44th Symposium on the Interface of Computing Science and Statistics 2013

(Interface 2013)

Orange, California, USA
4-6 April 2013

Editors:

Hesham El-Askary **Ed Wegman**

ISBN: 978-1-5108-0644-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Printed by Curran Associates, Inc. (2015)

For permission requests, please contact the Interface Foundation of North America
at the address below.

Interface Foundation of North America
PO Box 7460
Fairfax Station, VA 22039

Phone: (703) 993-1212
Fax: (703) 993-1700

interface@galaxy.gmu.edu

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Thursday April 4th 2013:
Sandhu Conference Center

7:30 am – 1:30 pm Registration
8:30 am – 10:00 am Technical Sessions

- **Astroinformatics: Learning from Data in the Astronomical Sciences 1**
  Chair: Kirk Borne, George Mason University

- **Data Science and Climate 1**
  Chair: Amy Braverman, JPL

10:00 am – 11:30 pm Technical Sessions

- **Learning from Data**
  Chair: James Gentle, George Mason University

- **From Large Earth Science Datasets to Compelling Scientific Results**
  Chair: Charles Ichoku, NASA GSFC

2. *Assessment of aerosol intercontinental transport with big data from EOS satellites....167*
Hongbin Yu, NASA GSFC
3. *Implications of Satellite Swath Width on Aerosol Optical Thickness Statistics....176*
Peter Colarco, NASA GSFC
4. *Improving our Understanding of Land-Use Change and Fire in Southeast Asia by Using Advanced Mathematical Techniques, Models, and Multiple Remotely Sensed Measurements in Tandem....187*
Jason Blake Cohen, National University of Singapore

11:30 pm – 1:15 pm Lunch Break and Poster Session
(Provided: Beckman Center 404)

1:15 pm – 1:45 pm **Official Opening, Introductions and Welcoming Notes**

- *Introductions* Hesham El-Askary, Interface 2013 Chair, Chapman University
- *Welcoming Remarks* James Doti, President Chapman University
- *Welcoming Remarks* Janeen Hill, Dean Schmid College of Science and Technology
- *Welcoming Remarks* Menas Kafatos, Director Center of Excellence in Earth Systems Modeling & Observations, on behalf of Chancellor, Chapman University

1:45 pm – 2:45 pm Panel Discussion by Experian: *The Future of Digital Wallets....202*
Moderator: Mark Kapczynski, VP, Strategy, Experian Consumer Services

1. *Panelist 1*: Dan Elvester, Director, Business Development, Experian Decision Analytics
2. *Panelist 2*: JP Voisinet, Strategy Associate, Experian Consumer Services
3. *Panelist 3*: Wences Casares, CEO, Lemon Wallet
4. *Panelist 4*: Andy Johnson, SVP Data, Experian Marketing Services
5. *Panelist 5*: Eric Haller, SVP, Data Lab, Experian

2:45 pm – 3:20 pm Keynote Speaker 1 "*Data Theory*": Edward Wegman, George Mason University
<u>Title:</u> *Big Data: Technology and Analysis....204*
3:20 pm – 3:30 pm Q&A
3:30 pm – 3:55 pm Afternoon Break
4:00 pm – 4:50 pm Keynote Speaker 2 "*Earth Systems Science Data*": Jack Kaye, NASA

	Title: *Earth System Science: Data Challenges Created by Observations and Models, and their use for Science and Applications....227*
4:50 pm – 5:00 pm	Q&A
5:00 pm – 5:50 pm	Keynote Speaker 3 "*Health Care Data Systems*": John Vigouroux, Frost Venture Partners **Title:** *Changing Role of Big Data in Healthcare....258*
5:50 pm – 6:00 pm	Q&A

(Beckman Center 404)

6:00 pm – 8:00 pm	Networking and Opening Reception

Friday April 5th 2013:

Sandhu Conference Center

(Data Theory and Earth Science Sessions)

Argyros Forum

(Health Care Sessions)

7:30 am – 8:00 am	Registration
8:00 am – 9:30 am	Technical Sessions

- **Astroinformatics: Learning from Data in the Astronomical Sciences 2**
  Chair: Kirk Borne, George Mason University

  1. *Things that go *Bang!* in the Night: Automated Classification of Transient Events in Sky Surveys....278*
     George Djorgovski, California Institute of Technology
  2. *Finding Rare Astronomical Objects Using Efficient Bayesian Networks....293*
     Ashish Mahabal, California Institute of Technology
  3. *Data Triage of Astronomical Transients and Variables: A Machine Learning Approach....308*
     Umaa Rebbapragada, Jet Propulsion Laboratory
  4. *Observations-driven PCA Models for seasonal Changes on Jupiter and Saturn....324*
     Padma Yanamandra-Fisher, Space Science Institute

