Curriculum Vitae

Sangmi Lee Pallickara

Computer Science Department,
1873 Campus Delivery,
Fort Collins, CO 80523-1873
Email : Sangmi.Pallickara@colostate.edu
URL: http://www.cs.colostate.edu/~sangmi

Curriculum Vitae Highlights

| Research

- Jack Cochran Family Professorship, 2018 – 2021
- IEEE TCSC Award for *Excellence in Scalable Computing* (Mid-Career Researcher), 2018
  - This is an international award given by the IEEE's Technical Committee on Scalable Computing
- National Science Foundation CAREER Award, 2016
- 4 Best Paper awards at IEEE and ACM forums; 1 runner up for a Best Paper award.
- Research funding from federal agencies
  - o NSF, NIFA, DHS, DOE ARPA-E, and EDF

| Teaching & Advising

- Curriculum Development: Creator of the Big Data track at CSU. This includes two 400-level capstone courses, a graduate course
- Educational Grants from Amazon and Google
- Supervised 4 Ph.D. dissertations, 18 M.S. theses, 4 M.S. projects, and 3 undergraduate honors theses.
- My advisees are recipients of prestigious, competitive awards including IEEE TCSC Outstanding Ph.D. Dissertation award, Anita Reed Graduate award, 6 Best Paper awards, and 1 runner up for the Best Paper award.
- Primary advisor of 3 NSF Research Experiences for Undergraduate (REU) students at CSU

| Outreach

- Founding director of the Summer STEM Camp for High School and Middle School Students in Fort Collins area (SWiFT Summer STEM Program, 2016 –). The camp focuses on exposing students to computing and STEM related applications to encourage pursuing careers in STEM and majoring in Computer Science.
- Faculty mentor for Girls Who Code: 2021

| Academic & Professional Service

- Associate Editor, *IEEE Transactions of Parallel and Distributed Systems*, 2019 – present
- Editor, Editorial Board, *Journal of Future Generation Computer Systems.* Publisher: Elsevier. 2021 – present
- Editorial Board, *Journal of Big Data.* Publisher: Springer New York/Heidelberg, Germany. 2013 – 2023
- Technical Program Committee of 38 IEEE and ACM conferences/symposium/workshops.

1 Research Interests

My research interests are in the area of Big Data for the sciences; specifically, issues related to geospatial AI, scalable storage, analytics, metadata, and visualization.

2 Education

Aug 2003	**Ph.D.** Computer Science, *Florida State University*, Tallahassee, FL
May 2000	**M.S.** Computer and Information Science, *Syracuse University*, Syracuse, NY
Feb 1993	**B.S.** Physics, *Sookmyung University*, Seoul, Korea (Minor in Computer Science)

3 Awards

Cochran Family Professorship. 2018 ~ 2021
This is the first endowed professorship in the College of Natural Sciences and I am the first recipient of this honor.

IEEE TCSC Award for *Excellence in Scalable Computing* (Mid-Career Researcher). 2018.
    This is an international award given by the IEEE's Technical Committee on Scalable Computing.

National Science Foundation CAREER Award, 2016.

Best Paper Award. the IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT) with Saptashwa Mitra *(graduate advisee)*, Matthew Young *(graduate advisee)*, 2023.

Best Paper Award. The IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT) with Menuka Warushavithana *(graduate advisee)*, Kassidy Barram*(graduate advisee)*, Saptashwa Mitra*(graduate advisee)*, Caleb Carlson*(graduate advisee)*, 2023

Best Paper Award. *ACM International Conference on Big Data Computing, Applications and Technologies with Sam Armstrong (graduate advisee) and Kevin Bruhwiler (graduate advisee)*, Auckland, New Zealand, 2019

Best Paper Award. *IEEE International Conference on Cluster Computing (CLUSTER) with Saptashwa Mitra (graduate advisee), Paahuni Khandelwal (graduate advisee), and Shrideep Pallickara*, Albuquerque, NM, 2019

Shortlisted for the Best Paper Award. *IEEE International Conference on Big Data and Cloud Computing (BDCloud2018)* [27% acceptance rate] *with Maxwell Roselius (graduate advisee)*, Melbourne, Australia, 2018

Best Paper Award. *IEEE/ACM Symposium on Big Data Computing* [22% acceptance rate] with Jared Koontz (graduate advisee), and Matthew Malensek (graduate advisee). 2014.

Best Paper Award. *IEEE/ACM Conference on Utility and Cloud Computing* [27% acceptance rate] with Matthew Malensek (graduate advisee) and Shrideep Pallickara. 2012.

4   Professional Experience

July 22 –	Professor, *Computer Science Department*, Colorado State University
Dec 18 – Dec 21	Cochran Family Professor, *College of Natural Sciences, Colorado State University*
July 17 – Jun 22	Associate Professor, *Computer Science Department*, Colorado State University
Jan 13 – Jun 17	Assistant Professor, *Computer Science Department*, Colorado State University
Jan 10 – Dec 12	Research Scientist, *Computer Science Department*, Colorado State University.
Aug 07 – Dec 09	*Research Technologies*, Indiana University.
Sep 04 – Jul 07	Post Doctoral Fellow, *Department of Computer Science*, Indiana University With Professor Beth Plale.
Sep 03 – Aug 04	Post Doctoral Fellow, *Community Grids Lab*, Indiana University.
May 00 – Aug 03	Research Assistant, *Department of Computer Science*, Florida State University.
Sep 01 – Aug 04	Visiting Research Associate, *Community Grids Lab*, Indiana University.
May 99 – Apr 00	Research Assistant, Northeast Parallel Architectures Center, *Department of Electrical Engineering & Computer Science*, Syracuse University.

5   Research Funding

My research is funded by the National Science Foundation, the Advanced Research Projects Agency-Energy (ARPA-E), the Department of Homeland Security, the National Institute of Food and Agriculture, the Environmental Defense Fund, Colorado School of Public Health, Google, Amazon and HP.

[R19]   CPS: Medium: Making Every Drop Count: Accounting for Spatiotemporal Variability of Water Needs for Proactive Scheduling of Variable Rate Irrigation Systems. **Sangmi Lee Pallickara** (PI) with Allan A. Andales, Jeffrey Niemann, Shrideep Pallickara, and F. Jay Breidt. National Science Foundation. $1,199,846. 08/2023-07/2026.

[R18]   National Science Foundation AI Institute: AI-CLIMATE (AI Institute for Climate-Land Interactions, Mitigation, Adaptation, Tradeoffs and Economy), PI: Shashi Shekhar. CSU Investigators: Keith Paustain (Co-PI) with Francesca Cotrufo, Patrick Keys, Nathan Mueller, Stephen Ogle, **Sangmi Lee Pallickara**, Shrideep Pallickara, and Yao Zhang. NSF and NIFA: $20,000,000. 8/2023-7/2028.

[R17]   FACT: A Scalable Infrastructure for High-precision Evapotranspiration Estimation and Effective Farm-level Decision Making. $495,585. 7/2020-6/2023. **Sangmi Lee Pallickara** (PI) with Allan A. Andales, F. Jay Breidt, Gabriel Senay, Shrideep Pallickara, and Darin Schulte. National Institute of Food and Agriculture.

[R16]   Frameworks: Collaborative Proposal: Software Infrastructure for Transformative Urban Sustainability Research. $3,081,355. 10/2019-9/2024. **Sangmi Lee Pallickara** (Co-PI) with Shrideep Pallickara (PI), Mazdak Arabi, Jay Breidt, Sudipto Ghosh, Mikhail Chester, Amir AghaKouchak, and Claire Welty. National Science Foundation.

[R15]   REU Supplement: Shrideep Pallickara (PI). National Science Foundation. Co-PIs: Mazdak Arabi, Sangmi Lee Pallickara, Jay Breidt, Sudipto Ghosh $16,000 4/2020–4/2021.

[R14]   REU Supplement: Shrideep Pallickara (PI). National Science Foundation. Co-PIs: Mazdak Arabi, Sangmi Lee Pallickara, Jay Breidt, Sudipto Ghosh $16,000 4/2023–4/2024.

[R13]   Cochran Family Professorship. Cochran Family Foundation $60,000. 2018-2021. **Sangmi Lee Pallickara** (PI).

[R12]   Root genetics in the Field to Understand Drought Adaptation and Carbon Sequestration, $6,019,238, 7/2017-7/2021. **Sangmi Lee Pallickara** (Co-PI) with John McKay (PI), Chris Turner (Co-PI), Randy Bartels (Co-PI), Francesca Cotrufo (Co-PI), Thomas Borch (Co-PI), Andy French (Co-PI), Keith Paustian (Co-PI), Antin Parker (Co-PI), Patrick Schnable (Co-PI), Chris Topp(Co-PI), Matthew Wallenstein(Co-PI), and Jianming Yu (Co-PI), the Advanced Research Projects Agency-Energy (ARPA-E), Department of Energy.

[R11]   CAREER: A Framework for Ad Hoc Model Construction in Data Streaming Environments. $491,243 5/2016-5/2021. **Sangmi Lee Pallickara** (PI). Faculty Early Career Development (CAREER) Award. National Science Foundation. ACI-1553685.

[R10]   REU Supplement:: Sangmi Pallickara (PI). National Science Foundation. $16,000 7/2020–7/2021

[R9]   Methane Mapping at Colorado State University: Phase-3, $77,553 10/2016 - 9/2017. **Sangmi Lee Pallickara** (Co-PI) with Joe von Fischer (PI), Environmental Defense Fund.

[R8]   Integration and Analeytics of Public Health Big Data for Identifying Risk Factors for Obesity. $19,850 7/2016-6/2017. **Sangmi Lee Pallickara** (Co-PI) with Kaigang Li (PI). Colorado School of Public Health.

[R7]   Geo for Google Cloud Credits Beta Program. $11,000. Google. 2016. **Sangmi Lee Pallickara** (PI).

[R6]   Assessing Epidemiological and Economic Impacts of Countermeasures and Vaccination Strategies in Disease Outbreaks at the National Scale. $2,046,914 9/2015–9/2018. **Sangmi Lee Pallickara** (Co-PI) with Shrideep Pallickara (PI), Jay Breidt, Neil Harvey, Mike Sanderson, and Dustin Pendell. Science and Technology Directorate. Department of Homeland Security.

