Team-Based Online Multidisciplinary Education on Big Data + High-Performance Computing + Atmospheric Sciences

Jianwu Wang[1], Matthias K. Gobbert[2], Zhibo Zhang[3], and Aryya Gangopadhyay[1]

[1] Dept. of Information Systems, University of Maryland, Baltimore County, Baltimore, MD 21250, USA, {jianwu, gangopad}@umbc.edu
[2] Dept. of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA, gobbert@umbc.edu
[3] Dept. of Physics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA, zhibo.zhang@umbc.edu

Abstract. Given the context of many institutions moving to online instruction due to the COVID-19 pandemic in 2020, we share our experiences of an online team-based multidisciplinary education program on big data + high performance computing (HPC) + atmospheric sciences (cybertraining.umbc.edu). This program focuses on how to apply big data and high-performance computing techniques to atmospheric sciences. The program uses both an instructional phase with lectures and team-based homework in all three areas and a multi-disciplinary research experience culminating in a technical report and oral presentation. The paper discusses how our online education program can achieve the same learning objectives as face-to-face instruction via pedagogy and communication methods including flipped classroom, online synchronous meetings and online asynchronous discussion forum.

Keywords: Online Education, Big Data, High-Performance Computing, Atmospheric Sciences, Multidisciplinary Education

1 Introduction

Next to theory and experimentation, computation has become the third pillar [1] and data-driven science has become the fourth pillar of the scientific discovery process [2] for many disciplines and critical to their research advances, such as bioinformatics, physics, computational chemistry, and mechanical engineering. It demands requirements on a course explaining how data and computation related techniques can help scientific discovery. Yet such a "Data + Computing + X" course is often missing in current curriculum design.

In 2017, the National Science Foundation (NSF) published the solicitation "Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining)" designed to address this national need. This program continues currently with solicitation number NSF 19-524. The four authors of this paper from three departments across two academic colleges at UMBC joined in response and proposed the UMBC CyberTraining initiative to create the nationwide online team-based training program “Big Data + HPC + Atmospheric Sciences” (cybertraining.umbc.edu) for students in three disciplines (Computing, Mathematics, and Physics) to foster multidisciplinary research and education using advanced cyberinfrastructure (CI) resources and techniques. The course teaches participants how to apply knowledge and skills of high-performance computing (HPC) and big data to solve challenges in Atmospheric Sciences. We focus on the application area of atmospheric physics and within it radiative transfer in clouds and global climate modeling, since these topics are important, pose computational challenges, and offer opportunities for big data techniques to demonstrate their impacts. The NSF funded our proposal in the inaugural year 2017 (OAC–1730250) for training programs conducted in 2018, 2019, and 2020.

Our program is now in its third year, and this paper reports on our experiences in conducting such training online and team-based with participants ranging from undergraduates (NSF-funded through an REU Supplement in Year 3), graduate students, post-docs/non-TT faculty, and TT (tenure-track) junior faculty. We specifically describe how to practically create the necessary training material, chiefly the tapings of lectures for later asynchronous online delivery of contents and homework, during Year 1, and how to accomplish this in an institutionally supportive environment, but without the type of resources an institution with an institutional focus on online teaching would have. Thus, we wish to share our experiences to regular faculty, who might want to add aspects of online teaching to their repertoire. We believe that this is extremely timely information in 2020, where many of us were forced into online teaching with next-to-no notice and no training because of the COVID-19 pandemic.

The rest of the paper is organized as follows. In Section 2, we explain how our graduate-level course on “Big Data + HPC + Atmospheric Sciences” was designed. Section 3 discusses how we recruited participants from applicants all over the U.S. Section 4 focuses on the creation of our online teaching program. Section 5 discusses some challenges we faced and our solutions to them. The paper concludes in Section 6.

