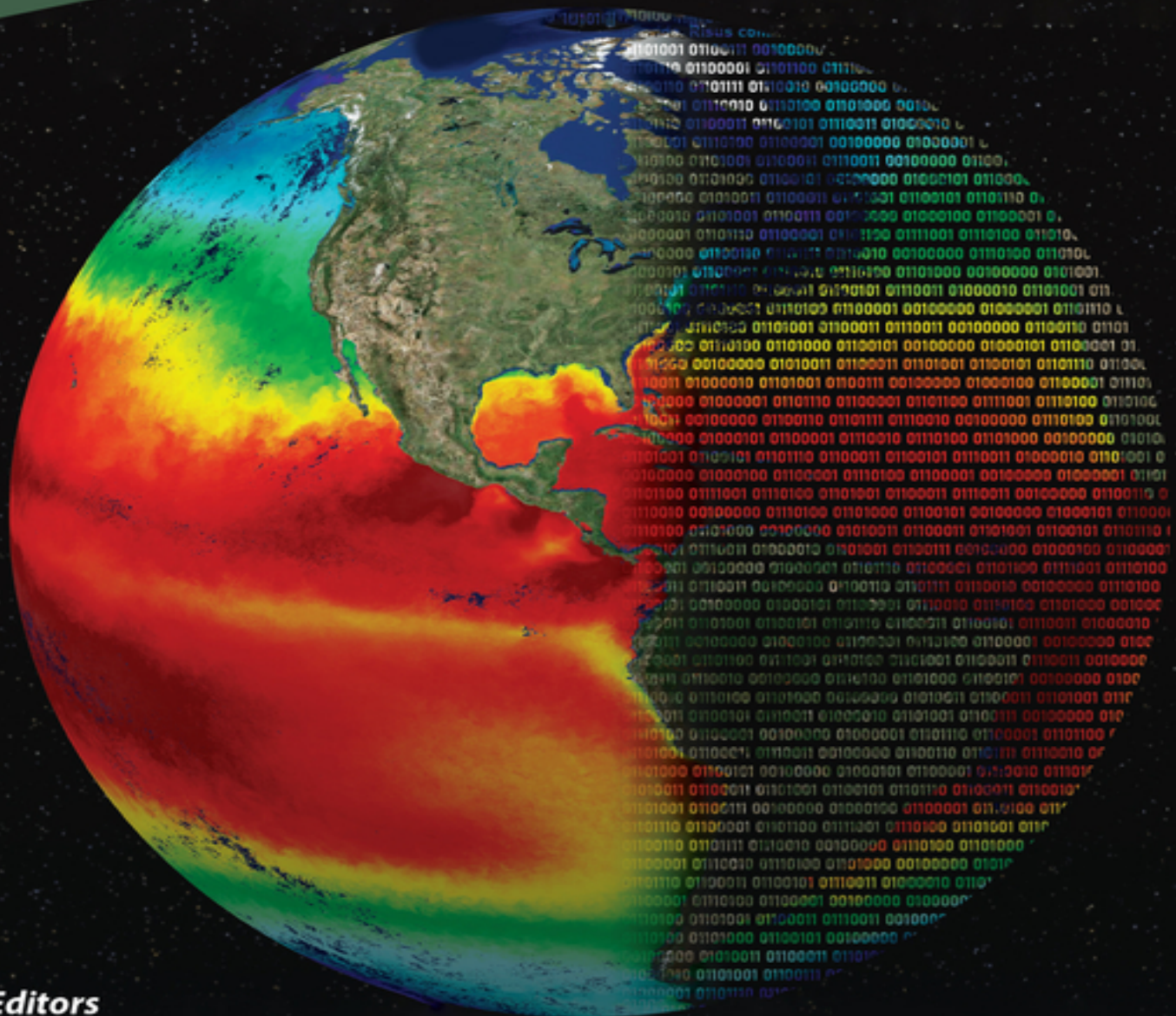


SPECIAL PUBLICATIONS SERIES

AGU
ADVANCING EARTH
AND SPACE SCIENCE

Big Data Analytics in Earth, Atmospheric, and Ocean Sciences



Editors
Thomas Huang
Tiffany C. Vance
Christopher Lynnes

WILEY

Table of Contents

[Cover](#)

[Title Page](#)

[Copyright](#)

[List of Contributors](#)

