p. 88 Our Dynamic Future

p. 103 Technical Debt in Computational Science

p. 108 Astronomical Images before the Internet

Computing in SCIENCE & ENGINEERING

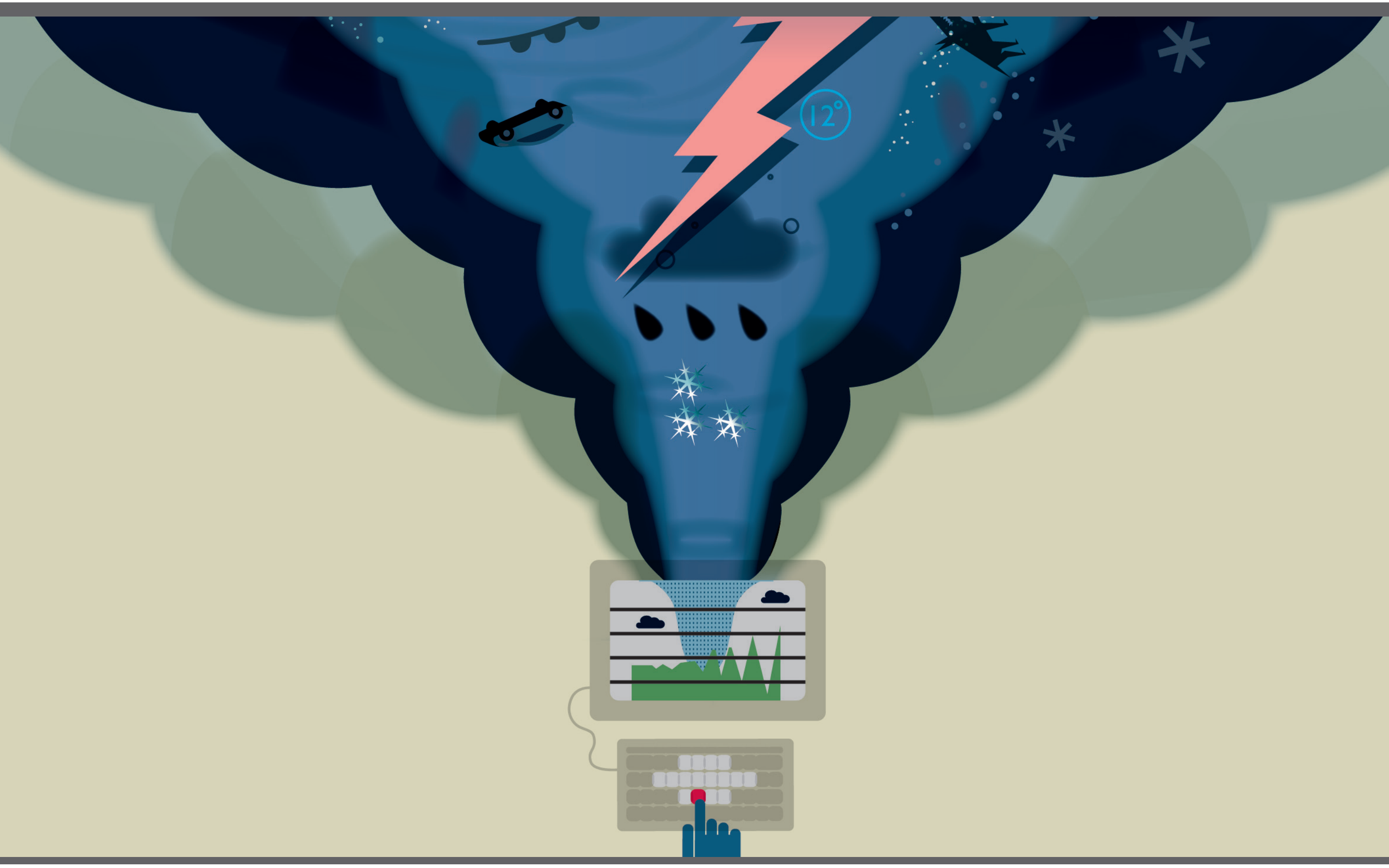
Vol. 17, No. 6 | November/December 2015

Computing & Climate

IEEE

AIP
cise.aip.org

IEEE computer society
www.computer.org/cise/



Computing and Climate

James H. Faghmous | Ichan School of Medicine at Mt. Sinai and Arnhold Global Health Institute
Vipin Kumar and Shashi Shekhar | University of Minnesota

Global climate change and its impact on natural resources, infrastructure, and population health are some of the greatest 21st century challenges. Although the scientific evidences of anthropogenic climate change are overwhelming,[1] greater uncertainties remain as to how ecosystems, infrastructure, and global water-food-energy security will respond to a dramatically altered climate.[2,3]

