

# **Faculty Dossier**

**April 4, 2012**

**Nagiza F. Samatova**

**Associate Professor**

Department of Computer Science

North Carolina State University

**Senior Research Scientist**

Computer Science and Mathematics Division

Oak Ridge National Laboratory

Voice: (919) 513-7575, Fax: (919) 515-7925

E-mail: [samatova@csc.ncsu.edu](mailto:samatova@csc.ncsu.edu)

## B. Brief Resume

### 1. Education

1993, PhD, Mathematics, COMPUTATIONAL CENTER OF RUSSIAN ACADEMY OF SCIENCES, MOSCOW  
*“On the complexity of specification and combinatorial irregularity of convex 3-polyhedrons,”*  
 under Prof. Abdimajid K. Pulatov

1991, BS, Applied Mathematics, TASHKENT STATE UNIVERSITY, UZBEKISTAN

1998, MS, Computer Science, UNIVERSITY OF TENNESSEE, KNOXVILLE

### 2. Professional Experience (Since 2000)

NORTH CAROLINA STATE UNIVERSITY, Department of Computer Science  
 Associate Professor, 2007-present

OAK RIDGE NATIONAL LABORATORY, Computer Science and Mathematics Division  
 Senior Research Scientist, Joint Faculty Appointee, 2007-present  
 Senior Research Scientist, 2005-2007; Research Scientist, 2000-2004

### 3. Scholarly and Creative Activities

Books	Career Total	Total since joining NCSU	Total over 7/12-6/13	Total over moving PTR	Submitted <sup>†</sup>
Edited+Authored Books	2	1	1	1	1
Refereed Book Chapters	4	3	2	3	2

<sup>†</sup>Included in Career Total, Total since joining NCSU, Total Over 7/11-6/12, or Total over moving PTR

Publications	Career Total	Total since joining NCSU	Total over 7/12-6/13	Total over moving PTR	Submitted <sup>†</sup>
Refereed Journal Articles	53	31	8	30	5
Invited Journal Articles	7	5	5	5	
Refereed Conference Articles	73	45	12	40	2
Invited Conference Articles	8	3	1	2	
Refereed Workshop Papers	14	9	1	8	
Featured Articles	3	1			
Invited Workshop Roadmaps	18	5		1	
Patents	3	2		1	
Tutorials	2				
Technical Reports	9	4		3	

Presentations	Career Total	Total since joining NCSU	Total over 7/12-6/13	Total over moving PTR
Panels	3	3		
Invited/Contributed	192	63	11	47

Funding	Career Total	Total since joining NCSU	Total over 7/12-6/13	Total over moving PTR	Submitted <sup>†</sup>
Grants (NCSU)	\$6,297,797	\$6,297,797		\$3,669,797	\$5,842,000
Grants (All, approx.)	\$60,000,000				

<b>Mentoring</b>	<b>Career Total</b>	<b>Total since joining NCSU</b>	<b>Total over 7/12-6/13</b>	<b>Total over moving PTR</b>	<b>In Progress<sup>†</sup></b>
Research Scientist	4	2			
Postdoc	5	1			
Postmaster+IT Specialist	7				
PhD chair/co-chair	9	8	1	7	15
MS chair/co-chair	2	2	1	1	
PhD committee member	11	9	5	8	
MS committee member	3	3	1	2	

<b>Courses Created in a Significant Way</b>	<b>Career Total</b>	<b>Total since joining NCSU</b>	<b>Total over 7/12-6/13</b>	<b>Total over moving PTR</b>
Undergraduate	4	4	2	3
Graduate	10	5	1	4

<b>Courses Taught</b>	<b>Career Total<sup>†</sup></b>	<b>Total since joining NCSU</b>	<b>Total over 7/12-6/13</b>	<b>Total over moving PTR</b>
Regular undergrad (3 CR, $\leq 100$ students)	3	3	1	2
Large undergrad (3 CR, $\geq 100$ students)	1	1	1	1
Regular grad (3 CR, $\leq 100$ students)	5	5	1	4

<sup>†</sup>Included only Since joining NCSU

<b>Other</b>	<b>Career Total</b>	<b>Total since joining NCSU</b>	<b>Total over 7/12-6/13</b>	<b>Total over moving PTR</b>
Software Packages	28	17	15	15

#### 4. Membership in Professional Organizations MEMBER, ACM Computing and IEEE

#### 5. Scholarly and Professional Honors (Since 2007)

DISTINGUISHED PAPER AWARD, The 17th Int'l European Conference on Parallel and Distributed Computing (EuroPar), 2011; BEST STUDENT PAPER AWARD, The NASA Conference on Intelligent Data Understanding (CIDU), 2011; OUTSTANDING MENTOR AWARD, UT-Battelle, 2008; SIGNIFICANT EVENT AWARD, *Sustained Excellence in Mentoring*, Battelle, 2007; AWARDS NIGHT NOMINATION, EXCEPTIONAL COMMUNITY OUT-REACH, UT-BATTELLE, 2007; YWCA TRIBUTE TO WOMEN FINALIST IN EDUCATION, Knoxville, 2007.

#### 6. Professional Service On Campus (Selected)

COE Research Committee, 2009-present; Graduate Program Oversight Committee, 2010-2011; Graduate Students Admission Committee, 2010-2011; CSC 333 Course Coordinator, 2009-present; Computing Infrastructure Committee, 2007-present.

#### 7. Professional Service Off Campus (Selected)

Regular NSF and DOE proposal (including CAREER) evaluation panelist (5-7 panels a year); Co-chair of DOE/NSF Mathematics for Analysis of Petascale Data (2008), DOE Genomics:GTL Program Systems Biology Network/Knowledgebase (2008), DOE/OASCR Mathematics for Peta-scale Data Analysis & Computation (2007), DOE Workshop on Modeling and Simulation at the Exascale (2007).

### C. Candidate Statement

**Research Innovations and Impact:** To date, I have made technical contributions to foundational research in “Big Data” analytics as well as its applications to mining scientific data. My accomplishments in the area of “Big Data” analytics have been recognized through invitations to the White House Big Data Event in March 2012, the National Academy of Sciences symposium on Science, Innovation, and Partnership for Sustainability Solutions in May 2012, invited talks to the DOE Office of Science ASCAC/BERAC Advisory Boards (2007, 2008, 2012), DOE and NSF News Releases, and DOE-featured articles (*SciDAC Reviews*, 06/07a/b). My broad impact has been read into the Congressional records (2006) and recognized by the *The New York Times* and *Science* magazine (2007, 2010). I have served as a co-Director of the Center for Data-intensive computing in biology (2004-2008), the area leader for the DOE SciDAC Data Management Center (2001-2011), the DHS Biodefence Knowledge Center (2003-2007), the DOE Genomes-to-Life Center (2001-2004), and the NSF Expeditions in Computing (2011-present).

As of August, 2012, my research accomplishments have been recorded in 146 peer-reviewed publications including 53 journal papers and 74 conference papers. Since coming to NCSU in August 2007, I have published 86 peer-reviewed publications including 33 journal, 41 conference, 8 workshop papers, 1 textbook (co-editor), and 3 book chapters; with an annual average of more than 17 publications, and with 1,200 citations since 2007; more than 28 software libraries with thousands of downloads (<http://freescience.org/cs>); and their applications to solving challenging problems in bioenergy and climate change.

Big Data Analytics. My *pioneering work on parallel R* (pR) [e.g., *ICPDCS* 05, *IPDPS* 07, *J. Parallel Distrib. Comput.* 11] is distributed across nearly 30 mirror sites around the world; **RScalLAPACK’s RPM package is part of Linux distributions.** I have *significantly advanced database management for “Big” scientific data.* I proposed a transformative shift from the traditional indexing of data to the indexing of information about data compression. Our ISABELA-QA [*Supercomputing* 11] and ALACRITY [*DEXA* 12] scientific database management systems for data analytics have reduced, by an order of magnitude, the storage overhead and query response time [*IPDPS* 12, *ICPP* 12, *Supercomputing* 12]. Our ISABELA and ISOBAR are the most efficient-to-date, data compression technologies [*ICDE* 12, *HPDC* 12, *ICDM* 11, *EuroPar* 11]. ISABELA [*EuroPar* 11] received the *Distinguished Paper Award*.

Graph Theory, Graph Algorithms, Graph Mining. To deal with noisy and uncertain data, **I have pioneered graph perturbation theory** that enabled an order of magnitude speed-ups [*J. Theor. Comput. Sci.* 10, *ACM KDD-UD* 10, *IPDPS* 11]. Likewise, to enable fast mining of graph data, I have introduced **graph theory with knowledge priors** (DENSE algorithm, [*BMC Systems Biology* 11/12, *J. Data Mining and Knowledge Discovery* 12]). I have **pioneered the area of comparative analysis of the ensembles of networks** by introducing contrast-based graph mining, generalizing the concept of graph anomaly detection, and enabling multiple network alignment [*J. of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies* 11, *ICDM* 10, *J. Theor. Comput. Sci.* 09, *J. Comb. Optim.* 11].

Data-driven Discovery from Scientific Data. I led a research team to develop a computer program PRO-RATA to automate the entire data analysis process for both stable-isotope-labeled [*J. Anal. Chem.*, 06a/b] and label-free [*J. Proteome Res.* 06] quantitative shotgun proteomics and VONODE for *de novo* sequencing [*BMC Bioinformatics* 10], with more than 300 citations. Using these tools, addressed important bioenergy and bioremediation problems [*Mol. Cell. Proteomics* 08, *ISME J.* 09/11, *Nature Biotechnology* 09, *PLoS ONE* 09, *PLoS ONE* 12]). We provided an unprecedented accuracy in seasonal hurricane activity forecasts (92% vs. 65% by traditional methods) as well as 90% accuracy and 10-15 day lead-time forecasting the end-game of hurricane tracks [*IJCAI* 11a/b, *J. Data Mining and Knowledge Discovery* 12]. I contributed to automatic discovery of climate tele-connections that has been recognized through the *Best Paper Award*

[CIDU 2012, *J. of Stat. Analysis and Data Mining* 12].

**Teaching Effectiveness and Instructional Development:** I have been consistently providing excellent instructions in classes, with both student and peer evaluations supporting this view. The graduate and undergraduate students in my CSC 522/422 course have written and co-edited with me a textbook titled “*Practical Graph Mining with R*” that is under the publication contract with the Chapman & Hall/CRC Press under the Data Mining and Knowledge Discovery Series, with all the proceeds from the sale of the textbook going to the NC State Department of Computer Science. *Writing the textbook entirely by the students* is just one example of non-traditional ways of my teaching among various ‘brain-teasers,’ team competitions, the ‘ugliest proof,’ ‘speed-track problem solving,’ and Jeopardy games. I consistently received high evaluations across all the evaluation criteria. The peer faculty evaluation by Prof. Savage, who has extensive work experience with NC State, has complemented to my class by saying that it was “*the best class I have ever attended as a peer reviewer.*”

**Advising and Mentoring:** Defending my PhD in Moscow at the age of 22, I became a strong believer in an early scientific career and have devoted my life to promoting high-quality science among high school students, undergraduates, graduates, and postdocs. I have supervised 24 PhD, 2 MS, 8 undergraduate, and 11 high-school students, 4 research scientists, 5 Postdoctoral fellows, and 23 internships. While at NC State (from August 2007), I have graduated a total of 8 PhD students: 6 PhDs as Chair, 1 PhD as Co-Chair (with Dr. Bahler), and 1 PhD student from the university of South Florida as Co-Chair. *Four NCSU graduated PhD students are currently on the faculty at universities.* I am currently advising and sponsoring 15 PhD students as Chair (4 are expected to graduate in 2012-2013 AY).

**Outreach and Broader Impact:** I have mentored a total of *16 female researchers* including 2 postdocs, 1 post-MS, 2 PhD, 2 MS, 5 undergraduates, and 4 high school students. I have also mentored *8 researchers from underrepresented minority groups* including 6 students and 3 faculty. I received multiple Diversity Awards from the ORAU-ORNL RAM Program. A high school team that I mentored placed 1<sup>st</sup> (*\$109,000 award*) as national team finalists in the 2006-07 Siemens Competition in Math, Science and Technology. Another team of high school sophomore and senior students, placed 2<sup>nd</sup> (*\$50,000 scholarship award*) as national team finalists in the 2009-10 Siemens Competition. Yet another high school team placed 4<sup>th</sup> (*\$39,000 award*) as national team finalists in the 2005-06 Siemens competition. *These accomplishments have been read into the Congressional Record.* Senator Lamar Alexander commented on the floor of the Senate during his speech in support of the National Competitiveness Investment Act [Congressional Record, Senate, Page S11588, Dec. 8, 2006] commented that “Part of the reason these three students succeeded is they were able to connect with the work and expertise at the Oak Ridge National Laboratory [and their] lead adviser, Dr. Nagiza F. Samatova ...”

**Research Funding:** I have been awarded and have managed, as a PI or co-PI, more than \$60M in research grants sponsored by DOE, DHS, NSF, and UT-Battelle since 2001. Since joining NCSU (August, 2007), I have been awarded \$6.3M to NC State alone (thus, maintaining more than \$1.2M new NCSU funds annually), with another \$1.5M to Oak Ridge National Laboratory as part of my joint appointment.

**Professional Service:** I was invited to serve on the DOE Office of Advanced Scientific Computing Research (OASCR) Panel of distinguished members to identify computational science breakthroughs across the Office of Science’s research programs. I have served on the DOE Joint ASCAC/BERAC Subcommittee on Modeling and Simulation; co-chaired several DOE, DHS and NSF workshops on roadmaps for future R&D activities in computational sciences and served on a number of scientific conference program committees and DOE/DHS/NSF proposal review committees.

# I. TEACHING AND MENTORING OF UNDERGRADUATE AND GRADUATE STUDENTS

## A. Teaching Effectiveness

### 1. Courses Taught

Due to my joint faculty appointment with the Oak Ridge National Laboratory, I am expected to teach 2 courses annually. Also, I volunteer teaching an extra section of the course. In addition, I often buy out of teaching a course. Since I came as a joint faculty between NCSU and Oak Ridge National Laboratory (ORNL), I have taught the following courses:

- |   |                    |
|---|--------------------|
| 1. Spring 2013, CSC 333 (601) (16 students) | <i>New in 2012</i> |
| 2. Fall 2012, CSC 707 (19 students)         | <i>New in 2012</i> |
| 3. Fall 2012, CSC 333 (001) (76 students)   | <i>New in 2012</i> |
| 4. Fall 2012, CSC 333 (601) (12 students)   | <i>New in 2012</i> |
| 5. Fall 2011, CSC 522 (39 students)         |                    |
| 6. Fall 2011, CSC 422 (22 students)         |                    |
| 7. Fall 2011, CSC 333 (001) (71 students)   |                    |
| 8. Fall 2011, CSC 333 (002) (38 students)   |                    |
| 9. Fall 2010, CSC 707 (17 students)         |                    |
| 10. Spring 2010, CSC 707 (14 students)      |                    |
| 11. Fall 2009, CSC 522 (44 students)        |                    |
| 12. Fall 2009, CSC 422 (6 students)         |                    |
| 13. Spring 2009, CSC 707 (22 students)      |                    |
| 14. Fall 2008, CSC 333 (71 students)        |                    |

### 2.a. Student Evaluations

At the end of each semester, NCSU Computer Science students rate the effectiveness of their teachers on a 1 – 5 scale, where 1 is poor and 5 is truly outstanding. My teaching evaluations are listed in the tables below. When I teach a course, I give my TAs an opportunity to co-teach with me, not simply grade the assignments. This way, I am training the students to be effective teachers via giving them opportunities to lecture, create exam, quiz, and homework assignments, offer recitation sections, etc. I let them learn their mistakes and also observe their teaching in class and provide the feedback on how to improve their teaching skills. The scores are often the reflection of this paired teaching and may often vary depending on the experience of the TAs that I may have.

I have received high evaluations despite the fact that students perceive my courses as presenting significantly more difficulty than average. I have received a number of informal complements both from the students and the faculty that the students really enjoy the courses, while learning quite a lot of new material. I enormously enjoy teaching and my exposure to such bright and motivated students in the classes.

#### **Anonymous Students' Feedback: CSC 333 Fall 2008**

Table 1: Teaching Evaluations at NCSU in 2008

COURSE		CSC 707		CSC 333	
SEMESTER		SPRING 08		FALL 08	
ENROLLED		14		71	
QUESTION		MEAN	DEPT	MEAN	DEPT
1.	The instructor stated course objectives/outcomes	<b>4.6</b>	4.31	<b>4.6</b>	4.3
2.	The instructor was receptive to students outside the classroom	<b>4.57</b>	4.16	<b>4.7</b>	4.1
3.	The instructor explained difficult material well	<b>4.6</b>	4.31	<b>4.4</b>	3.8
4.	The instructor was enthusiastic about teaching the course	<b>4.79</b>	4.27	<b>4.8</b>	4.0
5.	The instructor was prepared for class	<b>4.79</b>	4.28	<b>4.7</b>	4.2
6.	The instructor effectively used instructional technology	<b>4.36</b>	4.19	<b>4.5</b>	4.2
7.	The instructor consistently treated students with respect	<b>4.36</b>	4.39	<b>4.8</b>	4.3
8.	Overall, the instructor was an effective teacher	<b>4.57</b>	4.02	<b>4.6</b>	4.0
9.	The course readings were valuable aids to learning	<b>4.07</b>	4.00	<b>3.5</b>	3.9
10.	The course assignments were valuable aids to learning	<b>4.57</b>	4.13	<b>4.4</b>	4.1
11.	This course was intellectually challenging and stimulating	<b>4.71</b>	4.1	<b>4.6</b>	3.9
12.	This course improved my knowledge of the subject	<b>4.86</b>	4.22	<b>4.6</b>	4.1
13.	Overall, this course was excellent	<b>4.5</b>	3.96	<b>4.3</b>	3.7

- Dr. Samatova is one of the best professors I have had at NCSU and specifically in Computer Science.
- Dr. Samatova is a GREAT teacher. She is more than willing to help students, and does so with a smile. She will stay on a subject until she feels like most of the class understands it. She has a gift for explaining difficult concepts in easy terms (e.g. she explained one-to-one and onto functions with clever metaphors, and the encoding of Turing Machines with easy rules). Another strength: she relates individual topics to the bigger picture of Automata and Complexity Theory. e.g. How do DFAs, NFAs, PDAs, and TMs fit together? How do Regular, Context Free, Recursive, and Recursively Enumerable languages fit together? She is down-right entertaining, too. I remember the first day of class, I was in stitches because she made me laugh so hard. Dr. Samatova genuinely wants her students to understand the class topics. She is an amazing professor and a real catch for the NCSU Computer Science Department!
- She is so excited about teaching the material, and also so excited about the student learning the material. She communicates it so well in class, that I generally did not have to read the book or her lecture notes (which were always promptly posted online). Her office hours were: any time. She was so receptive to students (unlike some teachers who find a visit during their office hours to be a pain). She was funny, which kept people's attention during the class. She is also incredibly knowledgeable, which allows her to explain things from viewpoints that make the most sense. She did an excellent job of teaching visually, not just with text or words. I know NC State is currently trying to do what Georgia Tech did and shoot up in the college rankings; if that is going to happen you can't let Dr Samatova go back to Oak Ridge!!! This course covered difficult material and I picked up on it like I never thought I could. I honestly feel renewed confidence in my intelligence because I learned such difficult material so quickly.
- I loved this course. I hadn't realized until I took this course how much I enjoyed learning about the theories behind Computer Science. I enjoyed Discrete Math, but had thought that was just

due to the teacher. This class has led me to realize how much I enjoy the theories.

- I have ever met a teacher that was as prepared as you were with the perfect blend of teaching with repetition. Submitted HW to check to see if we are understanding the material, great ...GREAT idea! Constantly asking questions to see if we are keeping up. You are making sure no one is left behind.
- I wanted to say how much I am enjoying the class. I appreciate the clarify of your explanations particularly of the more subtle concepts, such as the pumping lemma.
- The course is rigorously theoretical – to be expected from a theory course. Dr. Samatova alleviated this somewhat by allowing class members to make short presentations on the applications of various topics covered in class, which helped the rest of us see that learning these strictly theoretical topics was not completely futile.
- Friendly, always willing to help - Open-door policy with regard to office hours - Tried to make sure everyone understood current concepts before moving on.

#### **Anonymous Students' Feedback: CSC 707 Spring 2008**

- I love this course!
- In a word, she is an excellent teacher.
- Well-organized and -delivered lectures; use of competitions was interesting and engaging; instructor would often review difficult topics if the class did not seem to be catching on. Homework grading was unusual.
- The informal atmosphere of the class was one of its strongest points- students were able to reinforce each others' understanding of the material, or discuss different approaches to a problem before learning the "right" one. This made it much more enjoyable.
- I really enjoyed your enthusiasm for this material. I don't think I will ever forget the image of you with your back to the board arbitrarily dividing strings for the pumping lemma. You were always engaged and excited about the class. Your bonus points kept everyone trying (harder than they might have), and provided an alternate environment to the traditional lecture. I liked them.
- Dr. Samatova was exceptionally dedicated to our learning. Her main focus during the semester was to make sure that we were learning the material. She also focused on practical learning of skills to use the theory of computation in our future work environments. She was very responsive to forum posts, e-mails, and student office visits about core material. Dr. Samatova occasionally had minor errors in her slides, which she promptly fixed.
- This was an excellent course, especially considering that Dr. Samatova was teaching it for the first time.

#### **Anonymous Students' Feedback: CSC 422/522 Fall 2009**

- It's a great class we have done a great project. This is a great help.
- Technical knowledge of the instructor is awesome.
- Gave us a new experience by writing a book.
- A highly motivated and innovative professor.
- Made us to think more like a researcher.Great experience.



Table 2: Teaching Evaluations at NCSU in 2009

COURSE		CSC 707		CSC 422/522	
SEMESTER		SPRING 09		FALL 09	
ENROLLED		22		50	
QUESTION		MEAN	DEPT	MEAN	DEPT
1.	The instructor stated course objectives/outcomes	<b>4.7</b>	4.3	4.0	4.2
2.	The instructor was receptive to students outside the classroom	<b>4.6</b>	4.3	<b>4.7</b>	4.1
3.	The instructor explained difficult material well	<b>4.3</b>	4.0	<b>4.2</b>	3.8
4.	The instructor was enthusiastic about teaching the course	<b>4.9</b>	4.4	<b>4.4</b>	4.1
5.	The instructor was prepared for class	<b>4.8</b>	4.3	<b>4.5</b>	4.2
6.	The instructor effectively used instructional technology	<b>4.6</b>	4.3	<b>4.3</b>	4.2
7.	The instructor consistently treated students with respect	<b>4.8</b>	4.5	<b>4.6</b>	4.3
8.	Overall, the instructor was an effective teacher	<b>4.7</b>	4.1	<b>4.2</b>	3.9
9.	The course readings were valuable aids to learning	<b>4.1</b>	4.1	<b>3.9</b>	3.9
10.	The course assignments were valuable aids to learning	<b>4.4</b>	4.2	3.6	4.0
11.	This course was intellectually challenging and stimulating	<b>4.8</b>	4.2	<b>4.0</b>	3.9
12.	This course improved my knowledge of the subject	<b>4.6</b>	4.3	3.7	4.1
13.	Overall, this course was excellent	<b>4.5</b>	4.0	<b>3.7</b>	3.7

- Extremely inspiring and motivating. Makes very complex algorithms sound very simple after explaining them.
- Excellent instructor - works with students to improve understanding. Teaches with passion; excellent knowledge of the material. Enthusiastic, high standards.

#### Anonymous Students' Feedback: CSC 707 Spring 2009

- This was a very challenging class which was well-prepared and well-presented. The instructor explained the material very well and was receptive to questions, always making herself available when students had questions. There were no real weaknesses.
- The project component of the class was excellent, and I thought that there was a really good balance between homework, exams, and project work. Overall this was an outstanding class, and I only wish that there was a sequel.
- It is one of the best courses I have taken in NCSU. I learned a lot from it although it is not an easy courses.
- Excellent approach to the subject, very knowledgeable. Course project helped to demonstrate practical applications in a highly theoretical course.
- The course is excellent. We need more theory courses like this.
- She did an excellent job of engaging students which really helped given the difficulty of the material. She provided us with in-class excercises that were extremely useful in learning. Overall, excellent.
- Professor Nagiza is a passionate teacher. She is very knowledgeable in her area of teaching. She explained difficult materials very well, so I really learnt a lot from this course. I think professor Nagiza is an excellent instructor.

Table 3: Teaching Evaluations at NCSU in 2010

COURSE		<b>CSC 707</b>	
SEMESTER		FALL 10	
ENROLLED		<b>17</b>	
QUESTION		MEAN	DEPT
1.	The instructor stated course objectives/outcomes	<b>4.9</b>	4.3
2.	The instructor was receptive to students outside the classroom	<b>4.9</b>	4.3
3.	The instructor explained difficult material well	<b>4.9</b>	4.0
4.	The instructor was enthusiastic about teaching the course	<b>5.0</b>	4.3
5.	The instructor was prepared for class	<b>4.9</b>	4.3
6.	The instructor effectively used instructional technology	<b>4.7</b>	4.2
7.	The instructor consistently treated students with respect	<b>4.9</b>	4.4
8.	Overall, the instructor was an effective teacher	<b>5.0</b>	4.1
9.	The course readings were valuable aids to learning	<b>4.6</b>	4.2
10.	The course assignments were valuable aids to learning	<b>4.9</b>	4.2
11.	This course was intellectually challenging and stimulating	<b>4.9</b>	4.2
12.	This course improved my knowledge of the subject	<b>5.0</b>	4.3
13.	Overall, this course was excellent	<b>5.0</b>	4.1

#### Anonymous Students' Feedback: CSC 707 Fall 2010

- By far the best instructor I have had at NCSU. Automata is such a difficult and at times unappealing subject, yet Dr. Samatova made it interesting and explained things so well that I honestly feel that I finally learned this topic.
- What is taught by her is very interesting and very intellectually challenging. She also has a very good method of teaching, and I have to say that she probably is the best professor who has taught me so far.
- The class is always very interactive; with students coming to the board and solving questions and in-class exercises.
- This is a very difficult course, and the material definitely stretched my brain!
- I was surprised at how "hands-on" the course was. I had never taken a theory course that was taught like this, so it came as a great and pleasant surprise. I think this is a great way of teaching a "hard" course.
- CSC 707 is the hard course, but I don't have that feelings. Dr. Samatova is one of the greatest instructors. She is really good at explaining difficult things into a simple way.
- The instructor and TA felt more like a team than in my other courses, where the TA's kind of feel like the worker ants that just do what they are told and do little to mix in with the students in the course, which often creates an invisible barrier between instructors and students. Overall this has been my favorite course at NC State so far, as it pushed me to learn without forcing me to memorize, as it is a more practical measure of learning.