- **Data Science and Climate 2**
  Chair: Amy Braverman, JPL
  1. *Statistical Downscaling of Two-Dimensional 10m Wind Fields....338*
     Mark Nakamura, University of California, Los Angeles
  2. *Understanding Climate Change - A Data Driven Approach....349*
     Vipin Kumar, University of Minnesota
  3. *A Software Architecture for the Systematic Analysis of Climate Science Data....376*
     Dan Crichton, Jet Propulsion Laboratory

- **Healthcare Process Reform**
  Chair: Arnold Goodman, Collaborative Data Solutions
  1. *Process Analysis of Healthcare to Better Manage Stakeholder Goals, Funding, Engineering, Delivery, Culture, Big Data and Accountability....N/A*
     Arnold Goodman, Collaborative Data Solutions
  2. *Benchmarking Healthcare Provider Performance: Some Statistical Considerations....N/A*
     Susan Paddock, RAND Corporation
  3. *Make More Accurate Genomic Conclusions from Sequence Findings by Reducing a Basic Uncertainty in Associating Genes with Diseases....N/A*
     Arnold Goodman, Collaborative Data Solutions

9:30 am – 11:00 am Technical Sessions

- **JCGS Highlights at the Interface Auditorium**
  Chair: Richard Levine, JCGS Editor, San Diego State University
  1. *Splitting Methods for Convex Clustering....391*
     Eric Chi, University of California, Los Angeles
  2. *Tapered Covariance: Bayesian Estimation and Asymptotics....412*
     Benjamin Shaby, University of California, Berkeley
  3. *A Discussion of Graphical Inference....N/A*
     Heike Hofmann, Iowa State University

- **Analysis of Big Data for Atmospheric Aerosol Research**
  Chair: Olga Kalashnikova, Jet Propulsion Laboratory
  Co-Chair: Michael Garay, Jet Propulsion Laboratory
  1. *Scientific Discovery and Anomaly Detection in Large Aerosol Data Sets....433*
     Kiri Wagstaff, Jet Propulsion Laboratory
  2. *Giovanni-4: the next generation of an online tool for satellite data visualization, exploration and inter-comparison....444*
     Christopher Lynnes, NASA/GSFC
  3. *Spatial Statistical Data Fusion for Remote Sensing Applications....460*
     Hai Nguyen, Jet Propulsion Laboratory

- **Healthcare Futures**
  Chair: Mark Kapczynski, VP, Strategy, Experian Consumer Services
  1. *Panelist 1*: Dan Johnson, President, Experian Healthcare
  2. *Panelist 2*: Anatoly Kvitnitsky, Strategy Associate, Experian Consumer Services
  3. *Panelist 3*: John Benson, CEO Verisys
  4. *Panelist 4*: Katie Vahle, CEO, CoPatient
  5. *Panelist 5:* Greg Jackson, Chief Data Officer, EverydayHealth

11:00 am – 11:30 pm Morning Break
11:30 pm – 1:00 pm Technical Sessions

- **Educating Data Scientists**
  Chair: Hadley Wickham, Rice University
  1. *Data Science: The Divide and Recombine (D&R) Illustration....473*
     William Cleveland, Purdue University
  2. *Teaching Statistics in the Data Deluge....482*
     Rob Gould, University of California, Los Angeles
  3. *Designing the Data Science Curriculum....N/A*
     Jeff Hammerbacher, Cloudera

- **Statistical Learning in Earth Systems Science**
  Convener: David A. Van Dyk, Imperial College London
  Chair: Richard Levine, JCGS Editor, San Diego State University
  1. *Co-clustering Spatial Data Using a Generalized Linear Mixed Model With Application to the Integrated Pest Management....497*
     Daniel Jeske, University of California, Riverside
  2. *Using state-space models for variance matrices to study climate patterns....518*
     Yaming Yu, University of California, Irvine
  3. *A Nonlinear Model for Predicting Plankton Dynamics....537*
     Barbara Bailey, San Diego State University

- **Application of Big Data and Analytics to the Healthcare Setting (Panel Discussion)**
  Chair: Janeen Hill, Chapman University
  1. *Panelist 1:* Paea LePendu, Stanford Center for Biomedical Informatics Research (BMIR)
  2. *Panelist 2:* Lorraine Fernandes, Information Management, IBM
  3. *Panelist 3:* Graham Nixon, Chief Health Informatics Officer, Veteran Affairs San Diego Healthcare System
  4. *Panelist 4:* Charles Boicey, Informatics Solutions Architect, UC Irvine Health

