

Springer Geography

The Springer Geography series seeks to publish a broad portfolio of scientific books, aiming at researchers, students, and everyone interested in geographical research. The series includes peer-reviewed monographs, edited volumes, textbooks, and conference proceedings. It covers the major topics in geography and geographical sciences including, but not limited to; Economic Geography, Landscape and Urban Planning, Urban Geography, Physical Geography and Environmental Geography.

Springer Geography—now indexed in Scopus.

More information about this series at <http://www.springer.com/series/10180>

Courage Kamusoko

Remote Sensing Image Classification in R

 Springer

Courage Kamusoko
Asia Air Survey Co., Ltd.
Kawasaki, Kanagawa, Japan

ISSN 2194-315X

ISSN 2194-3168 (electronic)

Springer Geography

ISBN 978-981-13-8011-2

ISBN 978-981-13-8012-9 (eBook)

<https://doi.org/10.1007/978-981-13-8012-9>

© Springer Nature Singapore Pte Ltd. 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*To my girls Ai and Reina, who provide
unending inspiration!!*

Preface

This workbook is an introduction to remote sensing image processing and classification in R. The decision to write this workbook is based on my experience during my graduate and post-graduate studies, university teaching, and work as a consultant in remote sensing and modeling land use/cover changes. The aim of the workbook is to assist people who cannot afford commercial software and those who want to implement machine learning classifiers for remote sensing image processing and classification in R. There are many excellent introductory and advanced books on R, and remote sensing. However, there are few books that guide students, researchers, university teachers or other remote sensing practitioners for practical implementation of remote sensing in R. Therefore, this workbook is designed to be concise and practical in nature not as a complete guide to image processing and classification in R. That is, it is more of a desktop reference workbook, which introduces R so that one can immediately start using the software platform and R packages for image processing and classification. Although R has an initial steep learning curve, it is worth investing in R because it is free and has many packages, which might not be available in commercial software.

I also want to make it clear that while R is a good software or platform for graphics, statistical analysis, and machine learning given the number of packages available, it might not be appropriate for the analysis of massive remotely-sensed data. This is because the data being analyzed in R is held in memory. However, there are many innovations in big data analytics such as parallel computing technologies in R or even cloud computing. Alternatively, other programming or

scripting languages such as Python, Perl, and Ruby can be used depending on the nature of the problem and data availability.

Who should use this workbook?

The workbook is for undergraduate and graduate students in remote sensing and geographic information science or other related disciplines. While I assume no prior knowledge of R, the basic understanding of remote sensing is required. The workbook is also aimed at university teaching staff, researchers, or anyone interested in remote sensing image processing and classification. In addition, consultants or other people who are familiar with remote sensing but have limited experience in R can use this book to quickly test machine learning classifiers on small data sets.

How is this workbook organized?

This workbook is organized into five chapters. Chapter 1 introduces remote sensing digital image processing in R, which is subdivided into six sections. Section 1.1 presents a brief background on remote sensing image processing and classification, while Sect. 1.2 provides a brief overview of R and RStudio. Section 1.3 describes the data and test site, while Sect. 1.4 provides a quick guide to R. Finally, Sects. 1.5 and 1.6 provide the summary and additional exercises.

Chapter 2 covers pre-processing. Section 2.1 provides a brief background on pre-processing. Next, Sects. 2.2 and 2.3 provide tutorial exercises 1 and 2, which focus on displaying Landsat 5 Thematic Mapper (TM) imagery, and radiometric correction and reprojection. Finally, Sects. 2.4 and 2.5 provide the summary and additional exercises.

Chapter 3 focuses on image transformation. Section 3.1 provides a brief background on image transformation. Next, Sects. 3.2 and 3.3 focus on computing vegetation and texture indices. The final Sects. 3.4 and 3.5 provide the summary and additional exercises, respectively.

Chapter 4 covers image classification. Sect. 4.1 provides an overview of remote sensing image classification. Next, Sects. 4.2 and 4.3

focus on single date and multirate image classification, respectively. Finally, Sects. 4.4 and 4.5 provide the summary and additional exercises, respectively.

