

Advanced Nonparametric Smoothing Methods

BY ARMAND VLADAU, COMMUNICATIONS OFFICER
(FIMS INTERN, THE UNIVERSITY OF WESTERN ONTARIO)



Jeffrey Racine
McMaster researcher

Statisticians play a pivotal role in developing and improving methods of estimation and inference used in many disciplines, such as the economic and medical sciences.

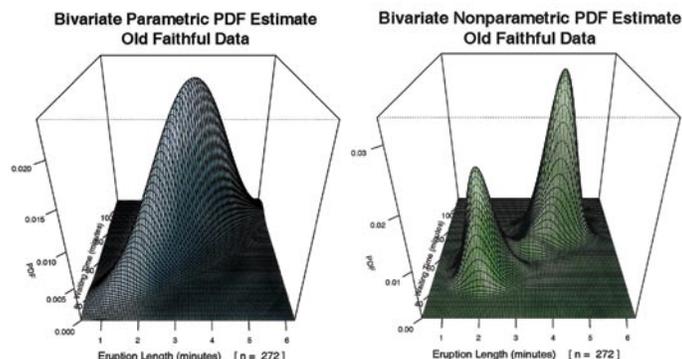
The field of nonparametric statistics continues to grow at an exponential rate. Many nonparametric smoothing

methods that were developed decades ago were computationally burdensome, severely limiting their application. Other nonparametric methods have only very recently been developed, facilitated directly by dramatic improvements in technology, specifically high performance computing resources such as those at SHARCNET. Dr. Jeffrey Racine, a Professor in McMaster University's Economics Department and in the Graduate Program in Statistics in the Department of Mathematics and Statistics, places a heavy emphasis on "the social importance of having a sound statistical basis for conclusions". He favors the methods that impose as few assumptions as possible on the relationship being studied, unlike more traditional "parametric" methods that often impose rigid and potentially false structure on (unknown) underlying relationships.

Dr. Jeffrey Racine is a tenured professor and holds a Senator William McMaster Research Chair at McMaster University.

He has recently co-authored an advanced textbook on nonparametric kernel techniques titled "Nonparametric Econometrics: Theory and Practice" (*Princeton University Press, 2007*) that explores the properties of nonparametric estimators when one is dealing with a mix of categorical and continuous datatypes, found in real-world data.

Nonparametric smoothing methods are statistical techniques that do not require a researcher to specify functional forms for objects being estimated. Instead, the data itself informs the resulting model in a particular manner. In a regression framework this approach is known as 'nonparametric regression' or 'nonparametric smoothing'. Such methods are becoming increasingly popular for applied data analysis because they are best suited for dealing with large data sets "for which the number", Dr. Racine explains, "of variables involved is manageable". These methods are often used after common parametric specifications are found to be unsuitable for the problem at hand, particularly when a parametric model yields no clues as to the direction in which to search for an improved parametric



in this issue:

- Message from the Chair of the Board
- Scientific Director's Message
- SHARCNET helps Team Canada Selection Panel
- SHARCNET Awards \$500k in Fellowships
- Labour Economics and Economics of Education: Human Capital Investment
- PASCO'07
- School's In for Summer
- Swimming with SHARCS: New Staff
- SHARCNET Visits Local Science Fair
- Tech Bytes: User Certification Program Q&A
- New User Certification Program

Message from the Chair of the Board



Ted Hewitt,
Chair of the Board

Since the last time I reported, SHARCNET has been a hub of activity, primarily due to the infusion of new operating funding from the Ministry of Research and Innovation's Ontario Research Fund (ORF). SHARCNET has been using these funds in a concerted outreach effort to the newer partners, working to connect the network to these sites, installing remote collaboration facilities, hiring additional technical staff, launching renewed research and training programs with a focus on non-traditional HPC disciplines, as well as promoting HPC to Ontario's youth.