2 The Big Data + HPC + Atmospheric Sciences Course

As illustrated in Figure 1, we believe there are a lot connections between big data, HPC and atmospheric sciences in terms of both education and research topics. So we designed our “Big Data + HPC + Atmospheric Sciences” course through the following innovative approaches: 1) it teaches students in atmospheric sciences how to implement and run parallel and big data programs at an HPC facility; 2) it teaches students in computing and applied mathematics how to solve atmospheric sciences challenges by applying their knowledge; 3) it provides distinctive learning outputs and homework to fit the background and interests of students in different disciplines; 4) it provides team-based frontier research projects where each team is composed with students in different disciplines so they can collaborate and contribute from their own research interests.

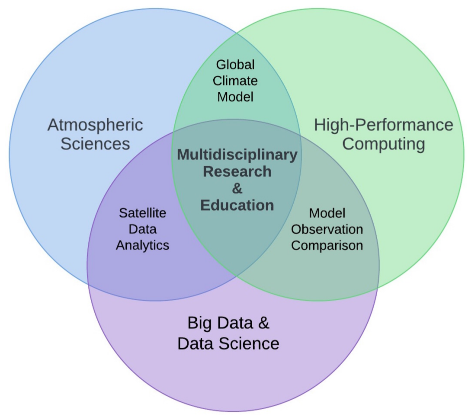

Fig. 1. Illustration of the connections between big data, HPC, and atmospheric sciences.

Our 15-module multidisciplinary course includes 1) customized course design for three disciplines with commonalities and differences; 2) data and computing techniques adoption for atmospheric sciences (three/four modules each for Data Science, HPC and Atmospheric Sciences); 3) identification of open challenges (including related open data) that can benefit from advanced CI resources and techniques; 4) five weeks long team-based project for frontier research challenges; 5) open source CI software implementation; 6) publications from the designed research projects. During a regular semester, the workload is equivalent to that of a three-credit course and we offer it to UMBC students as a cross-departmental special topic graduate course. The computing environment for the lectures and research projects is provided by our local HPC Facility at UMBC (hpcf.umbc.edu).

Table 1 lists the 15 modules of the course and it takes around three hours to teach each module. Details of each module are explained below.

Module 1: Introduction of Python/C, Linux and HPC environment. The first module explains the whole structure of the program and required basic knowledge for the program. It briefly goes through a programming language such as Python or C. It also introduces the hardware architecture, available software and basic usage of the UMBC HPCF environment (hpcf.umbc.edu).

Module 2: Numerical methods for partial differential equations (PDEs). This module explains the basics of partial differential equations, which is commonly used in physical models. It discusses the use of numerical methods for PDEs, which is one major driving force behind research in many other

Table 1. Modularized structure of our training program.

Module	Topic	Goal
1	Introduction of Python/C, Linux and HPC environment	Running their own jobs on HPC
2	Numerical methods for partial differential equations (PDE)	Model as PDE and solve them using numerical methods
3	Message Passing Interface (MPI)	Write MPI jobs and performance studies
4	Basics of earth-atmosphere radiative energy balance and global warming	Understand basic concepts and principles of radiative energy balance and global warming
5	Basics of radiative transfer simulation framework	Understand the basic physics underlying the transport of radiation in atmosphere
6	Global climate model (GCM) simulation and satellite observations	Understand the importance of GCM and satellite remote sensing
7	Introduction of big data	Understand the basics of big data and demo programs
8	Big data system: Hadoop/Spark	Write Hadoop/Spark jobs and run them on HPC
9	Big data machine learning	Write a machine learning program using Spark MLlib
10	Deep learning	Write a deep learning program
11	Project introduction	20 minutes project explanation from each team, including Q&A
12–14	Project progress report from each team and feedback	20 minutes report from each team including Q&A + rating
15	Final project presentation	Technical report, software and a final 30 minutes presentation from each team (by all team members) including Q&A

fields like numerical linear algebra, scientific computing, and the development of parallel computers. It covers the three basic PDE categories and their mathematical properties with examples. It discusses two large classes of methods: finite difference and finite element methods.