Table 4: Teaching Evaluations at NCSU in 2011

CSC COURSE	<b>333</b>	<b>333</b>	<b>422/522</b>	<b>All</b>
SECTION	002	001	001	N/A
SEMESTER	FALL 11	FALL 11	FALL 11	FALL 11
ENROLLED	<b>38</b>	<b>71</b>	<b>61</b>	<b>N/A</b>
QUESTION	MEAN	MEAN	MEAN	DEPT
The instructor stated course objectives/outcomes	4.4	4.1	4.5	4.4
The instructor was receptive to students outside the classroom	4.0	3.9	4.1	4.1
The instructor explained difficult material well	4.0	3.8	4.3	3.9
The instructor was enthusiastic about teaching the course	4.6	4.3	4.4	4.3
The instructor was prepared for class	4.4	4.4	4.5	4.4
The instructor gave useful feedback	4.0	3.8	4.0	4.0
The instructor consistently treated students with respect	4.3	3.9	4.2	4.3
Overall, the instructor was an effective teacher	4.3	3.9	4.3	4.0
The course readings were valuable aids to learning	3.4	3.3	3.8	3.8
The course assignments were valuable aids to learning	4.1	4.1	4.1	4.1
This course improved my knowledge of the subject	4.4	4.2	4.3	4.2
Overall, this course was excellent	3.9	3.7	3.9	3.8

#### Some Anonymous Students' Feedback: CSC 422/522 Fall 2011

- Excellent ! Just Excellent !
- Outstanding professor!
- Excellent course!
- She is one of the few professors out there that actually care about the students. The amount of effort she put into this course was admirable.
- Innovative style of teaching - Knows the subject very well - Is able to amicably answer the questions raised by students - Is humourous which makes the class fun to attend.
- Good insights into data mining technologies and algorithms - Good and ehxhaustive course materials
- This course has been very helpful to me as it helped me gain vital knowledge about data mining as a research area. I strongly recommend students to take this course.
- The strength of the instructor is that, she can very well teach Math concepts, which I think is a tough job! The instructor succeeded in imprinting the data mining concepts in the students' brain!

#### Some Anonymous Students' Feedback: CSC 333 Fall 2011

- I was very impressed with Dr. Samatova and the TAs. The material was very hard but they way they taught and the time they put into teaching us made this class easier and fun. I thought the most useful and beneficial parts of this class were the examples and in class exercises we did, the online forums, and the TA was excellent.
- The instructor was excellent in explaining difficult theoretical course material

- Dr. Samatova is a very good teacher and I was able to learn a lot under her instruction. Her ability to teach material to the class is exceptional.
- I enjoyed the material a lot. There are no weaknesses that I can think of. The class was very interactive and I really enjoyed the atmosphere.
- Tough material, but instructor taught it well.
- Dr. Samatova is VERY intelligent and knows how to break down a complicated topic so that a student can relate to. Highly motivated, and great enthusiasm for teaching.
- Really kept me motivated in this class, effective teaching!
- One of the most difficult courses Ive taken at NCSU, but definitely the most satisfying in terms of the quest for knowledge.

## 2.b. Peer Evaluations

### Peer Evaluation: CSC 707 January 2008

Peer evaluation of a CSC 707 lecture, conducted on January 28, 2008 by **Prof. Carla Savage** and **Prof. Rada Chirkova**:

We (Carla Savage and Rada Chirkova) attended Nagiza Samatova's course CSC 707, "Automata, Languages and Computability Theory" on Monday, January 28, 2008 for the purpose of conducting a peer evaluation of her teaching.

The lecture was on finite automata and regular languages, including closure properties and the pumping lemma. It began with an informative review of the last lecture - by individual students (!) with two facts contributed by each student. Dr. Samatova knows the names of the students, which made the review go quickly and smoothly.

Samatova used the computer projector to show extremely well-conceived slides, designed to lead the students through difficult proofs with illuminating visuals and carefully chosen examples. She is teaching the course for the first time and has developed all of these materials herself. The slides are available to the students on the course web page.

There are carefully designed weekly homeworks. Since the instructor has no TA for the class, she is grading all of the assignments herself, in addition to preparing the lecture materials.

### **The lecture was great!**

The students were lively and attentive throughout the class. Some comments and observations on how Dr. Samatova was able to achieve this:

- all students can see the names of the other students written on paper placards on the desks
- ask students questions during the lecture (by name)
- students seem relaxed and comfortable answering the instructor's questions, laughing at the instructor's jokes, attentive and nodding during the explanations of the lecture materials
- encourages questions from students during lecture
- very good examples on lecture slides, most of the examples are not from the textbook, nice writing of additional explanations on the whiteboard

A particularly innovative idea that the students enjoyed was to have fun competitions in class between student teams: a "speed track", 4-5 students per team, very nice setup, nice challenging questions to be solved quickly, good leads by the instructor to help the students succeed.

There is a well-designed web page for the course and the instructor is making use of all of the options available on Wolfware. Some specific comments and observations:

- online syllabus, including the lecture slides, good coverage of all the appropriate material
- clearly stated learning outcomes
- nice ideas for homework problems (e.g., submit the "ugliest" solution - as a way to make the students more discriminating in writing and reading proofs.)
- prompt and helpful responses to message-board questions
- encourages anonymous feedback from students

In summary, we found Dr. Samatova to be a dedicated and competent teacher, leading a stimulating, challenging, and enjoyable class.

One of us (Savage) would like to add: **this was the best class I have ever attended as a peer reviewer.**

#### **Peer Evaluation: CSC 333 September 2008**

Peer evaluation of a CSC 333 lecture, conducted on September 16, 2008 by **Prof. George Rouskas** and **Prof. Munindar Singh**:

This is a peer evaluation of Dr. Nagiza F. Samatova's teaching for Fall 2008 in CSC 333, *Automata, Languages and Computability Theory*. We attended Dr. Samatova's lecture of September 16, 2008. Dr. Samatova is teaching this course for the first time this semester, and it is also her first experience teaching undergraduates at NCSU.

**Relevance of course content.** CSC 333 is the second theory course offered in our program. It is the first course that involves the key concepts of theory of computation, which underlie much of subsequent theory. Thus it is an important course in the training of any computer scientist.

**Organization of course.** Dr. Samatova has organized the course in a series of three parts, each with its exam. The parts build on previous parts systematically. Dr. Samatova has created a web page for this course that is well organized and provides advance copies of lecture notes for students, along with homework assignments, solutions, sample tests, etc. Dr. Samatova also teaches a graduate course on the same subject. She has been adapting some of her materials from that class by increasing the emphasis on fundamental concepts and examples.

**Organization of lecture.** Dr. Samatova's lecture of September 16 was on nondeterministic finite state automata and regular languages. These are two of the main concepts in computing, and are used pervasively across the field. Dr. Samatova began with bookkeeping announcements, followed by a quick review of the previous lecture, followed by a *Speed Track* exercises, followed by the main discussion. In the Speed Track exercises, the students work in groups to be the first to answer simple questions from previous classes, and are rewarded with points. Students take these quite seriously, and it seems to be a wonderfully effective technique.

**Lecture delivery.** Dr. Samatova's delivery of the lecture was effective. She used slides (which she had already posted for her students) in addition to the whiteboard. Dr. Samatova knew her slides well and

was obviously highly prepared both on the topic and the presentation of it. She continually came up with examples and worked out examples on the whiteboard with the help of the students. Dr. Samatova asked questions of her students throughout the lecture. Importantly, some of the students asked her questions as well, which she answered effectively and nicely.

**Physical infrastructure.** This course is being held in EB2 1021, which is a large room with auditorium seating. A majority of the students placed themselves in the section directly in front of the projector screen that Dr. Samatova was using. They didn't all sit near the front, and a few sat in the far corner. However, we found that Dr. Samatova was easily audible in the far corner, where we sat. We should observe that some of the seats in this classroom are creaky and this can be distracting.

**Potential for improvement.** We noticed that the students in the far corner of the class were somewhat less engaged than those in the section facing the projector. More of them seemed to be using their computers for purposes other than the class; others around them were distracted as well. This was by no means uniform, but enough that it should be addressed. Although Dr. Samatova asked several questions throughout the lecture, we suggest she direct some questions to specific parts of the classroom, thus helping improve the attentivity of all students.

The Speed Track exercises by their very nature are quick. We noticed that some students who didn't "get" the exercises were left out. It might be worth taking a minute or two after each exercise to explain the answer to those who have fallen behind.

**Summary.** Dr. Samatova is an excellent teacher for a course that is notoriously difficult to present.

#### **Peer Evaluation: CSC 707 February 2009**

We (**Matt Stallmann** and **Steffen Heber**) visited Dr. Nagiza Samatova's course, Automata, Languages and Computability Theory (CSC 707) on Wednesday, February 4, Spring Semester 2009 to perform a peer evaluation of her teaching.

The class has a well-designed web page, providing information about class syllabus, assignments, and policies. Grading is based on homework, project work, and exams. At the day of our visit, 16 students attended the class. The class started and ended in time.

The main topic of the class was the Pumping Lemma. The class started with announcements about homework, class project, and midterm exam, followed by a warm-up exercise consisting of a series of short and precise questions. After each question, the students had 2 minutes time to find a solution individually, and then the answers were collected and discussed. Dr. Samatova encouraged the students to present their solutions; she only commented on the correctness, and provided additional follow-up questions. After this exercise, Dr. Samatova switched to a presentation in lecture style, and explained the Pumping Lemma using slides and black board. Her slides and handwriting were clear, and easy to read. In addition to explaining the material, Dr. Samatova was continuously asking questions and providing examples in order to deepen the students' understanding of the topic. After about 30 minutes, Dr. Samatova switched her teaching style again and performed a team exercise. The students did proofs based on the pumping lemma individually, and then corrected their teammates' solutions. At the end of class, Dr. Samatova presented a proof of closure properties of regular languages, foreshadowing the topic of the next session.

The lecture was interesting, well-organized, and easy to follow. Prof. Samatova changed the pace and the nature of activities multiple times during the class session, keeping the students engaged throughout. The students were very comfortable with her teaching; they readily participated in the exercises with minimal instructions and asked/answered questions without much prompting. Remarkably, although

the lecture topic was abstract and challenging, the entire class was highly interested and motivated, and participated very actively during the entire class period. Several students asked questions during class, which Dr. Samatova handled very well.

In summary, Prof. Samatova is an enthusiastic, innovative, and talented teacher. She has very good classroom management skills, and her contact to the students is exceptionally good.

**Peer Evaluation: CSC 333 on September 18, 2012 by Prof. Tiffany Barnes and Prof. Matthias Stallman**

*New in 2012*

Dr. Nagiza Samatova is teaching CSC 333 - Automata, Grammars, and Computability - during the Fall 2012 semester. Drs. Barnes and Stallmann attended her lecture on 9/18/2012, and also reviewed the course syllabus and other materials available through the web.

The class was held in 2213 EB-3, with approximately 60 students in attendance. Dr. Samatova began and ended the class on time.

The class began with an overview of two aspects of the course designed specifically to improve the communication skills of the students; Dr. Samatova has introduced significant innovations along these lines into her teaching of this particular course. In particular,

- Students do peer grading of one of the assignments - they are asked to evaluate the details of each other's mathematical proofs and are graded both on the appropriateness of their comments and the quality of their work, as graded by their peer and the TA. LaTeX scripts have been set up to facilitate this process (Dr. Samatova teaches LaTeX at the beginning of the semester and requires all submissions in LaTeX).
- Students are required to do a team project, which may be either a research paper on an application of theory (for example, machine translation or model-based testing in software engineering or a parser generated using tools based on finite automata and grammars).

Next came a Jeopardy game with review questions for an upcoming exam (and potential extra credit for the winners). Students collaborated to come up with answers to the, often difficult, questions. Though not all of the students contributed in the form of providing solutions, the game was fun and engaging and presented the students with key concepts. Dr. Samatova told us later that she uses three mechanisms for exam preparation: homework problems, Moodle quizzes, and Jeopardy. Questions on the closed-book exams are taken directly from these sources, giving students strong incentive to participate fully in all three.

The remainder of the class, about 25 minutes, was spent lecturing about finite automata. Rather than being a continuous lecture, it was broken up by a 5-minute partner exercise in the middle. Because the class also has a distance education component, students were asked to come to the front to present their solutions (this was also true for the Jeopardy game, when solutions required more than just a few words). Clearly, all aspects of the class were designed to mitigate the effects of the short attention spans prevalent among students.

Dr. Samatova's web site was rich in information useful to both students and colleagues: LaTeX tools, Jeopardy quizzes, an example of what a peer-graded paper should look like, in addition to the usual syllabus, schedule, lecture slides, and forums.

In short, Dr. Samatova is an engaging and entertaining instructor with excellent grasp of the subject matter and a vast collection of ideas, methods, and tools that make her class interesting and exciting for the students, no mean feat, given the abstract theoretical nature of the material.

## **Peer Evaluation: CSC 707 on September 18, 2012 by Prof. Kristy Boyer and Prof. Jon Doyle**

*New in 2012*

Dr. Nagiza Samatova is teaching CSC 707, Automata, Languages and Computability Theory during the Fall 2012 semester. **Dr. Kristy Boyer** and **Dr. Jon Doyle** attended her lecture on September 18, 2012, and also reviewed the course syllabus and other materials available through the course web page.

The class was held in Engineering Building II, room 1220. Dr. Samatova began and ended the class on time. The class consisted primarily of a review and problem examples on NP-completeness and reduction in preparation for an upcoming examination. The lecture portion of the class focused on complexity theory, self-reduction and optimization.

Dr. Samatova began the class with a reminder of the upcoming exam, a review of the peer grading process, and a discussion of the class project that was to be completed in teams of up to five people. She then engaged students in a Jeopardy-style game that they played in groups. During this game, students were faced with challenging problems relating to material covered previously in the class. They worked out solutions in groups, with the first group to finish presenting their solution on the board. Dr. Samatova supplemented and corrected the students' solutions on the board when necessary and asked compelling follow-up questions. She was effective in helping students solve the problems and providing clear explanations.

Dr. Samatova presented the new lecture material with clearly-organized overhead slides projected from a computer, and supplemented the slides with written material on the whiteboard. The use of the whiteboard to write out difficult concepts in real time was more effective than presenting the material on the slides alone. Students were engaged and taking notes during the lecture.

The course web site, organized in Moodle, provides a copy of the syllabus, course policies, class notes, homeworks, project information, and quizzes. The course syllabus contains the material required by the university. Grading includes submitted homework, automatically graded homework, quizzes, lecture assignments, three exams, peer grading, a project, class participation, and class attendance. The weight for each assignment is clearly posted. Due dates for assignments are posted on Moodle.

In summary, Dr. Samatova is knowledgeable on the subject matter and teaches in an engaging and effective manner.

## **B. Instructional Development**

### **Course Development**

I have developed two related courses in the core area of Theoretical Computer Science (Automata, Languages and Computability Theory). These courses are an integral part of the Bachelor, Master and PhD of Computer Science degrees. I also extensively revised the structure of the senior undergraduate and graduate course on Automated Learning and Data Analysis.

In these courses, I strived to achieve a good balance between theory and practice, and I developed substantial problem solving components that provide students with effective hands-on experience.

#### 1. CSC 707 – AUTOMATA, LANGUAGES AND COMPUTABILITY THEORY – Graduate

This is an advanced graduate course on the theory of computation and complexity. There are more than 15 students enrolled in the graduate course in contrast to 4-6 students in previous enrollments before my joining NCSU.

CSC707 is theoretically quite a challenging course. Since the students have quite a heterogenous background in mathematics, especially in formal proofs, teaching the course requires a very special



treatment of that issue to make the course effective across-the-board. I developed (from scratch) the course materials including lecture slides, lecture notes, practice exams, quizzes, solutions, discussions, etc. Particularly, I have exercised numerous non-traditional ways of teaching including various 'brain-teasers,' team competitions, the 'ugliest proof' and 'speed-track problem solving' competitions. In spite of such a highly interactive methodology, I managed to cover not only the typical set of topics required for this course but also to bring up more advanced topics on fixed parameter tractability, approximability, and real-world applications. I am particularly proud of the students' performance of a very high (close to 90 out of 100) class grade average on quizzes and exams, for which an average score of 60 or less has been reported by other instructors and/or universities. The difficulties of the exam materials tested are comparable.

## 2. CSC 333 – AUTOMATA, LANGUAGES AND COMPUTABILITY THEORY – Undergraduate

This is a core undergraduate course on the theory of computation and complexity. The course material has been adapted for the undergraduate level from a related course (CSC 707) that I created for the advanced graduate level.

CSC333 is theoretically a very challenging course. Computer science major students at the undergraduate level typically do not have advanced background in mathematics, especially, in formal proofs, abstraction, and deductive logic, which are the required skills for mastering the course objectives. Although I have taught a similar course at the graduate level with approximately 15 students, teaching this course at the undergraduate level with more than 70 students is an enormous challenge. I had to make significant adaptations to the course material. Lecture slides have been enriched with numerous examples, practicing an iterative cycle from an example to abstraction via mathematical notation and back. To help students better relate theory with practice, group projects with students' presentations on how specific computational models can be used to solve real problems in security, computational biology, compilers, computational reasoning, etc. have been introduced. Homeworks and their grading scheme have been set up in a way to improve the efficacy of students-teacher communication. I received several comments from the students who found this scheme quite effective—(a) frequent homeworks (twice a week) helped them re-emphasize lecture materials via practice, (b) the binary grading scheme (full credit or no credit) for Tuesday's homeworks reduced students' stress level and made them feel comfortable about sharing their mistakes in their solutions so that common patterns in students' misunderstanding of the material could be quickly addressed in a subsequent lecture(s), (c) grading of randomly selected Thursday's homeworks encouraged them to be rigorous in their proofs/reasoning. Also, I experimented with an open door policy for the office hours and many students found this policy extremely helpful – they were able to catch their lack of understanding of some of the material in a timely manner rather than letting it accumulate like a snow-ball. I spent the first 3 weeks of the course material on the core mathematical foundation that the course relies on. The goal was to bring student up to the same/similar page and somewhat level-off their heterogenous background in mathematics. We practiced very abstract concepts and proof techniques related to set theory, countability and non-countability, closure properties of countable sets, induction and proof-by diagonalization. Many students commented to me that they had a quantum jump to the level of sophisticated proofs they have mastered. Prof. George Rouskas informally commented to me that although this material is touched in the pre-requisite courses, most students will likely have significant difficulty in solving these types of problems. To my pleasant surprise, the students performed with a class average of 90 on the quiz, which was comparable in difficulty to the graduate level course.

## Teaching Outside NCSU

Fall 1993, Discrete Mathematics (Undergraduate), Tashkent State University  
Fall 1993, Special Topics in 3-polytopes (Graduate Seminar), Tashkent State University  
Spring 1994, Mathematical Logic (Undergraduate), Tashkent State University  
Spring 1994, Coding Theory (Graduate Seminar), Tashkent State University  
Fall 1994, Discrete Mathematics (Undergraduate), Tashkent State University  
Fall 1994, Graph Theory & Combinatorics (Graduate), Tashkent State University  
Fall 1994, Special Topics in 3-polytopes (Graduate Seminar), Tashkent State University  
Fall 1994, Discrete Math & its Applications (Graduate Seminar), Tashkent State University  
Spring 1995, Coding Theory (Graduate Seminar), Tashkent State University  
Spring 1995, Linear Programming/Optimization (Undergraduate), Tashkent State University  
Fall 1995, Discrete Mathematics (Undergraduate), Tashkent State University  
Fall 1995, Graph Theory & Combinatorics (Graduate), Tashkent State University  
Fall 2002, LSD 695: Advanced Computat. Methods in Bioinformatics (Graduate GST), U. of Tennessee (with Dr. Y. Xu)  
Fall 2003, CS 594: Algorithmic Methods for Bioinformatics (Graduate CS), U. of Tennessee (with Prof. M. Langston)  
Spring 2004, LSD 510: Bioinformatics (Graduate GST), U. of Tennessee (with Prof. Guo)

### C. Mentoring Activities

I am actively advising undergraduate and graduate students on their research activities. Besides directing doctoral and Masters theses, I have advised and supervised Research Scientists, Postdoctoral and Postmaster Fellows, IT Specialists, Summer Internship Students.

1. Research Scientist Supervisor for KARPINETS, TATIANA, ORNL, 2006-2007
2. Research Scientist Supervisor for KORA, GURUPRASAD, ORNL, 2006-2009
3. Research Scientist Supervisor for PAN, CHONGLE, ORNL, 2006-2011
4. Research Scientist Supervisor for PARK, BYUN-HOON, ORNL, 2005-2007
5. Postdoctoral Advisor for TIAN, WENHONG, NCSU, 2007-2008
6. Postdoctoral Advisor for BALDWIN, NICOLE, ORNL, 2004-2005
7. Postdoctoral Advisor for PARK, BYUN-HOON, ORNL, 2002-2005
8. Postdoctoral Advisor for TIAN, WENHONG, ORNL, 2007-2008
9. Postdoctoral Advisor for ZANG, BING, ORNL 2005-2006
10. Postmaster Fellow Advisor for KRISHNAMURTHY, RAMYA, ORNL, 2003-2007
11. Postmaster Fellow Advisor for SYMONS, CHRISTOPHER, ORNL, 2004-2007
12. IT Specialist Supervisor for CHANDRAMOHAN, PRAVEEN, ORNL 2003-2005
13. IT Specialist Supervisor for BAUER, DAVID, ORNL, 2001-2004
14. IT Specialist Supervisor for KORA, GURUPRASAD, ORNL, 2002-2006
15. IT Specialist Supervisor for MUNAVALLI, RAJESH, ORNL, 2003-2005
16. IT Specialist Supervisor for WATKINS, IAN, ORNL 2001-2005
17. PhD Dissertation Committee Member for FU, HAIZHOU, May 2013 *New in 2012*
18. PhD Dissertation Committee Member for PEARCE, RYAN, March 2013 *New in 2012*
19. PhD Dissertation Committee Member for GARLAND, GENEVIEVE, January 2013 *New in 2012*
20. PhD Dissertation Committee Member for STEFANSKI, DOUGLAS, December 2012 *New in 2012*
21. PhD Dissertation Committee Member for ZHANG, YONGPENG, Septemeber 2012 *New in 2012*
22. PhD Dissertation Committee Member for TETTEH, ISAAC, March 2012
23. PhD Dissertation Committee Member for MOUALLEM, PIERRE, July, 2011
24. PhD Dissertation Committee Member for LIN, HESHAN, March, 2009
25. PhD Dissertation Committee Member for LI, JIANTIAN, December, 2008
26. PhD Dissertation Committee Member for RAZUMOVSKAYA, YEVGENIYA, October 2004, UTK
27. PhD Dissertation Committee Member for XU, QIN, November 2004, UTK
28. PhD Dissertation co-Mentor for ZHANG, YUN, August 2008, UTK
29. PhD Dissertation co-Mentor for MCCOLLUM, MIKE, 2006, UTK
30. PhD Dissertation co-Mentor for ABU-KHZAM, FAISAL, 2003, UTK
31. PhD Proposal Committee Member for RAVINDRA, PADMASHREE, February, 2013 *New in 2012*
32. PhD Proposal Committee Member for PEARCE, RYAN, October 2012 *New in 2012*

33. PhD Proposal Committee Member for CLEARFIELD, RAPHAEL, February 2012 *New in 2012*
34. PhD Proposal Committee Member for STEFANSKI, DOUGLAS, April 2012
35. PhD Proposal Committee Member for FU, HAIZHOU, August 2011
36. PhD Proposal Committee Member for TETTEH, ISAAC, March 2011
37. PhD Proposal Committee Member for ZHANG, YONGPENG, May, 2011
38. PhD Dissertation Committee Member for ZHANG, YONGPENG, May 2011
39. PhD Proposal Committee Member for MOUALLEM, PIERRE, May, 2010
40. PhD Proposal Committee Member for GARLAND, GENEVIEVE, January, 2010
41. PhD Proposal Committee Member for LI, JIANGTIAN, May, 2008
42. PhD Written Qualifier Examination Member for GONG, ZHENHUAN, October, 2011
43. PhD Written Qualifier Examination Chair for DOKUZOGLU, NAZLI, December, 2010
44. PhD Written Qualifier Examination Chair for RAVINDRAN, SRINATH, October, 2009
45. PhD Written Qualifier Examination Member for ZHANG, ZHE, May, 2008
46. PhD Written Qualifier Examination Member for TANEJA, KUNAL, November, 2008
47. PhD Written Qualifier Examination Chair for COX, DAVID, November 2007
48. MS Thesis Committee Member for MATTHEW SESSOMS, August, 2012 *New in 2012*
49. MS Thesis Committee Member for CHANDRA, HARSHA GIRISH, July 21, 2008
50. MS Thesis Committee Member for WHEELER, BENJAMIN, Defended on May 6, 2008
51. Undergraduate Honors Advisor for NEIL SHAH, Spring, 2013 *New in 2012*
52. Undergraduate Honors Advisor for for SRINIVAS, SHASHANK, Spring, 2013 *New in 2012*
53. Undergraduate Honors Advisor for FETANAT, AFROUZ, Fall, 2012 *New in 2012*
54. High School Math and Sciences Theses Advisor for SHAH, NEIL, 2009, Greensboro High School
55. High School Math and Sciences Theses Advisor for SHPANSKAYA, YEKATERINA, 2009, Math & Science Home School, Raleigh
56. High School Math and Sciences Theses Advisor for ARCANGELLI, STEVEN, 2007, Oak Ridge High School
57. High School Math and Sciences Theses Advisor for GU, ALICE, 2007, Oak Ridge High School
58. High School Math and Sciences Theses Advisor for HORTON, SCOTT, 2007, Oak Ridge High School
59. High School Math and Sciences Theses Advisor for MOLONY, SCOTT, 2007, Oak Ridge High School
60. High School Math and Sciences Theses Advisor for UMAR, TARIK, 2006, Oak Ridge High School
61. High School Math and Sciences Theses Advisor for BRENT, PATRICIA, 2006, Oak Ridge High School
62. High School Math and Sciences Theses Advisor for GRABENSTEIN, NICHOLAS, 2006, Oak Ridge High School
63. Summer Internship Mentor for SHAH, NEIL, High School Student, 2009-2010
64. Summer Internship Mentor for SHPANSKAYA, YEKATERINA, High School Student, 2009

65. Summer Internship Mentor for JALLOUK, ANDREW, High School Student, 2006-2007
66. Summer Internship Mentor for HORTON, SCOTT, High School Student, 2006-2007
67. Summer Internship Mentor for ARCANGELLI, STEVEN, High School Student, 2006-2007
68. Summer Internship Mentor for KOTHE, BRETT, Undergraduate Student, 2007
69. Summer Internship Mentor for BRENT, PATRICIA, High School Student, 2005-2007
70. Summer Internship Mentor for GU, ALICE, High School Student, 2006
71. Summer Internship Mentor for LI, JIANGTIAN, PhD Student at NCSU, 2005-2006
72. Summer Internship Mentor for FRIDMAN, MAX, Undergraduate Student at UTK, 2006
73. Summer Internship Mentor for UMAR, TARIK, High School Student, 2005
74. Summer Internship Mentor for GRABENSTEIN, NICHOLAS, High School Student, 2005
75. Summer Internship Mentor for OZISIKYILMAZ, BERKIN, PhD Student at NWU, 2005
76. Summer Internship Mentor for LIN, HESHAN, PhD Student at NCSU, 2004
77. Summer Internship Mentor for BURTON, JEFFREY, Undergraduate Student, 2004
78. Summer Internship Mentor for BRAIMAN, AVITAL, Undergraduate Student at UTK, 2003
79. Summer Internship Mentor for BAUER, DAVID, Undergraduate Student at Georgia Tech, 2001-2002
80. Summer Internship Mentor for HESPEN, JENNIFER, Undergraduate Student, 2001-2002
81. Summer Internship Mentor for QU, YONGMING, PhD Student, Iowa State U., 2001
82. Summer Internship Mentor for MIRONOVA, SVETLANA, MS Student at UTK, 2001
83. Minority Summer Internship Mentor for ROCHA, ANDREA, PhD Student USF, 2007-2010
84. Minority Summer Internship Mentor for MCGARITY, JESSICA, 2006-2007
85. Minority Summer Internship Mentor for WALKER, KATHRYN, 2006
86. Minority Summer Internship Mentor for ASKIA, RASHIDA, 2004
87. Minority Summer Internship Mentor for LESLIE, TRAYVON, 2004

#### **D. Master's and Doctoral Theses Directed and Being Directed**

I have graduated seven PhD students and two master thesis students since joining NCSU. I am currently advising fifteen (15) PhD students on their thesis work.