[Preface](#)

[1 AN INTRODUCTION TO BIG DATA ANALYTICS](#)

[1.1 Overview](#)

[1.2 Definitions](#)

[1.3 Example Problems](#)

[1.4 Big Data Analysis Concepts](#)

[1.5 Technology and Tools](#)

[1.6 Challenges](#)

[1.7 Summary](#)

[References](#)

[Part I: Big Data Analytics Architecture](#)

[2 INTRODUCTION TO BIG DATA ANALYTICS
ARCHITECTURE](#)

[References](#)

[3 Scaling Big Earth Science Data Systems Via
Cloud Computing](#)

[3.1 Introduction](#)

[3.2 Key Concepts of Science Data Systems
\(SDSes\)](#)

[3.3 Increasing Data Processing, Volumes, and
Rates](#)

[3.4 Cloud Concepts for SDSes](#)

[3.5 Architecture Components of Cloud-Based SDS](#)

[3.6 Considerations for Multi-cloud and Hybrid SDS](#)

[3.7 Cloud Economics](#)

[3.8 Large-Scaling Considerations](#)

[3.9 Example of Cloud SDSes](#)

[3.10 Conclusion](#)

[References](#)

[4 NOAA Open Data Dissemination \(Formerly NOAA Big Data Project/Program\)](#)

[4.1 Obstacles to the Public's Use of NOAA Environmental Data](#)

[4.2 Public Access of NOAA Data Creates Challenges for the Agency](#)

[4.3 The Vision for NOAA's "Oddball" Approach to Big Data](#)

[4.4 A NOAA Cooperative Institute Data Broker provides Research and Operational Agility](#)

[4.5 Public-Private Partnerships Provide the Pipeline](#)

[4.6 BDP Exceeds Expectations and Evolves into Enterprise Operations](#)

[4.7 Engaging Users in the Cloud](#)

[4.8 Challenges and Opportunities](#)

[4.9 Vision for the Future](#)

[Acknowledgments](#)

[References](#)

[5 A Data Cube Architecture for Cloud-Based Earth Observation Analytics](#)

[5.1 Introduction](#)

[5.2 Open Data Cube for the Cloud Design](#)

[5.3 S3 Array I/O Performance](#)

[5.4 Discussion and Conclusion](#)

[References](#)

[6 Open Source Exploratory Analysis of Big Earth Data With NEXUS](#)

[6.1 Introduction](#)

[6.2 Architecture](#)

[6.3 Deployment Architecture](#)

[6.4 Benchmarking and Studies](#)

[6.5 Analytics Collaborative Framework](#)

[6.6 Federated Analytics Collaborative Systems](#)

[6.7 Conclusion](#)

[References](#)

[7 Benchmark Comparison of Cloud Analytics Methods Applied to Earth Observations](#)

[7.1 Introduction](#)

[7.2 Experimental Setup](#)

[7.3 AODS Candidates](#)

[7.4 Experimental Results](#)

[7.5 Conclusions](#)

[References](#)

[Part II: Analysis Methods for Big Earth Data](#)

[8 Introduction to Analysis Methods for Big Earth Data](#)

[References](#)

[9 Spatial Statistics for Big Data Analytics in the Ocean and Atmosphere: Perspectives, Challenges,](#)

and Opportunities

9.1 Spatial Data and Spatial Statistics

9.2 What Constitutes Big Spatial Data?

9.3 Statistical Implications of the Four Vs of Big Spatial Data

9.4 Challenges to the Statistical Analysis of Big Spatial Data

9.5 Opportunities in Spatial Analysis of Big Data

9.6 Conclusion

References

10 Giving Scientists Back Their Flow: Analyzing Big Geoscience Data Sets in the Cloud

10.1 Introduction

10.2 Where's the Opportunity?

10.3 The Future

Reference

11 The Distributed Oceanographic Match-Up Service

11.1 Introduction

11.2 DOMS Capabilities

11.3 System Architecture

11.4 Workflow

11.5 Future Development

11.6 Acknowledgments

Availability Statement

References

Part III: Big Earth Data Applications

12 Introduction to Big Earth Data Applications

References

13 Topological Methods for Pattern Detection in Climate Data

13.1 Introduction

13.2 Topological Methods for Pattern Detection

13.3 Case Study: Atmospheric Rivers Detection

13.4 Conclusions and Recommendations

Acknowledgments

References

14 Exploring Large Scale Data Analysis and Visualization for Atmospheric Radiation Measurement Data Discovery Using NoSQL Technologies

