UMBC

AN HONORS UNIVERSITY IN MARYLAND

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

Jianwu Wang (PI), *Information Systems, UMBC*

Matthias K. Gobbert, *Mathematics and Statistics, UMBC*

Zhibo Zhang, *Physics, UMBC*

Aryya Gangopadhyay, *Information Systems, UMBC*

cybertraining.umbc.edu

The 16th International Conference on Frontiers in Education: Computer Science & Computer Engineering (FECS'20), Las Vegas, NV, July 27-30, 2020

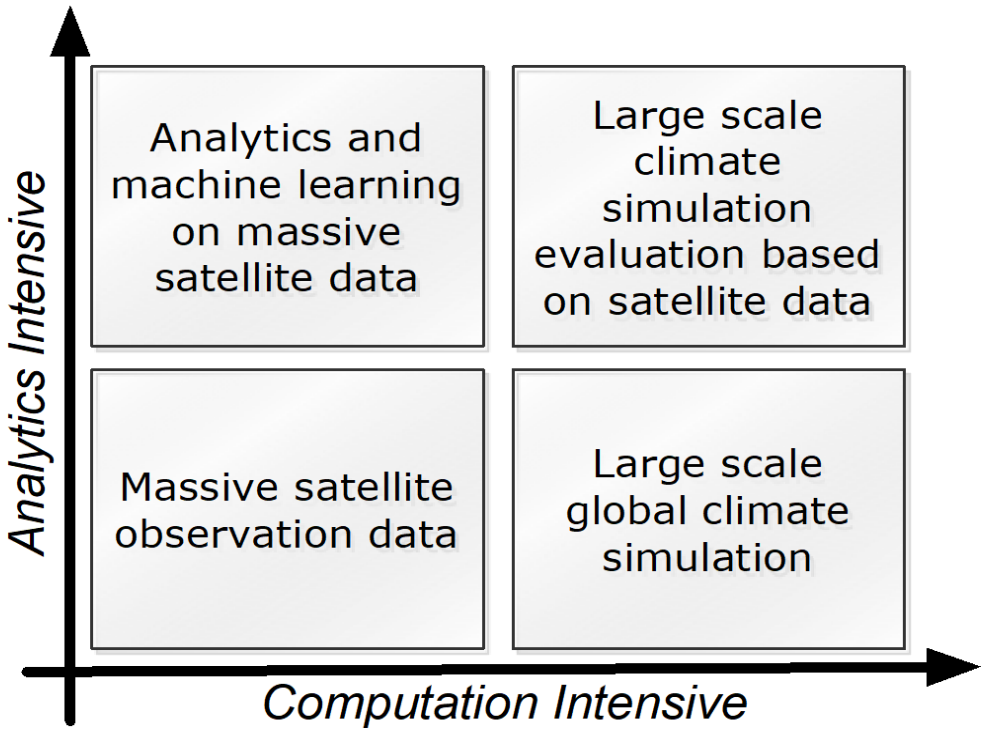
Acknowledgments: NSF (CyberTraining, MRI), HPCF, UMBC

NSF CISE Solicitation in 2017 (now 19-524): "Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining)"

- National need for training *"Big Data + High-Performance Computing + X"*
- *Interdisciplinary training* for students from each domain area is needed to accomplish *multidisciplinary projects* in real life.
- We proposed team-based training in teams of 3 with one from each area.
- We leverage advanced cyberinfrastructure by online training in Years 2 and 3 to participants from around the nation.
- In addition to *graduate students*, the NSF allows the inclusion of junior researchers including *post-docs and junior faculty*.
- Our funding in inaugural year 2017 for trainings in 2018, 2019, 2020

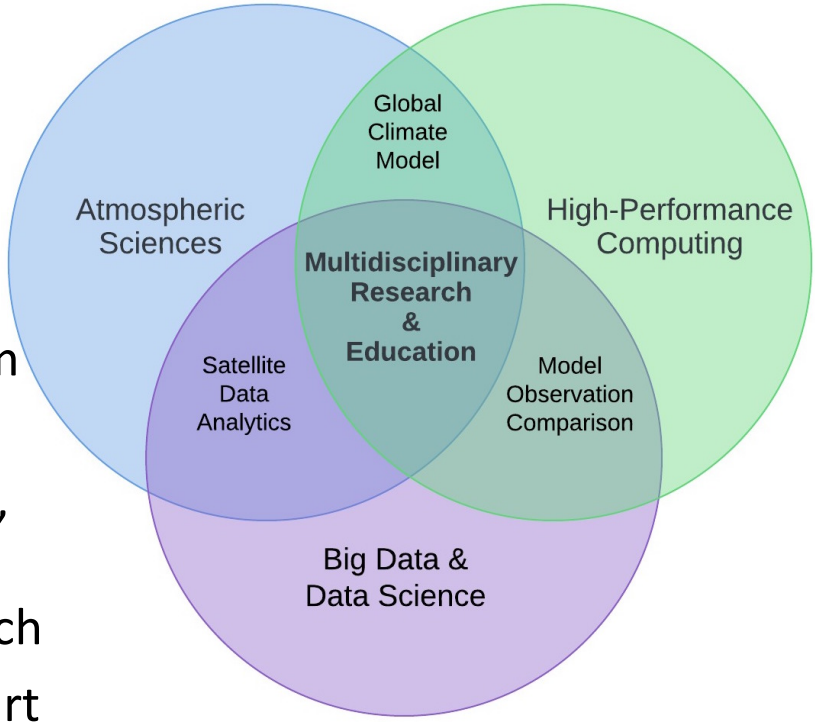
HPC and Big Data Requirements for Area X = Atmospheric Sciences