[R5]   Methane Mapping at Colorado State University: Database Supporting Scalable Analysis & Visualization of Multidimensional, Time-Series Gas Leak Data. $68,246 2015-2016, **Sangmi Lee Pallickara** (PI) with Joe von Fischer (Co-PI), Environmental Defense Fund.

[R4]   Coalition for Development and Implementation of Sensor Systems. $180,000 2014-2016, **Sangmi Lee Pallickara** (Co-PI) with David Dandy (PI), Ken Reardon, Chuck Henry, Melissa Reynolds, Anura Jayasumana, Tom Chen, Rick Lyons, Jennifer Mueller, Lori Peek, and John Volckens, The CSU Catalyst for Innovative Partnerships program.

[R3]   *Big Data Analytics of Epidemic Outbreaks.* $1,176,860 9/2013-7/2017. **Sangmi Lee Pallickara** (Co-PI) with Shrideep Pallickara (PI), Neil Harvey (Co-PI), and Jay Breidt (Co-PI). LONG RANGE PROGRAM, Science and Technology Directorate. Department of Homeland Security.

[R2]   *Fast and Accurate Approximations of Data Distributions in Voluminous Datasets.* $2,500 2013 **Sangmi Lee Pallickara** (PI), Undergraduate research project, Hewlett Packard Inc.

[R1]   *Exploring Performance Hotspots for Query Evaluation Algorithms Over Voluminous Time-Series Data. Infrastructure Grant* (Access to 2,000 virtual machines, $7,500) 6/2013- 6/2015, **Sangmi Lee Pallickara** (PI), Amazon Web Services in Education Research Grant Award. Amazon Inc.

Research Partnerships

[P3]   Soil Carbon Consortium, Soil Carbon Solution Center, Colorado State University [Link], 5/2023 – Present.

[P2]   Metrics, Management, and Monitoring, Foundation of Food and Agriculture Research [Link], 12/2022-Present.

[P1]   Center for Exascale Spatial Data Analytics and Computing (XSD) [Link], 1/2023 – Present.

Education Grants

[E5]   Google Cloud Platform Education Award (for the graduate course, CS535 Big Data, $1,700), 9/2016 – 12/2016. Google Cloud Platform, **Sangmi Lee Pallickara** (PI).

[E4]   AWS Gift (for the senior undergraduate course, CS435 Introduction to Big Data, $4,550), 1/2016 – 5/2016 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI).

[E3]   AWS Gift (for the graduate course, CS535 Big Data, $1,950), 8/2015 – 12/2015 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI).

[E2]   *AWS Faculty Educator Award* (for the senior undergraduate course, *CS480A2 Introduction to Big Data*, $7,000), 1/2015 – 5/2015 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI), Amazon Inc.

[E1]   *AWS Faculty Educator Award* (for the graduate course, *CS581 Big Data,* $3,500), 8/2014 – 12/2014 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI), Amazon Inc.

6   Publications

Listings with the symbol * identify my graduate advisees. Manuscripts where I am the senior author are indicated by §.

6.1   Journals

[J24]   Saptashwa Mitra*, Maxwell Roselius*, Pedro Andrade-Sanchez, John K. McKay, and **Sangmi Lee Pallickara**,
§   RADIX+: High-Throughput Georeferencing and Data Ingestion over Voluminous and Fast-Evolving Phenotyping Sensor Data, Concurrency and Computation: Practice and Experience (CCPE), John-Wiley, 2023.

[J23]   Chris Fisher, Stephen Leisz, Diana Wall, Melinda Laituri, Geoffrey Henebry, Damian Evans, Juan Carlos Fernandez- Diaz, Shrideep Pallickara, **Sangmi Lee Pallickara**, Thomas Garrison, Francisco Estrada-Belli, Eduardo Neves, Kathryn Reese-Taylor, Rachel Opitz, Thomas Lovejoy, William Sarni, Rodrigo Solinis, Grace Ellis, Creating an Earth Archive. (To appear) Proceedings of the National Academy of Sciences. 2022.

[J22] Sam Armstrong*, Paahuni Khandelwal*, Dhruv Padalia*, Gabriel Senay, Darin Schultz, Allan Andales, Jay
§   Breidt, Shrideep Pallickara, and Sangmi Lee Pallickara. Attention-Based Convolutional Capsules for
     Evapotranspiration Estimation at Scale. Environmental Modeling & Software, Elsevier. Vol. 152, 105366, 2022.

[J21] Daniel Rammer*, Thilina Buddhika, Matthew Malensek*, Shrideep Pallickara, and **Sangmi Lee Pallickara**.
§   Enabling Fast Exploratory Analyses Over Voluminous Spatiotemp,oral Data Using Analytical Engines. IEEE
     Transactions on Big Data. Vol. 8 (1) pp. 213-228. 2022. [Impact Factor: 5.67]

[J20] Thilina Buddhika, Matthew Malensek*, Shrideep Pallickara, and **Sangmi Lee Pallickara**. Living on the Edge:
§   Data Transmission, Storage, and Analytics in Continuous Sensing Environments. ACM Transactions on
     Internet of Things. Vol. 2 (3), pp 1-31. 2021.

[J19] Thilina Buddhika, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Pebbles: Leveraging Sketches for Processing
     Voluminous, High Velocity Data Streams. *IEEE Transactions on Parallel and Distributed Systems*. Vol. 32(8):
     2005-2020. 2021. [Impact Factor: 4.64]

[J18] Matthew Malensek*, Walid Budgaga*, Ryan Stern, Shrideep Pallickara, and **Sangmi Lee Pallickara**. Trident:
§   Distributed Storage, Analysis, and Exploration of Multidimensional Phenomena. *IEEE Transactions on Big Data*.
     Vol. 5 (2) pp 252 265. 2019. [Impact Factor: 5.67]

[J17] Naman Shah*, Matthew Malensek*, Harshil Shah*, Shrideep Pallickara, and **Sangmi Lee Pallickara**. Scalable
     Network Analytics for Characterization of Outbreak Influence in Voluminous Epidemiology Datasets.
§   Concurrency and Computation: Practice & Experience. John-Wiley, Vol. 31 (7), 2019. [Impact Factor: 1.133]

[J16] Katherine E. Boehle, Erin Doan*, Sadie Henry*, J. Ross Beveridge, **Sangmi Lee Pallickara**, Charles S. Henry,
     Single Board Computing System for Automated Colorimetric Analysis on Low-Cost Analytical Devices.
     Analytical Methods, Royal Society of Chemistry, 10, pp 5282—5290, 2018. [Impact Factor: 2.073]

[J15] Thilina Buddhika, Matthew Malensek*, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Synopsis: A Distributed
     Sketch over Voluminous Spatiotemporal Observational Streams. *IEEE Transactions on Knowledge and Data
     Engineering*. Vol. 29 (11) pp 2552-2566. 2017. [Impact Factor: 3.438]

[J14] Matthew Malensek*, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Hermes: Federating Fog and Cloud
     Nodes to Support Query Evaluations in Continuous Sensing Environments. *IEEE Cloud Computing*. Vol. 4(2)
     pp 54-62. 2017. [Impact Factor: 1.86]

[J13] Walid Budgaga*, Matthew Malensek*, **Sangmi Lee Pallickara**, and Shrideep Pallickara. A Framework for
§   Scalable Real-Time Anomaly Detection over Voluminous, Geospatial Data Streams. *Concurrency and
     Computation: Practice & Experience*. Vol. 29(12) pp 1-16. John-Wiley. 2017. [Impact Factor: 1.133]

[J12] Matthew Malensek*, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Fast, Ad Hoc Query Evaluations over
§   Multidimensional Geospatial Datasets. *IEEE Transactions on Cloud Computing*. Vol. 5(1) pp 28-42 2017.
     [Impact Factor: 3.77]

[J11] Cameron Tolooee*, Matthew Malensek*, and **Sangmi Lee Pallickara,** A Scalable Framework for Continuous
§   Query Evaluations over Multidimensional, Scientific Datasets. *Concurrency and Computation: Practice and
     Experience*. 28(8): pp. 2546-2563. 2016. [Impact Factor: 1.133]

[J10]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Analytic Queries over Geospatial Time-Series Data using Distributed Hash Tables. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 28(6): pp.1408-1422. 2016. [Impact Factor: 3.438]

[J9]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Autonomous Data Management and
§   Federation to Support High-throughput Query Evaluations over Voluminous Datasets. *IEEE Cloud* Computing. Vol. 3 (3): pp. 40-49. 2016. [Impact Factor: 1.86]

[J8]   Walid Budgaga*****, Matthew Malensek*****, **Sangmi Lee Pallickara**, Neil Harvey, Jay Breidt, and Shrideep Pallickara. Predictive Analytics Using Statistical, Learning, and Ensemble Methods to Support Real- Time Exploration of Discrete Event Simulations. *Future Generation Computer Systems*. Elsevier. Volume 56, March, pp. 360–374. 2016. [Impact Factor: 3.997]

[J7]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Minerva: Proactive Disk Scheduling for QoS in Multi-Tier, Multi-Tenant Cloud Environments. *IEEE Internet Computing*. Vol. 20 (3): pp. 19-27. 2016. [Impact Factor: 1.521]

[J6]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Geometry and Proximity Constrained
§   Query Evaluations over Large Geospatial Datasets Using Distributed Hash Tables. *IEEE Computing in Science and Engineering (CiSE)*. Special Issue on Extreme Data. Vol. 16(4): pp. 53-60. 2014. [Impact Factor: 2.074]

[J5]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. "Exploiting Geospatial and Chronological Characteristics in Data Streams to Enable Efficient Storage and Retrievals," *Future Generation Computer Systems*. Vol. 29(4): pp. 1049-1061. Elsevier. 2013. [Impact Factor: 3.997]

[J4]   **Sangmi Lee Pallickara**, Shrideep Pallickara, and Milija Zupanski, "Enabling Efficient Data Search and Subsetting of Large-scale Atmospheric Datasets". *Future Generation Computer Systems*, Vol. 28(1): pp. 112-118. Elsevier. 2012. [Impact Factor: 3.997]

[J3]   Marlon Pierce, Xiaoming Gao, **Sangmi Lee Pallickara**, Zhenhua Gau, and Geoffrey Fox, "QuakeSim Portal and Services: New Approaches to Science Gateway Development Techniques," *Concurrency & Computation: Practice & Experience*. 22(12): pp.1732-1749. 2010.