Module 3: Message Passing Interface (MPI). This module explains how to write MPI programs which is one of most common approach to build portable and scalable parallel scientific applications. It covers basic MPI commands such as MPI_Send and MPI_Recv, collective communication commands like MPI_Bcast, MPI_Reduce/MPI_Allreduce, and MPI_Gather/MPI_Scatter. It also explains how to write MPI programs in both C and Python (through mpi4py).

Module 4: Basics of earth-atmosphere radiative energy balance and global warming. This module explains the basic concepts and principles that control the radiative energy balance of earth-atmosphere system, and its implications to climate. The module starts with the fundamental physics, such as black-body radiation, followed by zero-order radiative energy balance between incoming solar radiation and outgoing terrestrial longwave radiation. The module ends with discussion of what kinds of roles the greenhouse gases, aerosols and clouds play in the radiative energy budget.

Module 5: Basics of radiative transfer simulation framework. Following the previous module, this module introduces the fundamental physical principles that control the transport of radiation (i.e., visible and infrared light) in our atmosphere. The module also includes the introduction of Monte-Carlo method and it application to radiative transfer.

Module 6: Global climate model (GCM) simulation and satellite observations. This module starts with an introduction to the basic concepts and principles of numerical climate simulations, followed by explaining the importance of evaluating climate simulations and why satellite remote sensing products are invaluable for climate model evaluation. Basic concepts and principle underlying satellite remote sensing are also introduced in this module.

Module 7: Introduction of big data. This module explains the basic concepts of Data Science, including generic lifecycle and different stages of data analytics, such as acquisition, cleaning, preprocessing, integration, aggregation, analysis, modeling and interpretation. It explains the basics of big data, including its 5V characteristics. It starts with the challenges and bottleneck of many applications when dealing with large volume of data. It also covers unique features and challenges for climate/atmospheric data.

Module 8: Big data systems: Hadoop and Spark. This module covers how to use two popular big data systems namely Hadoop and Spark. It explains how Hadoop Distributed File System (HDFS) can achieve data partitioning, and fault tolerance and cluster management and job scheduling in Hadoop/Spark. For Spark, it explains resilient distributed datasets (RDD), RDD transformations (map, join, cogroup, etc.) and actions (count, collection, foreach, etc.), lazy evaluation.

Module 9: Big data machine learning. This module explains how to conduct machine learning tasks in the above module in a scalable approach through Spark MLlib. Main techniques/concepts include DataFrame-based MLlib API vs. RDD-based MLlib API, ML pipelines, Transformer, Estimator and Parameter.

Module 10: Deep learning. This module covers deep learning using TensorFlow and Keras. It covers the basics of deep learning such as the network structure, activation functions, optimization, and backpropagation. Specific deep learning models such as convolutional neural networks, recurrent neural networks, and long short-term memory (LSTM) are covered with examples.

Module 11: Project introduction. Each team presents the basics of the research project they will work on in the following five weeks. It covers the background, required techniques, suggested phases and major tasks, expected outputs, output evaluation metrics and challenges to each discipline.

Modules 12–14: Project progress report from each team and feedback from instructors. These three modules are weekly project progress updates and discussions. Since most teams have three members, every member will be a presenter for the reports. All instructors and other teams discuss the progress, perform peer-review, provide feedback and give ratings.

Module 15: Final project presentation. The final module is the final project presentation and final CI software program and technical report delivery. Each team gives a talk on the problems to be solved, the approaches taken, demonstration of developed software program, the experiments and results, and contributions of each member. All instructors and other teams provide feedback and give ratings and suggestions for future work.

3 Recruitment, Applicants and Participants

Recruiting for the program used most effectively mailing lists in all three areas that the disciplines are housed in, with a flyer attached. That flyer was also distributed at relevant conferences if faculty or our graduate students attended. The flyers pointed to the program webpage cybertraining.umbc.edu that had program information as well as the link for application. For the three program years, we received 18, 94, and 100 applications, respectively, for the 15 funded participant slots. The recruiting was local and thus the number of applicants was limited in Year 1, since we conducted the training face-to-face, so participants had to be able to travel to UMBC every Friday afternoon.