- HPC requirements for cloud simulation in numerical global climate models (GCMs)
- Big Data analytics requirements for evaluation of GCM using multi-decadal satellite observations



CyberTraining in "Big Data + HPC + Atmospheric Sciences"

- CyberTraining in big data applied to atmospheric sciences as application area and using high-performance computing as indispensable tool
- The training consists of instruction in all three areas, followed by faculty-guided project research in a multidisciplinary team of participants from each area
- Participating graduate students, post-docs, and junior faculty from around the nation will be exposed to multidisciplinary research
- Thorough program evaluation from the start



Training Modules

Module	Topic
1	Introduction of Python/C, Linux and HPC environment
2	Numerical methods for Partial Differential Equations (PDE)
3	Message passing interface (MPI)
4	Basics of earth-atmosphere radiative energy balance and global warming
5	Basics of radiative transfer simulation framework
6	GCM simulation and satellite observations
7	Introduction of and Big Data
8	Big Data system: Hadoop/Spark
9	Big Data Machine learning
10	Deep learning
11-15	Multidisciplinary team-based research project with deliverables: technical report, taped final slide presentation, code on Github

Participant Profile

	Undergraduate students	Graduate students	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 (REU)	11	4	4	25	14	8
Total	6	34	10	8	58	27	18

Applicants in Year 1: 18 locally (attend UMBC on Friday afternoon),
Applicants in Years 2 and 3: 94 and 100, respectively, nationwide

Flipped Classroom Educational Model for Online Learning in Years 2 and 3

- Flipped classroom educational model
  - Taped lectures that each participant views asynchronously
  - Asynchronous communication via an online discussion forum
  - Team based communication on homework
  - Weekly synchronous homework presentation and discussion
- Team-based research project educational model
  - Each team has member from each domain area.
  - Each team demographically diverse, but on same educational level to avoid undue differences in leadership experience
  - Each team is assigned its own research topic, mentor, and RA
  - “Team building by homework” in the first 10 weeks establishes communications in team, while beginning work on research.
  - Weekly synchronous presentation and discussion

Discussion: Is Flipped Classroom a Good Teaching Method for Online Instruction?

- Goal is to use valuable synchronous class time for student engagement!
- Participants watch taped lecture on their time asynchronously
- One more step beyond “flipped classroom”: Teams work on homework asynchronously during week, then the synchronous class time is used for homework presentations!
- By doing so, interaction between teams, exposure to other teams’ solutions, and experience with online tools.
- All this experience is valuable during the research portion of the CyberTraining!

Lessons: How to Feasibly Transition to Online Teaching and Learning?

- We used local face-to-face instruction in Year 1 to make the transition; also, faculty had not team taught with each other and are extremely busy research-active faculty.
- Institution has (excellent, motivated, creative!) AV Support, but: there is no mission to be “online university”!
- By now (2020), tools like Panopto and Blackboard Collaborate exist (and campuses have licenses) to self-tape lectures, namely screen capture with voice-over; closed captioning is available.
- Thus, accomplish transition as part of a face-to-face class or in ‘studio’ at home office. Feasible for all faculty!

Discussion: How to Involve Undergraduate Students in a Program Designed for Advanced Graduate Students?

- The undergraduate students are local and can start the training one semester earlier than others
- They start with their own topic out of the 10 instructional modules, so each can later contribute to team
- The teams of undergraduates started on the homework and were able to get a head-start of several weeks of homework submissions
- 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

Is it Possible to Conduct a Publishable Research Project in Five Weeks?

- The faculty mentor identifies possible projects before the whole course starts so that each team can pick from them if they cannot come up their own project quickly
- We encourage teams to discuss a possible project they plan to do early on, starting really when getting to know each other
- We also pose only a soft deadline for the completion of the research project presentation, since each project is unique
- Even the program/course technically finished at 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 and conference/journal papers

Products: Presentation, Publications, and Code

- Oral presentation developed during Weeks 11 to 14 for final taped presentation in Week 15.
- Code posted publicly on Github; others on webpage
- Publications: Each team tech. report in HPCF technical report series, 5 + 5 + 8 = 18 total
- Conference and journal publications: 14 to date, including 2 educational papers by the PI/co-PIs, plus 1 MS thesis and 2 senior theses, this data as of Summer 2020 (updated from paper)

Conclusions

- 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.
- The teams can involve additional researchers from outside the PI/co-PIs group, thus spreading the work, accomplishing work for this client, and invigorating multidisciplinary research at the institution.
- The possibility of publications makes the project valuable to tenure-track faculty as well as dramatically strengthens their portfolio
- The anonymous feedback from participants were also overwhelmingly positive. It reflects 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.