[J2]   Beth Plale, Dennis Gannon, Yi Huang, Gopi Kandaswamy, **Sangmi Lee Pallickara**, and Aleksander Slominski, "Cooperating Services for Managing Data Driven Computational Experimentation," *IEEE Computing in Science and Engineering (CiSE)*, (Vol. 7, No. 5) pp. 34-43. 2005. [Impact Factor: 2.074]

[J1]   Geffrey Fox, Sung-Hoon Ko, Marlon Pierce, Ozgur Balsoy, Jungkee Kim, **Sangmi Lee**, Kangseok Kim, Sangyoon Oh, Xi Rao, Mustafa Varank, Hasan Bulut, Guruhan Gunduz, Xiahong Qui, Shrideep Pallickara, Ahmet Uyar, "Grid Services for Earthquake Science," *Concurrency and Computation: Practice and Experience in ACES Special Issue*, 14(6-7): pp. 371-393, 2002. [Impact Factor: 1.133]

6.2   Book Chapters

[B3]   **Sangmi Lee Pallickara**, Matthew Malensek*****, and Shrideep Pallickara. "On the Processing of Extreme Scale Datasets in the Geosciences," Handbook of Data Intensive Computing. Springer. 2012.

[B2]   **Sangmi Lee Pallickara**, Shrideep Pallickara and Marlon Pierce, "Scientific Data Management in the Cloud: A Survey of Technologies, Approaches and Challenges," Chapter 22: pp.517-534, Handbook of Cloud Computing. Springer. ISBN: 978-1-4419-6523-3. 2010.

[B1]   Geoffrey Fox, Dennis Gannon, Sung-Hoon Ko, **Sangmi Lee**, Shrideep Pallickara, Marlon Pierce, Xiaohong Qiu, Xi Rao, Ahmet Uyar, Minjun Wang, Wenjun Wu Book chapter on "Peer-to-Peer Grids,"

Grid Computing: Making The Global Infrastructure a Reality, John Wiley & Sons, ISBN 0470853190, 2003

6.3   Refereed Conference Proceedings

[C61]   Menuka Warushavithana*, Kassidy Barram, Saptashwa Mitra*, Caleb Carlson*, Sudipto Ghosh, Jay Breidt, **Sangmi Lee Pallickara,** Shrideep Pallickara, A Framework for Profiling Spatial Variability in the Performance of Classification Models, In proceedings of the IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), 2023. ** Best Paper Award (out of three best papers)

[C60]   Saptashwa Mitra*, Matthew Young, **Sangmi Lee Pallickara,** Shrideep Pallickara, Rubiks: Rapid Explorations and Summarization over High Dimensional Spatiotemporal Datasets, In proceedings of the IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), 2023. ** Best Paper Award (out of three best papers)

[C59]   Matthew Young, **Sangmi Lee Pallickara,** Shrideep Pallickara, aQua: A Framework for Spatiotemporal Analysis and Visualizations of Water Quality Data at Scale, In proceeding of the IEEE International Conference on Big Data (IEEE BigData), Sorento, Italy, 2023.

[C58]   Abdul Matin*, Paahuni Khandelwal*, Shrideep Pallickara, **Sangmi Lee Pallickara**, Discern: Leveraging Knowledge Distillation to Generate High Resolution Soil Moisture Estimation from Coarse Satellite Data, In proceeding of the IEEE International Conference on Big Data (IEEE BigData), Sorento, Italy, 2023.

[C57]   Ethan Seefried, James Yost, Jack Fitzgerald, **Sangmi Lee Pallickara,** and Nathan Blanchard, Paying Attention to Wildfire: Using U-Net with Attention Blocks on Multimodal Data for Next Day Prediction. Proceedings of the 25th ACM International Conference on Multimodal Interaction, Paris, France. 2023

[C56]   Saptashwa Mitra*, Paahuni Khandelwal*, Shrideep Pallickara and **Sangmi Lee Pallickara**. Argus: Rapid Tracking of Wildfires from Unlabeled Satellite Images. Proceedings of the International Conference on Cloud Computing (CLOUD). Chicago, USA 2023. [21.6% acceptance rate]

[C55]   Paahuni Khandelwal*, Menuka Warushavithana*, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Enabling Fast, Effective Visualization of Voluminous Gridded Spatial Datasets. Proceedings of the 23rd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing. (CCGrid 2023). Bangalore, India. [20% acceptance rate]

[C54]   Saptashwa Mitra*, Menuka Warushavithana*, Mazdak Arabi, Jay Breidt, **Sangmi Lee Pallickara,** and Shrideep Pallickara.   Alleviating Resource Requirements for Spatial Deep Learning Workloads. Proceedings of the 22nd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing. (CCGrid 2022). pp 452-462. Taormina (Messina), Italy. 2022.

[C53]   Pierce Smith, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Griddle: Effective Query Support over Voluminous Gridded Spatial Datasets. Proceedings of the IEEE International Conference on Big Data (IEEE BigData). Osaka, Japan. 2022. [18.6% acceptance rate].

[C52]   Caleb Carlson*, Menuka Warushavithana*, Saptashwa Mitra*, Kassidy Barram, Sudipto Ghosh, Jay Breidt, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Resource Efficient Profiling of Spatial Variability in Performance of Regression Models. Proceedings of the IEEE International Conference on Big Data (IEEE BigData). Osaka, Japan. 2022. [18.6% acceptance rate]

[C51] Abdul Matin*****, Samuel Armstrong*****, Saptashwa Mitra*****, Shrideep Pallickara, and **Sangmi Lee Pallickara**, Rapid Betweenness Centrality Estimates for Transportation Networks using Capsule Networks, Transportation Networks using Capsule Networks, IEEE Transdisciplinary AI (TransAI), 2022.

§

[C50] Paahuni Khandelwal*****, Samuel Armstrong*****, Abdul Matin*****, Shrideep Pallickara, **Sangmi Lee Pallickara**. CloudNet: A Deep Learning Approach for Mitigating Occlusions in Landsat-8 Imagery using Data Coalescence. Proceedings of the IEEE eScience Conference (eScience). 2022

§

[C49] Saptashwa Mitra*****, Daniel Rammer*****, Shrideep Pallickara, **Sangmi Lee Pallickara**. Glance: A Generative Approach to Interactive Visualization of Voluminous Satellite Imagery. Proceedings of the IEEE International Conference on Big Data (IEEE BigData). pp 359-367. 2021. [19% acceptance rate].

§

[C48] Menuka Warushavithana*****, Caleb Carlson*****, Saptashwa Mitra*****, Daniel Rammer*****, Mazdak Arabi, Jay Breidt, **Sangmi Lee Pallickara**, and Shrideep Pallickara , Distributed Orchestration of Regression Models Over Administrative Boundaries. Proceedings of the IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK. pp 80-90. 2021.

[C47] Menuka Warushavithana*****, Saptashwa Mitra*****, Mazdak Arabi, Jay Breidt, **Sangmi Lee Pallickara**, and Shrideep Pallickara, A Transfer Learning Scheme for Time Series Forecasting Using Facebook Prophet. Proceedings of the 2021 IEEE International Conference on Cluster Computing (CLUSTER). pp 809-810, 2021.

[C46] Saptashwa Mitra*****, Daniel Rammer*****, Shrideep Pallickara, **Sangmi Lee Pallickara**. A Generative Approach to Visualizing Satellite Data. Proceedings of the 2021 IEEE International Conference on Cluster Computing (CLUSTER), pp 815-816, 2021.

§

[C45] Paahuni Khandelwal*****, Daniel Rammer*****, Shrideep Pallickara, **Sangmi Lee Pallickara**. Mind the Gap: Generating Imputations for Satellite Data Collections at Myriad Spatiotemporal Scopes. Proceedings of the 21st *IEEE/ACM international Symposium on Cluster, Cloud and Internet Computing* (CCGrid). pp 92-102, 2021. Melbourne, Australia [26% acceptance rate]

§

[C44] Kevin Bruhwiler*****, Paahuni Khandelwal*****, Daniel Rammer*****, Samuel Armstrong*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Lightweight, Embeddings Based Storage and Model Construction Over Satellite Data Collections. Proceedings of the *IEEE International Conference on Big Data* (IEEE BigData). Atlanta, USA. 2020. [15.5% acceptance rate]

[C43] Daniel Rammer*****, **Sangmi Lee Pallickara**, Shrideep Pallickara. Towards Timely, Resource- Efficient Analyses Through Spatially-Aware Constructs within Spark. Proceedings of the IEEE/ACM Conference on Utility and Cloud Computing. Leicester, UK. 2020. [31% acceptance rate]

[C42] Daniel Rammer*****, Kevin Bruhwiler*****, Paahuni Khandelwal*****, Samuel Armstrong*****, Shrideep Pallickara, **Sangmi Lee Pallickara**, Small is beautiful: Distributed Orchestration of Spatial Deep Learning Workloads, Proceedings of the *IEEE/ACM Conference on Utility and Cloud Computing*, Leicester, UK 2020 [31% Acceptance Rate]

§

[C41] Kevin Bruhwiler*****, Thilina Buddhika, Shrideep Pallickara and **Sangmi Lee Pallickara**. Iris: Amortized, Resource Efficient Visualizations of Voluminous Spatiotemporal Datasets. Proceedings of the *IEEE/ACM International Conference on Big Data Computing, Applications and Technologies.* Leicester, UK. 2020.

§

[C40] Walid Budgaga*****, Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Concerto: Leveraging Ensembles for Timely, Accurate Model Training Over Voluminous Datasets. Proceedings of

the *IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*. Leicester, UK. 2020.