14.1 Introduction

14.2 Software and Workflow

14.3 Hardware Architecture

14.4 Applications

14.5 Conclusions

Acknowledgments

References

15 Demonstrating Condensed Massive Satellite Data Sets for Rapid Data Exploration: The MODIS Land Surface Temperatures of Antarctica

15.1 Introduction

15.2 Data

15.3 Methods

15.4 Results

15.5 Conclusions

Acknowledgments

Availability Statement

References

16 Developing Big Data Infrastructure for Analyzing AIS Vessel Tracking Data on a Global Scale

16.1 Introduction

16.2 Background

16.3 Use Case: Producing Heat Maps of Vessel Traffic using AIS Data

16.4 Data Processing Overview

16.5 Future Work

16.6 Conclusions

References

17 Future of Big Earth Data Analytics

17.1 Introduction

17.2 How Data Get Bigger

17.3 The Evolution of Analytics Algorithms

17.4 Analytics Architectures

17.5 Conclusions

References

Index

End User License Agreement

List of Tables

Chapter 1

Table 1.1 Terms for understanding general concepts

Chapter 5

[Table 5.1 Different operating characteristics between HPC and cloud deploym...](#)

[Table 5.2 EC2 instance types](#)

[Table 5.3 Storage types](#)

Chapter 7

[Table 7.1 Hardware systems supporting the AODS architectures](#)

Chapter 9

[Table 9.1 Examples of spatial algorithms and their associated time complexi...](#)

Chapter 13

[Table 13.1 List of data sources used in the experiments](#)

[Table 13.2 Classification accuracy score of the SVM classifier for 3 hourly...](#)

[Table 13.3 Precision and sensitivity scores calculated for all data sets li...](#)

List of Illustrations

Chapter 1

[Figure 1.1 Leveraging feature binning technology to see geographic trends be...](#)

[Figure 1.2 Ridesharing pick up locations in midtown Manhattan. In the southe...](#)

[Figure 1.3 Tornado hotspots \(+\) and reported start points across the United ...](#)

[Figure 1.4 Spatiotemporal clustering \(DBSCAN...](#)

[Figure 1.5 Cloud service models \(IaaS, PaaS, and SaaS\) \(Chou, 2018\).](#)

Chapter 3

[Figure 3.1 Fundamental systems within the end-to-end Earth Observing Systems...](#)

[Figure 3.2 Key concepts of a typical SDS of data ingest from GDS, processing...](#)

[Figure 3.3 On-premise SDS components.](#)

[Figure 3.4 NISAR processing per month of mission \(forward processing + bulk ...](#)

[Figure 3.5 The relevant service layers of an SDS, their key area of focus \(l...](#)

[Figure 3.6 A high-level SDS functional architecture.](#)

[Figure 3.7 The trade-off spectrum of on-premise to all-in cloud-native, and ...](#)

[Figure 3.8 Software layer stack of the hybrid SDS.](#)

[Figure 3.9 Cloud storage tiers of AWS from hot \(fast and expensive\) to cold ...](#)

[Figure 3.10 SMAP SDS architecture.](#)

[Figure 3.11 NISAR GDS/SDS/DAAC data processing workflow.](#)

Chapter 4

[Figure 4.1 The actual total archive volume \(active and secure copies\) of dat...](#)

[Figure 4.2 BDP's data distribution scheme is led by a one-way transfer of a ...](#)

Chapter 5

[Figure 5.1 Open data cube notional architecture.](#)

[Figure 5.2 Execution flow for a data request on S3.](#)

[Figure 5.3 The ODC Cloud execution engine.](#)

[Figure 5.4 Ingest timings \(time in log scale\).](#)

[Figure 5.5 Time taken for data retrieval \(time in log scale\).](#)

Chapter 6

[Figure 6.2 Using an RDD framework to generate 30-year time series.](#)

[Figure 6.3 NEXUS data tiling architecture. Source: Nexus.](#)

[Figure 6.4 Apache Extensible Data Gateway Environment \(EDGE\) architecture.](#)

[Figure 6.5 NEXUS system architecture.](#)

[Figure 6.6 NEXUS Serverless data ingestion and processing workflow subsystem...](#)