All of this work fits within the larger provincial HPC umbrella, driven by the Ontario High Performance Computing Council, which was formed in 2006 to provide effective oversight of the province's academic HPC capacity. This group has been working steadily over the last year generating a provincial HPC research and development capacity plan which we hope to release in the coming months. Assisting in this effort is the HPC Coordinating Committee, consisting of the Scientific Directors from each of the three provincial consortia, SHARCNET, SciNet and HPCVL. This Coordinating Committee will play a key role in the implementation of the plan, coordinating utilization of resources, and in reporting.

On the national front, the creation of Compute Canada is well underway. Compute Canada is a one-of-a-kind initiative and represents a major step forward in building HPC capacity and collaboration across Canada. It also fulfills one of the highest priority recommendations from the C3.ca-sponsored Long Range Plan for HPC.

As was earlier reported, SHARCNET and the Shanghai Supercomputer Center are building an international relationship and along with Hugh Couchman, Scientific Director, and Baolai Ge, HPTC Consultant, I intend to travel to Pudong, Shanghai at the end of October for a formal signing of a collaboration document between our two organizations.

More locally, there have been some additional changes to the SHARCNET Board of Directors over the summer months. Dr. Allison Sekuler, Associate Vice-President, Research, takes over for Dr. Mamdouh Shoukri as McMaster's Board representative. As most of you may already know, Mamdouh has taken the post of President and Vice-Chancellor for York University. Dr. George Dixon has assumed the role of Board representative for the University of Waterloo, replacing Dr. Alan George as their Vice-President, Research. I wish to thank both Mamdouh and Alan for their valued contributions and support of SHARCNET.

This Fall is shaping up to be even busier for SHARCNET as we continue to roll out renewed programs and engage with our newer partners. Users should watch the SHARCNET website for details of SHARCNET news, events and training opportunities.

W. E. (Ted) Hewitt, Ph.D
Chair, SHARCNET Board of Directors and Vice-President
(Research & International Relations), The University of Western Ontario

“The creation of Compute Canada is well underway. Compute Canada is a one-of-a-kind initiative and represents a major step forward in building HPC capacity and collaboration across Canada.”

-Ted Hewitt

Scientific Director's Message

I am pleased to present the Summer edition of SHARC Bytes and provide an update on what's been happening at SHARCNET over the last several months.

The SHARCNET Summer School Program, inaugurated in June as a week long series of workshops and courses, was a huge success and we look forward to offering it again next summer. (This new program replaces the Fall Workshop series.) The SHARCNET literacy seminars, which are held over AccessGrid every other Monday have been very well received and we will continue to offer these sessions throughout the year to enable new users to quickly get up to speed on our systems. We have also been thrilled with the steadily increasing use of our AccessGrid rooms which allow for increased research training and collaboration across sites. SHARCNET provides a wide and varied range of training opportunities for the user community, and I encourage you to watch the SHARCNET website for upcoming events.

We have also made significant headway with our high school outreach program, and plan to launch our "supercomputer-in-a-box" initiative this Fall. The "supercomputer-in-a-box" is a portable, mini-supercomputer on wheels, loaded with HPC demo applications that will be used for presentations at high schools, conferences and community events. The intent is to be able to demonstrate in a tangible way why HPC is so important in modern society. If you would be interested in having this demonstration as a part of an activity that you are planning, please contact us.

SHARCNET recently conducted a user satisfaction survey soliciting feedback on all aspects of our organization and operations. The results of the survey are being used to further enhance SHARCNET's services and operations and refine outreach efforts. The survey has already led to the recent development of in-house SHARCNET tools for Site Leaders to detect those users who seem to be struggling in the use of SHARCNET systems, so they can follow-up with these users directly. Thanks to the user community for providing such useful feedback. The results of the survey are available via the SHARCNET webportal for those who are interested.

Since I last reported, we have completed Round VI of the SHARCNET Fellowships and Dedicated Resources Programs and have recently launched Round VII. The Chairs Program was also re-introduced a few months ago, and Chair proposals will be reviewed by the SHARCNET Board at their September meeting. These programs, which enable strategic faculty hiring in HPC and facilitate research and training of postdocs and students, have been tremendously successful and we look forward to continuing these very valuable programs.