Evolving Fields

The fields of climate science and climate change impacts rely heavily on computationally-intensive simulations as well as troves of data from Earth observations. Thus, virtually every aspect of computing has a timely opportunity to help tackle this global challenge. The computer science community is beginning to engage climate science in notable ways, and there's increasing interest at the intersection of computing and climate. For example, the US National Center for Atmospheric Research (NCAR) has developed an open source Earth system model, with millions of lines of software code, known as the Community Earth System Model (CESM) to simulate various Earth system processes.[4]

Additionally, in 2010, the US National Science Foundation's largest grant for basic computer science research—the Expeditions in Computing program (www.nsf.gov/news/news_summ.jsp?cntn_id=117560)—funded a multi-institution grant led by the University of Minnesota to develop new data science methods

The fields of climate science and climate change impacts rely heavily on computationally-intensive simulations as well as troves of data from Earth observations. Thus, virtually every aspect of computing has a timely opportunity to help tackle this global challenge.

to better understand climate change and its impacts (http://climatechange.cs.umn.edu). We're also seeing nascent computer science communities forming around climate change such as the Climate Informatics workshop, now in its fifth installment (https://www2.cisl.ucar.edu/events/ci2015), and the Computational Sustainability research tracks in some of the largest computer science conferences such as the Association for the Advancement of Artificial Intelligence (AAAI; www.aaai.org/Conferences/AAAI/2016/aaai16computational.php) and the International Joint Conferences on Artificial Intelligence (IJCAI; http://ijcai-15.org/index.php/computational-sustainability-track).

In This Issue

Given the momentous challenge of climate change, this special issue showcases several innovations at the intersection of computing and climate as well as notable research opportunities. The articles we present here focus specifically on data-driven research, given that climate science is experiencing rapidly growing data and presenting unique challenges to traditional data science methods.[5]

The first article, "Climate Computing: The State of Play" by V. Balaji, provides a concise tutorial of global climate models to a broad computer science audience, with an emphasis on the major challenges facing these simulations.

The second article, "A Guide to Earth Science Data: Summary and Research Challenges" by Karpante and Liess, introduces the computer science community to the rich types of climate science data available for data-driven research. The article also provides data resources to help getting started with climate research. Finally, the authors highlight data-centric opportunities that future research could address.

In "Scalable Multivariate Time-Series Models for Climate Informatics," Yan Liu demonstrates how scalable multivariate time-series models can be used for a wide range of climate applications. The author demonstrates several data-driven models such as Granger graphical models and low-rank tensor learning on significant climate science questions with a goal of developing predictive analytics toolboxes for climate science.

The next three articles focus on machine learning applications to climate data. In "Identifying Physical Interactions from Climate Data: Challenges and Opportunities," Imme Ebert-Ubhoff and Yi Deng highlight a new research direction of tracking information flow in climate data by learning the structure of probabilistic graphical models directly from the data. The authors provide a general overview and discuss some challenges and opportunities for this emerging area. In "A Multitask Learning View on the Earth System Model Ensemble," André Ricardo Gonçalves, Fernando Von Zuben, and Arindam Banerjee present a machine learning application to analyzing Earth system model ensembles. The majority of climate change studies rely on mathematical simulations that are sensitive to parameters and initial conditions. To address such sensitivity, projections are based on an ensemble of realizations of (the same or different) Earth system models. However, there's no consensus on how to combine such models. The authors present a promising method based on multitask learning to combine the output of these ensembles. In "Can Topic Modeling Shed Light on Climate Extremes?," Cheng Tang and Claire Monteleoni demonstrate a potential use of a traditional machine learning method—probabilistic topic models—to automatically identify extreme climate events, such as droughts and floods, in data.

Finally, in "Climate Adaptation Informatics: Water Stress on Power Production," Auroop Ganguly and his colleagues highlight a case study of climate adaptation by looking at water stress on US thermoelectric power production. To properly manage energy production, we must better understand how stressors such as water availability will evolve in a changing climate. The authors highlight some of the major difficulties in projecting future water availability as well as ways to infer future water stress on US power production.

This special issue on computing and climate can be viewed as a call to action for the broader computer science community to get involved in this very exciting area of societal importance. Although the issue focuses primarily on data-driven opportunities for computer science contributions, we hope that future research will broaden this scope. For example, specialized computer architectures could be developed to improve the performance of complex climate simulations or novel software engineering practices could be introduced to handle the uncertain and stochastic nature of the earth system.