[C39] Sam Armstrong*****, Kevin Bruhwiler*****, and **Sangmi Lee Pallickara**, Rapid, Progressive Sub-Graph
§ Explorations for Interactive Visual Analytics over Large-Scale Graph Datasets, Proceedings of the *IEEE/ACM International Conference on Big Data Computing, Application, and Technology*, Auckland, New Zealand, 2019. [27.7 % acceptance rate] ** Best Paper Award

[C38] Daniel Rammer*****, Sangmi Lee Pallickara, and Shrideep Pallickara. ATLAS: A Distributed File System for Spatiotemporal Data. Proceedings of the *IEEE/ACM Conference on Utility and Cloud Computing*. Auckland, New Zealand. 2019. [29% acceptance rate]

[C37] Saptashwa Mitra*****, Paahuni Khandelwal*****, Shrideep Pallickara, and **Sangmi Lee Pallickara,** STASH: Fast Hierarchical Aggregation Queries for Effective Visual Spatiotemporal Explorations, Proceedings of the
§ *IEEE International Conference on Cluster Computing* (CLUSTER), Albuquerque, New Mexico, 2019. [27% acceptance rate]
** Best Paper Award

[C36] Bibek Shrestha*****, Saptashwa Mitra*****, and **Sangmi Lee Pallickara**, STRETCH: In-memory Storage With Autoscaling For Cluster Computing, Proceedings of the *IEEE International Conference on Cloud*
§ *Computing* (IEEE CLOUD), Milan, Italy, 2019. [20.8% acceptance rate]

[C35] Daniel Rammer*****, Walid Budgaga*****, Thilina Buddhika, Shrideep Pallickara, and **Sangmi Lee Pallickara**.
§ Alleviating I/O Inefficiencies to Enable Effective Model Training Over Voluminous, High-Dimensional Datasets. Proceedings of the IEEE International Conference on Big Data (IEEE BigData). Seattle, USA. 2018. [18.9% acceptance rate]

[C34] Saptashwa Mitra*****, and **Sangmi Lee Pallickara**, Confluence: Adaptive Spatiotemporal Data Integration Using Distributed Query Relaxation Over Heterogeneous Observational Datasets, *Proceedings of the*
§ *IEEE/ACM Conference on Utility and Cloud Computing (UCC)*, Zurich, Switzerland 2018. [20.5% acceptance rate]

[C33] Max Roselius***** and **Sangmi Lee Pallickara**, Enabling High-throughput Georeferencing for Phenotype Monitoring over Voluminous Observational Data, *Proceedings of the IEEE International Conference on*
§ *Big Data and Cloud Computing (BDCloud2018)*, Melbourne, Australia, 2018. [27% acceptance rate] **
Finalist for Best Paper Award

[C32] Saptashwa Mitra*****, Yu Qiu*****, Haley Moss, Kaigang Li, and **Sangmi Lee Pallickara**, Effective Integration
§ of Geotagged, Ancillary Longitudinal Survey Datasets to Improve Adulthood Obesity Predictive Models, *IEEE Big Data Science and Engineering (IEEE BigDataSE)*. New York, USA, 2018.

[C31] Johnson Charles Kachikaran Arulswamy*****, and **Sangmi Lee Pallickara,** Columbus: Enabling Scalable
§ Scientific Workflows for Fast Evolving Spatio-Temporal Sensor Data. *Proceedings of the 14th IEEE International Conference of Service Computing (IEEE SCC)*, pp.9-18. Honolulu, Hawaii, USA, 2017 [20% acceptance rate]

[C30] Naman Shah*****, Harshil Shah*****, Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara.
§ Network Analysis for Identifying and Characterizing Disease Outbreak Influence from Voluminous

Epidemiology Data. Proceedings of the *IEEE International Conference on Big Data* (*IEEE BigData*). Washington D.C., USA. 2016.   [18.68% acceptance rate]

[C29]
§
Cameron Tolooee*****, **Sangmi Lee Pallickara** and Asa Ben-Hur. Mendel: A Distributed Storage Framework for Similarity Searching over Sequencing Data. Proceedings of the 30ᵗʰ IEEE International Parallel & Distributed Processing Symposium (IPDPS). pp. 790-799. Chicago, USA, 2016.   [23% acceptance rate].

[C28]
Matthew Malensek*****, **Sangmi Lee Pallickara** and Shrideep Pallickara, Alleviation of Disk I/O Contention in Virtualized Settings for Data-Intensive Computing. Proceedings of the *IEEE/ACM International Symposium on Big Data Computing*. pp. 1-10. Cyprus. 2015. [16% acceptance rate]

[C27]
§
Jared Koontz*****, Matthew Malensek*****, and **Sangmi Lee Pallickara**. GeoLens: Enabling Interactive Visual Analytics over Large-scale, Multidimensional Geospatial Datasets. Proceedings of the *IEEE/ACM Symposium on Big Data Computing*. pp. 35-44. London, UK, 2014.   [22% acceptance rate]
** Best Paper Award

[C26]
Matthew Malensek*****, Walid Budgaga*****, **Sangmi Lee Pallickara**, Neil Harvey and Shrideep Pallickara, Using Distributed Analytics to Enable Real-Time Exploration of Discrete Event Simulations. *Proceedings of the IEEE/ACM International Conference on Utility and Cloud Computing*. pp. 49-58, London, UK, 2014. [19% acceptance rate]

[C25]
§
Cameron Tolooee*****, Matthew Malensek*****, and **Sangmi Lee Pallickara**. A Framework for Managing Continuous Query Evaluations over Voluminous, Multidimensional Datasets. Proceedings of *the IEEE Cloud and Autonomic Computing Conference*. London, UK. 2014.

[C24]
§
Matthew Malensek*****, **Sangmi Lee Pallickara** and Shrideep Pallickara. Polygon-Based Query Evaluation over Geospatial Data Using Distributed Hash Tables. Proceedings of the *IEEE/ACM* Conference on Utility and Cloud Computing. pp. 219-226. Dresden, Germany. 2013. [24% acceptance rate]

[C23]
§
Matthew Malensek*****, **Sangmi Lee Pallickara** and Shrideep Pallickara. "Autonomously Improving Query Evaluations over Multidimensional Data in Distributed Hash Tables." Proceedings of *the ACM Cloud and Autonomic Computing Conference*. Miami, USA. 2013. [35% acceptance rate]

[C22]
Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Expressive Query Support for Multidimensional Data in Distributed Hash Tables. Proceedings of *the IEEE/ACM Conference on Utility and Cloud Computing*. pp. 31-38. Chicago, USA. 2012.   [27% acceptance rate]
** Best Paper Award

[C21]
Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. "Galileo: A Framework for Distributed Storage of High-Throughput Data Streams," Proceedings of the *IEEE/ACM Conference On Utility and Cloud Computing*. pp. 17-24. Melbourne, Australia. 2011. [26.7% acceptance rate]

[C20]
§
Sangmi Lee Pallickara, Matthew Malensek***** and Shrideep Pallickara. "Enabling Access to Time-Series, Geospatial Data for On Demand Visualization," Proceedings of the *IEEE Symposium on Large-Scale Data Analysis and Visualization*, Providence, Rhode Island. 2011.

[C19]
§
Sangmi Lee Pallickara, Shrideep Pallickara, Milija Zupanski, and Stephen Sullivan, "Efficient Metadata Generation to Enable Interactive Data Discovery over Large-scale Scientific Data Collections," Proceedings of the IEEE International Conference on Cloud Computing Technology and Science, Indianapolis. November 2010 [24.9% acceptance rate]

[C18]
Sangmi Lee Pallickara, Marlon Pierce, Chin Hua Kong, and Qunfeng Dong, "Enabling Large Scale Scientific Computations for Expressed Sequence Tag Sequencing over Grid and Cloud Computing Clusters," Proceedings of the PPAM 2009 Eighth International Conference on Parallel Processing and Applied Mathematics, Wroclaw, Poland, September 13 – 16, 2009

[C17] **Sangmi Lee Pallickara** and Marlon Pierce, "SWARM: Scheduling Large-scale Jobs over the Loosely-Coupled HPC Clusters," Proceedings of the IEEE International Conference on e-Science. Indianapolis, 2008, December 7-12, 2008. [29.1% acceptance rate]

[C16] Marlon Pierce, **Sangmi Lee Pallickara**, Yu Ma, Mike Lowe, Qunfeng Dong and Samy Meroueh, "Indiana University TeraGrid Gateway Support," Proceedings of The *International Conference for High Performance Computing, Networking, Storage and Analysis* (collocated with *SuperComputing* '08), November, 2008.

[C15] Yogesh Simmhan, **Sangmi Lee Pallickara**, Nithya N. Vijayakumar, and Beth Plale, "Data Management in Dynamic Environment Driven Computational Science," Proceedings of The *IFIP International Federation for Information Processing*, Vol. 239, Grid-Based Problem Solving Environments, P.W. Gaffney and J.C.T. Pool. Eds. Spring Boston. pp.317-333, 2007

[C14] **Sangmi Lee Pallickara**, Beth Plale, Liang Fang, Dennis Gannon, "End-to-End Trustworthy Data Access in Data-Oriented Scientific Computing," Proceedings of The *IEEE Cluster Computing and Grid (CCGRID)* 2006: 395-400. [23.7% acceptance rate]

[C13] Yiming Sun, Scott Jensen, **Sangmi Lee Pallickara**, and Beth Plale, "Personal Workspace for Large-scale Data-driven Computational Experimentation," Proceedings of *the 7th IEEE/ACM International Conference on Grid Computing (Grid'06)*, Barcelona. 2006. [20.2% acceptance rate]

[C12] Dennis Gannon, Beth Plale, Marcus Christie, Liang Fang, Yi Huang, Scott Jensen, Gopi Kandaswamy, Suresh Marru, **Sangmi Lee Pallickara**, Satoshi Shirasuna, Yogesh Simmhan, Alex Slominski, and Yiming. Sun, "Service Oriented Architectures for Science Gateways on Grid Systems," Proceedings of The *International Conference on Service Oriented Computing 2005*, B. Benatallah, F. Casati, P. Traverso (Eds.), LNCS 3826, pp. 21-32, 2005. Springer-Verlag Berlin Heidelberg 2005

[C11] **Sangmi Lee Pallickara**, Beth Plale, Scott Jensen, Yiming Sun, Short Paper: "Monitoring Access to Stateful Resources in Grid Environments," Proceedings of the *IEEE International Conference on Services Computing*, Orlando, Florida, pp. 343-346, 2005.