[Figure 6.7 NEXUS Automated deployment on AWS.](#)

[Figure 6.8 NEXUS performance compared with GIOVANNI and AWS EMR.](#)

[Figure 6.9 On-the-fly multivariate analysis of Hurricane Katrina](#)

[Figure 6.10 The Apache Science Data Analytics Platform \(SDAP\), an Analytics ...](#)

[Figure 6.11 Federated Analytics Centers architecture. Source: Nexus.](#)

Chapter 7

[Figure 7.1 Area-averaged time series of MODIS/Terra aerosol optical depth fr...](#)

[Figure 7.1 Elapsed time for computations of area-averaged time series of MOD...](#)

[Figure 7.3 External spatiotemporal index of data chunks across the HDFS used...](#)

[Figure 7.4 Depiction of overall ClimateSpark architecture, which leverages t...](#)

[Figure 7.5 High-level architecture of NEXUS. Data ingest is handled by an Ex...](#)

[Figure 7.6 Elapsed time results for computation of area-averaged time series...](#)

Chapter 8

[Figure 8.1 Model-based analysis.](#)

Chapter 9

[Figure 9.1 Number of active Argo floats \(1997-2021\).](#)

[Figure 9.2 Time complexity of algorithms.](#)

[Figure 9.3 Storage requirements for high dimensional data.](#)

[Figure 9.4 Three-dimensional visualization of Ecological Marine Units \(EMU\)....](#)

[Figure 9.5 Mann-Kendall statistic for trends in maximum daily temperature \(1...](#)

Chapter 10

[Figure 10.1 The typical workflow of a scientist. They highlight a research q...](#)

[Figure 10.2 The progression of compute capabilities. Cloud computing is set ...](#)

Chapter 11

[Figure 11.1 DOMS supports distributed access to subsets of *in situ* data host...](#)

[Figure 11.2 An example of satellite to *in situ* data collocation in three dim...](#)

[Figure 11.3 Overall architecture of first DOMS prototype. The primary servic...](#)

[Figure 11.4 NEXUS data tiling architecture.](#)
[Source: Nexus.](#)

[Figure 11.5 Giovanni versus NEXUS performance \(time in seconds\) on area aver...](#)

[Figure 11.6 DOMS's workflow in the prototype GUI guides the user from select...](#)

[Figure 11.7 Sample analytical graphics showing nearest matched pairs between...](#)

Chapter 13

[Figure 13.1 Sample images of two weather patterns having distinguishable str...](#)

[Figure 13.2 The block diagram illustrating the extreme weather pattern detec...](#)

[Figure 13.3 A toy example of three connected components \(\$C_0\$, \$C_1\$, \$C_2\$ \) in the ...](#)

[Figure 13.4 A toy example of evolution plot describing the changes of the co...](#)

[Figure 13.5 An illustration of two-class data that are separable in a high-d...](#)

[Figure 13.6 Sample images illustrating AR detection problem. The upper row s...](#)

[Figure 13.7 Examples of normalized evolution plots of averaged \(red curves\)...](#)

Chapter 14

[Figure 14.1 Cassandra cluster ring with replication factor \(RF\) of 3.](#)

[Figure 14.2 Apache Spark cluster and workflow.](#)

[Figure 14.3 Overview of the big data software architecture workflow.](#)

[Figure 14.4 Screenshot of LASSO Bundle Browser found at <https://www.adc.arm...>](#)

[Figure 14.5 Screenshot of the web application created to simultaneously plot...](#)

[Figure 14.6 Interactive time series plot of different variables from ARMBE d...](#)

[Figure 14.7 Interactive parallel coordinates plot of different variables fro...](#)

[Figure 14.8 Apache Spark application result of ARMBE conditional querying fo...](#)

Chapter 15

[Figure 15.1 A flowchart of the data set condensation process.](#)

[Figure 15.2 The 5th_\(a\) and 95th_\(b\) percentiles for Antarctic land surface ...](#)

[Figure 15.3 Example statistical baseline images of Antarctic land surface te...](#)

[Figure 15.4 An example anomaly database schema. The Anomalies table stores a...](#)

[Figure 15.5 Baseline sample populations at each grid cell in the Antarctic l...](#)

[Figure 15.6 The most extreme cold temperatures in Antarctica occur along the...](#)