References

1. T F. Stocker, Q. Dahe, and G.-K. Plattner, "Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)," Cambridge Univ. Press, 2013; www.climatechange2013.org/report/full-report.
2. C. Ash et al., eds., "Natural Systems in Change Climates," *Science*, Aug. 2013; www.sciencemag.org/site/special/climate2013/index.xhtml.
3. C. Field et al., "Climate Change 2014: Impacts, Adaptation, and Vulnerability. Working Group II Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)," Cambridge Univ. Press, 2014; www.ipcc.ch/report/ar5/wg2.
4. J.W. Hurrell, M.M. Holland, and P.R. Gent, "The Community Earth System Model: A Framework for Collaborative Research," *Bulletin Am. Meteorological Soc.*, vol. 94, no. 9, 2013, pp. 1339–1360.
5. J.H. Faghmous and V. Kumar, "A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science," *Big Data*, vol. 2, no. 3, 2014, pp. 155–163.

James H. Faghmous is an assistant professor at the Icahn School of Medicine at Mount Sinai and CTO of the Arnhold Global Health Institute. His research interests lie at the intersection of big data and scientific discovery as well as spatiotemporal data mining. Faghmous earned a PhD in computer science from the University of Minnesota, where he received the Best Dissertation Award in physical sciences and engineering for his work at the intersection of data mining and climate change. Contact him at james.faghmous@mssm.edu.

Vipin Kumar is a regents professor at the University of Minnesota, where he holds the William Norris Endowed Chair in the Department of Computer Science and Engineering. He's also lead PI of a five-year, $10 million project ("Understanding Climate Change—A Data Driven Approach") funded by the NSF's Expeditions in Computing program aimed at pushing the boundaries of computer science research. Kumar's foundational research in data mining and its applications to scientific data was honored by the ACM SIGKDD 2012 Innovation Award. Contact him at kumar@cs.umn.edu.

Shashi Shekhar is a McKnight Distinguished University Professor at the University of Minnesota. He co-edited an Encyclopedia of GIS and co-authored a Spatial Databases textbook. His research interests include spatial data mining, spatial databases, and geographic information science. Shekhar received the IEEE CS Technical Achievement Award and the UCGIS Education Award and was elected an IEEE and AAAS Fellow. Contact him at shekhar@cs.umn.edu. ■

November/December 2015, Vol. 17, No. 6

Computing in SCIENCE & ENGINEERING

COMPUTING AND CLIMATE

6 **Guest Editors' Introduction**
Computing and Climate
James H. Faghmous, Vipin Kumar, and Shashi Shekhar

9 **Climate Computing: The State of Play**
V. Balaji
Climate models represent a large variety of both processes on different time and space scales and natural stochastic variability, with computation- and data-intensive simulations needed to extract signals of climate change. Scientific trends are driving toward higher resolution, greater complexity, and larger ensembles.

14 **A Guide to Earth Science Data: Summary and Research Challenges**
Anuj Karpatne and Stefan Liess
Recent growth in the scale and variety of Earth science data has provided unprecedented opportunities to big data analytics research for understanding the Earth's physical processes. But Earth science datasets exhibit some unique characteristics (such as adherence to physical properties and spatiotemporal constraints) that present challenges to traditional data-centric approaches.

19 **Scalable Multivariate Time-Series Models for Climate Informatics**
Yan Liu
Climate data not only have a massive scale but are also of high dimension and complex dependency structures, making the analysis task extremely challenging. Recent advances in machine learning can help researchers tackle a series of problems in climate data analysis, such as climate change attribution, spatiotemporal analysis, and extreme value time-series analysis.

27 **Identifying Physical Interactions from Climate Data: Challenges and Opportunities**
Imme Ebert-Uphoff and Yi Deng
Recent research has shown the potential of using probabilistic graphical models to identify and visualize interactions in the Earth's climate system. Studying the resulting pathways is of great interest to scientists because it helps them learn subtle details about the underlying dynamical mechanisms governing our planet's climate.

35 **A Multitask Learning View on the Earth System Model Ensemble**
André Ricardo Gonçalves, Fernando J. Von Zuben, and Arindam Banerjee
Earth system models (ESMs) are the primary mechanisms for obtaining projections under different climate change scenarios. Researchers use ensembles of climate models to gain better accuracy and reduce uncertainty. A multitask learning-based method can build ESM ensembles for all regions jointly to improve predictions for individual ones.

43 **Can Topic Modeling Shed Light on Climate Extremes?**
Cheng Tang and Claire Monteleoni
Understanding changes in climate extremes is an urgent challenge. Topic modeling techniques from natural language processing can help scientists learn climate patterns from data. Recent work extracts global climate patterns from multivariate climate data, modeling relations between variables via latent topics and discovering the probability of each climate topic appearing at different geographical locations.