The participants were selected competitively to form multidisciplinary teams of three participants with generally one participant from each area. The admission of participants was based on demographic information collected in a web form, a CV, a thorough personal statement, and for students with at least two letters of recommendation; upon acceptance of a student, we collected an explicit support from the student's advisor to ensure that the student was allowed to commit time to the program. The personal statement was asked to address specifically why the participant is interested in interdisciplinary research, how participation will promote his/her career goals, and how he/she can contribute to a team of participants from each discipline. The main selection criteria were: 1) how much the applicant can benefit from the training program; 2) how much the application's background is aligned with the program; 3) balanceness among home institutes of applicants. We gave preferences to applicants from underrepresented communities including historically black colleges and universities (HBCUs) and applicants from institutes that have no major HPC facilities. We also strived to make the teams demographically diverse, e.g., with respect to gender of the participants, but kept each team at a relatively consistent educational level. This means that we grouped graduate students of similar class standing together in a team as well as grouped the post-docs/faculty together. This avoids undue difference in leadership experience between the members and also allows to tailor the research project for each team slightly to an appropriate level for each team. This turned out to be a critically useful decision particularly in 2020, when the workload of many post-docs/faculty changed dramatically as their home institutions moved suddenly to online instruction, all while they had small children at home in many cases. It was noticeable that this change affected the graduate students relatively less negatively than the post-docs and faculty (including us faculty on this program ourselves).

The material is at the level of an advanced graduate course, and most participants were graduate students, but as permitted by this NSF program some can also be post-doctoral researchers or junior faculty. For all three groups, participating can have significant impact on their career in vastly expanding horizons from their own disciplines to two others. After an initial face-to-face course in Year 1 to develop the material, as explained in the next section in more detail, the training in the following years is completely online with participants working together remotely from anywhere in the nation. In this way, this training is available to participants who do not have local access to this kind of material. Another purpose of the face-to-face training in Year 1 was to create a pool of former participants, some of whom could be recruited to work as graduate assistants in Years 2 and 3.

Table 2. Profile of participants for our training program.

	under-graduates	gradates	postdocs & non-TT faculty	TT faculty	total participants	female participants	total teams
Year 1	0	9	4	3	16	7	5
Year 2	0	14	2	1	17	6	5
Year 3	6	11	4	4	25	14	8
Total	6	34	10	8	58	27	18

Table 2 summarizes the basic profile of the participants for our program over the three years. We can see 1) most participants are graduate students since we believe graduate students are still in their early years of their research career and the offering of multidisciplinary education would have bigger impacts on their future career growth; 2) we try to address the under-representation of female researchers in STEM disciplines by having relatively equal number of female participants (27) and male participants (31). Some additional participants not eligible for NSF funding (not graduate students or post-docs/faculty) were included without support. An additional benefit for local participants was the three-credit special-topics graduate course.

4 Creation of the Online Training

The fundamental goals of the proposed training were (i) the combination of teams with participants from three disciplines together and (ii) to conduct this training online with participants from around the nation. The multi-disciplinary nature of the work naturally gives rise to the use of team-based pedagogy.