[Figure 15.7 A profile of minimum temperature \(solid gray line\) and elevation...](#)

[Figure 15.8 A systematic data error in the MOD11A1 data set. The image shows...](#)

Chapter 16

[Figure 16.1 Technologies used to process AIS data.](#)

[Figure 16.2 AIS vessel traffic data processing pipeline.](#)

[Figure 16.3 Voyages are created by connecting straight lines between pings: ...](#)

[Figure 16.4 Algorithm for generating vessel voyages. The exact numbers used,...](#)

[Figure 16.5 These example images show filtered vessel traffic. Image \(a\) sho...](#)

[Figure 16.6 Heat maps are generated by overlaying a grid and then counting t...](#)

[Figure 16.7 \(a\) Total vessel traffic and \(b\) unique vessel traffic over one ...](#)

[Figure 16.8 Sample results for processing U.S. Coast Guard terrestrial data ...](#)

[Figure 16.9 Sample results for processing Marine Exchange of Alaska data for...](#)

Chapter 17

[Figure 17.1 Growth of Cumulative EOSDIS Archives since 2000, shown on a semi...](#)

Special Publications 77

BIG DATA ANALYTICS IN EARTH, ATMOSPHERIC, AND OCEAN SCIENCES

**Thomas Huang
Tiffany C. Vance
Christopher Lynnes**

Editors

This Work is a co-publication of the American Geophysical Union and John Wiley and Sons, Inc.

This Work is a co-publication of the American Geophysical Union and John Wiley and Sons, Inc.

This edition first published 2023

© 2023 American Geophysical Union

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

Published under the aegis of the AGU Publications Committee

Matthew Giampoala, Vice President, Publications

Carol Frost, Chair, Publications Committee

For details about the American Geophysical Union visit us at www.agu.org.

The rights of Thomas Huang, Tiffany C. Vance, and Christopher Lynnes to be identified as the editors of this work has been asserted in accordance with law.

Registered Office

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office

111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between

when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Huang, Thomas (Technologist), editor. | Vance, C. Tiffany, editor.

| Lynnes, Christopher, editor.

Title: Big data analytics in earth, atmospheric, and ocean sciences /

Thomas Huang, Vance, C. Tiffany, Christopher Lynnes, editors.

Description: Hoboken, NJ : Wiley-American Geophysical Union, 2023. |

Includes bibliographical references and index.

Identifiers: LCCN 2022020168 (print) | LCCN 2022020169 (ebook) | ISBN

9781119467571 (cloth) | ISBN 9781119467564 (adobe pdf) | ISBN
9781119467533 (epub)

Subjects: LCSH: Earth sciences--Data processing. | Atmospheric

science--Data processing. | Marine sciences--Data processing. | Big
data.

Classification: LCC QE48.8 .B54 2022 (print) | LCC QE48.8 (ebook) | DDC

550.0285/57--dc23/eng20220722

LC record available at <https://lcn.loc.gov/2022020168>

LC ebook record available at <https://lcn.loc.gov/2022020169>

Cover Design: Wiley

Cover Images: Courtesy of Kate Culpepper with design elements provided by Esri, HERE, Garmin, FAO, NOAA, USGS, EPA | Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

List of Contributors

Edward M. Armstrong

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Alberto Arribas

Microsoft
Reading, UK; and
Department of Meteorology
University of Reading
Reading, UK

Jessica Austin

Axiom Data Science, LLC
Anchorage, Alaska, USA

Rob Bochenek

Axiom Data Science, LLC
Anchorage, Alaska, USA

Mark A. Bourassa

Center for Ocean-Atmospheric Prediction Studies, and
Department of Earth, Ocean, and Atmospheric Science
Florida State University
Tallahassee, Florida, USA