1:00 pm – 2:30 pm Lunch Break and Poster session
(Provided: Beckman Center 404)

2:30 pm – 4:00 pm Technical Sessions

- **Cloud Computing and Big Data**
  Chair: Michael Fahy, Chapman University
  1. *Rc^2: R and SAS collaboration in the cloud....N/A*
     James Harner, West Virginia University
  2. *Cloud-Enabled Processing & Exploitation of Remotely Sensed Data....N/A*
     Michael Limcaco, Amazon Web Services
  3. *Computing Techniques for Big Data....552*
     Gayn B. Winters, Technical Program Manager, Technology Consultants

- **Climate Data Analysis: From Satellites to Climate Models**
  Chair: Robert Allen, University of California, Riverside
  1. *Heterogeneous Warming Agents and Widening of the Tropical Belt....571*
     Robert Allen, University of California, Riverside
  2. *The Future of Model Evaluation....586*
     Charlie Zender, University of California, Irvine
  3. *Statistical correction of satellite cloud data for climate change studies....592*
     Joel Norris, Scripps Institution of Oceanography, University of California, San Diego
  4. *The NASA-Unified WRF: Observation-Driven Modeling System for Studying Regional Aerosol-Cloud-Precipitation-Land Surface Processes and Interactions....615*
     Toshihisa Matsui, NASA GSFC

- **Big Data Challenges in Bioinformatics and Medical Informatics**
  Organizer: Nadim Alkharouf, Towson University
  Chair: Ian Misner, Towson University
  1. *Big Data Challenges in the Sequencing of plant pathogen Genomes....N/A*
     Ian Misner, Towson University
  2. *Speeding-Up Codon Analysis on the Cloud with Local MapReduce Aggregation....N/A*
     Atanas Radenski & Louis Ehwerhemuepha, Chapman University
  3. *Translational Bioinformatics and Quantitative Biology in Cancer Research: From Big Data Molecular Analysis of Human Disease to Personalized Medicine....N/A*
     Gennady Verkhivker, Chapman University

(Beckman Center 404)

4:15 pm – 5:15 pm Panel Discussion by NEXUS IS: *Applied Analytics in the Enterprise....N/A*
Moderator: Kevin Griffith, NEXUS IS

1. *Panelist 1:* Colin McNamara, Chief Cloud Architect
2. *Panelist 2:* Paul Caracciolo, Chief Healthcare Officer

5:15 pm – 7:00 pm (NEXUS IS Sponsored Reception)

Saturday April 6th 2013:

Sandhu Conference Center

7:30 am – 8:00 am Registration
8:00 am – 9:30 am Technical Sessions

- **Random Solutions to Big Problems**
  Co-Chairs: Eric C. Chi, University of California, Los Angeles and Miles Lopes, University of California, Berkeley
  1. *Implementing Randomized Matrix Algorithms in Parallel and Distributed Environments....N/A*
     Michael Mahoney, Stanford
  2. *A More Powerful Two-Sample Test in High Dimensions using Random Projection....N/A*
     Miles Lopes, University of California, Berkeley
  3. *Perturb-and-MAP Random Fields: The Interplay Between Random Sampling and Optimization....N/A*
     George Papandreou, University of California, Los Angeles

- **From Large Earth Science Datasets to Compelling Scientific Results**
  Chair: Mian Chin, NASA GSFC
  1. *Where does it come from, where does it go? Constraining sources and distributions of atmospheric pollutants with satellite data....N/A*
     Daven Henze, University of Colorado
  2. *Satellite and Model Data to Support Air Quality Management....634*
     Tracey Holloway, University of Wisconsin, Madison
  3. *Trends in Extreme United States Temperatures....648*
     Jaechoul Lee, Boise State University

9:30 am – 11:00 am Technical Sessions

- **Visualization of Big Data**
  Chair: Juergen Symanzik, Utah State University
  1. *Interactive Visualization of Big Data....N/A*
     Simon Urbanek, AT&T Research Labs
  2. *Divide and Recombine (D&R) for Comprehensive Visualization of Large Complex Data at Their Finest Granularity....N/A*
     William Cleveland, Purdue University
  3. *Visualization of "Big Data" in Molecular and Genomic Sciences....N/A*
     Nicholas Lewin-Koh, Genentech

- **From Large Earth Science Datasets to Compelling Scientific Results**
  Chair: Hesham El-Askary, Chapman University

11:00 pm – 11:30 pm Morning Break
11:30 pm – 1:00 pm Technical Sessions

- **Data Science Methods**
  Chair: Edward Wegman, George Mason University

- **Massive Data Challenges in Numerical Weather Modeling**
  Chair: Menas Kafatos, Chapman University

Additional Paper