##### **Doctoral Students Graduated (Total 9)**

1. CHEN, ZHENGZHANG, NCSU, May 2012 (Postdoc at Northwestern University)
2. CHEN, WEN BIN, NCSU, October 2010 (Faculty in Guangzhou University, China)
3. HENDRIX, WILLIAM, NCSU, November 2010 (Faculty at Northwestern University, USA)
4. SCHMIDT, MATTHEW, NCSU, December 2010 (Scientist at Lincoln Lab, MIT, USA)
5. PANSOMBUT, TATDOW (ANN), NCSU, April 2010 (Faculty at Prince of Songkla University, Pattani Campus, Pattani, Thailand)
6. ROCHA, ANDREA, USF, April 2010 (Postdoc at Oak Ridge National Laboratory, USA)
7. AWEKAR, AMIT, NCSU, December 2009 (Faculty at IIT, Guwahati, Assam, India)

8. BREIMYER, PAUL, NCSU, July 2009 (Scientist at Lincoln Lab, MIT, USA)
9. PAN, CHONGLE, UTK, Knoxville, 2006 (Scientist Oak Ridge National Laboratory, USA)

**MS Thesis Students Graduated (Total 2)**

1. ARKATKAR, ISHA, December, 2011
2. MOHAN, CHANDRA, November, 2008 (Microsoft Co.)

**Doctoral Students Currently Supported/Advised (Total 15)**

**RA**—Research Assistant supported by the research grant; **TA**—Teaching Assistant supported by the Department; **IR**—Independent Research study student

1. BELLO LANDER, GONZALO, Spring 2012-present (RA, IR)
2. BOYUKA, DREW, Spring 2012-present (RA, IR)
3. GONG, ZHENHUAN, Fall 2008-present, Passed PhD oral exam in Fall 2012 (RA, IR)
4. GONZALEZ, DOEL, Spring 2011-present, Passed Written PhD exam in Fall 2012 (RA, IR)
5. HARENBERG, STEVEN, Fall 2012-present (RA, IR)
6. HARLALKA, JITENDRA, Fall 2011-present (RA, IR)
7. JENKINS, JONATHAN, Fall 2010-present, Passed Written PhD exam in Spring 2011 (RA, IR)
8. LAKSHMINARASIMHAN, SRIRAM, Fall 2009-present, Passed Written PhD exam in Spring 2011 (RA, IR)
9. PADMANABHAN, KANCHANA, Fall 2009-present, Passed Written PhD exam in Spring 2011 (RA, IR)
10. PARUCHURI, AVANINDRA, Fall 2011-present (TA, IR)
11. PENDSE, SAURABH, Fall 2011-present, Passed Written PhD exam in Fall 2011 (RA, IR)
12. SEYE, ROBERT, Fall 2012-present (TA, IR)
13. SCHENDEL, ERIC, Fall 2010-present, Passed Written PhD exam in Spring 2011 (RA, IR)
14. TANG, HOJUN, Fall 2012-present (RA, IR)
15. ZOU, XIAOCHENG, Fall 2011-present (RA, IR)

## II. SCHOLARSHIP IN THE REALMS OF FACULTY RESPONSIBILITY

### A. Scholarly Accomplishments

#### Awards and Honors

- BEST PAPER AWARD NOMINATION, The 22nd International ACM Symposium on High Performance Parallel and Distributed Computing, New York City, USA, June, 2012 *New in 2012*
- BEST PAPER AWARD NOMINATION, The 21st International ACM Symposium on High Performance Parallel and Distributed Computing, Delft, The Netherlands, June 18-22, 2012 *New in 2012*
- IEEE PRE-COLLEGE AWARD NOMINATION, October, 2012 *New in 2012*
- DISTINGUISHED PAPER AWARD, 17th International European Conference on Parallel and Distributed Computing (EuroPar), Bordeaux, France, August 29–September 2, 2011
- BEST STUDENT PAPER AWARD, NASA Conference on Intelligent Data Understanding (CIDU), Sunnyvale, California, October 19-21, 2011
- SIGNIFICANT EVENT AWARD, *Sustained Excellence in Mentoring that Leads to Multiple National Students Awards*, Battelle, 2007
- 2007 NOMINATION FOR UT-BATTELLE AWARDS NIGHT, EXCEPTIONAL COMMUNITY OUT-REACH AWARD, Battelle, 2007
- 2007 YWCA TRIBUTE TO WOMEN FINALIST IN EDUCATION AND ADMINISTRATIVE PROFESSIONAL CATEGORY, Knoxville, 2007
- OUTSTANDING MENTOR AWARD, UT-Battelle, 2008
- INCENTIVIZED PERFORMANCE AWARD (IPA), UT-Battelle, 2006
- 2002 NOMINATION FOR UT-BATTELLE AWARDS NIGHT, SCIENTIFIC ACCOMPLISHMENT, EARLY CAREER, UT-Battelle, 2002
- DISTINGUISHED PERFORMANCE EVALUATION, UT-Battelle, 2001-2007
- SOFTWARE AGENT RESEARCH AWARD, *Virtual Information Processing Agent Research (VIPAR) Project, U.S. Pacific Command (USPACOM)*, Lockheed-Martin, 2001
- APPRECIATION FOR EXCEPTIONAL WORK ON VIPAR PROJECT, Advanced Computing Technologies, ORNL, 2001.
- THE FREEDOM SUPPORT ACT GRADUATE FELLOWSHIP, United States of America, 1996
- V. I. LENIN GRADUATE FELLOWSHIP, Tashkent State University, 1991-1993
- AL HOREZMI SCHOLARSHIP, Tashkent State University, 1988-1991
- 2<sup>nd</sup> PLACE WINNER, STATE (UZBEKISTAN) MATHEMATICS COMPETITION FOR HIGH SCHOOL STUDENTS, Tashkent, Uzbekistan, 1986
- 1<sup>st</sup> PLACE WINNER, STATE (UZBEKISTAN) CLASSIC GUITAR COMPETITION, Tashkent, Uzbekistan, 1984
- 1<sup>st</sup> PLACE WINNER, STATE (UZBEKISTAN) RHYME RECITATION COMPETITION, Tashkent, Uzbekistan, 1980

## Publications Summary

The Chancellor has asked faculty to identify the extent of their contributions in their various publications. Each publication listed below is coded with one or more letters defined as follows:

- C – Corresponding author
- L – Lead author
- E – Equal contribution as the other authors
- M – Minor contribution as an author

## Submitted Conference Publications

1. [CE] BOYUKA II, D., GONG, Z., LAKSHMINARASIMHAN, S., SCHENDEL, E., JENKINS, J., PODHORSZKI, N., LIU, Q., KLASKY, S., AND SAMATOVA, N.F., “Transparent In Situ Data Transform in ADIOS.” *2013 International Conference for High Performance Computing, Networking, Storage and Analysis (SC '13)*, November, 2013 (Submitted). *New in 2012*
2. [CE] E. R. SCHENDEL, S. HARENBERG, H. TANG, V. VISHWANATH, M. E. PAPKA, AND N. F. SAMATOVA, “A Generic High-performance Method for Deinterleaving Scientific Data.” *19th International European Conference on Parallel and Distributed Computing (Euro-Par)*, Aachen, Germany, August 2013 (Submitted). *New in 2012*

## Invited Refereed Journal Publications

3. [CE] JENKINS, J., ARKATKAR, I., LAKSHMINARASIMHAN, S., BOYUKA II, D.A., SCHENDEL, E.R., SHAH, N., ETHIER, S., CHANG, CS., JCHEN, J., KOLLA, H., KLASKY, S., ROSS, R., AND SAMATOVA, N.F., “ALACRITY: Analytics-driven Lossless Data Compression for Rapid In-situ Indexing, Storing, and Querying.” *LNCS Transactions on Large-Scale Data- and Knowledge-Centered Systems (TLDKS)*, 2013 (accepted). *New in 2012*
4. [E] KAWALE, J., LIESS, S., KUMAR, A., STEINBACH, M., SNYDER, P., KUMAR, V., GANGULY, A., SAMATOVA, N.F., SEMAZZI, F., “A Graph Based Approach to Find Teleconnections in Climate Data.” *Stat. Analysis and Data Mining*, 2012 (accepted). *New in 2012*
5. [CE] E. R. SCHENDEL, S. V. PENDSE, J. JENKINS, D. A. BOYUKA II, Z. GONG, S. LAKSHMINARASIMHAN, Q. LIU, S. KLASKY, R. ROSS, N. F. SAMATOVA, “ISOBAR Hybrid Compression-I/O Interleaving for Large-scale Parallel I/O Optimization.” *Cluster Computing*, 2013 (accepted). *New in 2012*
6. [CE] LAKSHMINARASIMHAN, S., SHAH, N., ETHIER, S., KU, S-H., CHANG, C.S., KLASKY, S., LATHAM, R., ROSS, R., AND SAMATOVA, N.F., “ISABELA: Effective in-situ compression of spatio-temporal data.” *Concurrency and Computation: Practice and Experience*, 25(4), 524-540, 2013, doi:10.1002/cpe.2887. *New in 2012*
7. [CE] PADMANABHAN, K., WILSON, K., ROCHA, A., MIHELICIC, J.R., AND SAMATOVA, N.F., “In silico identification of phenotype related modules.” *Proteome Science*, 10(1):S2, 10.1186/1477-5956-10-S1-S2 *New in 2012*
8. [CL] SAMATOVA NF, OSTROUCHOV G, GEIST A, MELECHKO AV., “RACHET: An efficient cover-based merging of clustering hierarchies from distributed datasets.” *Distrib. Parallel Databases*, vol. 11, no. 2, pp. 157-180, Mar 2002, PMID: 16119272.
9. [CL] SAMATOVA NF, POTOK TE, LEUZE MR, “Vector space model for the generalized parts grouping problem.” *Robot. Comput.-Integr. Manuf.*, vol. 17, no. 1-2, pp. 73-80, Feb-Apr 2001.



## Refereed Journal Publications

10. [CE] SEMAZZI, F., WANIHA, P., BELLO, G., HARLALKA, J., SMITH, K., CHEN, Z., KUMAR, V., AND SAMATOVA, N.F., “The Role of the Greater Horn of Africa in Controlling Atlantic Hurricane Variability.” *Science* (Submitted). New in 2012
11. [E] GOWTHAM ATLURI; KANCHANA PADMANABHAN; GANG FANG; JEFFREY R PETRELLA; MICHAEL STEINBACH; KELVIN O LIM; NAGIZA F SAMATOVA; ANGUS McDONALD III; VIPIN KUMAR, “Complex Biomarker Discovery in Neuroimaging Data: Finding a Needle in a Haystack.” *NeuroImage: Clinical* (Submitted). New in 2012
12. [CE] GONZALEZ II, D.L., TETTEH, I.K., PENDSE, S.V., ANGUS, M., SRINIVAS, S., PADMANABHAN, K., YU, J., SEMAZZI, F., KUMAR, V., F. SAMATOVA, N.F., “On the Data-driven Inference of Causal Networks of Key Dynamic Factors Affecting West African Climate.” *Journal of Applied Meteorology and Climatology* (Submitted). New in 2012
13. [CE] JENKINS, J., DINAN, J., BALAJI, P., PETERKA, T., SAMATOVA, N.F., AND THAKUR, R., “MPI Derived Datatypes Processing on Noncontiguous GPU-resident Data.” *IEEE Transactions on Parallel and Distributed Systems (TPDS)* (Submitted). New in 2012
14. [CE] TIAN, W. AND SAMATOVA, N.F., ‘A Multiple Alignment Method for Large Interaction Networks by Fast Identification of Maximal Conserved Patterns.’ *Algorithms for Molecular Biology*, 2013 (Submitted). New in 2012
15. [CE] CHEN, Z., HENDRIX, W., GUAN, H., TETTEH, I.K., CHOUDHARY, A., SEMAZZI, F., AND SAMATOVA, N.F., “Discovery of extreme events-related communities in contrasting groups of physical system networks.” *Data Mining and Knowledge Discovery*, 2012, pp. 1-34, DOI: 10.1007/s10618-012-0289-3. New in 2012
16. [CE] TIAN, W. AND SAMATOVA, N.F., “Global Alignment of Pairwise Protein Interaction Networks For Maximal Common Conserved Patterns.” *International Journal of Genomics*, 2013, doi:10.1155/2013/670623. New in 2012
17. [E] MELESHKO, S. V., SAMATOVA, N. F., MELECHKO, A. V., “Group analysis of the thin film dewetting equation.” *International Journal of Non-Linear Mechanics*, 2012, 47(1), pp. 9-13. New in 2012
18. [CE] SCHMIDT, M.C., ROCHA, A.M., PADMANABHAN, K., SHPANSKAYA, Y., BANFIELD, J., SCOTT, K., MIHELICIC, J.R., AND SAMATOVA, N.F. “NIBBS-search for fast and accurate prediction of phenotype-biased metabolic systems.” *PLoS Computational Biology*, 2012 (DOI: 10.1371/journal.pcbi.1002490).
19. [CE] PADMANABHAN, K., WANG, K., AND SAMATOVA, N.F., “Functional annotation of hierarchical modularity.” *PLoS ONE*, 2012 (DOI:10.1371/journal.pone.0033744).
20. [CE] CHEN, Z., PADMANABHAN, K., ROCHA, A.M., SHPANSKAYA, Y., MIHELICIC, J.R., AND SAMATOVA, N.F., “SPICE: Discovery of phenotype-determining component interplays.” *BMC Systems Biology*, 2012, 6(1):40 doi:10.1186/1752-0509-6-40.
21. [CE] CHEN, Z., HENDRIX, W., AND SAMATOVA, N.F., “Community-based anomaly detection in evolutionary networks.” *Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies*, 2011 (DOI: 10.1007/s10844-011-0183-2).
22. [CE] SCHMIDT, M.C., ROCHA, A.M., PADMANABHAN, K., CHEN, Z., SCOTT, K., MIHELICIC, J.R., AND SAMATOVA, N.F., “Efficient alpha,beta-motif finder for identification of phenotype-related functional modules.” *BMC Bioinformatics*, 2011, 12:440.

23. [CE] HENDRIX, W., ROCHA, A.M., PADMANABHAN, K., CHOUDHARY, A., SCOTT, K., MIHELICIC, J.R., AND SAMATOVA, N.F., "DENSE: efficient and prior knowledge-driven discovery of phenotype-associated protein functional modules." *BMC Systems Biology*, 2011, 5:172.
24. [CE] HENDRIX, W., SCHMIDT, M., BREIMYER, P., SAMATOVA, N.F., "Theoretical underpinnings for maximal clique enumeration on perturbed graphs." *Theor. Comput. Sci.*, 411(26-28): 2520-2536 (2010), doi: <http://dx.doi.org/10.1016/j.tcs.2010.03.011>.
25. [CE] CHEN, W., SCHMIDT, M., SAMATOVA, N.F., "On the parameterized complexity of the Multi-MCT and Multi-MCST problems." *J. Comb. Optim.*, 21(2): 151-158, 2011, doi 10.1007/s10878-009-9220-2.
26. [M] KARPINETS, T.V., ROMINE, M.F., SCHMOYER, D.D., KORA, G., SYED, M.H., LEUZE, M.R., SERRES, M.H., PARK, B.H., SAMATOVA, N.F. AND UBERBACHER, E.C., "Shewanella knowledgebase: integration of the experimental data and computational predictions suggests a biological role for transcription of intergenic regions." *Database*, 2010 Jul 6;2010:baq012, PMID: 20627862.
27. [M] BELNAP, C.P., PAN, C., DENEFF, V.J., SAMATOVA, N.F., HETTICH, R.L., BANFIELD, J.F., "Quantitative proteomic analyses of the response of acidophilic microbial communities to different pH conditions." *International Society for Microbial Ecology Journal (ISME J.)*, 2011 Jan 13, PMID: 21228889.
28. [M] LI, J., MA, X., YOGINATH, S., KORA, G., SAMATOVA, N.F., "Transparent runtime parallelization of the R scripting language." *J. Parallel Distrib. Comput.*, 71(2): 157-168 (2011).
29. [M] LIN, H., MA, X., FENG, W., SAMATOVA, N.F., "Coordinating computation and I/O in massively parallel sequence search." *IEEE Transactions on Parallel and Distributed Systems*, 22(4): 529-543, 2011, doi <http://doi.ieeecomputersociety.org/10.1109/TPDS.2010.101>.
30. [M] BELNAP, C.P., PAN, C., VERBERKMOES, N.C., POWER, M.E., SAMATOVA, N.F., CARVER, R.L., HETTICH, R.L., BANFIELD, J.F., "Cultivation and quantitative proteomic analyses of acidophilic microbial communities." *International Society for Microbial Ecology Journal (ISME J.)*, 4(4):520-30, 2009, PMID: 2003306 (**Impact factor**: 5.029).
31. [M] YANG, S., PAPPAS K.M., HAUSER, L.J., LAND, M.L., CHEN, G.-L., HURST, G.B., PAN, C., KOUVELIS V.N., PELLETIER, D.A., KLINGEMAN, D.M., CHANG, Y.-J., SAMATOVA, N.F., AND BROWN, S.D., "Improved genome annotation for *Zymomonas mobilis*." *Nature Biotechnology*, 27(10):893-4, 2009, PMID: 19816441 (**Impact factor**: 22.297).
32. [CE] PAN, C., PARK, B.H., McDONALD, H.W., BANFIELD, J.F., VERBERKMOES, N.C., HETTICH, R.L., SAMATOVA, N.F., "A high-throughput de novo sequencing approach for shotgun proteomics using high-resolution tandem mass spectrometry." *BMC Bioinformatics*, Mar 5;11(18):118, 2010, PMID: 20205730, (**Impact factor**: 3.78).
33. [CE] BREIMYER P., GREEN N., KUMAR V., SAMATOVA, N.F., "BioDEAL: community generation of biological annotations." *BMC Med Inform Decis Mak.*, 2009 Nov 3;9 Suppl 1:S5 PMID: 19891799.
34. [M] RAMAN B, PAN C, HURST GB, RODRIGUEZ M JR, MCKEOWN CK, LANKFORD PK, SAMATOVA NF, MIELENZ JR, "Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis." *PLoS One*, 4(4):e5271, 2009, PMID: 19384422.
35. [CE] KARPINETS TV, PELLETIER DA, PAN C, UBERBACHER EC, MELNICHENKO GV, HETTICH RL, SAMATOVA NF, "Phenotype fingerprinting suggests the involvement of single-genotype

- consortia in degradation of aromatic compounds by *Rhodopseudomonas palustris*.” *PLoS One*, 4(2):e4615, 2009, PMID: 19242537.
36. [CE] CHANG, C. S., KU, S., DIAMOND, P. H., LIN, Z., PARKER, S., HAHM, T. S., SAMATOVA, N., “Compressed ion temperature gradient turbulence in diverted tokamak edge.” *PHYSICS OF PLASMAS*, 16(5), Article Number: 056108, 2009 (**Impact factor**: 2.427).
  37. [CE] SCHMIDT, M.C., SAMATOVA, N.F., THOMAS, K., PARK, B.-H., “A Scalable, parallel algorithm for maximal clique enumeration.” *J. of Parallel and Distributed Computing*, 69(4), pp. 417-428, April, 2009.
  38. [CE] CHEN, W., SCHMIDT, M.C., SAMATOVA, N.F., “On parameterized complexity of the Multi-MCS problem.” *Theoretical Computer Science*, Volume 410, Issues 21-23, 17 May 2009, Pages 2024-2032.
  39. [E] PAN, C., ODA, Y., LANKFORD, P.K., ZHANG, B., SAMATOVA, N.F., PELLETIER, D.A., HARWOOD, C.S., HETTICH, R.L., “Characterization of anaerobic catabolism of *p*-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics.” *Mol Cell Proteomics*, vol. 7, no. 5, pp. 938-48, 2008, PMID: 18156135 (**Impact factor**: 9.62).
  40. [CE] ZHANG, B., PARK, B.H., KARPINETS, T., SAMATOVA, N.F., “From pull-down data to protein interaction networks and complexes with biological relevance.” *Bioinformatics*, vol. 24, no. 7, pp. 979-86, 2008, PMID: 18304937 (**Impact factor**: 4.9).
  41. [E] PARK BH, OSTROUCHOV G, SAMATOVA NF., “Sampling streaming data with replacement.” *Comput. Stat. Data Anal.*, vol. 52, no. 2, pp. 750-762, Oct 2007, PMID: 18304937 (**Impact factor**: 0.93).
  42. [CE] SISNEROS, R., JONES, C., HUANG, J., GAO, J., PARK, B.H., SAMATOVA, N.F., “A multi-level cache model for run-time optimization of remote visualization.” *IEEE Trans Vis Comput Graph*, vol. 13, no. 5, pp. 991-1003, Sep-Oct 2007, PMID: 17622682 (**Impact factor**: 1.8).
  43. [M] KARPINETS T, GREENWOOD DJ, POGRIBNY I, SAMATOVA N., “Bacterial stationary-state mutagenesis and Mammalian tumorigenesis as stress-induced cellular adaptations and the role of epigenetics.” *Curr Genomics*, vol. 7, no. 8, pp. 481-96, Nov 2006, PMID: 18369407 (**Impact factor**: 0.79).
  44. [CE] ZHANG B, VERBERKMOES NC, LANGSTON MA, UBERBACHER E, HETTICH RL, SAMATOVA NF., “Detecting differential and correlated protein expression in label-free shotgun proteomics.” *J Proteome Res.*, vol. 5, no. 11, pp. 2909-18, Nov 2006, PMID: 17081042 (**Impact factor**: 5.2).
  45. [CE] PAN C, KORA G, McDONALD WH, TABB DL, VERBERKMOES NC, HURST GB, PELLETIER DA, SAMATOVA NF, HETTICH RL., “ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation.” *Anal Chem.*, vol. 78, no. 20, pp. 7121-31, Oct 2006, PMID: 17037911 (**Impact factor**: 5.45).
  46. [CE] PAN C, KORA G, TABB DL, PELLETIER DA, McDONALD WH, HURST GB, HETTICH RL, SAMATOVA NF., “Robust estimation of peptide abundance ratios and rigorous scoring of their variability and bias in quantitative shotgun proteomics.” *Anal Chem.*, vol. 78, no. 20, pp. 7110-20, Oct 2006, PMID: 17037910 (**Impact factor**: 5.45).
  47. [E] AUSTIN DW, ALLEN MS, MCCOLLUM JM, DAR RD, WILGUS JR, SAYLER GS, SAMATOVA NF, COX CD, SIMPSON ML., “Gene network shaping of inherent noise spectra.” *Nature*, vol. 439, no. 7076, pp. 608-11, Feb 2006, PMID: 16452980 (**Impact factor**: 26.7).