But how to implement the training online leaves some choices. For instance, a fundamental decision to take is if online training should be completely asynchronous, or if only each team would work synchronously on their own time. We felt that this approach would deprive the teams from experiencing a whole-cohort feeling and we also wanted to foster communication skills. Therefore, the online training includes weekly synchronous meetings on Friday afternoons (Eastern time) that are conducted via Webex or Zoom. It is in principle possible to hold lectures online synchronously. However, it is not the most effective use of valuable synchronous time. Therefore, we use a flipped-classroom educational model [3], in which the contents are delivered via taped lectures that each participant views asynchronously in their own time. This model applies particularly during the first 10 modules, which constitute the instructional portion of the training. During that time, each team then communicates amongst themselves to organize the work on the team-based homework. State-of-the-art collaborative and communication tools are used throughout, thus providing deep exposure to skills vital in today's job market. This homework is due by the end of Thursday, so that the faculty can score it on Friday morning. In the synchronous Friday afternoon meeting, each team presents their homework solution to the whole cohort. The goal of the presentations of homework is to familiarize each participant with online presentation and the underlying goal of the team-based homework is to have the teammates gel together. This preparation pays off during the research training phase in Modules 11 to 15, when teammates now know each other, know each other's strengths, have experience with all communication and presentation technology, and can now present research updates effectively every week, culminating in a complete formal talk like at a conference in the final synchronous meeting. Additionally, the finished technical report from each team is published in the publication series of the UMBC High Performance Computing Facility (hpcf.umbc.edu).

The above describes the pedagogical techniques that we use in the fully online trainings in Year 2 and 3 (2019 and 2020). We feel that it is useful to explain in more detail, how we used Year 1 to conduct a transition from traditional (non-flipped) face-to-face teaching to flipped-classroom online training. Particularly, we wish to communicate here that the approach described in the following is a realistic one for busy research-active faculty with little or no long-term support by technological staff, such as is true at many institutions that do not have an explicit online teaching mission.

To accomplish the transition, we conducted the training in Year 1 (spring semester 2018) face-to-face in a team-taught three-credit course held on campus. This is realistic for workload of the instructors and to give enough time for coordination (preparation during fall 2017 and during the semester itself) among the instructors, who had not taught together before. These synchronous class meetings were traditional lectures and were taped, and these tapes form the basis of the online asynchronous contents delivery in the following years. These tapings in 2018 were done in an instructional classroom with a camera and initial support from AV staff that set up the equipment. One of the graduate assistants funded by the NSF grant operated the camera, so that the instructor could focus on a normal lecture delivery. We observe that by now (2020), other tools are more widespread, specifically in-cloud software such as Panopto or Blackboard Collaborate that can facilitate taping of lecture delivery within the instructor's own laptop, also during live lecture, thus an assistant would not necessarily be needed any more at all.

All work is conducted in a multidisciplinary team with participants from each area. In the first 10 modules consisting of instruction in all three areas, team building is achieved by homework. In the final 5 modules, each team applies the material learned immediately to a small research project, culminating in a technical report and a project presentation, by the end of the 15-module program. During the research phase, each team is mentored by a faculty member and supported by a graduate research assistant (RA), who is often a Ph.D. student of the faculty member. During the instructional phase of the first 10 modules, the instructor for each topical area is supported by a teaching assistant (TA) from that area. Many of the TAs and RAs are the same graduate students, but they do not need to be; in particular, some research projects were supported by additional graduate students not paid through this grant. In a similar way, some research projects were collaborations with other researchers, some of whom provided minor and some major leadership and mentorship. In this way, a program like this can be very exciting and invigorating for the research enterprise and even jumpstart new collaborations.

The description so far makes a separation of instruction and research, but we learned the lesson from Years 1 and 2 that it is in fact beneficial to overlap them in time. This means that the research mentor of each team is assigned and known to each team from the start. This mentor reaches out to the team already in the first week, thus giving the team an additional contact person throughout. The mentor can then learn if the team has already some available time and, for instance, start the research by making readings available during the first 10 weeks. In most cases, each mentor holds a weekly separate research project meeting with each team he mentors to discuss research topic, agenda and methodologies. This serves to get a head-start on the research phase and hit the ground running in Week 11. We used this approach in Year 3 to great success in that several teams have already first results in Week 12 that they could report in the synchronous meeting and get feedback from all other participants, earlier than in past years.

