, but most classes using the text will require supplementary materials such as those just mentioned as well as relevant research papers.

For readers with a database-centric view of data mining, *Principles of Data Mining* will probably disappoint. There is little or no coverage of topics such as data warehousing, data cubes, OLAP, data preprocessing, and data mining query languages and standards. For readers with a more statistics-centric view, but little by way of graduate-level statistics education, this text will serve as a terrific resource. In particular, *Principles of Data Mining* should find an important niche in computer science and electrical engineering graduate programs.

REFERENCES

[1] R. Olshen, *A conversation with Leo Breiman*, Statist. Sci., 16 (2001), pp. 184–198.

[2] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan–Kaufmann, San Francisco, 2001.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York, 2001.

[4] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-Plus*, 3rd ed., Springer-Verlag, New York, 1999.

David Madigan
Rutgers University

Vortex Methods: Theory and Practice. *By Georges-Henri Cottet and Petros Koumoutsakos.* Cambridge University Press, Cambridge, UK, 2000. $59.95. xiii+313 pp., hardcover. ISBN 0-521-62186-0.

Gaining quantitative insight into the mechanisms by which the convection, diffusion, stretching, and generation of vorticity interact is a prerequisite for understanding a wide range of problems in fluid mechanics. This realization prompted the early, inviscid point vortex calculations by Rosenhead [1], and it motivated the first attempts to model viscous effects by means of random walk methods (Chorin [2]). These investigations impressively demonstrate the naturally appealing features of vortex methods, namely that they require computational elements only "where the action (i.e., vorticity) is," that they avoid the problem of having to specify boundary conditions at infinity, and that they do not suffer from numerical dissipation in the same way as many grid-based algorithms. During the 1980s and 1990s the theoretical underpinnings of vortex methods matured rapidly. Simultaneously, their range of applications increased substantially, mostly due to the development in the research groups at Ecole Polytechnique and Caltech of accurate ways to handle viscous flows and the corresponding boundary conditions (Raviart [3], Koumoutsakos and Leonard [4]). Furthermore, the emergence of fast multipole algorithms (Greengard and Rokhlin [5]) has resulted in considerably more efficient vortex algorithms that are able to simulate more complex flows and to employ much finer resolutions. Both authors of this timely book have been key contributors to this rapid evolution of vortex methods, and hence they are in a unique position to present the current state of the art of the field.

The book focuses on vortex methods for the incompressible, constant density Euler and Navier–Stokes equations, and it is structured accordingly. After a brief review of the underlying mathematical and physical principles, the conceptually simplest form of inviscid vortex methods in two dimensions is examined. This provides the opportunity to introduce the basic concepts of smoothing and cut-off, which are