48. [CE] McCOLLUM JM, PETERSON GD, COX CD, SIMPSON ML, SAMATOVA NF., “The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior.” *Comput Biol Chem.*, vol. 30, no. 1, pp. 39-49, Feb 2006, PMID: 16321569 (**Impact factor**: 2.14).
49. [CE] YU GX, PARK BH, CHANDRAMOHAN P, GEIST A, SAMATOVA NF., “An evolution-based analysis scheme to identify CO<sub>2</sub>/O<sub>2</sub> specificity-determining factors for Ribulose 1,5-bisphosphate carboxylase/oxygenase.” *Protein Eng Des Sel.*, vol. 18, no. 22, pp. 589-96, Oct 2005, PMID: 16246824 (**Impact factor**: 3.0).
50. [CE] YU GX, PARK BH, CHANDRAMOHAN P, MUNAVALLI R, GEIST A, SAMATOVA NF., “In silico discovery of enzyme-substrate specificity-determining residue clusters.” *J Mol Biol.* , vol. 352, no. 5, pp. 1105-17, Oct 2005, PMID: 16140329 (**Impact factor**: 4.9).
51. [E] OSTROUCHOV G, SAMATOVA NF., “On FastMap and the convex hull of multivariate data: toward fast and robust dimension reduction.” *IEEE Trans Pattern Anal Mach Intell.*, vol. 27, no. 8, pp. 1340-3, Aug 2005, PMID: 16119272 (**Impact factor**: 4.3).
52. [CE] SUTERS WH, ABU-KHZAM FN, ZHANG Y, SYMONS CT, SAMATOVA NF, LANGSTON MA, “A new approach and faster exact methods for the maximum common subgraph problem.” *Lecture Notes in Computer Science*, vol. 3595, pp. 717-727, 2005.
53. [E] HEFFELFINGER GS, MARTINO A, GORIN A, XU Y, RINTOUL MD 3RD, GEIST A, AL-HASHIMI HM, DAVIDSON GS, FAULON JL, FRINK LJ, HAALAND DM, HART WE, JAKOBSSON E, LANE T, LI M, LOCASCIO P, OLKEN F, OLMAN V, PALENIK B, PLIMPTON SJ, ROE DC, SAMATOVA NF, SHAH M, SHOSHANI A, STRAUSS CE, THOMAS EV, TIMLIN JA, XU D., “Carbon sequestration in *Synechococcus Sp.*: from molecular machines to hierarchical modeling.” *OMICS*, vol. 6, no. 4, pp. 305-30, 2002, PMID: 12626091.
54. [M] MELECHKO AV, SIMKIN MV, SAMATOVA NF, BRAUN J, PLUMMER EW , “Complex structural phase transition in a defect-populated two-dimensional system.” *PHYSICAL REVIEW B*, vol. 64, no. 23, pp. 157-180, Dec 2001 (**Impact factor**: 3.1).
55. [CL] SAMATOVA NF, “On the nonhomogeneity index of graphs of polytopes.” *Diskr. Mat.*, vol. 6, no. 3, pp. 89-93, 1994, Moscow, Russia.
56. [CL] SAMATOVA NF, “On the nonhomogeneity index of graphs of polytopes.” *Discrete Mathematics and Applications*, vol. 4, no. 4, pp. 371-375, 1994, Moscow, Russia.
57. [CL] SAMATOVA NF, PULATOV AK., “On the measure of combinatorial instability of the specification of a convex polyhedron in  $R^3$ .” *Diskr. Mat.*, vol. 5, no. 3, pp. 150-156, 1993, Moscow, Russia.
58. [CL] SAMATOVA NF, PULATOV AK., “The estimation of the measure of combinatorial instability of a convex 3-polytope under limited deformations.” *UzNIINTI* (Deposited), 1993, Tashkent, Uzbekistan, N 1798 - Uz93.
59. [CL] SAMATOVA NF, “On the complexity of specification and combinatorial irregularity of convex 3-polyhedrons.” *Thesis for a Candidate’s degree* (Computing Center of Russian Academy of Sciences, 1993, Moscow, Russia).
60. [CL] SAMATOVA NF, “On the measure of combinatorial unsteadiness of a 3-polytope under linear deformations.” *UzNIINTI* (Deposited), 1993, Tashkent, Uzbekistan, N 1799 - Uz93.
61. [CL] SAMATOVA NF, SHIROKOVA OB, PULATOV AK., “Maximum values of some characteristics for vertex cover in a 3-polytope.” *Methods of discrete analysis in graph theory and complexity*, vol. 52, pp. , 1992, Novosibirsk, Russia.

62. [L] PULATOV AK, SAMATOVA NF, “Complexity of the specification of a convex polyhedron in  $R^3$ .” *Diskr. Mat.*, vol. 3, no. 2, pp. 148-156, 1991.

#### Invited Refereed Conference Publications

63. [CL] LAKSHMINARASIMHAN, S., KUMAR, P., LIAO, W., CHOUDHARY, A., KUMAR, V., AND SAMATOVA, N.F., “On the Path to Sustainable, Scalable, and Energy-efficient Data Analytics: Challenges, Promises, and Future Directions.” 2012 International Green Computing Conference (IGCC), San Jose, CA, U. S. A, 4-8 June 2012. New in 2012
64. [CL] SAMATOVA, N.F., BREIMYER, P., HENDRIX, W., SCHMIDT, M., AND RHYNE, T.-M., “An Outlook into Ultra-Scale Visualization Large-Scale Biological Data.” *Proceedings of the Supercomputing 2008 Conference, workshop on Ultrascale Visualization*, November 15-21, 2008, Austin, Texas, USA
65. [CL] SAMATOVA, N.F., SCHMIDT, M., HENDRIX, W., BREIMYER, P., THOMAS, K., PARK, B.-H., “Coupling graph perturbation theory with scalable parallel algorithms for large-scale enumeration of maximal cliques in biological graphs.” *Journal of Physics: Conference Series – SciDAC 2008*, v. 125, July 13-17, 2008, Seattle, Washington, USA
66. [CE] PARK BH, SAMATOVA NF, KARPINETS T, JALLOUK J, MOLONY S, HORTON S, ARCAN- GELI S, “Data-driven, data-intensive computing for modelling and analysis of biological networks: application to bioethanol production.” *Journal of Physics: Conference Series – SciDAC 2007*, v. 78, June 24-28, 2007, Boston, Massachusetts, USA
67. [E] SHOSHANI, A., ALTINTAS, I., CHOUDHARY, A., CRITCHLOW, T., KAMATH, C., LUDSCHER, B., NIEPLOCHA, J., PARKER, S., ROSS, R., SAMATOVA, N.F., VOUK, M., “SDM Center Technologies for Accelerating Scientific Discoveries.” *Journal of Physics: Conference Series – SciDAC 2007*, v. 78, June 24-28, 2007, Boston, Massachusetts, USA
68. [CL] SAMATOVA, N.F., BRANSTETTER, M., GANGULY, A.R., HETTICH, R., KHAN, S., KORA, G., LI, J., MA, X., PAN, C., SHOSHANI, A., S. YOGINATH, “High performance statistical computing with parallel R: Applications to biology and climate.” *Journal of Physics: Conference Series – SciDAC 2006*, v. 46, p. 505-509, 2006, Boston, Massachusetts, USA
69. [E] OSTROUCHOV G, SAMATOVA NF, “Embedding methods and robust statistics for dimension reduction.” *COMPSTAT 2004 Proceedings in Computational Statistics*, Physica-Verlag, A Springer Company, 2004, p359-370.
70. [E] BETHEL W, ABRAM G, SHARF J, FRANK R, AHRENS J, SAMATOVA NF, MILLER M, “Interoperability of visualization software and data models is not an achievable goal.” In *Proceedings of the IEEE Visualization*, Seattle, Washington, October 19-24, 2003, p. 607-610.

#### Refereed Conference Publications

71. [CE] LAKSHMINARASIMHAN, S., BOYUKA II, D.A., PENDSE, S.V., ZOU, X., JENKINS, J., VIS- MANATH, V., PAPKA, M.E. AND SAMATOVA, N.F., “Scalable In Situ Scientific Data Encoding for Analytical Query Processing.” *ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC)*, New York City, June, 2013 (**Nominated for Best Paper Award**). New in 2012
72. [CE] E. R. SCHENDEL, S. V. PENDSE, J. JENKINS, D. A. BOYUKA II, Z. GONG, S. LAKSHMI- NARASIMHAN, Q. LIU, S. KLASKY, R. ROSS, N. F. SAMATOVA, “ISOBAR Hybrid Compression- I/O Interleaving for Large-scale Parallel I/O Optimization.” *The 21st International ACM Sym- posium on High Performance Parallel and Distributed Computing*, Delft, The Netherlands, pp. 6172, June 18-22, 2012 (**Nominated for Best Paper Award**). New in 2012

73. [CE] JENKINS, J., SCHENDEL, E., LAKSHMINARASIMHAN, S., BOYUKA, D.A., ROGERS, T., ETHIER, S., ROSS, R., KLASKY, S., AND SAMATOVA, N.F., “Byte-precision Level of Detail Processing for Variable Precision Analysis”. *ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, Salt Lake City, Utah, U.S.A., pg. 1-11, November, 2012. New in 2012
74. [E] JENKINS, J., DINAN, J., BALAJI, P., SAMATOVA, N.F., AND THAKUR, R., “Enabling fast, non-contiguous GPU data movement in hybrid MPI+GPU environments.” *IEEE International Conference on Cluster Computing (Cluster)*, Beijing, China, pg. 468-476, September, 2012. New in 2012
75. [CE] JENKINS, J., ARKATKAR, I., LAKSHMINARASIMHAN, S., SHAH, N., SCHENDEL, E.R., ETHIER, S., CHANG, C.S., CHEN, J., KOLLA, H., KLASKY, S., ROSS, R. AND SAMATOVA, N.F., “Analytics-driven Lossless Data Compression for Rapid In-situ Indexing, Storing, and Querying.” *23rd International Conference on Database and Expert Systems Applications (DEXA)*, Vienna, Austria, pg. 16-30, 3-6 September, 2012. New in 2012
76. [CE] GONG, Z., ROGERS, T., JENKINS, J., KOLLA, H., ETHIER, S., CHEN, J., ROSS, R., KLASKY, S., AND SAMATOVA, N.F., “MLOC: Multi-level layout optimization framework for compressed scientific data exploration with heterogeneous access patterns.” *41st International Conference on Parallel Processing (ICPP 2012)*, Pittsburgh, PA, pp. 239-248, September 10-13, 2012. New in 2012
77. [CE] GONG, Z., BOYUKA II, D.A., ZOU, X., LIU, Q., PODHORSZKI, N., KLASKY, S., MA, X., AND SAMATOVA, N.F., “PARLO: PARallel Run-time Layout Optimization for Scientific Data Explorations with Heterogeneous Access Patterns.” *13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, Delft, The Netherlands, May, 2013. New in 2012
78. [CE] SHAH, N., SCHENDEL, E., LAKSHMINARASIMHAN, S., PENDSE, S.V., ROGERS, T., AND SAMATOVA, N.F., “Improving I/O Throughput with PRIMACY: Preconditioning ID-Mapper for Compressing Incompressibility.” *IEEE International Conference on Cluster Computing (Cluster)*, Beijing, China, September 2012. New in 2012
79. [E] CHEN, Z., AGRAWAL, A., LIAO, W-K., SAMATOVA, N.F. AND CHOUDHARY, A., “Forecast Oriented Classification of Spatio-Temporal Extreme Events.” *Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI 2013)*, Beijing, China, August, 2013. New in 2012
80. [CE] CHEN, Z., JENKINS, J., RAO, J., CHOUDHARY, A., SEMAZZI, F., MELECHKO, A.V., KUMAR, V. AND SAMATOVA, N.F., “Automatic Detection and Correction of Multi-class Classification Errors Using System Whole-part Relationships.” *13th SIAM International Conference on Data Mining (SDM 2013)*, Austin, Texas, USA, May 2-4, 2013. New in 2012
81. [CE] JIN, Y., LAKSHMINARASIMHAN, S., SHAH, N., GONG, Z., CHANG, C.S., CHEN, J., ETHIER S., KOLLA, H., KU, S., KLASKY, S., LATHAM, R., ROSS, R., SCHUCHARDT, K., AND SAMATOVA, N.F., “S-preconditioner for multi-fold data reduction with guaranteed user-controlled accuracy.” *IEEE International Conference on Data Mining (ICDM 2011)*, December 11-14, 2011, Vancouver, Canada.
82. [CE] SCHENDEL, E.R., JIN, Y., SHAH, N., CHEN, J., CHANG, C.S., KU, S-H., ETHIER, S., KLASKY, S., LATHAM, R., ROSS, R., AND SAMATOVA, N.F., “ISOBAR preconditioner for effective and high-throughput lossless data compression.” *The 28<sup>th</sup> IEEE International Conference on Data Engineering (ICDE 2012)*, April 1-5, 2012, Washington DC.
83. [CE] SENCAN, H., CHEN, Z., HENDRIX, W., PANSOMBUT, T., SEMAZZI, F., CHOUDHARY, A., KUMAR, V., MELECHKO, A. V., AND SAMATOVA, N. F., “Classification of Emerging Extreme

- Event Tracks in Multi-Variate Spatio-Temporal Physical Systems Using Dynamic Network Structures: Application to Hurricane Track Prediction.” *Proc. of the 22nd Int’l Joint Conference on Artificial Intelligence (IJCAI 2011)*, July 16-22, 2011, Barcelona, Spain.
84. [CE] PANSOMBUT, T., HENDRIX, W., GAO, Z., HARRISON, B.E., AND SAMATOVA, N. F., “Biclustering-Driven Ensemble of Bayesian Belief Network Classifiers for Underdetermined Problems.” *Proc. of the 22nd Int’l Joint Conference on Artificial Intelligence (IJCAI 2011)*, July 16-22, 2011, Barcelona, Spain.
  85. [CE] GONG, Z., LAKSHMINARASIMHAN, S., JENKINS, J., KOLLA, H., ETHIER S., CHEN, J., ROSS, R., KLASKY, S., AND SAMATOVA, N.F., “Multi-level layout optimization for efficient spatio-temporal queries on ISABELA-compressed data.” *The 26<sup>th</sup> IEEE International Parallel & Distributed Processing Symposium (IPDPS 2012)*, May 21-25, 2012, Shanghai, China.
  86. [CE] SCHENDEL, E.R., PENDSE, S.V., JENKINS, J., BOYUKA II, D.A., GONG, Z., LAKSHMINARASIMHAN, S., LIU, Q., KLASKY, S., ROSS, R., AND SAMATOVA, N.F., “ISOBAR hybrid compression-I/O interleaving for large-scale parallel I/O optimization.” *The 21st International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC 2012)*, Delft, the Netherlands, June 20-22, 2012.
  87. [CE] PENDSE, S.V., TETTEH, I.K., SEMAZZI, F., KUMAR, V., AND SAMATOVA, N.F., “Data-driven, semi-automatic inference of phenomenological physical models: Application to Eastern Sahel rainfall.” *The 2012 SIAM International Conference on Data Data Mining*, April 26-28, 2012, Anaheim, CA, US.
  88. [CE] LAKSHMINARASIMHAN, S., SHAH, N., ETHIER, S., KLASKY, S., LATHAM, R., ROSS, R., AND SAMATOVA, N.F., “Compressing the Incompressible with ISABELA: In-situ Reduction of Spatio-Temporal Data.” *Euro-Par*, August 29 - September 2, 2011, Bordeaux, France (**Distinguished paper award**).
  89. [CE] JENKINS, J., ARKATKAR, I., OWENS, J.D., CHOUDHARY, A., SAMATOVA, N.F., “Lessons Learned from Exploring the Backtracking Paradigm on the GPU.” *Euro-Par*, August 29 - September 2, 2011, Bordeaux, France.
  90. [CE] LAKSHMINARASIMHAN,S., JENKINS, J., ARKATKAR, I., GONG, Z., KOLLA, H., KU, S-H., ETHIER, S., CHEN, J., CHANG, C.S., KLASKY,S., LATHAM, R., ROSS, R., AND SAMATOVA, N.F., “ISABELA-QA:Query-driven Analytics with ISABELA-compressed Extreme-Scale ScientificData.” *Supercomputing (SC 2011)*, Novemer 12 - 18, 2011, Seattle, Washington.
  91. [CE] WILSON, K., ROCHA, A., PADMANABHAN, K., WANG, K., CHEN, Z., JIN, Y., MIHELICIC, J.R., AND SAMATOVA, N.F., “Detecting pathway cross-talks by analyzing conserved functional modules across multiple phenotype-expressing organisms.” *2011 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2011)*, pp. 443 - 449, 12-15 Nov 2011, Atlanta, GA.
  92. [E] KAWALE, J., LIESS, S., KUMAR, A., GANGULY, A., STEINBACH, M., SAMATOVA, N.F., SEMAZZI, F., SNYDER, P., AND KUMAR. V. “Data guided discovery of dynamic climate dipoles.” *NASA Conference on Intelligent Data Understanding (CIDU 2012)*, pp. 30-44, October 19-21, 2011, Sunnyvale, California (**Best Student Paper Award**).
  93. [CE] ARKATKAR, I., JENKINS, J., LAKSHMINARASIMHAN, S., SHAH, N., SCHENDEL, E.R., ETHIER, S., CHANG, CS, CHEN, J., KOLLA, H., KLASKY, S., ROSS, R., AND SAMATOVA, N.F., “Analytics-driven lossless data compression for rapid in-situ indexing, storing, and querying.” *The 23rd International Conference on Database and Expert Systems Applications (DEXA 2012)*, Vienna, Austria, September 3-7, 2012.

94. [CE] HENDRIX, W., ROCHA, A., ELMORE, M., TRIEN, J. AND SAMATOVA, N.F., “Discovery of enriched biological motifs using knowledge priors with application to biohydrogen production.” *The 2009 International Conference on Bioinformatics & Computational Biology (BIOCOMP 2010)*, pp. 17-23.
95. [CE] TIAN, W., SCHMIDT, M.C., CHEN, W. AND SAMATOVA, N.F., “Multiple alignment of large interaction networks by fast identifying maximal conserved patterns.” *The 2009 International Conference on Bioinformatics & Computational Biology (BIOCOMP 2010)*, pp. 178-184.
96. [CE] AWEKAR, A. AND SAMATOVA, N.F., “Parallel all pairs similarity search.” *The 2010 International Conference on Information and Knowledge Engineering (IKE 2010)*.
97. [CE] PANSOMBUT, T., HARRISON, B., BAHLER, D.R. AND SAMATOVA, N.F., “Improving the performance of Bayesian Belief Network classifiers via Decision Tree based feature selection.” *The 2010 International Conference on Information and Knowledge Engineering (IKE 2010)*, pp. 49-56.
98. [CE] SCHMIDT, M.C. AND SAMATOVA, N.F., “An algorithm for the discovery of phenotype related metabolic pathways.” *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2009)*, pp. 60-65.
99. [CE] CHEN, W., SCHMIDT, M.C., TIAN, W. AND SAMATOVA, N.F., “A fast, accurate algorithm for identifying functional modules through pairwise local alignment of protein interaction networks.” *The 2009 International Conference on Bioinformatics & Computational Biology (BIOCOMP 2009)*, pp. 816-821.
100. [CE] BREIMYER, P., HENDRIX, W., KORA, G. AND SAMATOVA, N.F., “pR: Lightweight, easy-to-use middleware to plugin parallel analytical computing with R.” *The 2009 International Conference on Information and Knowledge Engineering (IKE 2009)*, pp. 667-673.
101. [CE] BREIMYER, P., KORA, G., HENDRIX, W. AND SAMATOVA, N.F., “Web-enabled R for large-scale collaborative data mining: A survey.” *The 2009 International Conference on Information and Knowledge Engineering (IKE 2009)*, pp. 615-621.
102. [CE] BREIMYER, P., KORA, G., HENDRIX, W., SHAH, N. AND SAMATOVA, N.F., “pR: Automatic parallelization of data-parallel statistical computing codes for R in hybrid multi-node and multi-core environments.” *The 2009 IADIS International Conference Applied Computing (IADIS AC 2009)*, pp. 28-32.
103. [CE] AWEKAR, A., SAMATOVA, N.F., “Fast matching for all pairs similarity search.” *IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 295-300, September 15-18, 2009 in Milano, Italy.
104. [CE] AWEKAR, A., SAMATOVA, N.F. AND BREIMYER, P., “Incremental all pairs similarity search for varying similarity thresholds with reduced I/O overhead.” *The 2009 International Conference on Information and Knowledge Engineering (IKE 2009)*, pp. 687-693.
105. [CE] TIAN, W., SAMATOVA, N.F., “Pairwise Alignment of Interaction Networks by Fast Identification of Maximal Conserved Patterns.” *Pac Symp Biocomput (PSB2009)*, 2009:99-110, The Big Island of Hawaii, January 5-9, 2009, PMID: 19209698.
106. [CE] BREIMYER, P., GREEN, N., KUMAR, V., SAMATOVA, N.F., “BioDEAL: Biological data-evidence-annotation linkage system.” *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2008)*, Philadelphia, PA, USA, Nov. 7-9, 2008.



107. [CE] GREEN, N., BREIMYER, P., KUMAR, V., SAMATOVA, N.F., “WebBANC: Building Semantically-Rich Annotated Corpora from Web User Annotations of Minority Languages.” *Proceedings of the 17<sup>th</sup> Nordic Conference on Computational Linguistics (NODALIDA 2008)*, Odense, Denmark, May 14-16, 2009.
108. [CE] BRENT, P., GREEN, N., BREIMYER, P., KRISHNAMURTHY, R., SAMATOVA, N.F., “Systematic Evaluation of Convergence Criteria in Iterative Training for NLP.” *Proceedings of the 22<sup>nd</sup> International FLAIRS Conference (FLAIRS-22)*, Sanibel Island, Florida, USA, May 19-21, 2009.
109. [CE] PARK BH, ZHANG B, KARPINETS T, SAMATOVA NF, “Multi-stage Framework to Infer Protein Functional Modules from Mass Spectrometry Pull-Down Data with Assessment of Biological Relevance.” *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007)*, pp. 223-229, ISBN: 978-0-7695-3031-4 doi: 10.1109/BIBM.2007.52, Nov 2-4, 2007.
110. [E] ABU-KHZAM, F.N., RIZK, M.A., ABDALLAH, D.S., SAMATOVA N.F., “The Buffered Work-Pool Approach for Search-Tree Based Optimization Algorithms.” *Proceedings of the 7<sup>th</sup> International Conference on Parallel Processing and Applied Mathematics (PPAM 2007)*, Gdansk, POLAND, September 9-12, 2007.
111. [M] LIN H., MA X., LI J., YU T., SAMATOVA N.F., “Adaptive Request Scheduling for Parallel Scientific Web Services.” *Proceedings of the 20<sup>th</sup> International Conference on Scientific and Statistical Database Management (SSDBM '08)*, Hong Kong, Jul 9-11, 2008.
112. [CE] PARK BH, SAMATOVA NF, MUNAVALLI R, KRISHNAMURTHY R, KETTANI H, GEIST A, “Rapid and Robust Ranking of Text Documents in a Dynamically Changing Corpus.” *Proceedings of the 6<sup>th</sup> ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-08), Databases and Data Mining Track*, pp. , Doha, Qatar, Mar 31 - Apr 4, 2008
113. [CE] THOMAS K, SAMATOVA NF, SCHMIDT M, PARK BH, “Efficient Scaling Up of Parallel Graph Algorithms for Genome-Scale Biological Problems on Cray XT.” *Proceedings of the Cray User Group (CUG-08) Conference, Crossing the Boundaries*, Helsinki, May 5-8, 2008.
114. [CE] ABU-KHZAM FN, SAMATOVA NF, RIZK MA, LANGSTON MA., “The Maximum Common Subgraph Problem: Faster Solutions via Vertex Cover.” *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications (AICCSA07)*, pp. 367 - 373, 13-16 May 2007, Amman, Jordan, doi: 10.1109/AICCSA.2007.370907.
115. [CE] MA, X.; LI, J.; SAMATOVA, N.F., “Automatic Parallelization of Scripting Languages: Toward Transparent Desktop Parallel Computing.” *Proceedings of IEEE/ACS International Conference on Parallel and Distributed Processing Symposium (IPDPS 2007)*, pp. 1-6, 26-30 March 2007, doi: 10.1109/IPDPS.2007.370488.
116. [E] AUSTIN, D.; ALLEN, M.; MCCOLLUM, J.M.; DAR, R.D.; SAYLER, G.S.; SAMATOVA, N.F.; COX, C.D.; SIMPSON, M., “Gene network shaping of inherent noise spectra.” *Proceedings of the Bio Micro and Nanosystems Conference (BMN 2006)*, pp. 11-11, 15-18 Jan. 2006, doi: 10.1109/BMN.2006.330897.
117. [E] COX, C.D.; MCCOLLUM, J.M.; DAR, R.D.; AUSTIN, D.; ALLEN, M.S.; SAMATOVA, N.F.; SAYLER, G.S.; SIMPSON, M.L., “Calibration of a stochastic model of gene expression including feedback and extrinsic noise sources.” *Proceedings of the Bio Micro and Nanosystems Conference (BMN 2006)*, pp. 94-94, 15-18 Jan. 2006, doi: 10.1109/BMN.2006.330904.

118. [CE] SYMONS, C.T.; SAMATOVA, N.F.; KRISHNAMURTHY, R.; PARK, B.H.; TARIK UMAR; BUTTLER, D.; CRITCHLOW, T.; HYSOM, D., "Multi-Criterion Active Learning in Conditional Random Fields." *Proceedings of 8th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2006)*, pp. 323-331, Nov 13-15, 2006, Washington D.C., doi: 10.1109/ICTAI.2006.90.
119. [CE] PARK BH, KARPINETS T, ZHANG B, SAMATOVA NF, "Common-target model for identification of protein interaction modules in pull-down experiments." *Proceedings of the Annual Meeting of the Institute for Biological Engineering*, Mar 30-April 1, St. Louis, Missouri.
120. [CE] PAN C, KORA G, TABB D, McDONALD WH, PELLETIER D, HURST G, HETTICH R, SAMATOVA NF, "ProRata: a software package with improved point and interval estimation of protein abundance ratio for quantitative shotgun proteomics." *Proceedings of the 54th American Society for Mass Spectrometry*, Seattle, Washington, May 28 - June 1, 2006.
121. [E] PAN C, LANKFORD PK, ZHANG B, ODA Y, SAMATOVA NF, HARWOOD CS, PELLETIER DA, HETTICH RL, "Characterization of aromatic compound degradation pathways and their regulation in *Rhodospseudomonas palustris* using stable isotope labeling quantitative proteomics." *Proceedings of the 54th American Society for Mass Spectrometry*, Seattle, Washington, May 28 - June 1, 2006.
122. [CE] ZHANG Y, ABU-KHZAM FN, BALDWIN NE, CHESLE EJ, LANGSTON MA, SAMATOVA NF, "Genome-Scale Computational Approaches to Memory-Intensive Applications in Systems Biology." *Proceedings of the ACM/IEEE Supercomputing 2005 Conference (SC 2005)*, pp. 12-12, 12-18 Nov 2005, Seattle, Washington, doi: 10.1109/SC.2005.29.
123. [CE] LIN H, MA X, CHANDRAMOHAN P, GEIST A, SAMATOVA NF, "Efficient Data Access for Parallel BLAST." *Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005)*, pp. 72, 04-08 April 2005, doi: 10.1109/IPDPS.2005.190.
124. [CE] YU G.X., PARK BH, CHANDRAMOHAN P, GEIST A, SAMATOVA NF, "In-silico prediction of surface residue clusters for enzyme-substrate specificity." *Proceedings of the IEEE Computational Systems Bioinformatics Conference (CSB 2004)*, pp. 696-697, 16-19 Aug. 2004, doi: 10.1109/CSB.2004.1332548.
125. [CE] YOGINATH S, SAMATOVA NF, BAUER D, KORA G, FANN G, GEIST A, "RScaLAPACK: High-performance parallel statistical computing with R and ScaLAPACK." *Proceedings of the 18th International Conference on Parallel and Distributed Computing Systems (ICPDCS-2005)*, Sep 12-14, 2005, Las Vegas, Nevada.
126. [M] ABU-KHZAM FN, BALDWIN NE, LANGSTON MA, SAMATOVA NF, "On the relative efficiency of maximal clique enumeration algorithms, with application to high-throughput computational biology." *Proceedings of the International Conference on Research Trends in Science and Technology*, Beirut, Lebanon, March, 2005.
127. [CE] SUTERS WH, ABU-KHZAM FN, ZHANG Y, SYMONS CT, SAMATOVA NF, AND LANGSTON MA, "A new approach and faster exact methods for the maximum common subgraph problem." *Proceedings of the Computing and Combinatorics (COCOON)*, p. 717-727, 2005.
128. [E] COX CD, MCCOLLUM JM, AUSTIN DW, DAR RD, ALLEN MS, SAMATOVA NF, SIMPSON ML, "Estimation of spectral properties and molecular biokinetic parameters from stochastic gene expression data." *Proceedings of the Foundations of Systems Biology and Engineering (FOSBE)*, Santa Barbara, CA, August 2005.