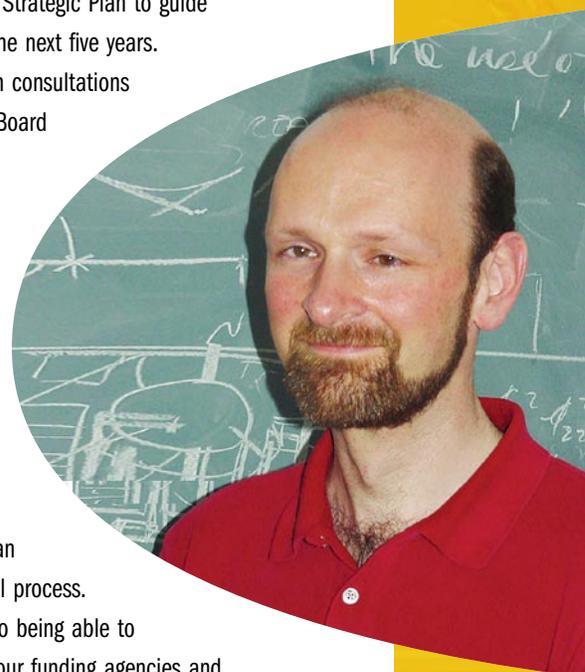
Following a meeting of its Strategic Council last December, SHARCNET has been developing a Strategic Plan to guide our focus and development over the next five years.

The plan has been drafted through consultations with the SHARCNET Site Leaders, Board members, management and the SHARCNET research community. Once approved, this plan will form the basis for SHARCNET's strategic directions over the next five years.

As the summer winds down, SHARCNET will be seeking new research and publications data from researchers as part of an annual SHARCNET account renewal process.

This information gathering is key to being able to provide up-to-date information to our funding agencies and to demonstrate the value of their investments. I encourage researchers to use this process to keep us posted on your research efforts and outcomes.

Hugh M.P. Couchman
SHARCNET Scientific Director
Fellow, Canadian Institute for Advanced Research
Professor, Physics and Astronomy, McMaster University



Hugh
Couchman,
Scientific Director

SHARCNET helps Team Canada Selection Panel



BY TED MEDSON, SELECTION PANEL, TEAM CANADA

SHARCNET Awards \$500k in Fellowships

Earlier this summer, SHARCNET was pleased to announce the results of the Round VI competition for Fellowships and Dedicated Resources. With over \$2.5m in funding requests from 120 applications, this was a particularly competitive round.

SHARCNET awarded \$525,312 over two years for 33 new Fellowships for visiting scientists, graduate, undergraduate and postdoctoral students. With matching contributions of \$440,084 provided by the researchers themselves, the total investment to increasing the number and quality of research personnel using high performance computing totals close to \$1m.

For details of individual awards, please visit: www.sharcnet.ca/research/roundVIresults.html

SHARCNET wishes to acknowledge **Hewlett Packard** and **Silicon Graphics** for their generous financial contributions which support this very important research program.



The Intel International Science & Engineering Fair is the Olympics of Science Fairs. Held every year in a city in the USA, it brings together over 1,500 young scientists from 40 countries to share their projects, explore some extraordinary science and compete for over four million dollars in prizes, scholarships and awards. The Intel ISEF 2007, presented by Agilent Technologies, was held May 13-19, 2007 in Albuquerque, New Mexico. The best 16 students in Canada were in attendance. Selecting the best young scientists from the pool of applicants is a challenging task indeed, and we are grateful for the support, technology and facilities that SHARCNET provided to allow this process to proceed so smoothly.

The selection panel of sixteen had professors from science and engineering, school teachers, and some of the most successful previous attendees at ISEF with many international awards between them. We met from 9:45 am until the final moment of decision at 5:00 pm. During this time, the panel interviewed all 32 of the final round of applicants, using SHARCNET technology. We used the information garnered from the five page reports, the PowerPoint presentations and the interviews to finally decide on the top sixteen students, twelve individuals and two pairs, who represented us in Albuquerque.