5 Discussion

Is flipped classroom a good teaching method for online instruction? We used face-to-face teaching in Year 1 and flipped online teaching in Years 2 and 3. Comparing their differences, we believe flipped classroom is suitable for online instruction. Our lecturing was most one-direction from instructor to students and the interaction between instructor and students was limited because students were busy following the lecture. By flipping classroom, students study the lecture videos ahead of time and practice their skills via homework before synchronous online discussion. An online Discussion Forum was available for questions throughout and actively monitored by the faculty and TA for the topical area. During the synchronous online meetings, each team presented their homework results and got feedback from instructors and other teams. By doing so, we had more interaction at our synchronous online discussions. Further, each team gets chances to learn how other teams solved the homework problems and compare differences among teams, which we could not do in Year 1's face-to-face teaching because most time was taken for lecturing and no time could be given for in-class homework discussion and presentation.

How to keep students engaged in online instruction? One challenge we often face for online instruction is student/participant engagement. Students might feel isolated because they do not see other students like regular face-to-face instruction. It is particularly challenging to this program since most participants are all over the country and each participant might be the only person from his/her institution involved in the program. Also most participants do not know anyone else before the program starts. We addressed the challenge by providing and facilitating various types of communications. The four main communication mechanisms we used are: 1) synchronous weekly online discussion where each team presents their homework or project progress and interact with other teams and instructors; 2) asynchronous web forum discussion for problems raised by students during studying lecture and working on homework and project so questions can answered be as soon as possible; 3) regular online meetings between an instructor and each team he mentors to discuss their progress and problems faces; 4) regular communications within each team to know each other better and collaborate on homework and project. One team reported they had three regular online meetings each week to discuss homework and research. Another team had over 3000 messages via Slack instant messaging (IM) software during the 15-week training period. A third team used WebEx Teams, since that software can hold meetings and save the chat across the whole duration of the program. Overall, we believe communication is the key to keep students engaged throughout the program.

Is it possible to complete a solid research project within five weeks? We admit five week is quite short for a solid research project. In actual implementation of our program/course, the instructors already have identified possible projects before the whole course started so that each team can pick from them if they cannot come up their own project quickly. Further, we encourage teams to discuss and define the project they plan to do early on. In most cases, each team mentor started to have regular weekly research project meetings with each team he mentors in around week 7. We chose this time point for two reasons: 1) team members have been collaborating with each others for a while via several modules and homework; 2) they all have some knowledge of atmospheric science by studying modules 4-6 to understand what could be an application challenge they can work on. We also did not pose hard deadline for the completion of the research project since each project is unique. Even the program/course technically finished after week 15, all teams were willing to continue working on their projects for a few weeks after in order to have good final technical reports. Many teams went on collaborating further to extend their reports to conference/journal papers.

How to involve undergraduate students in a program designed for advanced graduate students? In Year 3, we were successful in applying for REU Supplement support for six undergraduate students at our institution from the NSF. We recruited for these positions in August 2019 and admitted two students from each discipline in September 2019. We report on how it is possible to successfully integrate undergraduate students in a program that was conceived for advanced graduate students and junior faculty. The key was to start the training for these local students during the fall 2019 semester. Since the students had a full course load to start with, the spreading out of material is crucial. The students were grouped by department during fall 2019 with a faculty mentor from that home department. They started by learning about the topics out of the 10 instructional modules that are in their own area, thus when they joined a multi-disciplinary team, they had all something to contribute. We then during winter 2019-20 leveraged the fact that the lecture videos of the first 10 modules are available for asynchronous delivery. The two teams of undergraduates in fact started on the homework and were able to get a head-start of several weeks of homework submissions before the official start of the program. Using the time thus freed up during several weeks of instructions in Weeks 1 to 10, the undergraduate teams also started on research substantially earlier than Week 11. This concept is currently working, and the undergraduate teams have results on the same level as the more senior teams.