For more information on these and other computing topics, please visit the IEEE Computer Society Digital Library at www.computer.org/csdl.

Cover illustration: Andrew Baker
www.debutart.com/illustration/andrew-baker

STATEMENT OF PURPOSE

Computing in Science & Engineering (*CiSE*) aims to support and promote the emerging discipline of computational science and engineering and to foster the use of computers and computational techniques in scientific research and education. Every issue contains broad-interest theme articles, departments, news reports, and editorial comment. Collateral materials such as source code are made available electronically over the Internet. The intended audience comprises physical scientists, engineers, mathematicians, and others who would benefit from computational methodologies. All articles and technical notes in *CiSE* are peer-reviewed.

53 Climate Adaptation Informatics: Water Stress on Power Production

Auroop R. Ganguly, Devashish Kumar, Poulomi Ganguli, Geoffrey Short, and James Klausner

Resilience to nonstationarity and deep uncertainty is a prerequisite for decadal to century scale water security. Adaptation is urgent at near-term decadal horizons, when projections of stressors and vulnerability are typically more reliable but climate internal variability may preclude actionable insights. A case study of at-risk power production suggests that informed decisions are still possible.

DISTRIBUTED SYSTEMS

61 The Evolution of Global Scale Filesystems for Scientific Software Distribution

Jakob Blomer, Predrag Buncic, René Meusel, Gerardo Ganis, Igor Sfiligoi, and Douglas Thain

Delivering complex software across a worldwide distributed system is a major challenge in high-throughput scientific computing. To address this problem in high-energy physics, a global scale filesystem delivers software to hundreds of thousands of machines around the world.

COLUMNS

DEPARTMENTS

RESOURCES

Editorial: Unless otherwise stated, bylined articles, as well as product and service descriptions, reflect the author's or firm's opinion. Inclusion in *Computing in Science & Engineering* does not necessarily constitute endorsement by IEEE, the IEEE Computer Society, or the AIP. All submissions are subject to editing for style, clarity, and length. IEEE prohibits discrimination, harassment, and bullying. For more information, visit www.ieee.org/web/aboutus/whatis/policies/p9-26.html. **Circulation:** *Computing in Science & Engineering* (ISSN 1521-9615) is published bimonthly by the AIP and the IEEE Computer Society. IEEE Headquarters, Three Park Ave., 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Cir., Los Alamitos, CA 90720, phone +1 714 821 8380; IEEE Computer Society Headquarters, 2001 L St., Ste. 700, Washington, D.C., 20036; AIP Circulation and Fulfillment Department, 1NO1, 2 Huntington Quadrangle, Melville, NY, 11747-4502. Subscribe to *Computing in Science & Engineering* by visiting www.computer.org/cise. **Reuse Rights and Reprint Permissions:** Educational or personal use of this material is permitted without fee, provided such use: 1) is not made for profit; 2) includes this notice and a full citation to the original work on the first page of the copy; and 3) does not imply IEEE endorsement of any third-party products or services. Authors and their companies are permitted to post the accepted version of IEEE-copyrighted material on their own web servers without permission, provided that the IEEE copyright notice and a full citation to the original work appear on the first screen of the posted copy. An accepted manuscript is a version that has been revised by the author to incorporate review suggestions, but not the published version with copy-editing, proofreading and formatting added by IEEE. For more information, please go to: http://www.ieee.org/publications_standards/publications/rights/paperversionpolicy.html. Permission to reprint/republish this material for commercial, advertising, or promotional purposes or for creating new collective works for resale or redistribution must be obtained from IEEE by writing to the IEEE Intellectual Property Rights Office, 445 Hoes Lane, Piscataway, NJ 08854-4141 or pubs-permissions@ieee.org. Copyright © 2015 IEEE. All rights reserved. **Abstracting and Library Use:** Abstracting is permitted with credit to the source. Libraries are permitted to photocopy for private use of patrons, provided the per-copy fee indicated in the code at the bottom of the first page is paid through the Copyright Clearance Center, 222 Rosewood Dr., Danvers, MA 01923. **Postmaster:** Send undelivered copies and address changes to *Computing in Science & Engineering*, 445 Hoes Ln., Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and at additional mailing offices. Canadian GST #125634188. Canada Post Corporation (Canadian distribution) publications mail agreement number 40013885. Return undeliverable Canadian addresses to PO Box 122, Niagara Falls, ON L2E 6S8 Canada. Printed in the USA.