129. [M] THURMON BP, MCCOLLUM JM, PETERSON GD, COX CD, SAMATOVA NF, SAYLER GS, SIMPSON ML, "Accelerating exact stochastic simulation using reconfigurable computing." *Proceedings of the Engineering of Reconfigurable Systems and Algorithms (ERSA)*, June 27-30, 2005, Monte Carlo Resort, Las Vegas, Nevada, USA.
130. [CE] PAN C, CHANDRAMOHAN P, PARK BH, MUNAVALLI RV, SAMATOVA NF, "Prediction of residue-residue contacts in domain interface by co-evolution." *Proceedings of the American Society for Microbiology 105th General Meeting*, Atlanta, GA, June 5, 2005.
131. [E] PAN C, HURST G, McDONALD H, PELLETIER D, SAMATOVA NF, HETTICH R, "Novel algorithms for extracting abundance ratios of stable isotope labeled peptide pairs from selected ion chromatograms in proteome samples." *Proceedings of the 53rd American Society for Mass Spectrometry*, San Antonio, Texas, June 2005.
132. [CE] PARK BH, OSTROUCHOV G, SAMATOVA NF, "Reservoir-based random sampling from data stream." *Proceedings of the Fourth SIAM International Conference on Data Mining*, Orlando, FL, April, 2004.
133. [CE] YU GX, PARK BH, CHANDRAMOHAN P, GEIST A, SAMATOVA NF, "In-silico prediction of surface residue clusters for enzyme-substrate specificity." *Proceedings of the 2004 IEEE Computational Systems Bioinformatics (CSB2004)*, Stanford, CA, August 16-19, 2004, p. 696-697.
134. [M] PAN C, VERBERKMOES N, CHANDRAMOHAN P, SAMATOVA NF, HETTICH RL, "Integration of nanoscale HPLC-FTICRMS and HPLC-QIT for accurate mass measurements and high throughput MS/MS to achieve enhanced proteome characterization." *Proceeding of the 52nd American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics*, 2004.
135. [CE] YU, G.-X.; OSTROUCHOV, G.; GEIST, A.; SAMATOVA, N.F., "An SVM-based algorithm for identification of photosynthesis-specific genome features." *Proceedings of the IEEE Computational Systems Bioinformatics Conference (CSB 2003)*, pp. 235 - 243, 11-14 Aug. 2003, doi: 10.1109/CSB.2003.1227323, PMID: 16452798.
136. [CE] PARK, B.-H.; OSTROUCHOV, G.; YU, G.-X.; GEIST, A.; GORIN, A.; SAMATOVA, N.F., "Inference of protein-protein interactions by unlikely profile pair." *Proceedings of the Third IEEE International Conference on Data Mining (ICDM 2003)*, pp. 735-738, 19-22 Nov. 2003.
137. [CE] ABU-KHZAM FN, SAMATOVA NF, OSTROUCHOV G, LANGSTON MA, GEIST GA, "Distributed dimension reduction algorithms for widely dispersed dataa." *Proceedings of the Fourteenth IASTED International Conference on Parallel and Distributed Computing and Systems (IASTED PDCS 2002)*, p. 167-174, 2002, ACTA Press.
138. [E] POTOK, T.E.; ELMORE, M.T.; REED, J.W.; SAMATOVA, N.F., "An ontology-based HTML to XML conversion using intelligent agents." *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS 2002)*, pp. 1220 - 1229, 7-10 Jan 2002.
139. [CL] SAMATOVA, N.F., POTOK, T.E.; LEUZE, M.R., "A vector perturbation approach to the generalized aircraft spare parts grouping." *Proceedings of the 10th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM 2002)*, pp. 1017-1026.
140. [CE] QU Y, OSTROUCHOV G, SAMATOVA NF, GEIST A, "Principal component analysis for dimension reduction in massive distributed data sets." *Proceedings of the Second SIAM International Conference on Data Mining*, p4-9, April 2002.
141. [E] POTOK, T.E.; IVEZIC N.E.; SAMATOVA, N.F., "Agent-based architecture for flexible lean cell design, analysis and evaluation." *Proceedings of the 4th Design of Information Infrastructure*

*Systems for Manufacturing Conference (DIISM 2001)*, 2001, pp181-188, Melbourne, Victoria, Australia.

#### Invited Refereed Workshop Publications

142. [CL] SAMATOVA NF, YU GX, PARK BH, GEIST A, OSTROUCHOV G, “From genomics to functional proteomics: in silico approach.” *A mini-symposium on Parallel Computational Biology (in conjunction with SIAM Conference on Parallel Processing for Scientific Computing*, February 25-27, 2004, San Francisco, CA.

#### Refereed Workshop Publications

143. [CE] GONZALEZ, D.L., CHEN, Z., TETTEH, I.K., PANSOMBUT, T., SEMAZZI, F., KUMAR, V., MELECHKO, A.V. AND SAMATOVA, N.F., “Hierarchical Classifier-Regression Ensemble for Multi-Phase Non-Linear Dynamic System Response Prediction: Application to Climate Analysis.” *2012 IEEE 12th International Conference on Data Mining (ICDM)*, pp. 781-788, December 10, 2012. Brussels, Belgium. *New in 2012*
144. [CE] HENDRIX, W., KARPINETS, T., PARK, B.-H., SCHENDEL, E., CHOUDHARY, A., AND SAMATOVA, N.F., “Sensitive and Specific Identification of Protein Complexes in “Perturbed” Protein Interaction Networks from Noisy Pull-Down Data.” *Tenth IEEE International Workshop on High Performance Computational Biology (HiCOMB) held in conjunction with the International Parallel and Distributed Processing Symposium (IPDPS)*, Anchorage, AK, May 16, 2011.
145. [CE] CHEN, W., ROCHA, A.M., HENDRIX, W., SCHMIDT, M.C., SAMATOVA, N.F., “The Multiple Alignment Algorithm for Metabolic Pathways without Abstraction.” *ICDM Workshops, ICDMW 2010, The 10th IEEE International Conference on Data Mining Workshops*, pp. 669-678, December 14-17, 2010, Sydney, Australia.
146. [CE] CHEN, Z. WILSON, K.A., JIN, Y., HENDRIX, W., AND SAMATOVA, N.F., “Detecting and Tracking Community Dynamics in Evolutionary Networks.” *IEEE ICDM Workshop on Social Interactions Analysis and Services Providers*, pp.318-327, December 14-17, 2010, Sydney, Australia.
147. [CE] GREEN, N., BREIMYER, P., VINAY AND SAMATOVA, N.F., “PackPlay: Mining semantic data in collaborative games.” *LAW IV '10 Proceedings of the Fourth Linguistic Annotation Workshop, Association for Computational Linguistics (ACL)*, Stroudsburg, PA, USA, 2010, pp. 227-234, ISBN: 978-1-932432-72-5.
148. [CE] HENDRIX, W., SCHMIDT, M.C., BREIMYER, P. AND SAMATOVA, N.F., “On perturbation theory and an algorithm for maximal clique enumeration in uncertain and noisy graphs.” *The ACM KDD Workshop on Knowledge Discovery from Uncertain Data*, 2010, pp. 48-56.
149. [CE] AWEKAR, A., SAMATOVA, N.F., “Incremental all pairs similarity search for varying similarity thresholds.” *The ACM International Conference on Knowledge Discovery and Data Mining, Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, 2009, doi: <http://doi.acm.org/10.1145/1731011.1731012>.
150. [CE] AWEKAR, A., SAMATOVA, N.F., “Fast matching for all pairs similarity search.” *IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 295-300, September 15-18, 2009 in Milano, Italy.
151. [CE] PARK B.H., SCHMIDT M., THOMAS T., KARPINETS, T., AND SAMATOVA NF, “Parallel, Scalable, Memory-Efficient Backtracking for Combinatorial Modeling of Large-Scale Biological

Systems.“ *Proceedings of the 22nd IEEE International Parallel and Distributed Processing Symposium (IPDPS 2008), the 7th IEEE International Workshop on High Performance Computational Biology (HiCOMB 2008)*, pp. , Miami, Florida USA, Apr 14-18, 2008

152. [CE] PARK BH, MUNAVALLI R, GEIST A, SAMATOVA NF, “Analysis of interaction site predictions from separated data spaces.” *Proceedings of the SIAM Bioinformatics Workshop (in conjunction with the Fourth SIAM International Conference on Data Mining)*, Orlando, FL, April, 2004.
153. [CE] PARK, B.-H.; SAMATOVA, N.F., OSTROUCHOV, G.; GEIST, A., “Xmap: Fast dimension reduction algorithms for multivariate streamline data.” *Proceedings of the 6th International Workshop on High Performance Data Mining: Pervasive and Data Stream Mining (in conjunction with Third International SIAM Conference on Data Mining)*, San Francisco, CA May 1-3, 2003.
154. [CL] SAMATOVA, N.F.; GEIST, A.; OSTROUCHOV, G.; MELECHKO, A.V., “Parallel out-of-core algorithm for genome-scale enumeration of metabolic systemic pathways.” *Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS 2002), the 1st Workshop on High Performance Computational Biology (HiCOMB2002)*, pp. 185-192, 15-19 April 2002, Fort Lauderdale, Florida,.
155. [CL] SAMATOVA NF, OSTROUCHOV G, GEIST A, MELECHKO AV, “RACHET: A new algorithm for mining multi-dimensional distributed datasets.” *Proceedings of the SIAM Third Workshop on Mining Scientific Datasets*, Chicago, IL, April 2001.

### Books and Book Chapters

156. [CL] EDS. SAMATOVA, N.F., HENDRIX, W., CHAKRABORTY, A., JENKINS, J., PADMANABHAN, K., “Practical Graph Mining with R.” CRC Press/Taylor and Francis, 2012 (In Press). *New in 2012*
157. [CE] PADMANABHAN, K., LAKSHMINARASIMHAN, S., GONG, Z., JENKINS, J., SHAH, N., SCHENDEL, E., ARKATKAR, I., ROSS, R., KLASKY, S., AND SAMATOVA, N.F., “In situ analysis in support of exploratory scientific discovery in data intensive science.” in *Data Intensive Science*, T. Chritchlow and K. K. Van Dam (editors), CRC Press/Taylor and Francis, 2012 (In Press). *New in 2012*
158. [CL] SAMATOVA, N.F. AND HENDRIX, W., “Introduction to Graph Data Mining.” in *Practical Graph Mining with R*, Samatova, N.F., Hendrix, W., Chakraborty, A., Jenkins, J., Padmanabhan, K. (editors), CRC Press/Taylor and Francis, 2012 (In Press). *New in 2012*
159. [CL] SAMATOVA NF, BREIMYER P, KORA G, PAN P, YOGINATH S, “Parallel R for High Performance Analytics: Applications to Biology.” in *Scientific Data Management*, A. Shoshani and D. Rotem (editors), CRC Press/Taylor and Francis, 2010, ISBN 978-1-4200-6980-8.
160. [CE] PARK BH, DAM P, PAN P, XU Y, GEIST A, HEFFELFINGER G, SAMATOVA NF, “In silico recognition of protein-protein interactions: theory and applications.” in *Advanced Data Mining Technologies in Bioinformatics*, Hui-Hwang Hsu (editor), Idea Group Inc., 2006.
161. [E] PULATOV AK, SAMATOVA NF, *Convex 3-polytopes*, Reference, Tashkent, Uzbekistan, 1992.

### Patents

162. [E] POTOK TE, ELMORE MT, REED JW, TREADWELL JN, SAMATOVA NF, “Method for gathering and summarizing internet information.” *US Patent 7,693,903* , issued on April 6, 2010.

163. [E] POTOK TE, ELMORE MT, REED JW, TREADWELL JN, SAMATOVA NF, "Method for gathering and summarizing internet information." *US Patent 7,315,858*, issued on January 1, 2008.
164. [E] POTOK TE, ELMORE MT, REED JW, SAMATOVA NF, AND TREADWELL JN, "System for gathering and summarizing internet information." *US Patent 7,072,883*, issued July 4, 2006.

#### **Invited Featured Articles**

165. [CL] SAMATOVA NF, GORIN A, UBERBACHER E, KARPINETS T, PARK BH, PAN C, STRAATSMA TP, CANNON W, RESAT H, LINS RD, OEHMEN C, "BioPilot: Data-driven computing for biological systems." *SciDAC Review*, v. 5, p. 10-25, Fall 2007.
166. [CL] SAMATOVA NF, SYMONS C, "Mentoring scientists of the future." *SciDAC Review*, v. 5, p. 39-43, Spring 2007.
167. [E] SHOSHANI A, CRICHLow T, ROSS R, SAMATOVA NF, "SDM Center: From data To discovery." *SciDAC Review*, v. 2, p. 28-39, Fall 2006.

#### **Invited Workshop Research Roadmap Reports**

168. [E] "U.S. Department of Energy Office of Science Systems Biology Knowledgebase for a New Era in Biology: A Genomics:GTL Workshop Report," Report from the DOE Genomics:GTL Systems Biology Knowledgebase Workshop, Editors: S. Gregurick, J. Frederickson, R. Stevens, February, 2009.
169. [E] "Mathematics for Analysis of Petascale Data: Report on a Department of Energy Workshop," Report from the Joint NSF/DOE Workshop, Editors: P. Kegelmeyer, R. Calderbank, R. Chritchlow, L. Jameson, C. Kamath, J. Meza, N.F. Samatova, A. Wilson, June, 2008, <http://www.science.doe.gov/ascr/ProgramDocuments/PetascaleDataWorkshopReport.pdf>.
170. [E] "The DOE OASCR Report of the Panel on Recent Significant Advancements in Computational Science Breakthroughs 2008," Anthony Mezzacappa, Panel Chair, September, 2008, <http://www.er.doe.gov/ascr/ProgramDocuments/Docs/Breakthroughs2008.pdf>.
171. [E] "Toward the Development of Predictive Theory and Modeling in Biology," Report from the Joint ASCAC/BERAC Subcommittee on Modeling and Simulation for GTL, Editors: T. Zacharia, R. Stevens,...., N.F. Samatova, et. al., February 2008.
172. [E] "Modeling and Simulation at the Exascale for Energy and the Environment," DOE Report on the Advanced Scientific Computing Research Town Hall Meetings on Simulation and Modeling at the Exascale for Energy, Ecological Sustainability, and Global Security (E3), Editors: H. Simon, T. Zacharia, R. Stevens, January 2008.
173. [M] "NSF Petascale Computing in the Biological Sciences," Editors: A. Snavely, D.Bader, G. Jacobs, August 2006.
174. [E] "DHS Data Integration and Dissemination Workshop Report," Editors: E. Stephen, N.F. Samatova, et. al, November 2005.
175. [E] "DHS Data Science Workshop Report," Editors: T. Kolda, N. Samatova, et. al, September 2004.
176. [E] "DHS Advanced Scientific Computing Requirements Workshop Report," Editors: S. Ashby,...., N.F. Samatova, et. al, October 2003.

177. [E] “DOE Office of Science Data-Management Workshops,” Editors: R. Mount,..., N.F. Samatova, et. al, March - May 2004.
178. [E] “NITRD High End Computing Revitalization Task Force (HECRTF): High End Computing for Full-Context Analysis and Visualization: When the Experiment is Done,” White paper by Ostrouchov G. and N.F. Samatova, June 2003.
179. [M] “DOE Visualization Frameworks Requirements Workshop,” Editors: R. Stevens, M. Papka,..., N.F. Samatova, et. al, June 2003.
180. [E] “DOE Report on Three Genomes to Life Workshops: Data Infrastructure, Modeling and Simulation, and Protein Structure Prediction,” Editors: T. Zacharia, R. Mann,..., N.F. Samatova, et. al, June 2003.
181. [L] “DOE OS Advanced Scientific Computing Research Accomplishments: Gleaning insight from scientific simulation data,” White paper by N.F. Samatova and Ostrouchov G., 2003.
182. [M] “DOE Report on the Mathematics Workshop for the Genomes to Life Program,” Editors: D. Brown, J. Guckenheimer, E. Ng,..., N.F. Samatova, et. al, March 2002.
183. [M] “DOE Report on the Computer Science Workshop for the Genomes to Life Program,” Editors: R. Bair, G. Johnson, J. Houghton, P. Karp, R. Stevens, B. Gropp,..., N.F. Samatova, et. al, March 2002.
184. [L] “DOE Report on the Computational Infrastructure Workshop for the Genomes to Life Program,” Editors: G. Heffelfinger, A. Geist,..., N.F. Samatova, et. al, January 2002.
185. [M] “DOE Report on the Computational Biology Workshop for the Genomes to Life Program,” Editors: M. Colvin, R. Mann, ..., Samatova NF, et. al., August, 2001

### Technical Reports

186. [CE] CHEN, Z., PANSOMBUT, A., HENDRIX, W., GONZALEZ, D., SEMAZZI, F., CHOUDHARY, A., KUMAR, V., MELECHKO, A.V., AND SAMATOVA, N.F., “Forecaster: Forecast Oriented Feature Elimination-based Classification of Adverse Spatio-Temporal Extremes.” *NCSU TR 1840.2/2408, 2011.*
187. [CE] A. AWEKAR, SAMATOVA NF, BREIMYER P, “Incremental All Pairs Similarity Search for Varying Similarity Thresholds.” *May 26, 2009: NCSU TR-2009-13.*
188. [CE] A. AWEKAR, SAMATOVA NF, “Fast Matching for All Pairs Similarity Search.” *May 26, 2009: NCSU TR-2009-12.*
189. [M] LIN H, MA X, LI J, SAMATOVA NF, TING Y, “Processor and Data Scheduling for Online Parallel Sequence Database Servers.” *August 2007: NCSU TR-2007-23.*
190. [CE] LI J, MA X, YOGINATH S, KORA G, SAMATOVA NF, “Automatic, Transparent Runtime Parallelization of the R Scripting Language.” *January 2007: NCSU TR-2007-3.*
191. [CL] SAMATOVA NF, PARK BH, KRISHNAMURTHY R, MUNAVALLI R, SYMONS C, BUTTLER DJ, COTTOM T, CRITCHLOW TJ, SLEZAK T, “Information extraction from unstructured text for the Biodefense Knowledge Center.” *2005: United States. p. 18p., UCRL-CONF-213354.*
192. [E] SLEZAK T, CRITCHLOW T, HAZLETT S, SAMATOVA NF, CHANDRAMOHAN P, KRISHNAMURTHY R, “Design of the National Bioforensics library infrastructure.” *2004: United States. p. 22p., UCRL-TR-202217.*

193. [E] P. AGARWAL, J. AHRENS, D. BAUER, R. BLECK, P. CHANDRAMOHAN, T. DUNIGAN, G. FANN, M. GUEST, F. JAEGER, G. KORA, R. LATHAM, V. MEUNIER, L. OLIKER, R. ROSS, N. SAMATOVA, A. SHOSHANI, S. STUDHAM, B. SUMPTER, R. THAKUR, R. TOEDTE, J. S. VETTER, P. H. WORLEY, AND D. XU, "Evaluation of the SGI Altix 3700 at Oak Ridge National Laboratory." *ORNL Technical Report, November, 2004.*
194. [E] BURRIS RD, CHOLIA S, DUNIGAN TH, FOWLER FM, GLEICHER MK, HOLMES HH, JOHNSTON NE, MEYER N, MILLION DL, OSTROUCHOV G, SAMATOVA NF, "Probe project status & accomplishments." 2003, *RNL/TM-2003/140.*

### **Tutorials**

195. [CL] SAMATOVA, N.F., "Three-day Tutorial on Setting up a PC cluster for high performance computing." Presented at the *Joint Institute for Computational Science (JICS) Workshop Series* Summer, 1998, University of Tennessee.
196. [CL] SAMATOVA, N.F., "Three-day Tutorial on Getting supercomputing power from a build-it-yourself PC cluster." Presented at the *Joint Institute for Computational Science (JICS) Workshop Series* Summer, 1999, University of Tennessee.

### **Panels and Advisory Boards**

1. "DOE OBER Virtual Institute for Microbial Stress and Survival Center," External Advisory Board Meeting, February 11-12, 2009, Bethesda, MD.
2. "DOE OASCR Panel of Distinguished Members of the Computational Science Community to Identify Breakthroughs in Computational Science and Enabling Technologies across the Office of Sciences Scientific, Applied Mathematics, and Computer Science Research Programs," March – June, 2008.
3. "DOE Panel to Assess the Strategic Priorities of the Office of Advanced Scientific Computing Research Program," January, 2008.



## Invited Research and Programmatic Presentations (Sole and Contributed)

1. “Hurricanes,” NSF Site Review, Washington DC, October 15–17, 2012. *New in 2012*
2. “Accurate Forecasting of Adverse Spatio-Temporal Extreme Events,” Graph and Hypergraph Problems in Computational Science: Applications and Algorithms, July 14–21, 2012. *New in 2012*
3. “On the Path to Sustainable, Scalable, and Energy-efficient Data Analytics: Challenges, Promises, and Future Directions,” 2012 International Green Computing Conference (IGCC), San Jose, CA, U. S. A, 4-8 June 2012. *New in 2012*
4. “Scalable In Situ Scientific Data Encoding for Analytical Query Processing.” ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC), New York City, June, 2013 *New in 2012*
5. “Byte-precision Level of Detail Processing for Variable Precision Analysis,” ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), Salt Lake City, Utah, U.S.A., November, 2012. *New in 2012*
6. “Enabling fast, non-contiguous GPU data movement in hybrid MPI+GPU environments,” IEEE International Conference on Cluster Computing (Cluster), Beijing, China, September, 2012. *New in 2012*
7. “Analytics-driven Lossless Data Compression for Rapid In-situ Indexing, Storing, and Querying,” 23rd International Conference on Database and Expert Systems Applications (DEXA), Vienna, Austria, 3–6 September, 2012. *New in 2012*
8. “MLOC: Multi-level layout optimization framework for compressed scientific data exploration with heterogeneous access patterns,” The 41st International Conference on Parallel Processing (ICPP 2012), Pittsburgh, PA, September 10-13, 2012. *New in 2012*
9. “PARLO: PARallel Run-time Layout Optimization for Scientific Data Explorations with Heterogeneous Access Patterns.” 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Delft, The Netherlands, May, 2013. *New in 2012*
10. “Improving I/O Throughput with PRIMACY: Preconditioning ID-Mapper for Compressing Incompressibility,” IEEE International Conference on Cluster Computing (Cluster), Beijing, China, September 2012. *New in 2012*
11. “ISOBAR hybrid compression-I/O interleaving for large-scale parallel I/O optimization,” The 21st International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC 2012), Delft, the Netherlands, June 20-22, 2012. *New in 2012*
12. “End-game of the Hurricanes: Forecasting with Precision,” 2012 Strategic Advisory Board, Department of Computer Science, NC State University, April 12, 2012.
13. “Graph Theory and Climate Change,” The Spring 2012 Advanced Scientific Computing Advisory Committee meeting held March 27, 2012 at the American Geophysical Union - 2000 Florida Avenue, NW, Washington, DC.
14. “Understanding Climate Dynamics via Mining Climate Networks,” NSF Expeditions in Computing, Showcase visit at National Science Foundation, January 25, 2012.
15. “S-preconditioner for multi-fold data reduction with guaranteed user-controlled accuracy,” IEEE International Conference on Data Mining (ICDM 2011), December 14, 2011, Vancouver, Canada.
16. “ISOBAR preconditioner for effective and high-throughput lossless data compression,” The 28<sup>th</sup> IEEE International Conference on Data Engineering (ICDE 2012), April 1–5, 2012, Washington DC.

17. "Classification of Emerging Extreme Event Tracks in Multi-Variate Spatio-Temporal Physical Systems Using Dynamic Network Structures: Application to Hurricane Track Prediction," The 22nd Int'l Joint Conference on Artificial Intelligence (IJCAI 2011), July 16-22, 2011, Barcelona, Spain.
18. "Biclustering-Driven Ensemble of Bayesian Belief Network Classifiers for Underdetermined Problems," The 22nd Int'l Joint Conference on Artificial Intelligence (IJCAI 2011), July 16-22, 2011, Barcelona, Spain.
19. "Multi-level layout optimization for efficient spatio-temporal queries on ISABELA-compressed data," The 26<sup>th</sup> IEEE International Parallel & Distributed Processing Symposium (IPDPS 2012), May 21-25, 2012, Shanghai, China.
20. "Data-driven, semi-automatic inference of phenomenological physical models: Application to Eastern Sahel rainfall," The 2012 SIAM International Conference on Data Mining, April 26-28, 2012, Anaheim, CA, US.
21. "Lessons Learned from Exploring the Backtracking Paradigm on the GPU," Euro-par conference, September 1st, 2011, Bordeaux, France.
22. "ISABELA-QA: Query-driven Data Analytics over ISABELA-compressed Extreme-Scale Scientific Data," Supercomputing '11, November 15, 2011, Seattle, WA, USA.
23. "Scientific Data on the Path to Exascale: Lessons, Insights, and Predictions from 10+ Years on the Front Lines." Supercomputing '11, November 16, 2011, Seattle, WA, USA.
24. "Compressing the Incompressible with ISABELA: In-situ Reduction of Spatio-Temporal Data," Euro-par conference, September 1st, 2011, Bordeaux, France.
25. "Detecting Pathway Cross-talks by Analyzing Conserved Functional Modules across Multiple Phenotype-Expressing Organisms," IEEE International Conference Bioinformatics & Biomedicine (BIBM 2011), November 13th, 2011, Atlanta, GA.
26. "Data guided discovery of dynamic climate dipoles," NASA Conference on Intelligent Data Understanding (CIDU 2012), pp. 30-44, October 19-21, 2011, Sunnyvale, California.
27. "Computational Methodologies for Identification of Phenotype-specific Biological Processes in Microbial Communities," Genomic Sciences Contractor-Grantee Workshop, February 28th, 2012, Bethesda, MD.
28. "Detecting and Tracking Community Dynamics in Evolutionary Networks," The 10th IEEE International Conference on Data Mining Workshops: Social Interactions Analysis and Services Providers (SIASP), December 14th, 2010, Sydney, Australia.
29. "On perturbation theory and an algorithm for maximal clique enumeration in uncertain and noisy graphs." The ACM KDD Workshop on Knowledge Discovery from Uncertain Data, 2010.
30. "The Multiple Alignment Algorithm for Metabolic Pathways without Abstraction," The 10th IEEE International Conference on Data Mining Workshops: Biological Data Mining and its Applications, December 14th, 2010, Sydney, Australia.
31. "Approximation Algorithms for the Maximum Duo-preservation String Mapping Problem," The 5th IEEE International Conference on Future Information Technology, December 15th, 2010, Changsha, China.
32. "Sensitive and Specific Identification of Protein Complexes in "Perturbed" Protein Interaction Networks from Noisy Pull-Down Data." Tenth IEEE International Workshop on High Performance Computational Biology (HiCOMB) held in conjunction with the International Parallel and Distributed Processing Symposium (IPDPS), Anchorage, AK, May 16, 2011.