6 Conclusions

Both the National Strategic Computing Initiative [4] and the Federal Big Data Research and Development Strategic Plan [5] highlight the importance of workforce development on HPC and big data. In this paper, we present our efforts of creating a training program or graduate-level online course in big data applied to atmospheric sciences as application area and using HPC as indispensable tool. We outline a concrete procedure how to create the course and believe that this approach could also be used to create other courses for the “Computational and Data Science for All” educational ecosystem. This ongoing program already produced 10 technical reports, 10 peer-reviewed papers [6,7,8,9,10,11,12,13,14] and a M.S. thesis [15], and most of them are led by participants. The anonymous feedback from participants were also overwhelmingly positive. It reflects, to some extent, the success of our program in its offering of learning and research opportunity to the participants. Also, our experiences on online instruction would be particularly valuable to many instructors during the COVID-19 pandemic.

Acknowledgment

This work is supported in part by the U.S. National Science Foundation under the CyberTraining (OAC–1730250) and MRI (OAC–1726023) programs. The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF).

References

1. Computational Science: Ensuring America's Competitiveness. https://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf (2005)
2. National Academies of Sciences, E., Medicine, et al.: Future directions for NSF advanced computing infrastructure to support US science and engineering in 2017–2020. National Academies Press (2016)
3. Abeysekera, L., Dawson, P.: Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. Higher Education Research & Development **34**(1), 1–14 (2015)
4. The federal big data research and development strategic plan. https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=63 (2016)
5. Executive Order – Creating a National Strategic Computing Initiative. https://www.whitehouse.gov/the-press-office/2015/07/29/executive-order-creating-national-strategic-computing-initiative (2015)
6. Guo, P., Liu, C., Tang, Y., Wang, J.: Parallel gradient boosting based granger causality learning. In: 2019 IEEE International Conference on Big Data (Big Data), pp. 2845–2854. IEEE (2019)
7. Barajas, C., Guo, P., Mukherjee, L., Hoban, S., Wang, J., Jin, D., Gangopadhyay, A., Gobbert, M.K.: Benchmarking parallel k-means cloud type clustering from satellite data. In: International Symposium on Benchmarking, Measuring and Optimization, pp. 248–260. Springer (2018)
8. Barajas, C.A., Gobbert, M.K., Wang, J.: Performance benchmarking of data augmentation and deep learning for tornado prediction. In: 2019 IEEE International Conference on Big Data (Big Data), pp. 3607–3615. IEEE (2019)
9. Shi, P., Song, Q., Patwardhan, J., Zhang, Z., Wang, J., Gangopadhyay, A.: A hybrid algorithm for mineral dust detection using satellite data. In: 2019 15th International Conference on eScience (eScience), pp. 39–46. IEEE (2019)
10. Song, H., Wang, J., Tian, J., Huang, J., Zhang, Z.: Spatio-temporal climate data causality analytics-an analysis of enso's global impacts. In: Proceedings of the 8th International Workshop on Climate Informatics (CI2018) (2018)
11. Song, H., Tian, J., Huang, J., Guo, P., Zhang, Z., Wang, J.: Hybrid causality analysis of enso's global impacts on climate variables based on data-driven analytics and climate model simulation. Frontiers in Earth Science **7**, 233 (2019)
12. Zhang, W., Wang, J., Jin, D., Oreopoulos, L., Zhang, Z.: A deterministic self-organizing map approach and its application on satellite data based cloud type classification. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 2027–2034. IEEE (2018)
13. Wang, J., Gobbert, M.K., Zhang, Z., Gangopadhyay, A., Page, G.G.: Multidisciplinary education on big data+ hpc+ atmospheric sciences. In: Workshop on Education for High-Performance Computing (EduHPC-17) (2017)

14. Zhang, Z., Song, H., Ma, P.L., Larson, V., Wang, M., Dong, X., Wang, J.: Subgrid variations of the cloud water and droplet number concentration over tropical ocean: Satellite observations and implications for warm rain simulation in climate models. Atmospheric Chemistry and Physics **19**(PNNL-SA-136226) (2019)
15. Barajas, C.A.: An approach to tuning hyperparameters in parallel: A performance study using climate data. Master's thesis, Department of Mathematics and Statistics, University of Maryland, Baltimore County (2019)