[C10] Shrideep Pallickara, Geoffrey Fox, Beytullah Yildiz, **Sangmi Lee Pallickara**, Sima Patel and Damodar Yemme, "On the Costs for Reliable Messaging in Web/Grid Service Environments," Proceedings of the *IEEE International Conference on e-Science & Grid Computing*, Melbourne, Australia. 2005. [31.6% acceptance rate]

[C9] Shrideep Pallickara, Geoffrey Fox, and **Sangmi Lee Pallickara**, "An Analysis of Reliable Delivery Specifications for Web Services," Proceedings of the *IEEE ITCC Conference on Information Technology*. Las Vegas. 2005.

[C8] **Sangmi Lee** and Geoffrey Fox, "Wireless Reliable Messaging Protocol for Web Services (WS -WRM)," Proceedings of the *IEEE 2nd International Conference on Web Services (ICWS 2004)*, pp.350-357, San Diego. 2004. [28.7% acceptance rate]

[C7] Geoffrey Fox, Sunghoon Ko, Kangseok Kim, **Sangmi Lee**, Sangyoon Oh, "Universally Accessible Collaboration Frameworks for Ubiquitous Computing Environments," Proceedings of the *International Conference on Ubiquitous Computing (ICUC2003)*, Seoul Korea, 2003.

[C6] **Sangmi Lee**, Sunghoon Ko, Geoffrey Fox, "Adapting Content for Mobile Devices in Heterogeneous Collaboration Environments," Proceedings of the *International Conference on Wireless Networks (ICWN'03)*, pp. 211-217, 2003

[C5] Geoffrey Fox, Hasan Bulut, Kangseok Kim, Sung-Hoon Ko, **Sangmi Lee**, Sangyoon Oh, Shrideep Pallickara, Xiaohong Qiu, Ahmet Uyar, Minjun Wang, Wenjun Wu "Collaborative Web Services and Peer-to-Peer Grids," Proceedings of the *IEEE Collaborative Technologies Symposium (CTS'03)*, 2003.

[C4]   **Sangmi Lee**, Sunghoon Ko, Geoffrey Fox, Kangseok Kim, Sangyoon Oh, "A Web Service Approach to Universal Accessibility in Collaboration Services," Proceedings of the *International Conference on Web Services (ICWS'03)*, pp. 333-339, USA. 2003. [29.8% acceptance rate]

[C3]   Geoffrey Fox, Sung-Hoon Ko, Kangseok Kim, Sangyoon Oh, **Sangmi Lee**, "Integration of Hand-Held Devices into Collaborative Environments," Proceedings of the *International Conference on Internet Computing (IC'02)* pp.231-250, 2002. USA.

[C2]   **Sangmi Lee**, Geoffrey Fox , Sunghoon Ko, Minjun Wang, Xiaohong Qiu ,"Ubiquitous Access for Collaborative Information System Using SVG," Proceedings of the *SVG OPEN Conference*. 2002, Zurich, Switzerland

[C1]   Hasan Bulut, Geoffrey Fox, Dennis Gannon, Kangseok Kim, Sung-Hoon Ko, **Sangmi Lee**, Sangyoon Oh, Xi Rao, Shrideep Pallickara, Quinlin Pei, Marlon Pierce, Aleksander Slominski, Ahmet Uyar, Wenjun Wu, Choonhan Youn, "An Architecture for e-Science and its Implications," Proceedings of *the IEEE International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*. July, 2002.

6.4 Refereed Workshop Proceedings

[W6]   Menuka Warushavithana*, Saptashwa Mitra*, Mazdak Arabi, Jay Breidt, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Containerization of Model Fitting Workloads over Spatial Datasets. Big Spatial Data at the IEEE Big Data Conference. pp 3770-3779. 2021.

[W5]   **Sangmi Lee Pallickara** and Beth Plale, "Enabling End-to-End Trustworthiness in Data-Oriented Scientific Computing," Proceedings of *the Workshop on Web Services-based Grid Applications (WGSA'06)* in association with International Conference on Parallel Processing (ICPP-06). 2006.

[W4]   Scott Jensen, Beth Plale, Sangmi Lee Pallickara, and Yiming Sun, "A Hybrid XML-Relational Grid Metadata Catalog," Proceedings of the Workshop on Web Services-based Grid Applications (WGSA'06) in association with International Conference on Parallel Processing (ICPP-06). 2006.

[W3]   **Sangmi Lee Pallickara**, Beth Plale, Scott Jensen, and Yiming Sun, "Structure, sharing, and preservation of scientific experiment data," Proceedings of the *IEEE 3rd International Workshop on Challenges of Large Applications in Distributed Environments (CLADE)*, pp 107-114, 2005.

[W2]   Sangyoon Oh, **Sangmi Lee Pallickara**, Sunghoon Ko, Jai-Hoon Kim, Geoffrey Fox, "Cost Model and Adaptive Scheme for Publish/Subscribe Systems on Mobile Environments," Proceedings of the *International Workshop on Active and Programmable Grids Architectures and Components (APGAC05), Lecture Notes in Computer Science, Springer-Verlag* 3516, pp. 275-278, Atlanta, USA, 2005.

[W1]   Sangyoon Oh, **Sangmi Lee Pallickara**, Sunghoon Ko, Jai-Hoon Kim, Geoffrey Fox, Publish/Subscribe Systems on Node and Link Error Prone Mobile Environments," Proceedings of the *Wireless and Mobile Systems Workshop (ICCS 2005), Lecture Notes in Computer Science, Springer-Verlag*, 3515, pp. 576-584

6.5 Ph.D. Dissertation

Sangmi Lee. August 2003.
Title: A Modular Data Pipelining Architecture for Enabling Universal Accessibility in Peer to Peer Grids.
Advisors: Prof. Geoffrey Fox and Prof. Gordon Erlebacher

6.6 Technical Reports and Posters (Not Peer-Reviewed)

[T13] Paahuni Khandelwal*, **Sangmi Lee Pallickara**, Shrideep Pallickara, "A Physics-Guided Deep Learning Approach to Generate Daily 30m Soil Moisture Maps at the CONUS Scale", Graduate Research Showcase, Colorado State University, November 2023, The CNS Outstanding Scholar Award

[T12] Junhwan Kim*, Hermela Darebo*, Jackson Holden*, Tarun Sai Pamulapati*, Kunal Agarwal*, **Sangmi Lee Pallickara**, "Are Parking Lots Cooking Fort Collins? - Correlation Analysis of Parking Lots and Heat Island Effects using Satellite Imagery and Open Street Maps." Research Poster at Celebrating Undergraduate Research and Creativity (CURC), Colorado State University, April 2023, 2nd place in the Climate Change Research Track

[T11] Abdul Matin*, Sangmi Lee Pallickara, "Rapid Betweenness Centrality Estimates for Transportation Networks using Capsule Networks ", Graduate Research Showcase, Colorado State University, November 2022

[T10] Emma Hamilton*, Mandey Brown*, Meridith McCann*, Saptashwa Mitra*, **Sangmi Lee Pallickara,** "Recommendation Guided Immersive Visual Explorations using Random Forests", Research Poster at Celebrating Undergraduate Research and Creativity (CURC), Colorado State University, April 2022, 2nd place in the Data Science Research Track

[T9] Poornima Gunhalkar*, **Sangmi Lee Pallickara,** "Benchmarking of distributed data processing frameworks for large-scale real time image classification", Graduate Research Showcase, Colorado State University, November 2021

[T8] Laksheen Mendis*, **Sangmi Lee Pallickara**, "Embedding based Clustering of the Time Series Data", Graduate Research Showcase, Colorado State University, November 2021

[T7] Kevin Bruhwiler*, Philip Sharpp*, Nick Czarnecki*, Jim Xu*, Fawad Ahmed*, Saptashwa Mitra*, **Sangmi Lee Pallickara,** "Immersive Analytics for Traffic Analysis using Machine Learning Techniques", Research Poster at Celebrating Undergraduate Research and Creativity (CURC), Colorado State University, April 2018

[T6] Duck Keun Yang*, Zach Cutler*, Jared Koontz*, **Sangmi Lee Pallickara**, Adam Gaylord, Joe von Fischer, "Enabling Active Data Collection and Dissemination of Methane Concentrations," Research Poster at Celebrating Undergraduate Research and Creativity (CURC), Colorado State University, April 2015

[T5] Kong, C.H., **Sangmi Lee Pallickara**, and Marlon Pierce, "Fault Detection of TeraGrid Resources Using Inca," June, 2009. Poster at *TeraGrid* '09

[T4] **Sangmi Lee Pallickara**, Marlon Pierce, "Orienting Scientific Data Management to Harness the Data Cloud," Digital Science Center, Indiana University, Dec, 2008, Technical Report.

[T3] Geoffrey Fox, **Sangmi Lee**, Sung-Hoon Ko, Kangseok Kim, Sangyoon Oh, "CAROUSEL: Universally Accessible Web Service Architecture for Collaborative Applications," Community Grids Lab, Indiana University, 2002

[T2] Geoffrey Fox, Sung-Hoon Ko, Kangseok Kim, **Sangmi Lee**, Sangyoon Oh, "Status of Hand-Held Interfaces to the Garnet Collaborative Environment," Community Grids Laboratory, Indiana University, 2002

[T1]   **Sangmi Lee**, "Overview of the Virtual Collaborative Network (VNC)," Department of the Computer Science, Florida State University, 2000

7   Research Projects

Aug 23 –   HYDRO

We all depend on agriculture for sustenance. When compared to seafood and livestock, cropping systems provide the primary source of nutrition. Yields and productivity of cropping systems must grow to meet the demands of a growing population. Once seeds are available, a successful cropping season is determined by water. There are two sources for this: irrigation and precipitation. Irrigation water is a major input to agriculture, especially in semi-arid and arid regions. In a recent appraisal for the Soil and Water Resources Conservation Act, the USDA identified irrigation water conservation as a national need. Under-watering induces stresses and adversely impacts both crop growth and yields. Over-watering, on the other hand, leads to nutrient runoff, soil erosion, and water waste. Farms are also impacted by the adverse effects of droughts, variability in precipitation, and lengthening of the growing season. The proposed effort with its emphasis on water management and conservation represents an adaptation to the head winds often encountered at farms. The effort addresses the interrelated aspects of over-watering (soil erosion and nutrient runoff) and underwatering (adverse crop yields and stress) while ensuring sustainability and profitability of agricultural systems.