This year's Team Canada-ISEF included 8 ISEF veterans and 8 newcomers from Alberta, British Columbia, Manitoba, Ontario, Québec and Saskatchewan. Here are three of the fourteen projects selected:

- Ben Gulak & Jason Morrow, Milton & Hamilton, ON
Tango: Tomorrow's Transportation Solution
Ben and Jason have developed a unicycle that anyone can ride, using sophisticated stabilization systems.
- Adrian Veres, Montréal, QC
Design and Development of a Novel DNA Computer System for Advanced Genetic Analysis
- Sarah McCuskee, Regina, SK
Lightning in the Laboratory: Electromagnetic Radiation from Red Sprites

The full list of projects is available at: www.yf.ca/News_Releases/news03250701.aspx

Both the young scientists and the judging panel agreed that this new selection process was superior in several ways to previous years, and we look forward to continuing collaboration and support from the SHARCNET team in London for this important contribution towards the development of the next generation of Canadian scientists. For further information on the Intel ISEF visit: www.sciserv.org/isef/

Labour Economics and Economics of Education: Human Capital Investment

BY ARMAND VLADAU, COMMUNICATIONS OFFICER (FIMS
INTERN, THE UNIVERSITY OF WESTERN ONTARIO)

In order to ensure a flourishing and prosperous economic future, it's important for nations to investigate the factors determining a skilled and productive workforce. The study of 'human capital' refers specifically to empirically examining policies and early investment practices that influence the future productivity of the population—and therefore, the productivity of the nation. Recent studies have found that for approximately every dollar that is spent on early childhood 'interventions', a staggering nine dollars in benefits are returned to society. Ultimately, by examining the causal effects of education and training investments in the life of a person, we can predict their economic contribution to the nation.

Dr. Lance Lochner, a tenured professor at the University of Western Ontario, in the Department of Economics, is currently researching the casual relationships of (i) early development and the family, (ii) post-secondary education finance policy, and (iii) how post-school human capital formation manifests in the work place. Dr. Lochner's primary fields of studies are: Labour Economics, Public Economics, Economics of Education, and Economics of Crime.

Dr. Lochner's present research is primarily focused on the estimation and simulation of, as he says, "complex behavioral models that focus on human capital formation over the human lifecycle". Essentially, his studies are concerned with improving the existing understanding of human capital and economic dynamics through the use of complex, richer and, therefore, more realistic empirically estimated models.

In his research, Dr. Lochner explores how effectively new economic theories reflect actual borrowing, repayment and default incentives. By more accurately illustrating the influence of current higher education policy and methods, the financial problems associated with post-secondary education can be significantly improved, thereby decreasing personal debt and increasing national productivity.

His research also extends beyond early childhood and past the post-secondary lifecycle stage into a quantitative investigation of two important post-school skills called 'learning-by-doing' and 'learning-on-the-job'. Learning-by-doing refers to how a worker learns at work from the work itself. Learning-on-the-job, or more commonly known as 'on-the-job-training', is when the worker is educated at the place of employment. Research into these fields will attempt to clarify the causal factors that determine why some workers experience little wage growth, whereas others experience the doubling of their wages, over their lifecycle.

Dr. Lochner's research is based on the estimation and calculation of relatively complex behavioral models that require the submission of extremely large data sets and great computational power and speed. Dr. Lochner makes use of SHARCNET resources to estimate and calibrate models related to his research. Since SHARCENET is equipped with high performance computing technology, he is able to incorporate more constraints and decisions in the simulations, thereby creating richer models, ultimately more comprehensive and realistic simulations.

The implications of Dr. Lochner's research into new theories of human capital, in terms of early childhood investment, financial behavioral models at the levels of post-secondary borrowing/repayment and post-school learning combine, to formulate an infinitely more accurate representation of Labour Economic and Education Economic theories.

Dr. Lochner says that "understanding the processes by which individuals acquire skills and [as well as] the constraints affecting still investment decisions is... crucial to designing effective government policy and promoting productivity in Canada". Essentially, the heightened understanding of the three primary areas of human capital formation will not only lead to the improvement of financial deficiencies in a person's life, but it will also aid in the creation of radically better public policy.