Last but not least, Chap. 5 focuses on improving image classification. Section 5.1 presents an overview of improving image classification, while Sect. 5.2 provides a brief background on feature selection. Next Sect. 5.3 focuses on image classification using multiple predictor variables, while Sect. 5.4 focuses on image classification using multiple predictor variables and feature selection. Finally, Sects. 5.5 and 5.6 provide the summary and additional exercises. An attempt has been made to organize the workbook in a general sequence of topics. Therefore, I encourage you to read the workbook in sequence from Chap. 1.

Conventions used in this workbook

The R commands or scripts are written in Lucida Console font size 10 in italics, while the output from the R is written in Lucida Console font size 10. Note that long output from R code is omitted from the workbook to save space. In some cases, I use small font sizes in Lucida Console to show how the R output or results would appear. This is just for illustration purposes. Readers will of course see the whole R output when they execute the commands. The hash sign (#) at the start of a line of code indicates that it is a comment. Finally, all explanations are written in New Times Roman font size 12.

Data sets, R scripts, and online resources

Information about the data sets and some sample scripts used in this workbook are available. Furthermore, I provide additional online resources on R, R packages or remote sensing in the appendix.

Kawasaki, Japan

Courage Kamusoko

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Professor Yuji Murayama, who has been inviting me to deliver a “Special Lecture on Geoenvironmental Sciences”, with a focus on “Practical Machine Learning for Geospatial Processing and Modeling” at the Graduate School of Life and Environmental Sciences, University of Tsukuba in Japan. This workbook would not have been possible without his support. Second, I would like to thank students from the Graduate School of Life and Environmental Sciences in general, and students from the Spatial Information Science Laboratory at the University of Tsukuba in particular who have taken this lecture for years. I am truly indebted to the students because I learnt a lot about machine learning and R during the lectures.

I am also indebted to my colleague, Ms. Emiko Ariyasu from Asia Air Survey Co., Ltd. who read the first draft of the workbook. Her comments provided invaluable insights during the revision. I am also truly grateful to the Global Business Development Division management at Asia Air Survey Co., Ltd. who gave me permission to teach part-time at the University of Tsukuba. I extend my special thanks to emeritus professor Masamu Aniya, my academic advisor who provided critical support during my graduate and post-graduate days at the University of Tsukuba. At Springer (Japan), I would like to thank Yosuka Nishida, Asami Komada, and Arun Kumar, and anonymous reviewers.

Last but not least, this workbook would not have been possible without the continued support of my wife Yuko, and my girls Ai and Reina.

Contents

1 Remote Sensing Digital Image Processing in R	1
1.1 Introduction	1
1.1.1 Remote Sensing Digital Image Processing	1
1.1.2 Overview of Machine Learning	3
1.2 Overview of R	6
1.2.1 What Is R?	6
1.2.2 Installing R	6
1.2.3 RStudio	8
1.2.4 R Packages	9
1.2.5 Overview of Data Type, Structure, and Functions	10
1.2.6 Getting Help	11
1.2.7 Handling Errors and Warnings	12
1.2.8 Other Issues	13
1.3 Data and Test Site	13
1.3.1 Landsat Imagery	13
1.3.2 Reference Data Sets	14
1.3.3 Overview of Harare Metropolitan Province	15
1.4 Tutorial 1: A Quick Guide to R	16
1.4.1 Objectives	16
1.4.2 Overview of the Packages	16
1.4.3 R Basics	17
1.4.4 Procedure	17
1.5 Summary	20