The overarching objective of this project is to develop an end-to-end cyber-physical intelligence system that forecasts space-time crop water needs in a given field and implements variable rate irrigation strategies to optimize crop yield throughout the field. We instrument the field with a limited number of in-situ soil moisture content sensors; these in situ observations are complemented with remotely sensed data from radars and satellites. The effort includes design of novel AI (Artificial Intelligence) methods based on deep neural networks (DNN) to generate forecasts of water needs. These DNNs operate on multimodal, high-dimensional data to identify soil moisture deficits and variability in different parts of the field. The generated forecasts account for crop, soil type, precipitation events, and the crop growing phase. The project closes the loop between the sensing environment and actuation within the AI-guided cyber physical system. These projections are leveraged within a game theory based algorithm to inform precise actuations of the watering arm with prescription plans that control watering rates at the nozzle and zone level. The algorithm is adaptive and responsive to precipitation events, uncertainty in the forecasts, and the actuation overheads. This multifaceted research advances the science of cyber-physical systems by innovatively combining sensing environments, algorithmic game theory, scientific models and domain-science, and AI/DNNs.

Aug 20 –   FINEET
Jul 23   Accurate, high-resolution evapotranspiration (ET) estimates are critical for water resource management, particularly for sustainable agriculture. However, generating such ET estimates that inform daily decision-making at farms is challenging due to spatial variability in process- based ET model performance and the lack of high precision observations. This project harnesses publicly available remote sensing data, process-based ET models developed by the scientific community, and machine-learning algorithms to generate accurate, high-resolution ET estimates. This effort integrates multi-scale satellite observations and other earth observations to enhance the temporal and spatial resolutions of observational data using deep learning. Our project, refer to as FineET, combines process-based models and empirical models (based on deep learning and random forests) to generate a heterogeneous ensemble that is customized to the particular spatial region to produce accurate ET estimates. FineET also enables effective CONUS-scale visualizations of ET estimates allowing users to explore contributing features (or independent

variables) and estimates from diverse models. The proposed data integration, analytics, and forecasting infrastructure will enable end-to-end analysis of farm-level ET data from diagnostic to predictive analytics. Our long-term goal is to enhance this framework to support other remote sensing platform while also accounting for diverse environmental/ecological parameters that will lead to effective decision-making and extend the model to global scales. We combine expertise in computer science, agriculture, ecology, and statistics to create cyberinfrastructure tools that democratize access to cutting-edge developments in machine learning and big data to stakeholders in the agricultural community.
http://fineet.cs.colostate.edu

Oct 19 –   Sustain
The United States is highly urbanized with more than 80% of the population residing in cities. Cities draw from and impact natural resources and ecosystems while utilizing vast, expensive infrastructures to meet economic, social, and environmental needs. The NSF has invested in several strategic research efforts in the area of urban sustainability, all of which generate, collect, and manage large volumes of spatiotemporal data. Voluminous datasets are also made available by governmental agencies and NGOs in domains such as climate, ecology, health, and census. These data can spur exploration of new questions and hypotheses, particularly across traditionally disparate disciplines, and offer unprecedented opportunities for discovery and innovation. However, the data are encoded in diverse formats and managed using a multiplicity of data management frameworks all contributing to a Balkanization of the observational space that inhibits discovery. A scientist must reconcile not only the encoding and storage frameworks, but also negotiate authorizations to access the data. A consequence is that data are locked in institutional silos, each of which represents only a sliver of the observational space. This project, SUSTAIN (Software for Urban Sustainability to Tailor Analyses over Interconnected Networks), facilitates and accelerates discovery by significantly alleviating data-induced inefficiencies. This effort has deep, far-reaching impact. It transforms urban sustainability science by establishing a community of interdisciplinary researchers and catalyzing their collaborative capacity. SUSTAIN accomplishes these interconnected goals by enabling holistic visibility of the observational space, interactive visualizations of multidimensional information spaces using overlays, fast evaluation of expressive queries tailored to the needs of the discovery process, generation of custom exploratory datasets, and interoperation with diverse analyses software frameworks -- all leading to better science.
URL: http://urban-sustain.org

Dec 16 –   Subterra
Nov 19   America's vast terrestrial resources (over 520 million hectares of crop, range and forestland) are strategic assets essential for sustainable economic growth. While advances in technology have resulted in a ten-fold increase in crop productivity over the past hundred years, soil quality has declined, incurring a soil carbon debt equivalent to 65 parts per million (ppm) of atmospheric carbon dioxide (CO_2). The soil carbon debt also increases the need for costly nitrogen fertilizer, which has become the primary source of nitrous oxide (N_2O) emissions, a greenhouse gas. The soil carbon debt also impacts crop water use, increasing susceptibility to drought stress, which threatens future productivity. Given the scale of domestic (and global) agriculture resources, there is tremendous potential to reverse these trends by harnessing the photosynthetic bridge between atmospheric carbon, plants, microbes and soil. Development of new root-focused plant cultivars could dramatically and economically reduce atmospheric CO_2 concentrations while improving productivity, resilience and sustainability. To this end, projects in the ARPA-E Rhizosphere Observations Optimizing Terrestrial Sequestration (ROOTS) program seek to develop advanced technologies and crop cultivars that enable a 50 percent increase in soil carbon accumulation while reducing N_2O emissions by 50 percent and increasing water productivity by 25 percent.
URL: https://www.subterra.org/resources

May 16 –   SYNAPSE
Apr 22

Over the past decade there has been an exponential growth in data volumes driven in part by data streams generated by observational equipment and computer programs. Given the data volumes, it can be difficult to harness the data to understand phenomena and/or to make forecasts. Fitting models to the observational data is one way to accomplish this. A precursor to building such models is extracting features that track attribute statistics and changes over different inspection windows (segmentation of the time-series data streams). Models constructed using such features can then be used to predict what the outcome will be and when it is likely to happen. We consider models for both regression and classification. Challenges in ad hoc model creation stem from the combinatorically explosive number of ways in which models can be realized.

The research objective of this effort is to design an enabling infrastructure to support generation, assessment, and refinement of ad hoc models from voluminous, multidimensional, time-series observational data at scale. The proposed framework, Synapse, will be designed to support and simplify the naturally iterative and interactive model building process over voluminous streaming datasets. Modelers only need to specify a basic set of bootstrap parameters; the framework manages the complexities relating to: (1) how streams are dispersed, (2) how data accesses are managed, (3) coping with I/O and memory contentions, and (4) dispersion of training workloads and interference avoidance. The framework will include visualizations that allow modelers to quickly and effectively assess the quality of a multiplicity of models each possibly covering a different portion of the input feature space and to use these assessments to guide decisions about updates or replacements to models. The proposed approach will scale with increases in data volumes, the number of available data streams, model training workloads, and live model evaluations.

Nov 14 – GLEAN

July 19 The Glean project focuses on performing analytics at scale over Big Data. The datasets we consider are in the order of Petabytes and encompass billions of files representing trillions of observations, measurements, or simulation datapoints. Glean achieves this by combing innovations in large-scale storage systems, cloud computing, machine learning, and statistics. A particular focus of this effort is to perform analytics in real-time over streaming data representing time-series observations. URL: https://glean.cs.colostate.edu

Nov10 – GALILEO

Jan 19 This geospatial data storage system is designed to provide efficient access to time-varying geospatial datasets generated by networked observational instruments such as radars and satellites. The storage system is designed to enable large-scale visualizations and processing of geospatial datasets. URL: http://galileo.cs.colostate.edu

Jan 13 – GEOLENS

Aug 15 GEOLENS is an interactive visual analytics framework that supports fast and expressive interactions with voluminous geo-spatial datasets. It incorporates support for advanced visual analytics techniques such as brushing and linking.

Apr 10 – ADDS (ATMOSPHERIC DATA DISCOVERY NETWORK SERVICE)

Dec 12 This project investigates methods to efficiently discover data that are part of large, binary collections published by major data hosting services. ADDS crawls such servers and builds metadata collections that can be queried to access relevant portions of the dataset. The service supports the BUFR binary data format for observational datasets and the netCDF format that is used to encode simulation output data. This research is based on a collaborative effort with the Cooperative Institute for Research in Atmosphere (CIRA) at Colorado State University and the University Corporation for Atmospheric Research (UCAR) at Boulder.

Aug 07 –
Dec 09

TERAGRID project

This open scientific discovery infrastructure combines resources at eleven partner sites to create an integrated, persistent computational resource. Using high-performance network connections, the TERAGRID integrates high-performance computers, data resources and tools.

As part of my work on this project, I designed the SWARM meta scheduling system. SWARM leverages Condor and Globus technologies to harness traditional high throughput computing clusters. It is not unusual for these high-throughput jobs to run for several days at a time. SWARM is presently deployed in a gene sequencing application developed by the *Center for Genomics and Bioinformatics* at Indiana University.

Aug 07 –
Dec 09

POLAR GRID project

The NSF-funded POLAR GRID project aims to understand the impact of rising sea levels and how they relate to global climate change. As part of this effort I worked on building science gateways that support data collection, examination and analysis for ice sheet research that help scientists. My research within the project included issues relating to metadata and storage.

Sep 04 –
Jul 07

LEAD (Linked Environments for Atmospheric Discovery) project:

The NSF-funded LEAD project makes meteorological data, forecast models, and analysis and visualization tools available to anyone who wants to interactively explore the weather as it evolves http://lead.ou.edu. My research on this project has included work on data cataloging, transferring and storing schemes based on Service Oriented Architecture (SOA) principles. These issues were investigated within the context of the MYLEAD system, which I designed.