Lance Lochner
Western researcher

PASCO '07

BY BAOLAI GE, HPTC CONSULTANT, SHARCNET AND MARC MORENO MAZA, COMPUTER SCIENCE AND APPLIED MATHEMATICS, THE UNIVERSITY OF WESTERN ONTARIO

After 10 years of interruption, the International Workshop on Parallel Symbolic Computations (PASCO) was held for 3 days this summer at The University of Western Ontario. The pervasive ubiquity of parallel architectures, from SMPs to multi-core laptops, has naturally led to the renaissance of this meeting. The goal of the workshop was to stimulate the development of parallel algorithms and software for achieving high performance in symbolic computation at all scales from grids to personal computers.

The sessions, led by experts in high performance computing, tried to address the many HPC questions that the wondering computational scientists and engineers had on their minds. The first day started early with a quick introduction to SHARCNET's parallel computer infrastructure. This session gave the first-time audience a fresh start on how to use the resources on SHARCNET with over 8,000 processors. A few important messages went out to the audience: 1) getting a SHARCNET account is free! 2) user programs are run in batch mode and scheduled on a fareshare basis, and 3) support for programming and project consultation is available and is also free.

The two-hour, intensive HPC session that followed covered many areas including: modern computer architecture, basics in floating point arithmetic, concepts of software pipelining, cache friendly algorithms, parallel processing with message passing and multi-threaded programming for parallel and concurrent processing. The audience was walked through a number of working examples that demonstrated how to think in parallel and implement algorithms for distributed and shared memory environments using MPI and POSIX thread APIs in details. The tutorial also emphasized the fact that HPC is not just about using supercomputers but also the efficient use of desktops and workstations in day-to-day computing. The audience saw how a change of algorithms made a matrix-matrix multiplication program "fly" on a laptop. What amused the audience further is it ran 10 times faster with a modified algorithm on SHARCNET systems, and with the use of a highly optimized BLAS library, it could achieve an extra performance boost by a factor of 10! What about multi-core processors these days? Why does a code scale perfectly on a quad-core machine, but scales poorly sometimes, not always though, on an 128-way SMP shared memory machine? The tutorial also attempted to answer those questions as well.

By knowing a program language, and knowing the basics of message passing and threading, one is able to write code that works. But a problem that many people face is that this "spaghetti" code is hard to comprehend and modify. The two hour tutorial session on software design attempted to give the audience an overview of key software engineering principles to guide users in developing efficient code that is easier to maintain and grow.

The one day tutorial was wrapped up with an in-depth session on cache replacement algorithms. This nearly two hour seminar presented the recent research results of Prof. Hanan Lutfiyya and her graduate students from the Computer Science Department at The University of Western Ontario. For further details on PASCO'07, visit: www.orcca.on.ca/conferences/pasco2007/site/tutorials/tutorials.html



“Without the support of SHARCNET, these PASCO meetings would not have been possible.”

-Marc Moreno Maza

School's in for Summer!

BY DAVID MCCAUGHAN, HPTC CONSULTANT,
SHARCNET

The SHARCNET Fall Workshop has been an annual staple event for the user community for a number of years, and has seen numerous changes in response to attendee feedback. The most notable change was the shift to in-depth laboratory sessions from one of exclusively invited talks. Two minor issues remained for us: two or three days was insufficient for any breadth of material to be presented; and the timing of the workshop during the fall academic term was a barrier to attendance for a number of our users due to coursework, conferences and the like. 2007 saw the most significant change to date, relocating the workshop to the summer months, when people's schedules are generally more relaxed, and greatly expanding the instructional content and support offered to attendees by holding the workshop over a full week. This is how the SHARCNET Summer School was born.

McMaster University was our host this year, and over the week of June 18-22 we offered no fewer than eight in-depth, laboratory-based workshops consisting of a full two day session on MPI Programming for Clusters, and two days of parallel sessions on Fortran90 array processing, bash shell scripting, parallel debugging, visualization, and SMP programming using both OpenMP and pthreads. Parallel sessions maximized the efficiency with which attendees can target their individual interests, while a subsidized on-campus residence option and travel assistance minimized many barriers to attendance, particularly by graduate students to whom the workshop is primarily aimed, and who typically form the bulk of attendees. Laboratory sessions were delivered by SHARCNET technical staff, and were well received based on feedback from those in attendance.