33. "PackPlay: Mining semantic data in collaborative games." LAW IV '10 Proceedings of the Fourth Linguistic Annotation Workshop, Association for Computational Linguistics (ACL), Stroudsburg, PA, USA, 2010.
34. "Discovery of enriched biological motifs using knowledge priors with application to biohydrogen production," The 2009 International Conference on Bioinformatics & Computational Biology (BIOCOMP 2010).
35. "Multiple alignment of large interaction networks by fast identifying maximal conserved patterns," The 2009 International Conference on Bioinformatics & Computational Biology (BIOCOMP 2010).
36. "Parallel all pairs similarity search," The 2010 International Conference on Information and Knowledge Engineering (IKE 2010).
37. "Improving the performance of Bayesian Belief Network classifiers via Decision Tree based feature selection." The 2010 International Conference on Information and Knowledge Engineering (IKE 2010).
38. "An algorithm for the discovery of phenotype-related metabolic pathways," IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2009).
39. "A fast, accurate algorithm for identifying functional modules through pairwise local alignment of protein interaction networks." The 2009 International Conference on Bioinformatics & Computational Biology (BIOCOMP 2009).
40. "pR: Lightweight, easy-to-use middleware to plugin parallel analytical computing with R." The 2009 International Conference on Information and Knowledge Engineering (IKE 2009).
41. "Web-enabled R for large-scale collaborative data mining: A survey." The 2009 International Conference on Information and Knowledge Engineering (IKE 2009).
42. "pR: Automatic parallelization of data-parallel statistical computing codes for R in hybrid multi-node and multi-core environments," The 2009 IADIS International Conference Applied Computing (IADIS AC 2009).
43. "Fast matching for all pairs similarity search," IEEE/WIC/ACM International Conference on Web Intelligence, September 15-18, 2009, Milano, Italy.
44. "Incremental all pairs similarity search for varying similarity thresholds with reduced I/O overhead," The 2009 International Conference on Information and Knowledge Engineering (IKE 2009).
45. "pR: Enabling Automatic Parallelization of Data-Parallel Tasks and Interfacing Parallel Computing Libraries in R with Application to Fusion Reaction Simulations," International user! Conference, July 21st, 2010, Gaithersburg, MD.
46. "Incremental all pairs similarity search for varying similarity thresholds." The ACM International Conference on Knowledge Discovery and Data Mining, Proceedings of the 3rd Workshop on Social Network Mining and Analysis, 2009
47. "An Outlook into Future Visualization of Large-Scale Biological Data," Supercomputing 2008 Conference, Ultrascale Visualization Workshop, November 16, Austin, TX.
48. "Coupling graph perturbation theory with scalable parallel algorithms for large-scale enumeration of maximal cliques in biological graphs," Presented at the DOE SciDAC 2008 Conference, July 13-17, 2008, Seattle, WA.
49. "Automatic Parallelization with pR," International Association for Development of the Information Society (IADIS) Applied Computing, December 21st, 2009, Rome, Italy.

50. "Pairwise Alignment of Interaction Networks by Fast Identification of Maximal Conserved Patterns." Pac Symp Biocomput (PSB2009), The Big Island of Hawaii, January 5-9, 2009.
51. "BioDEAL: Biological data-evidence-annotation linkage system," IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2008), Philadelphia, PA, USA, Nov. 7-9, 2008.
52. "WebBANC: Building Semantically-Rich Annotated Corpora from Web User Annotations of Minority Languages." The 17<sup>th</sup> Nordic Conference on Computational Linguistics (NODALIDA 2008), Odense, Denmark, May 14-16, 2009.
53. "Systematic Evaluation of Convergence Criteria in Iterative Training for NLP." The 22<sup>nd</sup> International FLAIRS Conference (FLAIRS-22), Sanibel Island, Florida, USA, May 19-21, 2009.
54. "Lessons learned from collaborations with GTL researchers," Presented to the DOE ASCAC-BERAC Committee, October 4, 2007, San Francisco, CA.
55. "Towards Uncovering Simplicity from Complexity in Biological Systems: Finding the Dots, Connecting the Dots, Understanding the Dots in Petascale Data," DOE Mathematics PI Meeting, May 22-24, 2007, Livermore, CA.
56. "Scientific Data Management Center: Technologies, Applications, and Future Needs," DOE ASCR PI Meeting, Mar 31 – Apr 2, 2008, Denver, CO.
57. "Scientific Data Management Center: Technologies and Applications," 2008 GTL-PI and ASCR-BER Bioenergy Research Center Workshop, Feb. 13, 2008, Rockville, MD.
58. "The Shewanella Knowledgebase," Genomics:GTL Awardee Workshop VI and Metabolic Engineering Working Group, Interagency Conference on Metabolic Engineering, Bethesda, Maryland February 1013, 2008.
59. "Comprehensive Integration of Regulatory Data in the Shewanella Knowledgebase," Genomics:GTL Awardee Workshop VI and Metabolic Engineering Working Group, Interagency Conference on Metabolic Engineering, Bethesda, Maryland February 1013, 2008.
60. "The Center for Molecular and Cellular Systems: Biological Insights from Large Scale Protein-Protein Interaction Studies," Genomics:GTL Awardee Workshop VI and Metabolic Engineering Working Group, Interagency Conference on Metabolic Engineering, Bethesda, Maryland February 1013, 2008.
61. "Computational Tools for Modeling of Biological Pathways and Networks," UT Health Science Center Meeting, March 6, 2008, Oak Ridge, TN.
62. "Parallel, Scalable, Memory-Efficient Backtracking for Combinatorial Modeling of Large-Scale Biological Systems." The 22<sup>nd</sup> IEEE International Parallel and Distributed Processing Symposium (IPDPS 2008), the 7<sup>th</sup> IEEE International Workshop on High Performance Computational Biology (HiCOMB 2008), Miami, Florida USA, Apr 14-18, 2008.
63. "Multi-stage Framework to Infer Protein Functional Modules from Mass Spectrometry Pull-Down Data with Assessment of Biological Relevance." IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007), Nov 2-4, 2007.
64. "The Buffered Work-Pool Approach for Search-Tree Based Optimization Algorithms." The 7<sup>th</sup> International Conference on Parallel Processing and Applied Mathematics (PPAM 2007), Gdansk, Poland, September 9-12, 2007.
65. "Adaptive Request Scheduling for Parallel ScientificWeb Services." The 20<sup>th</sup> International Conference on Scientific and Statistical Database Management (SSDBM '08), Hong Kong, Jul 9-11, 2008.

66. "Rapid and Robust Ranking of Text Documents in a Dynamically Changing Corpus." The 6th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-08), Databases and Data Mining Track, Doha, Qatar, Mar 31 - Apr 4, 2008.
67. "Scientific Data Management Center," Supercomputing Conference, November 10-16, 2007, Reno, NV.
68. "BioPilot: Data-Intensive Computing for Complex Biological Systems," Supercomputing Conference, November 10-16, 2007, Reno, NV.
69. "Uncovering Simplicity from Complexity: Discovering Fundamental Laws of Biology," UTK Mathematics Colloquium, September 20, 2007, Knoxville, TN.
70. "Data Analytics at Petascale," UC-Davis Computer Science Seminar Series, March 13, 2007, Davis, CA.
71. "Data Analytics at Petascale: Accelerating Scientific Discovery," NCSU Computer Science Seminar Series, March 23, 2007, Raleigh, NC.
72. "Finding the Dots, Connecting the Dots, Understanding the Dots," WSU Computer Science Seminar Series, April 6, 2007, Pasco, WA.
73. "Common-target model for identification of protein interaction modules in pull-down experiments," Annual Meeting of the Institute for Biological Engineering, Mar 30-April 1, 2007, St. Louis, Missouri.
74. "Maximum common subgraph problem: faster solutions via vertex cover," ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), May 13-16, 2007, Amman, Jordan (F. Abu-Khzam)
75. "Integrated knowledge resource for the Shewanella Federation," Joint Genomics:GTL Awardee Workshop V and Metabolic Engineering 2007 and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop, Feb 11-14, 2007, Bethesda, MD.
76. "Quantitative shotgun proteomics with ProRata: application to anaerobic aromatic degradation in *Rhodospseudomonas palustris*," Joint Genomics:GTL Awardee Workshop V and Metabolic Engineering 2007 and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop, Feb 11-14, 2007, Bethesda, MD.
77. "Scientific Data Management Center," DOE SciDAC Organization Worksho, Feb 5-6, 2007, Atlanta.
78. "Uncovering Mathematical Principles of Biology: Breaking the curses of dimensionality, intractability, & noise," presented to Dr. Michael Strayer, Acting Director of the DOE's Office of Advanced Scientific Computing (OASCR), March 6, 2006, ORNL.
79. "Comparative systems biology using scalable graph-theoretical approaches," MIT Computational Research Seminar, November 3, 2006 (host: Prof. Alan Edelman, MIT).
80. "Scalable graph-theoretical approaches to biological networks analysis," SIAM Conference on Parallel Processing for Scientific Computing, February 22-26, 2006, San Francisco, CA.
81. "Efficient data handling in comparative genome analysis applications," SIAM Conference on Parallel Processing for Scientific Computing, February 22-26, 2006, San Francisco, CA.
82. "High performance statistical computing with parallel R: applications to biology and climate modeling," The 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06), Nov 13-15, 2006, Washington, DC.

83. "Multi-criterion active learning in conditional Random Fields," DOE SciDAC Conference, June 25-29, 2006, Denver, CO.
84. "BioPilot: data-intensive computing for complex biological systems," UC Mercede Seminar Series, April 19, 2006, UC Mercede (host: Prof. Mike Colvin).
85. "Computational algorithms and software tools for quantitative shotgun proteomics," UC Mercede Seminar Series, April 19, 2006, UC Mercede (host: Prof. Mike Colvin).
86. "BioPilot: data-intensive computing for complex biological systems," Supercomputing Conference, November 11-17, 2006, Tampa, FL.
87. "Scientific Data Management Center," Supercomputing Conference, November 11-17, 2006, Tampa, FL.
88. "Comparative systems biology using scalable graph-theoretical approaches," Supercomputing Conference, November 11-17, 2006, Tampa, FL.
89. "Advanced technologies for identifying protein-protein interactions," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
90. "The Microbial Interactome Database: an online system for identifying interactions between proteins of microbial species," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
91. "Computational algorithms and software tools for quantitative shotgun proteomics," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
92. "Integrated knowledge resource for the Shewanella Federation," Joint Genomics: GTL Contractor-Grantee Workshop IV and Metabolic Engineering Working Group Interagency Conference on Metabolic Engineering, Feb 12-15, 2006, Bethesda, MD.
93. "Automatic parallelization for statistical computing with pR," DOE SciDAC SDM All Hands Meeting, Dec 11-13, 2006, Berkeley.
94. "Semantic annotation: how to make it painless, fast, and precise?," DHS Advanced Scientific Computing (ASC) All Hands Meeting, Oct 10-11, Livermore.
95. "Shewanella Federation knowledge integration resource," DOE Genomics:GTL Shewanella Federation Reverse Site Visit, September 18, Chantilly, VA.
96. "ProRata: a software package with improved point and interval estimation of protein abundance ratio for quantitative shotgun proteomics," American Society for Mass Spectrometry (ASMS) Conference, May 30 - April 2, 2006, Seattle, WA.
97. "Characterization of aromatic compound degradation pathways in *Rhodospseudomonas palustris* with stable isotope labeling quantitative shotgun proteomics and microarray," American Society for Mass Spectrometry (ASMS) Conference, May 30 - April 2, 2006, Seattle, WA.
98. "Genome-scale computational approaches to memory-intensive applications in systems biology," Supercomputing Conference, November 13-18, 2005 Seattle, WA.
99. "Systems biology framework for *R. palustris*," Environmental Microbes Meeting, September 15-16, 2005 Madison, WI (host: Prof. Julie Mitchell).

100. "Differential proteomics of *Rhodospseudomonas palustris* under its versatile metabolism states," International Conference on Microbial Genomesg, September 15-16, 2005 Madison, WI.
101. "Novel algorithms for extracting abundance ratios of stable isotope labeled peptide pairs from selected ion chromatograms in proteome sample," American Society for Mass Spectrometry (ASMS) Conference, June 3 - 7, 2005, San Antonio, Texas.
102. "Statistical and graph theoretical approaches to semantic tagging of unstructured text for the Biodefence Knowledge Center," National Homeland Security R&D Conference, April 27-28, 2005, Boston.
103. "Statistical and graph theoretical approaches to semantic tagging of unstructured text for the Biodefence Knowledge Center," DHS TVTA Portfolio Offsite Conference, April 20-22, 2005, Las Vegas, NV.
104. "Advancing, integrating and deploying efficient statistical computing to high-throughput scientific applications," DOE SciDAC SDM All Hands Meeting, Oct 5-7, Raleigh, NC.
105. "Parallel R (pR) for high performance statistical computing," DOE SciDAC SDM All Hands Meeting, March 1-5, 2005, Salt Lake City, UT.
106. "Computer science and biology research at ORNL," Brown University Seminar, July 12-13, 2005 (hosts: Sam Fulkomer, Andy von Dam).
107. "Semantic tagging of unstructured text," DHS ASC PI Meeting, February 16, 2005, Livermore, CA.
108. "SciDAC II Plans," DOE SciDAC PI Meeting, Feb 17, 2005, Washington, DC (host: DOE OASCR Director, M. Strayer).
109. "Probabilistic weighting and semantic tagging of terms in free text," DHS TVTA Text Processing Group Meeting, June 20, 2005, Salt Lake City, Utah.
110. "Prediction of residue-residue contacts in domain interface by co-evolution," ASM 105th General Meeting, June 5, 2005, Atlanta, GA.
111. "Accelerating exact stochastic simulation using parallel supercomputing," Computational Methods in Systems Biology, April, 2005, Edinburgh, Scotland.
112. "A parallel implementation of Gillespie's exact stochastic simulation algorithm," International Symposium on Computational and Cellular Biology, March, 2005, Lenox, MA.
113. "RScaLAPACK: high-performance parallel statistical computing with R and ScaLAPACK," International Conference on Parallel and Distributed Computing Systems (PDCS-2005), September 12 - 14, 2005, Las Vegas, Nevada.
114. "The role of the SciDAC II in predictive DOE mission-driven systems biology," DOE SciDAC Conference, June 26-30, 2005, San Francisco, CA.
115. "Enabling microbial systems biology through advanced computing," Supercomputing Conference, November 13-18, Seattle, WA.
116. "Parallel R for high performance statistical computing," Supercomputing Conference, November 13-18, Seattle, WA.
117. "Computer science and biology research at ORNL," William & Marry University Seminar, May 15-16, 2005 (host: Prof. John von Rosendale).
118. "From biomolecular interactions to microbial transcriptional regulation of metabolism," DOE Sandia-ORNL Genomics:GTL Project Reverse Site Visit, May 5, 2005, Washington, DC.

119. "Integrating heterogeneous databases and tools for high throughput microbial analysis," GTL Contractor-Grantee Workshop, Feb 6-9, 2005, Washington, DC.
120. "Microbial Organism Encyclopedia," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Apr 13, 2005, Seattle, WA.
121. "Protein-protein interaction prediction by examining co-occurrence of protein domains," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Apr 13, 2005, Seattle, WA.
122. "From genomics to functional omics - in silico challenges and opportunities," NC State Genomic Sciences Seminar, March 23, 2005, Raleigh, NC (host: Prof. Xiasong Ma).
123. "Information extraction from unstructured text - concepts discovery, concepts mapping, relations inference," DHS ASC Program Review, Dec 9, 2005, Livermore, CA (host: DHS TVTA Program Manager, Sandy Landsburg).
124. "A new approach and faster exact methods for the maximum common subgraph problem," Computing and Combinatorics Conference (COCOON), August 16-19, 2005, Kunming, Yunnan.
125. "A parallel implementation of Gillespie's exact stochastic simulation algorithm," Third International Symposium on Computational Cell Biology, 2005, Lenox, MA.
126. "BioSpreadsheet: A biological model design, simulation, and analysis tool," Third International Symposium on Computational Cell Biology, 2005, Lenox, MA.
127. "Accelerating exact stochastic simulation using parallel supercomputing," International Conference on Computational Methods in Systems Biology (CMSB), 2005, Edinburgh, Scotland.
128. "Gene circuit inference by analysis of single cell reporter protein expression data," 6th Annual International Conference on Systems Biology, 2005, Boston, MA.
129. "Reservoir-based random sampling with replacement from a data stream," SIAM International Conference on Data Mining, April 22-24, 2004, Kissimmee, FL.
130. "Analysis of interaction site predictions from separated data spaces," SIAM International Conference on Data Mining, April 22-24, 2004, Kissimmee, FL.
131. "Surface patch ranking method identified cooperative substrate specificity residues in highly homologous enzymes," International Conference on Microbial Genomes, Sep 28-Oct 2, 2004 Durham, NC.
132. "In-silico prediction of surface residue clusters for enzyme-substrate specificity," IEEE Computational Systems Bioinformatics (CSB) Conference, Sep 28-Oct 2, 2004 Durham, NC.
133. "Surface patch ranking method identified cooperative substrate specificity residues in highly homologous enzymes," International Conference on Microbial Genomes, August 16-20, 2004, Stanford, CA.
134. "Data Intensive Analysis and Visualization," National Institute of Standards, May 6, 2004, Washington, DC.
135. "Embedding methods and robust statistics for dimension reduction," Symposium of the International Association for Statistical Computing (COMPSTAT-2004), August 23-27, 2004, Prague.
136. "Advancing data science to support DHS data intensive operational requirements," Supercomputing Conference, November 6-12, 2004, Pittsburgh, PA.
137. "Data intensive computing for complex biological systems," Supercomputing Conference, November 6-12, 2004, Pittsburgh, PA.



138. "Scientific Data Management Center - data mining and analysis," Supercomputing Conference, November 6-12, 2004, Pittsburgh, PA.
139. "SDM Center's Data Mining and Analysis," DOE SciDAC Conference, March 22-27, 2004, Charleston, SC.
140. "Systems level protein interactions annotation tools," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 20-21, 2004, San Diego, CA.
141. "An integrated computational environment for systems biology," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 20-21, 2004, San Diego, CA.
142. "Accelerating exact stochastic simulation using parallel supercomputing," DARPA PI Meeting, Oct 13, 2004, Vienna, VA.
143. "The Synechococcus Encyclopedia," Genomics:GTL Contractor-Grantee Workshop II, eb 29 - Mar 2, 2004, Washington, DC.
144. "Multi-resolution functional characterization of Synechococcus WH8102," Genomics:GTL Contractor-Grantee Workshop II, Feb 29 - Mar 2, 2004, Washington, DC.
145. "Parallel R: Intelligent task-level parallel execution in R," DOE SciDAC SDM All Hands Meeting, August 3-7, 2004, Berkeley, CA.
146. "PVTk: parallel netCDF reader and ROMIO geometry writer," DOE SciDAC SDM All Hands Meeting, August 3-7, 2004, Berkeley, CA.
147. "ASPECT Components," DOE SciDAC SDM All Hands Meeting, August 3-7, 2004, Berkeley, CA.
148. "Systems level protein interactions annotation tools," University of Georgia Systems Biology Seminar, Nov 21-22, 2004, Athens, GA (host: Prof. Ying Xu).
149. "From genomics to functional omics - in silico challenges and opportunities," SIAM minisymposium on parallel computational biology, February 26, 2004, San Francisco, CA.
150. "Statistical and graph theoretical approaches to semantic tagging of unstructured text for BKC," DHS Biodefence Knowledge Center (BKC) PI Meeting, February 22, 2004, Livermore, CA.
151. "Integration of HPLC-FTICR MS and HPLC-QIT MS2 to achieve enhanced proteome characterization," American Society for Mass Spectrometry (ASMS) Conference, May 23-27, 2004, Nashville, TN.
152. "An SVM-based algorithm for identification of photosynthesis-specific genomes features," IEEE Computational Systems Bioinformatics (CSB) Conference, August 11-14, 2003, Stanford, CA.
153. "Towards unraveling CO<sub>2</sub>/O<sub>2</sub> specificity in microbial ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO)," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 21-22, 2003, Berkeley, CA.
154. "Identification of protein-protein interaction sites," Sandia-ORNL Genomics:GTL Project All Hands Meeting, Oct 21-22, 2003, Berkeley, CA.
155. "Bioinformatics tools for characterization of protein-protein interactions," Sandia-ORNL Genomics:GTL Project All Hands Meeting, March 26-27, 2003, Salt Lake, UT.
156. "PICUPP: Protein Interaction Classification by Unlikely Profile Pair," Sandia-ORNL Genomics:GTL Project All Hands Meeting, March 26-27, 2003, Salt Lake, UT.
157. "Xmap: fast dimension reduction algorithms for multivariate streamline data," SIAM Conference on Data Mining, May 1-3, 2003, San Francisco, CA.

158. "Distributed data mining," Spring Research Conference on Statistics, Analysis and Visualization of Massive Simulation Data Sets, June 4-6, 2003, Dayton, Ohio.
159. "Can dimension reduction be fast and robust? FastMap and the Convex Hull of multivariate data," UTK Statistics Seminar, December 6, 2003, UTK.
160. "Interoperability of visualization software and data models is Not an achievable goal," IEEE Visualization, October 19-24, 2003, Seattle, Washington.
161. "Gleaning insight from scientific simulation data," DOE SciDAC Program Review, March 12, 2003, Washington DC.
162. "Advanced data science for homeland security," Supercomputing Conference, November 15-21, 2003, Phoenix.
163. "ASPECT: Adaptable Simulation Product Exploration and Control Tool," Supercomputing Conference, November 15-21, 2003, Phoenix.
164. "Sandia-ORNL Genomes to Life Project," Supercomputing Conference, November 15-21, 2003, Phoenix.
165. "Analysis of protein complexes from a fundamental understanding of protein binding domains and protein-protein interactions in *Synechococcus* WH8102," DOE Genomes to Life Contractor - Grantee Workshop I, Feb 9-12, 2003, Arlington, Virginia.
166. "Carbon sequestration in *Synechococcus* sp.: from molecular machines to hierarchical modeling," DOE Genomes to Life Contractor - Grantee Workshop I, Feb 9-12, 2003, Arlington, Virginia.
167. "Data exploration environment," DOE Scientific Data Management Workshop, Aug 28, 2003, Argonne, IL (host: DOE Program Manager, John von Rosendale, Arie Shoshani, LBL).
168. "Adaptable Simulation Product Exploration and Control Tool (ASPECT)," SciDAC Terascale Supernova Initiative All Hands Meeting, February, 2002 (host: Dr. Tony Mezzacappa).
169. "RACHET: petascale distributed data analysis suite," SPEEDUP Workshop on Distributed Supercomputing Data Intensive Computing, March 4-6, 2002, Leukerbad, Valais, Switzerland.
170. "CADS challenges in systems biology: a computer science perspective," NCSA Alliance All-Hands Meeting, May 8-10, 2002, Urbana.
171. "Parallel out-of-core algorithm for genome-scale enumeration of metabolic systemic pathways," First International Workshop on High Performance Computational Biology (HiCOMB), International Parallel and Distributed Processing Symposium (IPDPS-02), April 15, 2002, Fort Lauderdale, Florida.
172. "Advanced algorithms for computational biology," DOE Science and Technology Review, February 11, 2002, Oak Ridge, TN.
173. "Scientific Data Management Center," DOE Program Review, August 16, 2002, ORNL (presentation to DOE SciDAC program director, Dr. Alan Laub).
174. "Advanced algorithms for computational biology," DOE Program Review, August 16, 2002, ORNL (presentation to DOE OS program manager, Dr. Walter Stevens).
175. "ASPECT: Adaptable Simulation Product Exploration and Control Toolkit," SciDAC SDM All Hands Meeting, September 11-13, 2002, San Diego.
176. "Data mining and access pattern discovery," DOE SciDAC Conference, June 26-27, 2002, Argonne.
177. "High resolution tools and algorithms for genome scale analysis," Supercomputing Conference, Nov 16-21, 2002, Baltimore, MD.

178. "ASPECT - data stream monitoring tool," Supercomputing Conference, Nov 16-21, 2002, Baltimore, MD.
179. "Research in algorithms for scalable analysis of distributed and streaming data," Supercomputing Conference, Nov 16-21, 2002, Baltimore, MD.
180. "Medard W. Welch Award Lecture: Intertwined charge density wave and defect-ordering phase transitions in a 2-D system," IUUSTA 15th International Vacuum Congress (IVC-15), AVS 48th International Symposium (AVS-48), 11th International Conference on Solid Surfaces (ICSS-11), San Francisco, October 2002(invited talk by Prof. W. Plummer, UTK).
181. "Combining distributed local principal component analyses into a global analysis," C. Warren Neel Conference on Statistical Data Mining and Knowledge Discovery, June 22-25, 2002, Knoxville, TN.
182. "Annotation of DNA regulatory regions through the utilization of underlying recognition principles," C. Warren Neel Conference on Statistical Data Mining and Knowledge Discovery, June 22-25, 2002, Knoxville, TN.
183. "Multivariate analysis of massive distributed data sets," Spring Research Conference on Statistics in Industry and Technology, May 20-22, 2002, Ann Arbor, Michigan.
184. "Evolutionary analysis of enzymes," International Meeting on Microbial Genomes, September 8-12, 2002, Lake Arrowhead, CA.
185. "Resource and Location Aware Mining: Principal component analysis for dimension reduction in massive distributed data sets," 5th International Workshop on High Performance Data Mining (HPDM), April 13, 2002, Washington, DC.
186. "RACHET: a new algorithm for mining multi-dimensional distributed dataset," SIAM Third Workshop on Mining Scientific Datasets, Chicago, IL, April 5-7, 2001.
187. "Characterization of the solution space from metabolic flux balance analysis," UCSD Genetic Circuits Group Seminar, July 12, 2001 (host: Prof. Bernhard Palsson, UCSD).
188. "Multi-agent based high-dimensional cluster analysis," DOE Program Review, August 9, 2001 (presentation to DOE OS program manager, Dr. Steve Eckstrand).
189. "Multi-agent based high-dimensional cluster analysis," Supercomputing Conference, July 10, 2001, Gatlinburg, TN.
190. "Cracking computational complexity for genome-scale modeling of biochemical pathways," SciDAC SDM Kick-off Meeting, November 10-16, 2001, Denver, CO.
191. "RACHET: petascale distributed data analysis suit," SciDAC SDM Kick-off Meeting, November 10-16, 2001, Denver, CO.
192. "Predictive models of biochemical pathways and microbial behavior," DOE Genomes to Life (GTL) Workshop, August 7, 2001 (invited talk by Prof. B. Palsson, UCSD).