Sep 01 –
Aug 04

CAROUSEL

The CAROUSEL project focused on developing an environment for supporting ubiquitous accesses to real-time collaborative applications in Grid settings. Devices that were supported include portable devices, such as 3G SmartPhones and 801.11b equipped PDAs and conventional desktop PCs. As part of this project, I designed a data pipelining architecture that was formally verified using Perti Nets. I also developed a protocol for reliable communications between these pervasive devices in wireless settings.

8 Teaching

Department of Computer Science, Colorado State University
* Offered for the first time

Term	Course	Enrollment	Audience
Spring 2024	CS535: Big Data (with an online section)	25	Graduate students and AMP[1] students
Fall 2023	CS435: Introduction to Big Data (with two online sections)	90	Senior level undergraduate students and graduate students
Spring 2023	CS535: Big Data (with an online section)	35	Graduate students and AMP[2] students
Spring 2023	CS481-A5 Data Mining at Scale *	9	Senior level undergraduate students and graduate students
Fall 2022	CS435: Introduction to Big Data (with an online section)	75	Senior level undergraduate students and graduate students
Spring 2022	CS535: Big Data (with an online section)	35	Graduate Course
Fall 2021	CS435: Introduction to Big Data (with an online section)	85	Senior level undergraduate students and graduate students
Spring 2021	CS535: Big Data (with an online section)	35	Graduate Course
Fall 2020	CS435: Introduction to Big Data (with an online section)	58	Senior level undergraduate students and graduate students
Spring 2020	CS535: *Big Data (with an online section)*	50	Graduate Course
Fall 2019	CS435: Introduction to Big Data	60	Senior level undergraduate students and graduate students
Spring 2019	CS535: *Big Data*	26	Graduate Course
Fall 2018	CS435: Introduction to Big Data	55	Senior level undergraduate students and graduate students
Spring 2018	CS435: Introduction to Big Data	57	Senior level undergraduate students and graduate students
Fall 2017	CS535: *Big Data*	20	Graduate Course
Spring 2017	CS435: *Introduction to Big Data*	58	Senior level undergraduate students and graduate students
Fall 2016	CS535: *Big Data*	20	Graduate Course
Spring 2016	CS435: *Introduction to Big Data*	60	Senior level undergraduate students and graduate students
Fall 2015	CS535: *Big Data*	22	Graduate course
Spring 2015	CS480A: *Introduction to Big Data* *	39	Senior level undergraduate students and graduate students
Fall 2014	CS581/CS535: *Big Data*	25	Graduate course
Fall 2013	CS581: *Big Data* *	22	Graduate course

---

[1] Accelerated Master's Program
[2] Accelerated Master's Program

Spring 2013	CS200: *Algorithms and Data Structures*	79	Core undergraduate course
Spring 2013	CS480: *Principles of Data Management*	18	Senior level undergraduate students and graduate students
Fall 2012	CS200: *Algorithms and Data Structures*	100	Core undergraduate course
Spring 2012	CS480: *Principles of Data Management**	17	Senior undergraduate students
Spring 2012	CS200: *Algorithms and Data Structures*	57	Core undergraduate course
Fall 2011	CS200: *Algorithms and Data Structures*	72	Core undergraduate course

* Offered for the first time

Department of Computer Science, Indiana University

Co-Instructor	*Fundamentals of Computer Networks*, Spring 2006. (Undergraduate course)	
Instructor	Indiana University Science Summer Camp, Summer 2005. (High School level)	
Guest Instructor	*Distributed Systems*, Spring 2005. (Graduate course)	
Guest Instructor	*Fundamentals of Computer Networks*, Spring 2005. (Undergraduate course)	
Supervision of Graduate Students	Sep 08 –Dec 09	Chin Hua Kong: Fault detection in Grid Computing Environments
	Jul 05–Aug 06	Suba Periyasami: MVC based Portal interfaces for Web-based data sharing.
	Sep 05–Aug 06	Ning Liu: Trustworthy scientific data management
NSF-REU Program	Aug 06- Jul 07	Supervised 2 undergraduate students (Christina Hoffa and You-Wei Chaeh) as part of the NSF's REU program,

9   Educational Outreach Activities

Outreach

Founding Director	2016 – SWiFT STEM Education Outreach for Women High School Students, Fort Collins, Colorado (In-person summer camp canceled in 2020/21)
Faculty Mentor	2021 CSU Computer Science Summer Coding Camp https://www.cs.colostate.edu/codingcamp/#
Committee Member, Judge	2019 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University
	2018 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University
	2017 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University
	2015 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University
	2014 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University

	2006 Indiana Women in Computing (INWIC)
	2005 Women in Science Program (WISP) 7th Annual Research Day

Co-organizer, RAM Welcome	2012 Computer Science, Colorado State University.
Session Instructor	2005 ACM JETT (Java Engagement for Teacher Training), Indiana University
Assistant Session Instructor	2004 ACM JETT (Java Engagement for Teacher Training), Indiana University

---

SWiFT Education Outreach for Female High School Students [2016 -]
http://swift.cs.colostate.edu
Role: Creator and Director

Since 2016, I have been organizing a weeklong summer educational camp for about 12-14 rising 11th grade female students from High Schools in Fort Collins, Colorado. *SWiFT* (*Summer STEM Camp for Women in Fort Collins*) has been held in the Computer Science Department at Colorado State University. SWiFT is sponsored by the National Science Foundation, the Computer Science Department at CSU, and the CNS Jack Cochran Family Professorship.

The camp provides activity-based learning and hands-on experiences to demonstrate how mathematical concepts that students learn in school are applied to solve real-world problems using data science. The key objective of SWiFT is encouraging female High School students to explore careers in computer science by exposing them to a holistic environment of encompassing programming, team interactions, access to real-world examples, and applicability to mathematics and other sciences and engineering.

Distinctive approaches within SWiFT include the following aspects: (1) **Carefully selected camp cohorts**. Students are primarily recruited based on recommendations from their high school science and math teachers. We are targeting female students who are demonstrating good progress in math and/or sciences but have not programmed before. This provides a more supportive and focused learning environment to the camp attendees. (2) **Two-way review process of the camp curriculum.** SWiFT maintains 3-4 undergraduate instructors, as well as 2-3 graduate student mentors. In the Spring semester instructors and mentors begin to work on analyzing and constructing lesson plans. We review the curriculum with a science teacher in one of the High Schools to ensure coherence and alignment with the High School curriculum and to calibrate the curriculum's level of difficulty. (3) **Diverse set of activities.** Programming is an important part of computer science, but there is a lot more to immerse students in computer science. We emphasize team building activities, mentoring relationships, and exploring on-going research projects in computer science. Students also enjoy yoga, dance, and playing games between sessions. The college experience is reinforced with campus tours, research lab visits, and guest lectures by targeted faculty.

Girls Who Code
https://www.cs.colostate.edu/codingcamp/#/
Role: Faculty mentor

Girls Who Code is a student-led outreach activity in Computer Science Department. I was a faculty mentor for the summer camp held in Summer 2021 (June 14 – June 25). We had more than 20 students from High schools and Middle schools. As faculty mentor, my role was guiding Girls Who Code team to develop camp curriculum and logistics particularly for the online environment.

10 Academic Advising

Current Graduate Advisees

Paahuni Khandelwal, Ph.D.
Abdul Matin, Ph.D.
Menuka Warushavithana, Ph.D. [Co-advised with Shrideep Pallickara]

Tarun Sai Pamulapati, M.S.
Srivarshini Ksheerasagar, M.S.
Freddy Larrieu, MS [Co-advised with Shrideep Pallickara]

Current Undergraduate Research Advisee
Junhwan Kim
Hermela Darebo

Supervised Undergraduate Honors Thesis
Zachary Fuller (May 2023)

Erin Doan (May 2018)
Currently: Software Engineer, Lockheed Martin

Sadie Henry, B.S. (May 2017)
Currently: Project Manager, Microsoft Research

Supervised Graduate Degrees

Saptashwa Mitra
Degree Type: Ph.D.
Dissertation Title: Distributed In-memory Framework for Interactive Analytics over Voluminous Spatiotemporal Data
Date Completed: June 2023

Rushabh Shah
Degree Type: Master of Science, Plan B project option
Project Title: Integrated Training Data Generation using Image Augmentation
Date completed: May 2023

Caleb Carlson   [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Master of Science
Thesis Title: A Framework for Resource Efficient Profiling of Spatial Model Performance

Date Completed: October 2022

Abdul Matin
Degree Type: Master of Science
Thesis Title: Towards Interactive Betweenness Centrality Estimation for Transportation Network Using Capsule Network
Date Completed: October 2022

Laksheen Mendis
Degree Type: Master of Science
Thesis Title: Embedding based Clustering of Time Series Data Using Dynamic Time Warping
Date Completed: December 2021 (Graduated in Spring 2022)

Poornima Gunhalkar
Degree Type: Master of Science, Project option
Project Title: Benchmarking Distributed Stream Processing Frameworks For Real-Time Image Classification
Date Completed: October 2021

Menuka Warushavithana   [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Master of Science
Dissertation Title: Containerization of Model Fitting Workloads over Spatial Datasets
Date Completed: October 2021

Daniel Rammer   [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Doctor of Philosophy
Dissertation Title: Harnessing Spatiotemporal Data Characteristics to Facilitate Large-Scale Analytics Over Voluminous, High-Dimensional, Observational Datasets
Date Completed: August 2021
Currently: Union.AI

Budgaga Walid   [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Doctor of Philosophy
Dissertation Title: Leveraging Ensembles: Balancing Timeliness and Accuracy for Models Training over Voluminous Datasets
Date Completed: May 2020

Matthew Malensek   [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Doctor of Philosophy
Dissertation Title: Low-latency, Query-Driven Analytics over Voluminous Multidimensional, Spatiotemporal Datasets
Date Completed: August 2017
Currently: Assistant Professor, Computer Science – University of San Francisco

Sanket Mehrotra
Degree Type: Master of Science, Project option
Project Title: Runtime Speculation-based Model for Data Retrieval Guidelines for interactive Analytics

Date Completed: Degree requirements completed, to graduate in August 2021

Ryan Becwar
Degree Type: Master of Science, Project option
Project Title: Accelerated Georeferencing with Apache Spark
Date Completed: May 2021

