Preface

Surveys provide a major impetus in astronomy for understanding old and discovering new phenomena. With surveys we test hypotheses, improve statistics and attempt to put our understanding on a more secure basis or to extend it to new regimes. Through surveys we may also discover new types of objects, uncover new structures or observe new phenomena. Yet a survey alone is insufficient: what we do with the data is decisive, and the techniques we use could make the difference between discovery and non-discovery (or false discovery). Analysing the large amounts of complex data delivered by current and future surveys is a major challenge.

The articles in this book present various techniques which have been designed to effectively and efficiently analyse survey data in order to answer astronomical questions. Two themes in particular stand out. First, in almost every survey we need to classify objects or to estimate astrophysical parameters. This is often a prerequisite for assembling samples or modelling a population. But how do we best use the data for classification? Do we want discrete classification or continuous parameter estimation? What parameters can we estimate and with what accuracy? How do we train our classifiers? The second theme is that of discovery. Not all objects can be classified into our pre-defined schemes, and "outliers" may mark new discoveries. Likewise, uncovering relationships between objects (e.g. via clustering or parameter correlations) is a route to learning about astrophysical phenomena. How do we maximise the chances of discovery (without an unmanageable level of false positives)? How do we best find structure in multidimensional data? How can we distinguish significant features from background noise or contamination?

The contributions in this book were presented at the workshop *Classification and Discovery in Large Astronomical Surveys* held at the conference centre of the Max Planck Society in Ringberg Castle from 14-17 October 2008. It brought together 58 astronomers and computer scientists from 12 different countries and many different subfields to discuss methods for analysing and interpreting large, complex, astronomical data sets. Many articles present machine learning and statistical data analysis techniques and their application to problems ranging from stellar and galaxy classification, through time series analysis to clustering on large scales. This book reflects the outcome of our meeting, but by bringing together work from different fields within and beyond astronomy, I hope that it will also be a useful resource for other researchers.

I am grateful to my colleagues on the Scientific and Local Organising Committees for their help in organising what was a very interesting and inspiring meeting. On behalf of them and the other participants, I would like to thank the Deutsche Forschungsgemeinschaft (DFG), the Max Planck Society, the Schloessmann Stiftung and the Max Planck Institute for Astronomy for financial support for the workshop.

Coryn A.L. Bailer-Jones
Max Planck Institute for Astronomy, Heidelberg
October 2008