## Organizing Committees

1. *Session Co-Chair*, DOE Workshop on Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale, March 18-20, 2009, Washington DC
2. CO-CHAIR, *DOE/NSF Mathematics for Analysis of Petascale Data*, June 3–5, 2008, Rockville, Maryland
3. CO-CHAIR, *DOE Genomics:GTL Program Systems Biology Knowledgebase Workshop*, May 28 - May 31, 2008, Bethesda North, Washington D.C.
4. CO-CHAIR, *DOE/OASCR Mathematics for Peta-scale Data Analysis & Computation*, June 2007
5. TECHNICAL SESSION CHAIR, *DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment*, Apr 17 - 18, 2007, San Francisco, CA.
6. TECHNICAL SESSION CHAIR, *DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment*, May 17 - 18, 2007, Oak Ridge, TN.
7. TECHNICAL SESSION CHAIR, *DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment*, May 31 - June 1, 2007, Argonne, IL.
8. COMMITTEE MEMBER, *DOE Data Sharing & Integration Summit*, April 20, 2006, Berkeley.
9. COMMITTEE MEMBER, *DOE SciDAC Conference*, June 26-30, 2005, San Francisco, CA.
10. CHAIR, *DOE SciDAC Biology Summit*, August 8-9, 2005, Reston, VA.
11. CO-CHAIR, *DHS Threat Awareness Customer Data Integration & Dissemination*, November 1-2, 2005, Alexandria VA.
12. CO-CHAIR, *DHS Data Sciences for Homeland Security Information Management and Knowledge Discovery*, Sep. 22-23, 2004, Alexandria, VA.
13. COMMITTEE MEMBER, *Scientific Data Management*, April 20-22, Menlo Park, CA.
14. COMMITTEE MEMBER, *DOE Computational Infrastructure for the Genomes to Life Program*, July 22, 2002, Gaithersbrug, MD.

## Invited Workshop Participant

1. PARTICIPANT, *Extremely Large Scale Databases (XLDB) Workshop*, Oct. 25, 2007, Stanford, CA.
2. PARTICIPANT, *2008 GTL-PI and ASCR-BER Bioenergy Research Center Workshop*, Feb. 13, 2008, Rockville, MD.
3. PARTICIPANT, *DOE OASCR PI Meeting*, Mar 31 – Apr 2, 2008, Denver CO.
4. PARTICIPANT, *NSF Petascale Computing in the Biological Sciences*, Aug. 29-30, 2006.
5. SESSION CHAIR, *Workshop on Enabling Petascale Science and Engineering Applications*, Dec 9, 2005, GA Tech, Atlanta.
6. SESSION CHAIR, *DOE SciDAC II Planning Workshop on Accelerating Scientific Discovery in Experimental Science via Advanced Computing*, Aug 2-3, 2005, Argonne, IL.
7. SESSION CO-CHAIR, *DOE BER/ASCR GTL Deep Dive Technology*, June 14-16, 2004, Alexandria, VA.
8. SESSION CO-CHAIR, *DOE Scientific Data Management*, March 16-18, 2004, Menlo Park, CA.
9. PARTICIPANT, *DOE Scientific Data Management*, March 16-18, 2004, Menlo Park, CA.

10. PARTICIPANT, *DOE Scientific Data Management*, May 24-26, 2004, Chicago, IL.
11. SESSION CO-CHAIR, *DHS Advanced Scientific Computing Requirements*, Oct 8-9, 2003, Arlington, VA.
12. PARTICIPANT, *DOE/OBER GTL and Beyond Data Standards*, Sep 10-11, 2003, Berkeley, CA.
13. PARTICIPANT, *NITRD High End Computing Revitalization Task Force (HECRTF)*, June 15-18, 2003, Washington, DC.
14. PARTICIPANT, *DOE Visualization Frameworks Requirements*, June 2-3, 2003, Bethesda, MD.
15. PARTICIPANT, *DOE Neutron Science Software Initiative (NeSSI)*, Oct 13-15, 2003, Oak Ridge, TN.
16. PARTICIPANT, *DOE Genomes To Life Data Infrastructure*, July 22, 2003, Gaithersburg, MD.
17. PARTICIPANT, *DOE Mathematics for the Genomes-to-Life Program*, March 18-19, 2002, Gaithersburg, MD.
18. PARTICIPANT, *DOE Computer Science for GTL*, March 6-7, 2002, Gaithersburg, MD.
19. PARTICIPANT, *DOE First Genomes to Life (GTL) Computational Biology Workshop*, August 7-8, 2001.

#### Conference Program Committees

1. PC MEMBER, *International Conference on Parallel Processing (ICPP)*, September 22-25, 2009, Vienna, Austria.
2. PC MEMBER, *The 2007 IEEE International Symposium on Bioinformatics and Life Science Computing (BLSC)*, in conjunction with *The IEEE 21st International Conference on Advanced Information Networking and Applications (AINA-07)*, ay 21-23, 2007, Niagara Fall, Ontario, Canada.
3. PC MEMBER, *2nd IEEE Workshop on High Performance Computing in Medicine and Biology (HiPCoMB-2006)*, in conjunction with the *12th International Conference on Parallel and Distributed Systems (ICPADS 2005)*, April 18-20, Vienna, Austria.
4. PC MEMBER, *1st IEEE Workshop on High Performance Computing in Medicine and Biology (HiPCoMB-2005)*, in conjunction with the *11th International Conference on Parallel and Distributed Systems (ICPADS 2005)*, July 20-22, 2005, Fukuoka, Japan.
5. PC MEMBER, *16th International Conference on Scientific and Statistical Database Management (SSDBM)*, June 21-23, 2004, Santorini Island, Greece.
6. PC MEMBER, *Third SIAM International Conference on Data Mining, Workshop on Data Mining for Counter Terrorism and Security*, May 3, 2003, San Francisco, CA.
7. PC MEMBER, *IEEE Bioinformatics Conference*, August 11-14, 2003, Stanford, CA.
8. TECHNICAL PROGRAM CO-CHAIR, *C. Warren Neel Conference on Statistical Data Mining*, June 22-25, 2002, Knoxville, TN.
9. PC MEMBER, *16th International Conference on Scientific and Statistical Database Management (SSDBM)*, April 23-27, 2002, Baltimore, MD.
10. PC MEMBER, *CLUSTER: IEEE International Conference on Cluster Computing*, September 23-26, 2002, Chicago, IL.

**B. NCSU RADAR Report of Submitted 2012-2013 Grants: (Total Amount: \$5,842,000.00)**

- “Data Mining and Information Visualization for Climate Data,” DOE SBIR-I, October 2012, \$50,000, NCSU PI *New in 2012*
- “IGERT-CIF21: Transforming the Development of High-Performance Computing Applications,” NSF IGERT, August 2012, \$3,221,071, NCSU co-PI *New in 2012*
- “BIGDATA: Mid-Scale: DCM: DA: Collaborative Research: STARBID: Scalable Thinking and Analytics Research for Big Data,” NSF BIGDATA, July 2012, \$500,995, NCSU PI *New in 2012*
- “BIGDATA: Mid-Scale: DA: Collaborative Research: A Graph Theoretic Framework For Understanding the Neurologically Disordered Human Brain From Heterogeneous Biomedical Data,” NSF BIGDATA, July 2012, \$621,306, NCSU PI *New in 2012*
- “SKALE: Scalable Knowledge Acquisition and Learning Environment,” DARPA, \$1,999,950 *New in 2012*

**C. NCSU RADAR Report of Awarded Grants: (Total Amount Awarded: \$6,297,797.00)**

1. **Project ID:** (1419) 2012-1044  
**Title:** *Scalable Data Management, Analysis, and Visualization (SDAV) Institute*  
**PI:** Samatova, Nagiza  
**co-PI:** Melechko, Anatoli  
**Sponsor:** Department of Energy (DOE)  
**Status:** Awarded  
**Amount Awarded:** \$750,000  
**Period Proposed:** 02/15/2012 – 12/31/2016
2. **Project ID:** (1419) 2012-0922  
**Title:** *Scalable Statistical Computing for Physical Science Applications*  
**PI:** Samatova, Nagiza  
**co-PI:** Melechko, Anatoli  
**Sponsor:** Oak Ridge National Laboratories - UT-Battelle LLC (Prime-US Dept. of Energy (DOE))  
**Status:** Awarded  
**Amount Awarded:** \$354,646.00  
**Period Proposed:** 12/02/2011 through 06/30/2012
3. **Project ID:** (1419) 2012-2314  
**Title:** *Analytics-driven Efficient Indexing and Query Processing of Extreme Scale AMR Data*  
**PI:** Samatova, Nagiza  
**Sponsor:** National Science Foundation  
**Status:** Awarded  
**Amount Awarded:** \$149,999  
**Period Proposed:** 06/01/2012 through 05/31/2014
4. **Project ID:** (1419) 2010-1533  
**Title:** *Collaborative Research: Understanding Climate Change: A Data Driven Approach*  
**PIs:** Samatova, Nagiza and Semazzi, Frederick  
**Sponsor:** National Science Foundation (NSF)  
**Status:** Awarded  
**Amount Awarded:** \$1,815,739  
**Period Proposed:** 09/01/2010 – 08/30/2015

5. **Project ID:** (1419) 2010-2001  
**Title:** *Damsel: A Data Model Storage Library for Exascale Science*  
**PI:** Samatova, Nagiza  
**Sponsor:** Department of Energy (DOE)  
**Status:** Awarded  
**Amount Awarded:** \$329,998  
**Period Proposed:** 09/01/2010 – 08/30/2013
6. **Project ID:** (1419) 2011-0801  
**Title:** *Runtime System for I/O Staging in Support of In-Situ Processing of Extreme Scale Data*  
**PI:** Samatova, Nagiza  
**Sponsor:** Oak Ridge National Laboratories - UT-Battelle LLC (Prime–US Dept. of Energy (DOE))  
**Status:** Awarded  
**Amount Awarded:** \$247,414  
**Period Proposed:** 09/01/2010 – 08/30/2013
7. **Project ID:** (1419) 2011-0802  
**Title:** *Scalable and Power Efficient Data Analytics for Hybrid Exascale Systems*  
**PI:** Samatova, Nagiza  
**Sponsor:** Oak Ridge National Laboratories - UT-Battelle LLC (Prime–US Dept. of Energy (DOE))  
**Status:** Awarded  
**Amount Awarded:** \$364,944  
**Period Proposed:** 09/01/2010 – 08/30/2013
8. **Project ID:** (1419) 2008-0010  
**Title:** *Joint Faculty Agreement For Nagiza Samatova*  
**PIs:** Samatova, Nagiza  
**Sponsor:** Oak Ridge National Laboratories - UT-Battelle LLC (Prime–US Dept. of Energy)  
**Status:** Awarded (Appointment has been extended for 3 additional years)  
**Amount Awarded:** \$448,945 (total) and \$287,847 (renewal part)  
**Current Award Period:** 10/04/2007 through 08/31/2013
9. **Project ID:** (1419) 2009-0155  
**Title:** *Ultrascale Computational Modeling of Phenotype-Specific Metabolic Processes in Microbial Communities*  
**PIs:** Samatova, Nagiza and Melechko, Anatoli  
**Sponsor:** Oak Ridge National Laboratories - UT-Battelle LLC (Prime–US Dept. of Energy)  
**Status:** Awarded  
**Amount Awarded:** \$454,311  
**Period Proposed:** 01/15/2010 – 09/30/2012
10. **Project ID:** (1419) 2009-2678  
**Title:** *Scalable Statistical Computing For Physical Science Applications*  
**PIs:** Samatova, Nagiza and Melechko, Anatoli  
**Sponsor:** Oak Ridge National Laboratories - UT-Battelle LLC (Prime–US Dept. of Energy)  
**Status:** Awarded  
**Amount Awarded:** \$299,745.00  
**Period Proposed:** 10/05/2009 through 06/30/2011

11. **Project ID:** (1419) 2008-0221  
**Title:** *High-Performance Data Analytics with Demonstrations to DOE-Mission Applications*  
**PIs:** Samatova, Nagiza  
**Sponsor:** Oak Ridge National Laboratories - UT-Battelle LLC (Prime-US Dept. of Energy)  
**Status:** Awarded  
**Amount Awarded:** \$1,120,002  
**Current Award Period:** 10/04/2007 through 08/31/2013
12. **Project ID:** (1419) 2009-0985  
**Title:** *A Novel Method for Recommendation Systems Utilizing Computationally Structured Social Architectures*  
**PI:** Samatova, Nagiza  
**Sponsor:** Topiat, Inc.  
**Status:** Awarded  
**Amount Awarded:** \$15,024.00  
**Current Award Period:** 09/24/2009 through 01/31/2010
13. **Project ID:** (1419) 2008-1680  
**Title:** *Workshop on Mathematics for Peta-scale Data*  
**PIs:** Samatova, Nagiza  
**Sponsor:** National Science Foundation  
**Status:** Awarded  
**Amount Awarded:** \$20,000  
**Current Award Period:** 08/09/2007 through 08/08/2010

#### Past Sponsored Grants

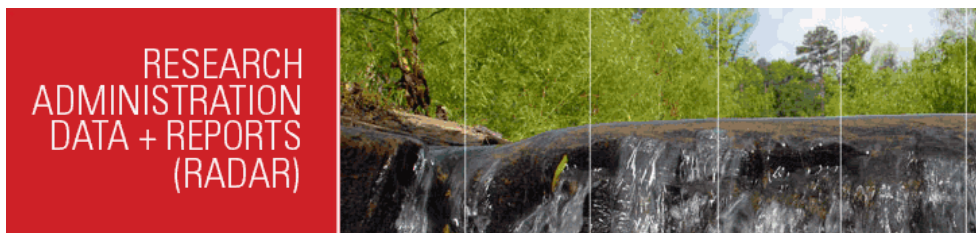
1. **Title:** *Scientific Data Management Technologies to Accelerate Fusion Scientific Discovery: A Strategic Application Partnership (SAP) Proposal*  
**Agency/Program:** DOE  
**PIs:** Arie Shoshani (LBL), Nagiza F. Samatova, Scott Klasky  
**Period:** 07/2008 - 03/2011  
**Amount:** \$600,000  
**Project Title in the NCSU RADAR System:** Joint Faculty Appointment  
**Comment:** NCSU funding was delivered as a subcontract from ORNL.
2. **Title:** *Data Intensive Computing for Complex Biological Systems*  
**Agency/Program:** DOE  
**PIs:** Nagiza F. Samatova and Tjerk Straatsma (PNNL)  
**Period:** 10/2007 - 9/2009  
**Amount:** \$1,500,000  
**In-coming to NCSU Amount:** \$400,000  
**Project Title in the NCSU RADAR System:** High-Performance Data Analytics with Demonstrations to DOE-Mission Applications
3. **Title:** *Workshop on Mathematics for Petascale Data*  
**Agency/Program:** NSF  
**PI:** Nagiza F. Samatova  
**Period:** 5/2008 - 9/2008  
**Amount:** \$20,000 (NCSU)



4. **Title:** *Data Intensive Computing for Complex Biological Systems*  
**Agency/Program:** DOE  
**PIs:** Nagiza F. Samatova and Tjerk Straatsma (PNNL)  
**Period:** 9/2004 - 9/2007  
**Amount:** \$2,700,000  
**In-Coming to NCSU Amount:** \$200,000  
**Project Title in the NCSU RADAR System:** High-Performance Data Analytics with Demonstrations to DOE-Mission Applications
  
5. **Title:** *DOE SciDAC Scientific Data Management Center*  
**Agency/Program:** DOE  
**ORNL PI:** Nagiza F. Samatova  
**Period:** 7/2006 - 6/2011  
**Amount:** \$2,700,000  
**In-coming to NCSU Amount:** \$1,500,000  
**Project Title in the NCSU RADAR System:** High-Performance Data Analytics with Demonstrations to DOE-Mission Applications  
**Comments:** \$520,000 out of \$1,500,000 NCSU funding has been recorded in RADAR under the ORNL-NCSU contract 'High-Performance Data Analytics with Demonstrations to DOE-Mission Applications' (Current). Only ORNL funding is reflected. The overall SciDAC SDM Center (PI: Arie Shoshani, LBL) budget is not reflected here.
  
6. **Title:** *Integrated Knowledge Resource for the Shewanella Federation*  
**Agency/Program:** DOE  
**PIs:** Edward Uberbacher (ORNL) and Nagiza Samatova  
**Period:** 10/2005 - 9/2008  
**Amount:** \$3,300,000  
**In-coming to NCSU Amount:** \$228,000  
**Project Title in the NCSU RADAR System:** Joint Faculty Agreement
  
7. **Title:** *Integrated Genome-Based Studies of Shewanella Ecophysiology*  
**Agency/Program:** DOE  
**PIs:** Edward Uberbacher (ORNL), Nagiza Samatova, etc.  
**Period:** 10/2006 - 9/2009  
**Amount:** \$1,100,000  
**In-coming to NCSU Amount:** \$80,000  
**Project Title in the NCSU RADAR System:** Joint Faculty Agreement  
**Comment:** The overall Shewanella Federation (PI: Jim Fredrickson, PNNL) budget is not reflected here. Only the portion of the funding for which the responsibilities of the PIs are counted are reflected.
  
8. **Title:** *Modeling Cellular Mechanisms for Efficient Bioethanol Production through Petascale Comparative Analysis of Biological Networks*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** Andrey Gorin (ORNL) and Nagiza Samatova  
**Period:** 10/2007 - 9/2008  
**Amount:** \$350,000  
**In-coming to NCSU Amount:** \$50,000  
**Project Title in the NCSU RADAR System:** Joint Faculty Agreement

9. **Title:** *Modeling Cellular Mechanisms for Efficient Bioethanol Production through Petascale Comparative Analysis of Biological Networks*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** Nagiza Samatova  
**Period:** 10/2006 - 9/2007  
**Amount:** \$400,000 (ORNL)
10. **Title:** *DOE SciDAC Scientific Data Management Center*  
**Agency/Program:** DOE  
**ORNL PI:** Nagiza F. Samatova  
**Period:** 7/2001 - 6/2006  
**Amount:** \$2,700,000 (ORNL)  
**Comments:** Only ORNL funding is reflected. The overall SciDAC SDM Center (PI: Arie Shoshani, LBL) budget is not reflected here.
11. **Title:** *Biodefense Knowledge Center (BKC)*  
**Agency/Program:** DHS/TVTA  
**PIs:** Nagiza Samatova  
**Period:** 04/2006 - 03/2007  
**Amount:** \$650,000
12. **Title:** *Natural Language Processing for Biological Text Mining*  
**Agency/Program:** DHS/ASC  
**PIs:** Nagiza Samatova  
**Period:** 10/2003 - 09/2006  
**Amount:** \$800,000
13. **Title:** *Large Scale Exploration of Protein Models for System Biology Applications*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** Andrey Gorin and Nagiza Samatova  
**Period:** 10/2005 - 9/2007  
**Amount:** \$385,000 (ORNL)
14. **Title:** *Systems biology framework for postgenomic microbiology*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** Edward Uberbacher, Andrey Gorin, and Nagiza Samatova  
**Period:** 10/2004 - 9/2006  
**Amount:** \$520,000 (ORNL)
15. **Title:** *National Bio-forensics Information Encyclopedia: Heterogeneous Data Integration*  
**Agency/Program:** DHS/ASC  
**PIs:** Nagiza Samatova  
**Period:** 10/2002 - 09/2003  
**Amount:** \$250,000
16. **Title:** *Carbon Sequestration of Synechococcus: From Molecular Machines to Hierarchical Modeling*  
**Agency/Program:** DOE/OASCR  
**PIs:** Grant Heffelfinger (SNL), Al Geist (ORNL)  
**co-PIs:** Nagiza Samatova, Andrey Gorin and others  
**Period:** 10/2002 - 09/2005  
**Amount:** \$9,000,000

17. **Title:** *SKALE: Scalable Clustering and Feature Extraction from Massive High-Dimensional Distributed Scientific Data (under the Probe project)*  
**Agency/Program:** DOE  
**PIs:** Nagiza Samatova  
**Period:** 01/2000 - 12/2003  
**Amount:** \$900,000
18. **Title:** *High-throughput Biological Data Analysis and Modeling Tools for Genomes To Life Facilities*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** Nagiza Samatova  
**Period:** 10/2002 - 09/2004  
**Amount:** \$500,000 (ORNL)
19. **Title:** *Biologically Driven Controlled Synthesis and Directed Assembly of Nanophase Inorganic Materials*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** Michael Simpson (ORNL) and Nagiza Samatova  
**Period:** 10/2003 - 09/2005  
**Amount:** \$350,000 (ORNL)
20. **Title:** *Scalable Tools for Petascale Distributed Data Analysis*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** George Ostrouchov (ORNL) and Nagiza Samatova  
**Period:** 10/2001 - 09/2003  
**Amount:** \$630,000 (ORNL)
21. **Title:** *PDQ: A PACI Petascale Data Quest*  
**Agency/Program:** NSF  
**PIs:** Nagiza Samatova  
**Period:** 06/2002-05/2005  
**Amount:** \$150,000  
**Comments:** Only ORNL funding is reflected. The overall project (PI: Ian Foster, ANL) budget is not reflected here.
22. **Title:** *Scientific Workspaces of the Future: Grid-based Visualization and Collaboration Services*  
**Agency/Program:** NSF  
**PIs:** Nagiza Samatova  
**Period:** 06/2002-05/2004  
**Amount:** \$100,000  
**Comments:** Only ORNL funding is reflected. The overall project (PI: Rick Stevens, ANL) budget is not reflected here.
23. **Title:** *Program Development for Computational Biology*  
**Agency/Program:** UT-Battelle ORNL/LDRD  
**PIs:** Nagiza Samatova  
**Period:** 10/2001 - 09/2006  
**Amount:** \$60,000 (ORNL)



### Search Results (13 projects)

investigator is Samatova, Nagiza; project is funded;

Project ID, Descending order

**If Project ID cell is shaded, project contains subcontracts**

Project ID	Title	Investigator(s)	PI Department	Sponsor	Status	Amount Awarded/ Contract Number	Current Award Period
(1419) 2012-2314 <b>Details</b>	Analytics-driven Efficient Indexing and Query Processing of Extreme Scale AMR Data	Samatova, Nagiza F.	Computer Science	National Science Foundation	Awarded	\$149,999.00 CCF-1240682 (49.0% - B - )	05/01/2012 through 04/30/2014
(1419) 2012-1044 <b>Details</b>	Scalable Data Management, Analysis, and Visualization (SDAV) Institute	Samatova, Nagiza F. Melechko, Anatoli V	Computer Science	US Dept. of Energy (DOE)	Awarded	\$300,000.00 DE-SC0007445 (49.0% - B - )	02/15/2012 through 02/14/2014
(1419) 2012-0922 <b>Details</b>	Scalable Statistical Computing For Physical Science Applications	Samatova, Nagiza F. Melechko, Anatoli V	Computer Science	Oak Ridge National Laboratories - UT-Battelle LLC (Prime--US Dept. of Energy (DOE))	Awarded	\$354,646.00 4000110308 (49.0% - B - )	12/02/2011 through 06/30/2013
(1419) 2011-0802 <b>Details</b>	Scalable and Power Efficient Data Analytics for Hybrid Exascale Systems	Samatova, Nagiza F.	Computer Science	Oak Ridge National Laboratories - UT-Battelle LLC (Prime--US Dept. of Energy (DOE))	Awarded	\$364,944.00 4000102031 (49.0% - B - )	01/31/2011 through 12/31/2013

Figure 1: The RADAR report showing the sponsored awards. A total of \$6,297,797.00 has been awarded since joining NCSU in Fall, 2007. Due to the incremental funding nature for some of the awards, the RADAR shows the Total Amount Awarded as \$5,783,303.

(1419) 2011-0801 <a href="#">Details</a>	Runtime System for I/O Staging in Support of In-Situ Processing of Extreme Scale Data	Samatova, Nagiza F.	Computer Science	Oak Ridge National Laboratories - UT-Battelle LLC (Prime--US Dept. of Energy (DOE))	Awarded	\$163,140.00 4000101770 (49.0% - B -)	01/31/2011 through 10/31/2012
(1419) 2010-2001 <a href="#">Details</a>	Damsel: A Data Model Storage Library for Exascale Science	Samatova, Nagiza F.	Computer Science	US Dept. of Energy (DOE)	Awarded	\$330,000.00 DE-SC0004935 (49.0% - B -)	09/01/2010 through 08/31/2013
(1419) 2010-1533 <a href="#">Details</a>	Collaborative Research: Understanding Climate Change: A Data Driven Approach	Samatova, Nagiza F. Semazzi, Fredrick H.	Computer Science	National Science Foundation	Awarded	\$1,815,739.00 IIS-1028746 (49.0% - B -)	09/01/2010 through 08/31/2015
(1419) 2009-2678 <a href="#">Details</a>	Scalable Statistical Computing For Physical Science Applications	Samatova, Nagiza F. Melechko, Anatoli V	Computer Science	Oak Ridge National Laboratories - UT-Battelle LLC (Prime--US Dept. of Energy (DOE))	Awarded	\$299,745.00 4000086391 (49.0% - B -)	10/05/2009 through 06/30/2013
(1419) 2009-0985 <a href="#">Details</a>	A Novel Method for Recommendation Systems Utilizing Computationally Structured Social Architectures	Samatova, Nagiza F.	Computer Science	Topiat, Inc.	Awarded	\$15,024.00 2009-0985 (49.0% - B -)	09/24/2009 through 01/31/2010
(1419) 2009-0155 <a href="#">Details</a>	Ultrascale Computational Modeling of Phenotype-Specific Metabolic Processes in Microbial Communities	Samatova, Nagiza F. Melechko, Anatoli V	Computer Science	Oak Ridge National Laboratories - UT-Battelle LLC (Prime--US Dept. of Energy (DOE))	Awarded	\$454,311.00 4000089282 (49.0% - B - X)	01/15/2010 through 01/14/2014
(1419) 2008-1680 <a href="#">Details</a>	Workshop on Mathematics for PetaScale Data	Samatova, Nagiza F.	Computer Science	National Science Foundation	Awarded	\$20,000.00 DMS-0829830 (0.0% - - A)	05/15/2008 through 04/30/2009
(1419) 2008-0221 <a href="#">Details</a>	High-Performance Data Analytics with Demonstrations to	Samatova, Nagiza F.	Computer Science	Oak Ridge National Laboratories -	Awarded	\$1,120,002.00 4000063955 (48.5% - B -)	10/04/2007 through 08/31/2012

Figure 2: The RADAR report showing the sponsored awards (cont.)