Kevin Bruhwiler   [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Master of Science
Thesis Title: Aperture: A System for Interactive Visualization of Voluminous Geospatial Data
Date Completed: December 2020
Currently: Ph.D. Student, University of Southern California

Kartik Khurana
Degree Type: Master of Science
Thesis Title: Enabling Autoscaling for In-Memory Storage In Cluster Computing Framework
Date Completed: May 2020
Currently: Software Engineer, Oracle

Aaron Pereira
Degree Type: Master of Science
Thesis Title: Towards Federated Learning over Large-scale Streaming Data
Date Completed: May 2020
Currently: Software Engineer, Azure, Microsoft

Bibek Raj Shrestha
Degree Type: Master of Science
Thesis Title: Enabling Autoscaling for In-Memory Storage In Cluster Computing Framework
Date Completed: May 2019
Currently: Software Engineer, Rally Health

Maxwell Roselius
Degree Type: Master of Science
Thesis Title: Toward effective high-throughput georeferencing over voluminous observational data in the domain of precision agriculture
Date Completed: December 2018
Currently: Software Engineer, Spectra Logic

Yu (Peter) Qui
Degree Type: Master of Science
Thesis Title: Questionnaire integration system based on question classification and short text semantic textual similarity
Date Completed: December 2018
Currently: Software Engineer, PayPal

Saptashwa Mitra
Degree Type: Master of Science
Thesis Title: Adaptive spatiotemporal data integration using distributed query relaxation over heterogeneous observational datasets
Date Completed: August 2018
Currently: Ph.D. student, Colorado State University

Chuanqi Huang

Degree Type: Master of Science
Thesis Title: Leveraging Structural-Context Similarity of Wikipedia Links to Predict Twitter User Locations
Date Completed: December 2017

Naman Rajivkumar Shah
Degree Type: Master of Science
Thesis Title: Determining Disease Outbreak Influence from Voluminous Epidemiology Data on Enhanced Distributed Graph-Parallel System
Date Completed: June 2017
Currently: Software engineer, Illumon

Johnson Charles Kachikaran Arulswamy
Degree Type: Master of Science
Thesis Title: A Locality-Aware Scientific Workflow Engine for Fast-Evolving Spatiotemporal Sensor Data
Date Completed: March 2017
Currently: Software engineer, Cisco

Harshil Shah   [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Master of Science
Thesis Title: Identification and Characterization of Super-spreaders from Voluminous Epidemiology Data
Date Completed: Dec 2016
Currently: Software engineer, Conviva

Cameron Tolooee
Degree Type: Master of Science
Thesis Title: On the Use of Locality Aware Distributed Hash Tables for Homology Searches Over Voluminous Biological Sequence Data
Date Completed: Dec. 2015
Currently: Hewlett Packard

Jared Koontz
Degree Type: Master of Science
Thesis Title: GeoLens: Enabling Interactive Visual Analytics Over Large Scale, Multidimensional Data Sets
Date Completed: May 2015
Currently: Software engineer, Intel

Walid Budgaga [Co-advised with Shrideep Pallickara, Primary Advisor: Shrideep Pallickara]
Degree Type: Master of Science
Thesis Title: A Framework for Real-time, Autonomous Anomaly Detection over Voluminous Time-Series Geospatial Data Streams
Date Completed: May 2014
Currently: Colorado State University

Ph.D. Dissertation Committee Memberships
Sachini Weerawardhana (December, 2021)
Thilina Buddhika (May, 2020)
Chengyu Fan (May, 2020)
Majdi Alnnfiai (May, 2019)
Ryan Stern (December, 2018)

Ph.D. Dissertation Committee External Committee Memberships

Statistics: Chien-Min Huang (current)
Physics: Matthew Judah (2019)
Electrical and Computer Engineering: Tushar Ganguli

Masters Thesis Committee Memberships
Adrian Esparza (2020)
Sitakanta Mishra (2018)
Darshan Wamshiker (2016)
Leo Vigneshwaran Sudalaikkan (2016)
Hanishsa Koneru (2015)
Amila Suriarachchi (2015)
Thilina Buddhika (2015)

Undergraduate Honors Theses Committee
Cassidy Skorczewski (2019)
Amanda Carbonari (2015)
Danielle Alexander (2013)

Masters Thesis External Committee Memberships
Electrical and Computer Engineering: Shibayan Chatterjee (2016)
Physics: Matthew Judah (May 2016)

Masters Project External Committee Memberships
Civil Engineering: Patrick Noe (2021)
Electrical and Computer Engineering: Marvin Antony Devadass (2015)
Electrical and Computer Engineering: Tushar Jagtap (Nov. 2014)

Past Undergraduate Research Advisees
Jackson Holden (Jan 2023 – May 2023)
Emma Hamilton (Jan 2022 – May 2022)
Meridith McCann (Jan 2022 – May 2022)
Mandey Brown (Jan 2022 – May 2022)
Kassidy Baram (Sept 2021 – April 2022): NSF REU
Thomas Lujan (Sept 2020 – May 2021): NSF REU
Ellie Martinez[3] (July 2021 – Jan 2021): NSF REU
Caleb Carlson (Jan 2020 ~ Aug 2020)
Nick Czarnecki (Jan 2018 – May 2018)
Jim Xu (Jan 2018 – May 2018)
Kevin Bruhwiler (Jan 2018 – May 2018)
Maxwell Roselius (July 2016 ~ Dec 2016)
Xuehao Hu (Jan 2015 – May 2016)
Duck Keun Yang (Jan 2015 – May 2016)
Zachary Cutler (Dec. 2014)
Cameron Tolooee (Dec 2013)
Li (Mike) Yang (Dec 2013)

---

[3] Ellie Martinez was one of the participants of the SWiFT summer camp in 2016.

11 Professional and Academic Service

Associate Editor
IEEE Transactions of Parallel and Distributed Systems, 2019 – present

Editor
Elsevier Journal of the Future Generation Computer Systems, 2021– present

Member of Editorial Board
Journal of Big Data. Publisher: Springer New York/Heidelberg, Germany. 2013 – 2023

Technical Program Committees: Conferences, Symposia, and Workshops

Track Chair, IEEE CCGrid, 2024

Member, Technical Program Committee: ACM The Platform for Advanced Scientific Computing (PASC), 2024

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), 2023

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications (BDCAT), 2023

Member, Technical Program Committee: IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2023

Member, Technical Program Committee: IEEE International Conference on Cloud Computing (CLOUD), 2023

Member, Technical Program Committee: IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2022

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications (BDCAT), 2022

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), 2022

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications, 2021

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), 2021

Member, Technical Program Committee: ACM Symposium on High Performance Parallel and Distributed Computing (ACM HPDC). 2021

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications, 2020

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), 2020

Member, Technical Program Committee: ACM Symposium on High Performance Parallel and Distributed Computing (ACM HPDC). 2020

Member, Technical Program Committee: IEEE International Parallel and Distributed Processing Symposium, New Orleans, Louisiana USA. 2020

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), Auckland, New Zealand, 2019

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications, Auckland, New Zealand, 2019

Member, Technical Program Committee: The 4th Joint International Workshop on Parallel Data Storage and Data Intensive Scalable Computing Systems (PDSW-DISCS), Held in conjunction with SC 19, the International Conference for High Performance Computing, Networking, Storage and Analysis, Denver, CO, USA 2019

Member, Technical Program Committee: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2019

Member, Technical Program Committee: IEEE International Parallel and Distributed Processing Symposium, Rio de Janeiro, Brazil, 2019

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), Zurich, Switzerland, 2018

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications, Zurich, Switzerland, 2018

Member, Technical Program Committee: The 3rd Joint International Workshop on Parallel Data Storage and Data Intensive Scalable Computing Systems (PDSW-DISCS), Held in conjunction with SC 18, the International Conference for High Performance Computing, Networking, Storage and Analysis, Dallas, TX, USA 2018

Member, Technical Program Committee: ACM Symposium on High Performance Parallel and Distributed Computing (ACM HPDC). Tempe, AZ, USA. 2018.

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications, Austin, TX, USA, 2017

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), Austin, TX, USA 2017

Member, Technical Program Committee: The 2nd Joint International Workshop on Parallel Data Storage and Data Intensive Scalable Computing Systems (PDSW-DISCS), Held in conjunction with SC 17, the International Conference for High Performance Computing, Networking, Storage and Analysis, Denver, CO, USA 2017

Member, Technical Program Committee: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2017.

Member, Technical Program Committee: ACM Symposium on High Performance Parallel and Distributed Computing

(HPDC). Washington D.C., USA. 2017.

Member, Technical Program Committee: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications, Shanghai, China, 2016.

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC), Shanghai, China, 2016.

Member, Technical Program Committee: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2016.

Member, Technical Program Committee: ACM Symposium on High Performance Parallel and Distributed Computing (HPDC). Kyoto, Japan. 2016.

Member, Technical Program Committee: ACM Symposium on High Performance Parallel and Distributed Computing (HPDC). Vancouver, Canada. 2015.

Member, Technical Program Committee: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2015.

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing (UCC). Cyprus. 2015

Member, Technical Program Committee: ACM Symposium on High Performance Parallel and Distributed Computing (HPDC). 2014.

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing. London, UK. 2014.

Member, Technical Program Committee: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2014.

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing. Dresden, Germany. 2013

Member, Technical Program Committee: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2013.

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing. Chicago, USA. 2012

Member, Technical Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing. Melbourne, Australia, 2011.

NSF Panel
2013, 2016, 2017, 2018, 2019, 2020, 2021

USDA Panel
2019

Journal Reviewer

I also review manuscripts for IEEE Transactions on Cloud Computing, IEEE Transactions on Big Data, IEEE Internet of Things Journal, IEEE Transactions on Service Computing, *Future Generation Computer Systems*, Journal of *Geographical Information Science*, International Journal of Geographical Information, Springer *Cluster Computing* and International Journal of *Distributed Sensor Networks*.

Departmental Service

Co-Chair, Graduate Program Committee, Computer Science Department (2018 – 2022)
Coordinator, Computer Science Colloquium, Computer Science Department (2017~2019)
Member, Graduate Program Committee, Computer Science Department (2014 – present)
Member, Operations Committee, Computer Science Department (2013 – 2014)