Jonathan Brannock

North Carolina Institute for Climate Studies

NOAA Cooperative Institute for Satellite Earth System Studies

North Carolina State University

Asheville, North Carolina, USA

Otis Brown

North Carolina Institute for Climate Studies

NOAA Cooperative Institute for Satellite Earth System Studies

North Carolina State University

Asheville, North Carolina, USA

Kevin A. Butler

Environmental Systems Research Institute

Redlands, California, USA

Nga T. Chung

NASA Jet Propulsion Laboratory

California Institute of Technology

Pasadena, California, USA

Thomas Cram

National Center for Atmospheric Research

Boulder, Colorado, USA

Jenny Dissen

North Carolina Institute for Climate Studies

NOAA Cooperative Institute for Satellite Earth System
Studies

North Carolina State University

Asheville, North Carolina, USA

Kyle Dumas

ARM Research Facility

Oak Ridge National Laboratory

Oak Ridge, Tennessee, USA

John-Marc Dunaway

Axiom Data Science, LLC

Anchorage, Alaska, USA

Jocelyn Elya

Center for Ocean-Atmospheric Prediction Studies

Florida State University

Tallahassee, Florida, USA

Eamon Ford

NASA Jet Propulsion Laboratory

California Institute of Technology

Pasadena, California, USA

David W. Gallaher

National Snow and Ice Data Center

Boulder, Colorado, USA

Kevin Michael Gill

NASA Jet Propulsion Laboratory

California Institute of Technology
Pasadena, California, USA

Glenn E. Grant

National Snow and Ice Data Center
Boulder, Colorado, USA

Frank R. Greguska III

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Mahabaleshwara Hegde

NASA Goddard Space Flight Center
Greenbelt, Maryland, USA

Alex Held

CSIRO Centre for Earth Observation
Canberra, ACT, Australia

Erik Hoel

Environmental Systems Research Institute
Redlands, California, USA

Benjamin Holt

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Hook Hua

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Thomas Huang

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Joseph C. Jacob

NASA Jet Propulsion Laboratory
Pasadena, California, USA

Zaihua Ji

National Center for Atmospheric Research
Boulder, Colorado, USA

Karthik Kashinath

Lawrence Berkeley National Laboratory
Berkeley, California, USA

Edward J. Kearns

First Street Foundation
Brooklyn, New York, USA

Bhargavi Krishna

ARM Research Facility
Oak Ridge National Laboratory
Oak Ridge, Tennessee, USA

Vitaliy Kurlin

Department of Computer Science
University of Liverpool
Liverpool, UK

Michael M. Little

NASA Goddard Space Flight Center
Greenbelt, Maryland, USA

Qin Lv

Department of Computer Science
University of Colorado
Boulder, Colorado, USA

Christopher Lynnes

NASA Goddard Space Flight Center (retd.)
Greenbelt, Maryland, USA

Gerald Manipon

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Theo McCaie

Met Office
Exeter, UK

Dmitriy Morozov

Lawrence Berkeley National Laboratory

Berkeley, California, USA

Grzegorz Muszynski

Lawrence Berkeley National Laboratory

Berkeley, California, USA; and

Department of Computer Science

University of Liverpool

Liverpool, UK

Matt Paget

CSIRO Centre for Earth Observation

Canberra, ACT, Australia

Tom Powell

Met Office

Exeter, UK

Giri Prakash

ARM Research Facility

Oak Ridge National Laboratory

Oak Ridge, Tennessee, USA

Prabhat Ram

Lawrence Berkeley National Laboratory

Berkeley, California, USA

Niall Robinson

Met Office

Exeter, UK; and

University of Exeter
Exeter, UK

Sujen Shah

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Adrienne Simonson

Office of the Chief Information Officer
National Oceanic and Atmospheric Administration
Asheville, North Carolina, USA

Shawn R. Smith

Center for Ocean-Atmospheric Prediction Studies
Florida State University
Tallahassee, Florida, USA

Kate Szura

Interactions LLC
Franklin, Massachusetts, USA

Ronnie Taib

CSIRO Data61
Sydney, NSW, Australia

Jacob Tomlinson

NVIDIA
Reading, UK

Vardis Tsontos

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Tiffany C. Vance

U.S. Integrated Ocean Observing System
National Oceanic and Atmospheric Administration
Silver Spring, Maryland, USA

Peter Wang

CSIRO Data61
Sydney, NSW, Australia

Michael Wehner

Lawrence Berkeley National Laboratory
Berkeley, California, USA

Brian D. Wilson

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Robert Woodcock

CSIRO Centre for Earth Observation
Canberra, ACT, Australia

Elizabeth Yam

NASA Jet Propulsion Laboratory

California Institute of Technology
Pasadena, California, USA

Chaowei Phil Yang

George Mason University
Fairfax, Virginia, USA

Alice Yepremyan

NASA Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

Hailiang Zhang

NASA Goddard Space Flight Center
Greenbelt, Maryland, USA

Preface

The seeds for this book were sown in sessions on Big Data Analytics, held at the 2016 Fall Meeting of the American Geophysical Union. At the time, Earth Science data were projected to rise by orders of magnitude in the coming decade, and the community was investigating a variety of emergent technologies and techniques to make the best use of the coming deluge. The chapters of this book are a representative, but by no means exhaustive, collection of those and similar investigations.