	DOE-Mission Applications			UT-Battelle LLC (Prime--US Dept. of Energy (DOE))			
(1419) 2008-0010 <a href="#">Details</a>	Joint Faculty Agreement For Nagiza Samatova	Samatova, Nagiza F. Vouk, Mladen A.	Computer Science	Oak Ridge National Laboratories - UT-Battelle LLC (Prime--US Dept. of Energy (DOE))	Awarded	\$395,753.00 (0.0% - - A)	4200000036-4000061675 08/09/2007 through 08/08/2013
Total Amount Awarded:						\$5,783,303.00	

[Create a New Query](#)  
[Return to Main Menu](#)

Figure 3: The RADAR report showing the sponsored awards (cont.)

### **Participation in Centers, Consortia and Other Organized Scholarly Efforts**

JOINT FACULTY APPOINTEE, Oak Ridge National Laboratory (ORNL), 2007-present

ASSOCIATE FACULTY MEMBER, Center for High Performance Simulation (CHiPS), 2007-present

AFFILIATED FACULTY, UT-ORNL Graduate School of Genome Science and Technology, 2002-present

RESEARCH AFFILIATE, UT-ORNL Joint Institute for Computational Sciences (JICS), 1997-present

AREA LEAD, DOE Scientific Data Management Center, 2001-present

AREA CO-LEAD, DOE Shewanella Federation, 2004-present

DIRECTOR, DOE BioPilot Center for Data Intensive Computing in Biology, 2004-present

AREA LEAD, DOE Sandia-ORNL Genomics:GTL Center, 2002-2004

### **Cross-Disciplinary Activities**

CO-ADVISOR, PhD student, Chongle Pan, with Bob Hettich (ORNL, Analytics Chemistry), 2004-2006

CO-ADVISOR, PhD student, Heshan Lin, with Xiasong Ma (NCSU, Systems), 2004-present

CO-ADVISOR, PhD student, Jiangtin Li, with Xiasong Ma (NCSU, Systems), 2005-present

CO-ADVISOR, PhD student, Andrea Rocha, with Peter Stroot (USF, Biology), 2006-present

### III. EXTENSION AND ENGAGEMENT WITH CONSTITUENCIES OUTSIDE THE UNIVERSITY

#### A. Accomplishments

##### Media Coverage

1. 'N.C. State researchers develop better tool for hurricane prediction,' Raleigh News and Observer, November 4, 2012 *New in 2012*
2. 'New model could improve hurricane forecasting,' Television interview with Raleigh WRAL on October 31, 2012 *New in 2012*
3. 'Researchers Devise More Accurate Method For Predicting Hurricane Activity,' the top "breaking news" story on the National Science Foundation's official news site, September 12, 2012 *New in 2012*
4. 'Researchers Devise More Accurate Method for Predicting Hurricane Activity,' NSF News, September, 2012 *New in 2012*
5. 'Researchers Devise More Accurate Method for Predicting Hurricane Activity,' Computing Research Highlights of the Week, September–October, 2012 *New in 2012*
6. 'New hurricane detection method increases predictability,' The Conversation, September, 2012 *New in 2012*
7. 'Researchers Devise More Accurate Method for Predicting Hurricane Activity,' eScienceNews, September, 2012 *New in 2012*
8. 'Refining hurricane forecasts by connecting the dots,' DOE Office of Science News Site, June 20, 2012
9. 'Young fusion scholar shines.' Raleigh girl, 15, attends NCSU - and eyes Duke, Midtown Raleigh News, 12/02/2009
10. 'Competitor doesn't act her age,' News and Observer, 11/30/2009
11. North Carolina Students Sweep Regional Finals of Nation's Premier High School Science Competition for Research in Biochemistry and Computer Science,' CSR Wire, 11/16/2009
12. 'The Nation's Best Math and Science Students Earn Top High School Honors and Strive for \$100,000 Scholarship,' PRNewswire, October 29, 2010
13. 'Raleigh teen finishes in second in Siemens Competition,' News and Observer, December 7, 2009
14. 'Student headed for science contest,' News and Observer, November 3, 2009
15. 'N.C. teens finish second in science contest,' News and Observer, December 7, 2009
16. 'ORNL-Mentored Students Named 2009 Siemens Regional Finalists,' DOE - Science - ASCR - Monthly News Roundup, November -December, 2009
17. 'Six who live or work in OR named YWCA Tribute to Women finalists,' Oak Ridger, August 2, 2007
18. 'Outstanding mentors and teachers honored at ORNL,' ORNL News Release, Feb. 14, 2008
19. 'U.S. Patents Awarded to Inventors in Tennessee,' Targeted News Service, Jan 5, 2008



20. 'Rising Stars,' Science 314, 1665e (2006)
21. 'National Competitive Investment Act, Senator Mr. Alexander, The Library of Congress, Congressional Record, Senate, Page S11588, Dec. 8, 2006
22. 'New Opportunities for Data and Information: Finding the Dots, Connecting the Dots, Understanding the Dots,' Presentation by Dr. Raymond L. Orbach, Director, Office of Science, at The Expanding Universe of Digital Data Collections Symposium, 2006 AAAS Annual Meeting, St. Louis, MO
23. Nation's Top Science Prize for High School Students Goes to Mathematics and Bioinformatics Research,' 2006-07 Siemens Competition in Math, Science & Technology News Release, Dec 6, 2006
24. 'ProRata TOOLbox,' Journal of Proteome Research, Vol. 5, No. 11, 2006
25. 'From Data to Discovery, SciDAC Review, Fall 2006
26. 'Gu semifinalist in Intel Science Talent Search,' Oak Ridger, January 24, 2007
27. 'Oak Ridge: Science making history,' Oak Ridger, January 23, 2007
28. 'ORHS team wins \$100K,' Oak Ridger, December 5, 2006
29. 'Oak Ridge's brightest on top,' Knoxville News Sentinel, December 5, 2006
30. '3 OR students: National science champs,' Knoxville News Sentinel, December 4, 2006
31. 'ORHS team wins regional Siemens contest,' Oak Ridger, November 9, 2006
32. Oak Ridge High students Siemens Competition finalists,' Oak Ridger, November 1, 2006
33. Local students take home \$100,000 prize in national science competition,' Wate 6, The News Radio Station, December 4, 2006
34. 'Massachusetts & Tennessee Teens Take Regional Title in Nation's Top High School Science Competition,' Siemens News Release, November 4, 2006
35. 'ORNL-Mentored Oak Ridge High School team wins Siemens Competition,' Oak Ridge, Dec.6, 2006
36. 'High School Bioenergy Champions' ORNL Review, 2007
37. 'Tennessee students and researcher win \$900,000 US for biofuel mathematical research (R&D),' Bioenergy and Biofuels, Dec 11, 2006
38. 'Oak Ridge's brightest on top,' Innovation Valley - Nanotech, December 5, 2006
39. 'Siemens Science & Technology Competition National Winners,' Cogito, December 5, 2006
40. 'Oak Ridge High School Team wins \$100K,' Rarity Community, December 5, 2006
41. 'Scientists' 'genetic noise' research published in Nature,' Oak Ridger, February 7, 2006
42. 'Gene Network Shaping of Inherent Noise Spectra,' Center for Nanophase Materials Center News, 2006
43. 'ORNL supercomputer helps trio excel, Computer Science and Mathematics PR, ORNL, Dec 5, 2006
44. 'Last-minute idea collects dividends,' Advanced Physics Forums, Nov 7, 2005

45. 'Nation's Top Science Prize for High School Students Goes to Mathematics and Bioinformatics Research,' Yahoo News, Dec 4, 2006
46. 'Nation's Top Science Prize for High School Students Goes to Mathematics and Bioinformatics Research,' Digital50, Dec 4, 2006
47. 'UT and ORNL scientists' genetic research published in journal,' The Oak Ridge Observer, Feb 9, 2006
48. 'Oak Ridge High School students to participate in national science competition,' Oak Ridge, TN, Dec. 6, 2005
49. 'ORNL leading effort to help harness power of Shewanella,' Oak Ridge, TN, Nov 11, 2005
50. 'ORHS team wins \$30,000 at Siemens Westinghouse competition,' Oak Ridger, December 6, 2005
51. 'ORHS team wins regional science competition, headed to nationals,' Oak Ridger, November 23, 2005
52. 'ORNL leading research efforts involving Shewanella microbe,' Oak Ridger, November 18, 2005
53. '3 from ORHS win regional science event,' Computer Science and Mathematics PR, ORNL, Nov 21, 2005
54. 'Three from ORHS Win Regional Science Event,' Tennessee Valley Corridor, November 11, 2005
55. 'Project leads students to high-profile academic competition,' Oak Ridge Observer, Issue 48, Vol 1, 2005
56. 'New Biological Discoveries through Data Intensive Computing,' Linux Electrons, Nov 13, 2004
57. 'New Biological Discoveries through Data Intensive Computing,' Supercomputing Online, Nov 3, 2004
58. 'Two DOE Labs Buck the Biocluster Trend to Test-Drive a Few Architecture Alternatives,' GenomeWeb, Volume 8, Number 47, December 6, 2004
59. 'HPC Beyond Clusters,' John S. McNeil, Genome Technology, March, 2004
60. 'Data intensive computing leads to biological discoveries,' Oak Ridge, TN, Dec 10, 2004
61. 'Computational Technology,' CCN Magazine, Dec 14, 2004
62. 'ORNL staffers honored for support of tech transfer initiatives,' Oak Ridger, May 29, 2003
63. 'Bandwidth Challenge Teams Push Networking Performance Envelope at SC2003 Conference - Sustained 23 Gigabits per Second Sets New Record,' Supercomputing 2003, Application Foundation Award
64. 'Bandwidth Challenge Teams Push Networking Performance Envelope at SC2003,' Supercomputing Online, Nov 25, 2003
65. 'Bandwidth Challenge Teams Push Networking Performance Envelope at SC2003,' Daily News and Information for the Global Grid Community, December 8, 2003: Vol 35, No 49
66. 'Developing Computer Tools for Scientists,' ORNL Review, Vol. 35, No. 1, 2002
67. 'Developing computer tools for scientists,' DOE OS Research News, 2002
68. 'Retaining and Retrieving Data More Effectively,' ORNL Review, Vol. 35, No. 1, 2002

69. 'Retaining and Retrieving Data More Effectively,' DOE OS Research News, 2002

## Public Service

- OAK RIDGE HIGH SCHOOL MATHEMATICS THESIS PROGRAM

During 2005-2007 I have mentored more than half a dozen senior high school students in their Mathematics Thesis Program. The students participation in this program resulted in a number of national level students' awards (see 'High School Math and Sciences Theses Directed' section)

- DOE RESEARCH ALLIANCE IN MATH AND SCIENCES (RAMS) PROGRAM

Since 2003, each summer I mentor two-to-three internship students from Black Colleges and Universities and Minority Educational Institutions.

- SUMMER STUDENTS TRAINING PROGRAMS Since 2005, each summer I teach various 1-4 week workshops for senior high school and undergraduate students:

Computational Systems Biology, 2007

Enrollment: 10 undergraduate from UC Mercede and 4 high school students from ORHS

Object-Oriented Programming, 2006

Statistical Theory: Basics, 2005-2006

Graph Theory and Algorithms, 2005-2006

Artificial Intelligence Methods, 2005

## B. Program Impacts

N/A

## IV. TECHNOLOGICAL AND MANAGERIAL INNOVATIONS

### A. Knowledge and Technology Transfer Accomplishments

1. VONODE, A *de novo* sequencing tool that exploits the potential of high-resolution MS/MS data by using a unique tag scoring function and a novel type of spectrum graphs (<http://compbio.ornl.gov/Vonode>).
2. NETWORK INSTANCE-BASED BIASED SUBGRAPH SEARCH (NIBBS), Tool for comparing hundreds of genome-scale metabolic networks to identify metabolic subsystems that are statistically biased toward phenotype-expressing organisms (<http://freescience.org/cs/NIBBS/>).
3.  $\alpha, \beta$ -MOTIF FINDER, Tool for identification of protein functional modules (e.g., metabolic subsystems, regulators, sensors, transporters, uncharacterized proteins) that are predicted to be related to the target phenotype (<http://freescience.org/cs/ABClique/>).
4. MCE-PARALLEL, A scalable, parallel algorithm for the NP-hard clique enumeration problem (<http://freescience.org/cs/parallelclique/>).
5. BI-CLUSTERING, Software for identification of phenotype-related modules and their putative cross-talks (<http://freescience.org/cs/bi-clustering/>).
6. DENSE ENRICHED SUBGRAPH ENUMERATION (DENSE), Tool for identifying protein functional modules that are “enriched” by query proteins representing knowledge *priors* (<http://freescience.org/cs/DENSE/>).
7. SYSTEM PHENOTYPE-RELATED INTERPLAYING COMPONENTS ENUMERATOR (SPICE) iteratively enumerates statistically significant and phenotype-relevant cellular subsystems and can be applied to both network data and gene-expression data (<http://freescience.org/cs/SPICE>).
8. HIERARCHICAL MODULARITY SCORE (HMS), Tool for functional annotation and functional enrichment analysis of hierarchically organized protein functional modules by taking into account an inherent design principle of biological networks—hierarchical modularity (<http://freescience.org/cs/HMS/>).
9. IN-SITU SORT-AND-B-SPLINE ERROR-BOUNDED LOSSY ABATEMENT (ISABELA), Library for in-situ, embarrassingly parallel  $B$ -spline based lossy compression of scientific floating point data with user-controlled accuracy bounds (software available upon request).
10. ISABELA-QA, A memory and storage light-weight parallel query processing engine over ISABELA-compressed scientific data capable of multi-core, multi-node, GPU executions (software available upon request).
11. ANALYTICS-DRIVEN LOSSLESS DATA COMPRESSION FOR RAPID IN-SITU INDEXING, sTORING, AND QUERYING (ALACRI<sup>2</sup>TY), Fast and memory light-weight query processing (and lossless compression) engine for scientific floating point data that is optimized for heterogeneous access pattern (software available upon request).
12. IN-SITU ORTHOGONAL BYTE AGGREGATE REDUCTION (ISOBAR) COMPRESSION, A preconditioner-based, high-throughput lossless compression technique for hard-to-compress scientific datasets (software available upon request).
13. PRM\_CAUSALITY, A tool for data-driven, semi-automatic inference of plausible phenomenological models ([http://freescience.org/cs/prm\\_causality/](http://freescience.org/cs/prm_causality/)).
14. FORECAST ORIENTED FEATURE ELIMINATION-BASED CLASSIFICATION OF ADVERSE SPATIO-TEMPORAL EXTREMES (FORECASTER), A forecast-oriented feature elimination-based ensemble of classifiers for robust forecasting of adverse spatio-temporal extremes (software available upon request).

15. FORECAST ERROR DETECTION AND CORRECTION (DETECTOR), A library for detecting and correcting prediction errors in extreme event forecasts (software available upon request).
16. ANOMALOUS COMMUNITY GENERATOR, A library for detecting predictive and phase-biased communities in contrasting groups of networks (software available upon request).
17. METABOLIC PATHWAY ALIGNMENT, Tool for evolutionary analysis of metabolic pathways via multiple metabolic pathway alignment (software available upon request).
18. PROTEIN NETWORK ALIGNMENT ALGORITHM, Tool for evolutionary analysis of protein functional association networks via alignment of multiple protein functional association networks (software available upon request).
19. VIPAR: VIRTUAL INFORMATION PROCESS RESEARCH AGENT, System for gathering and summarizing internet information.  
Impact: US Patent, Awarded on Jan 5, 2008 07072883 Cl. 707-3.
20. SHEWANELLA KNOWLEDGEBASE, Integrated Knowledgebase for DOE Shewanella Federation.  
Impact: A continuously growing knowledgebase used by a large community of users.
21. PRO-RATA: QUANTITATIVE SHOTGUN PROTEOMICS SOFTWARE, Open source software distributed from: <http://www.MSProRata.org>  
Impact: More than 1000s downloads since June 2006 till present; featured in the *Journal of Proteome Research*, Vol. 5, No. 11, 2006.
22. PR: PARALLEL R FOR HIGH PERFORMANCE STATISTICAL COMPUTING, Open source software distributed from R's CRAN: <http://cran.r-project.org/src/contrib/Descriptions/RScalAPACK.html>  
Impact: Distributed through more than 30 mirror sites across around 20 countries. RScalAPACK library is part of RedHat Linux distribution through RPM packages.
23. ORNL SEMANTIC TAGGER, Open source software distributed from [samatovan@ornl.gov](mailto:samatovan@ornl.gov)  
Impact: Used by the Department of Homeland Security (DHS) Biodefense Knowledge Center (BKC) and ADVISE program.
24. CORPUS INDEPENDENT KEYPHRASE EXTRACTION TOOL, Open source software distributed from [samatovan@ornl.gov](mailto:samatovan@ornl.gov)  
Impact: Deployed within the DHS BKC and ADVISE systems.
25. MPIBLAST-PIO: EFFICIENT PARALLELIZATION OF NCBI BLAST, Open source software distributed from: <http://mpiblast.lanl.gov/Downloads.Stable.html>  
Impact: The incorporation of performance optimizations in the NCSU-ORNL research prototype pioBLAST into the popular open-source mpiBLAST software.
26. PGRAPH: LIBRARY OF PARALLEL AND SCALABLE GRAPH ALGORITHMS, Open source software distributed from [samatova@csc.ncsu.edu](mailto:samatova@csc.ncsu.edu)  
Impact: Used by a number of computational biology projects, developed in collaboration with Cray Inc.
27. PARALLEL SATLEED: LIBRARY OF PARALLEL AND SCALABLE GRAPH ALGORITHMS, Parallel Symmetrized Automated Tensor Low Energy Electron Defraction  
Open source software distributed from <http://www.cs.utk.edu/~samatova/LEED/report.html>  
Impact: Used by multiple scientists internationally.
28. PESS, Parallel Exact Stochastic Simulator  
Open source software distributed from [samatovan@ornl.gov](mailto:samatovan@ornl.gov)

## B. Program Impacts

- My team’s pioneering work on parallel R (PR) has provided a broad community with an easy-to-use middleware between the popular R environment for statistical computing and scalable data analytics kernels through both data and task parallelism. Its constituent RScalAPACK and taskPR packages are distributed across nearly 30 mirror sites around the world from the official R’s CRAN software distribution website; RScalAPACK’s RPM package is also part of various Linux distributions, such as Fedora, RedHat, and Debian.
- My team’s data compression, indexing, and query processing technologies (ALACRITY, ISABELA and ISOBAR) have been selected to be part of production software (ADIOS, GLEAN) in the DOE computing facilities.
- My team’s computer program PRORATA automated the entire data analysis process for both stable-isotope-labeled and label-free quantitative shotgun proteomics, and has been used by thousands of users. It has enabled improved reconstruction of metabolic pathways in the hydrogen producing bacteria, *Rhodospseudomonas palustris*, in the acid mine drainage microbial community, and improved functional annotation of bioethanol producing bacteria, *Zymomonas mobilis* and *Clostridium thermocellum*.
- My team provided an unprecedented accuracy in seasonal hurricane activity forecasts (92% vs. 65% by traditional methods) as well as 90% accuracy and 10-15 day lead-time forecasting the end-game of hurricane tracks (land-hitting or offshore). This work received a lot of attention in national and international press including the top “breaking news” story on the National Science Foundation’s official Science360 news site, the DOE news web-site, various newspapers including *Raleigh News and Observer*, television highlights including *WRAL*, etc.
- My high school mentorship accomplishments have been read into the Congressional Record. Senator Lamar Alexander commented on the floor of the Senate during his speech in support of the National Competitiveness Investment Act [Congressional Record, Senate, Page S11588, Dec. 8, 2006] commented that “Part of the reason these three students succeeded is they were able to connect with the work and expertise at the Oak Ridge National Laboratory [and their] lead adviser, Dr. Nagiza F. Samatova ...” Numerous acknowledgements in the news, such as *The New York Times*, and in press, such as *Science* magazine, have drawn attention to the importance of early science education and the role my teams have played.

## V. Service to the University and Professional Societies

### Committee Appointments

#### COLLEGE OF ENGINEERING COMMITTEES

1. CoE Research Committee, 2009-present

Advised Associate Dean for Research and Graduate Programs on indentifying and promoting major research thrust areas for the College; encouraged and promoted cross-disciplinary research efforts; reviewed research policies and procedures and identified improvements to enhance faculty success in obtaining research funding; and assisted the Associate Dean in the review and oversight of CoE Research Centers.

#### DEPARTMENT OF COMPUTER SCIENCE COMMITTEES

1. Faculty Advisor, Graduate Student Association (GSA), 2008-present

2. Member, NSF-CPATH Team, 2010

3. Chair/Member, Undergraduate Course Assessment Committee:

- CSC 333 – Automata Theory – Prof. Robert Fornaro (chair), Prof. Robert Rodman, and Dr. Nagiza Samatova (course instructor)
- CSC 405 – Introduction to Computer Security – Prof. Peng Ning, Prof. Khaled Harfoush, Dr. Nagiza Samatova (chair), and Sammie Carter (course instructor)
- CSC 474 – Information Systems Security – Prof. Peng Ning (course instructor), Prof. Khaled Harfoush, Dr. Nagiza Samatova (chair), and Sammie Carter

4. Computing Infrastructure Committee, 2007-present

5. Planning for Increasing Research Funds Committee, 2007-2008

6. Assisting Faculty in Proposal Writing, 2007-present Provide regular assistance to faculty on funding opportunities and increasing the chances for winning the proposals.

7. Peer Evaluation Committee, 2007-present

Conducted teaching evaluations of

- Prof. Tao Xie (April, 2008)
- Prof. Steffen Heber (April, 2008)
- Prof. Kemafor Anyanwu (February, 2008)

8. Meeting with Faculty candidates, 2007-2008: Haixun Wang (IBM), Zhenchang Xing (U. Alberta, Canada), Ingolf Krueger (UCSD), Anotnio Roque (USC), Barbara Ryder (State Univ of New Jersey), Naren Ramakrishnan (Virginia Tech)

9. Giving lectures to new graduate students in Fall, 2007-present

10. Advising PhD students seeking faculty/job positions: Harini Ramapra, Sibir Mohan, Wenhong Tian, Nirmal Desai, Michael Gegick

## National and International Activities

I have been, and continue to be, involved in numerous national activities in various roles:

- DOE/DHS/NSF WORKSHOP ORGANIZING COMMITTEES

1. *Session Co-Chair*, DOE Workshop on Scientific Grand Challenges in Biological Sciences and the Role of Computing at the Extreme Scale, Aug 17-19, 2009, Chicago, IL
2. *Session Co-Chair*, The Scientific Discovery through Advanced Computing (SciDAC) 2009 Conference, June 14-18, 2009, San Diego, CA
3. *Session Co-Chair*, DOE Workshop on Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale, March 18-20, 2009, Washington DC
4. *Session Co-Chair*, DOE Genomics:GTL Systems Biology Knowledgebase Workshop, Fall-Spring, 2009 (writing report)
5. *Co-Chair*, DOE/NSF Mathematics for Analysis of Petascale Data, June 2008
6. *Co-Chair*, DOE Genomics:GTL Program Systems Biology Knowledgebase, May 2008
7. *Co-Chair*, OE/OASCR Mathematics for Peta-scale Data Analysis & Computation, June 2007
8. *Technical Session Chair*, DOE Workshop on Modeling and Simulation at the Exascale for Energy and the Environment, Apr, May, June 2007
9. *Chair*, DOE SciDAC Biology Summit, August 2005
10. *co-Chair*, DHS Threat Awareness Customer Data Integration & Dissemination, November 2005
11. *co-Chair*, DHS Data Sciences for Homeland Security Information Management and Knowledge Discovery, September 2004
12. *Committee Member*, DOE Data Sharing & Integration Summit, April 2006
13. *Committee Member*, DOE SciDAC Conference, June 2005
14. *Committee Member*, Scientific Data Management, April 2004
15. *Committee Member*, OE Computational Infrastructure for the Genomes to Life Program, July 2002

- GRANT PROPOSAL REVIEW PANELS:

1. DOE OASCR CAREER panel, 2013 *New in 2012*
2. NSF III-SMALL DB Misc panel, 2013 *New in 2012*
3. NSF SEES panel, 2012 *New in 2012*
4. DOE OBER CAREER panel, 2009
5. DOE OASCR CAREER panel, 2009
6. Pennsylvania Department of Health (PA DOH) Proposal Performance Reviewer, 2008-2010
7. NSF III Core panel, 2009
8. NSF CAREER panel, 2008
9. Army Research Office (ARO), 2009
10. DOE OBER SBIR/STTR panel, 2007



11. DOE OASCR SBIR/STTR panel, 2007
  12. DOE OBER SBIR/STTR DE-FG01-05ER05-15 solicitation, 2004
  13. DOE OASCR SBIR/STTR DE-FG01-05ER05-28 solicitation, 2004
  14. DOE Early Career Award Panel, 2003
  15. Chair of the External Review Panel for the DOE GTL VIMSS Center, 2006
  16. DHS Internal Proposal Review BAA 06-003 Panel, 2006
  17. DHS External Proposal Review BAA 06-003 Panel, 2006
  18. DHS Pre-proposal Review BAA 06-003 Panel, 2006
  19. NSF Proposal Reviewer for the NSF05-570 solicitation, 2005
  20. DOE OASCR unsolicited proposal reviewer, 2002-present
  21. ORNL Seed Money Proposal Reviewer, 2002-present
- JOURNAL REVIEWER:
    1. IEEE/ACM Transactions on Computational Biology and Bioinformatics
    2. BMC Bioinformatics
    3. Special Issue on Model Based Testing
    4. Journal of Bioinformatics and Computational Biology
    5. Bioinformatics
    6. ACM Computing Reviews
    7. IEEE Transactions on Pattern Analysis
  - TECHNICAL PROGRAM COMMITTEE MEMBER for several conferences and workshops:
    1. International Conference on Scientific and Statistical Database Management (SSDBM 2010)
    2. International Conference on Parallel Processing (ICPP), 2009
    3. IEEE International Symposium on Bioinformatics and Life Science Computing (BLSC), 2007
    4. IEEE Workshop on High Performance Computing in Medicine and Biology, 2005-2006  
International Conference on Scientific and Statistical Database Management (SSDBM), 2002, 2004
    5. SIAM International Conference on Data Mining, 2003
    6. IEEE Bioinformatics Conference, 2003
    7. C. Warren Neel Conference on Statistical Data Mining, 2002
    8. CLUSTER: IEEE International Conference on Cluster Computing, 2002
  - SESSION CHAIR at several conferences and workshops
  - PANEL ORGANIZER at several conferences and workshops
  - BOOKS REVIEWER:
    1. Elain Rich, Automata, Computability and Complexity, Prentice Hall, 2008
  - REVIEWER for the major journals and conferences in bioinformatics and data mining