- 1.6 Additional Exercises 21
- References 21
- 2 Pre-processing** 25
 - 2.1 Background 25
 - 2.1.1 Radiometric Correction 26
 - 2.2 Tutorial 1: Display Landsat 5 TM Imagery 28
 - 2.2.1 Objectives 28
 - 2.2.2 Overview of Packages 28
 - 2.2.3 Procedure 29
 - 2.3 Tutorial 2: Radiometric Correction and Reprojection 42
 - 2.3.1 Objectives 42
 - 2.3.2 Procedure 42
 - 2.4 Summary 65
 - 2.5 Additional Exercises 65
 - References 65
- 3 Image Transformation** 67
 - 3.1 Background 67
 - 3.1.1 Spectral Indices 68
 - 3.1.2 Texture Indices 68
 - 3.2 Tutorial 1: Vegetation Indices 69
 - 3.2.1 Objective 69
 - 3.2.2 Overview of the Packages 70
 - 3.2.3 Procedure 70
 - 3.3 Tutorial 2: Texture Analysis 72
 - 3.3.1 Objective 72
 - 3.3.2 Overview of the Packages 73
 - 3.3.3 Procedure 73
 - 3.4 Summary 78
 - 3.5 Additional Exercises 78
 - References 79
- 4 Image Classification** 81
 - 4.1 Overview of Image Classification 81
 - 4.1.1 k-Nearest Neighbors (KNN) 82
 - 4.1.2 Artificial Neural Networks (ANN) 83

- 4.1.3 Single Decision Trees (DT) 83
- 4.1.4 Support Vector Machines (SVM) 84
- 4.1.5 Random Forest (RF) 85
- 4.2 Tutorial 1: Single Date Image Classification 85
 - 4.2.1 Objectives 85
 - 4.2.2 Overview of Packages 86
 - 4.2.3 Procedure 87
 - 4.2.4 Summary for Tutorial 1 133
- 4.3 Tutorial 2: Multidate Landsat Image Classification 134
 - 4.3.1 Objective 134
 - 4.3.2 Procedure 135
 - 4.3.3 Summary for Tutorial 2 150
- 4.4 Summary 150
- 4.5 Additional Exercises 151
- References 151
- 5 Improving Image Classification** 155
 - 5.1 Overview 156
 - 5.2 Feature Selection 156
 - 5.2.1 Importance of Feature Selection 156
 - 5.2.2 Recursive Feature Elimination (RFE) 157
 - 5.3 Tutorial 1: Image Classification Using Multiple Data Sets 157
 - 5.3.1 Objective 157
 - 5.3.2 Procedure 157
 - 5.3.3 Summary for Tutorial 1 170
 - 5.4 Tutorial 2: Image Classification Using Multiple Data Sets with Feature Selection 170
 - 5.4.1 Objective 170
 - 5.4.2 Procedure 170
 - 5.4.3 Summary for Tutorial 2 180
 - 5.5 Summary 180
 - 5.6 Additional Exercises 181
 - References 181
- Appendix** 183
- Index** 185

Abbreviations and Acronyms

ANN	Artificial Neural Network
ART	Adaptive Resonance Theory
ASM	Angular Second Moment
COST	Cosine Estimation of Atmospheric Transmittance
CRAN	Comprehensive R Archive Network
DEM	Digital Elevation Model
DOS	Dark Object Subtraction
DN	Digital Number
DSG	Department of the Surveyor-General
DT	Decision Trees
EMR	Electromagnetic Radiation
FOSS	Free and Open Source Software
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information Systems
GLCM	Grey Level Co-occurrence Matrix
GUI	Graphical User Interface
IDE	Integrated Development Environment
KNN	k-Nearest Neighbors
MIR	Mid Infrared
MODTRAN	Moderate Resolution Atmospheric Radiance and Transmittance
ML	Machine Learning
MPL	Multi-Layer Perceptron
MSAVI	Modified Soil Adjusted Vegetation Index
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared

RF	Random Forests
RFE	Recursive Feature Elimination
SAVI	Soil Adjusted Vegetation Index
SDOS	Simple Dark Object Subtraction
SOM	Self-Organizing Feature Maps
SR	Spatial Resolution
SVM	Support Vector Machines
TIR	Thermal Infrared
TM	Thematic Mapper
UAV	Unmanned Aerial Vehicles
USGS	United States Geological Survey