Big Earth Data Analytics can be defined as the application of increasingly sophisticated tools for data analysis and display to the rapidly increasing volume of Earth science data to obtain information, and eventually insight. This combines two concepts: Big Earth Data and Data Analytics. Big Earth Data refers both to the volume of data sets and the combination of data from a variety of sources, in a variety of formats, and from a variety of disciplines. To get a sense of the volume, NOAA generates tens of terabytes of data a day from satellites, radars, ships, weather models, and other sources. The National Aeronautics and Space Administration (NASA) Earth Observation archives were growing by more than 30 TB per day in 2020 with daily growth expected to increase to 130 TB/day by 2024 as new satellites launch; and the European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological data archive adds 200 terabytes of new data daily. However, the data are "big" not only in their volume but in their varied formats, disciplines, structures, and formats. As such, they are disruptors to traditional analysis methods, and to the kinds of questions that can be asked by researchers. Data analytics are increasingly driven by the availability of high-

volume and heterogeneous data sets. Data size and complexity affect all aspects of data management and usage, requiring new approaches and tools. Despite the challenges to acquire, use, and analyze Big Earth Data, they are already being utilized extensively in climate, oceanographic, and biology related works. Easily available data lead to the ability to analyze longer scale records and patterns over large spatial domains.

Analyses of these data borrow both from traditional scientific analyses and from tools developed for business applications. These types of data analytics are developed by university and other research teams. They are increasingly becoming an area of interest to cloud providers and analytics companies. From Google's Earth Engine for analyzing Earth science data at scale, to the National Oceanic and Atmospheric Administration's (NOAA's) Big Data Program, big data about the Earth and their analysis are increasingly common. Amazon's Elastic MapReduce and SageMaker are common building blocks for cloud-based analysis and Galileo (a.k.a. Service Workbench) is Amazon's latest Web application for interactive analysis. Microsoft Azure ML Studio is another popular cloud-based data analysis solution. Big Earth Data analyses increasingly rely on cloud-based storage and processing capabilities as the volume of the data and the computing resources needed go beyond local resources.

This book is organized into three parts. It starts with the big picture, covering Big Data Analytics Architecture. This part begins with a chapter addressing the geospatial aspect of Big Earth Data from a variety of perspectives. This is followed by a chapter discussing the data management challenges posed by data at scale, particularly in the context of making them available for analysis. This is complemented by a chapter discussing the challenges of scaling up the analysis itself. The following chapters cover

large-scale projects such as NASA's Earth Exchange, which enables large scale data analysis in a supercomputing environment and the NOAA Big Data Project, which makes data sets available to end users via several cloud providers. Part I also includes chapters on architectures and fully realized systems, such as Data Cube, NEXUS and the Apache Science Data Analytics Platform, and a NoSQL based platform for exploring and analyzing in situ data.

The second part of the book, Analysis Methods for Big Earth Data, addresses some specific techniques to derive information and/or insight from big data, emphasizing the unique aspects of Earth Observations. Part II begins with two chapters on the use of geospatial statistics for analysis, followed by a chapter melding machine learning with geophysical constraints, and finally a chapter benchmarking different analytical methods for spatiotemporal analysis.

The third part of the book, Big Earth Data Applications, describes a few specific applications of big analysis techniques and platforms: weather and climate model analysis, atmospheric river patterns, Antarctic land surface temperatures extremes, satellite in situ match-ups of oceanographic data, and vessel tracking. This is clearly a small sample of existing applications; rather, the sample shows how some very different analysis methods can find diverse applications in the Earth sciences.

While the application of big Earth data analytics covers a range of applications, a number of common themes in the chapters of this book include (1) the role of the cloud, especially with ever increasing data sizes; (2) limitations and costs of using the cloud, including the unpredictability of costs and the high cost of data egress from the cloud; (3) techniques to maintain data integrity during file transfers; (4) efficiencies via partial reads from Web object storage;