“Relocating the workshop to the summer months, when people's schedules are generally more relaxed and expanding the instructional content over a full week, contributed to the success of the 2007 Summer School.”

-David McCaughan

Particularly noteworthy was the inclusion of an “HPC Day”, which was intended to provide some of the richness of invited talks, even for those not attending the Summer School itself. HPC Day attendees were treated to insightful talks from our invited speakers: programming challenges for the Cell processor (Christopher Anand, McMaster) and parallel computing using MATLAB (Mike Agostini, Mathworks Inc.). As an effort to foster communication among the user community itself, we pursued a call for brief research presentation from the user community (for which we offered Summer School attendance benefits), for which response was outstanding. Speakers from Waterloo, UOIT and McMaster stepped up and provided one of the most entertaining and engaging days of talks we've ever held, according to attendees. The user community benefits tremendously from cross-pollination of ideas and HPC day, and the Summer School itself, can only be described as a rousing success.

A huge thanks goes out to all the staff, researchers and speakers who made the SHARCNET Summer School 2007 a success, and we hope to engage even more of the user community as we pilot this annual event into the future; where one hopes things can only get better (we are working on some light scheduled social activities for SS 2008)! Additional information, including speakers, program and copies of slides and instructional materials can be found online at www.sharcnet.ca/Events/ss2007/

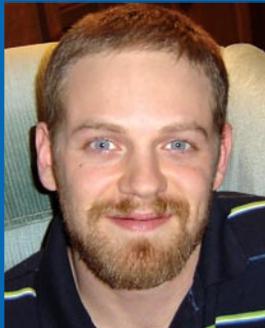




Adam Munro



Nick Chepurniy



Hugh Merz

Swimming with SHARCS: New Staff

We are pleased to welcome the following individuals to the SHARCNET team:

Adam Munro joined SHARCNET as a Systems Administrator at the University of Waterloo, on June 26th. Adam is a University of Waterloo graduate who majored in Computer Science. He spent the last four years developing and supporting IT solutions in large and small UNIX environments including Sunlife Financial and the Bank of Canada. Adam's initial SHARCNET focus areas will be in systems monitoring development and support, networking, and supporting the whale cluster.

Nick Chepurniy joined SHARCNET as a Technical Support Analyst at the University of Windsor, August 1st. Nick has a BSc (Mathematics), MSc (Fluid Mechanics) and a PhD (Applied Mechanics), all from McGill University. He has worked on scientific and engineering software applications in Canada, Spain, Alaska and Brazil. Since 1990 Nick has been working on High Performance Computing, initially for Vector Supercomputers (Cray and NEC) and on cluster systems in the last few years.

Hugh Merz joined SHARCNET as an HPC Technical Consultant based out of the University of Waterloo, August 1st as well. Hugh is a University of Waterloo graduate who majored in Honours Physics. He spent the last five years as a Scientific Parallel Programmer at the Canadian Institute for Theoretical Astrophysics at the University of Toronto, developing and supporting HPC research applications and administering beowulf clusters.

SHARCNET Visits Local Science Fair

SHARCNET is committed to an ambitious plan for outreach to Ontario students and was thrilled to be part of the London Science Fair exhibition held at Galleria Mall back in May. SHARCNET had a booth alongside dozens of student displays and we were delighted by the number of visitors – both students and parents alike – asking all sorts of questions about HPC and what it can do. Along with our “supercomputer-in-a-box”, SHARCNET plans to take its message on the road, visiting high schools, conferences and events to raise awareness about high performance computing.



model. Nonparametric methods are appealing because they relax the parametric assumptions imposed on the data generating process and allow the data to determine an appropriate model. Dr. Racine explains that “the issue with parametric models is that often they work well, but if you get the structure wrong then inference can be affected adversely”, consequently statisticians might look to nonparametric smoothing methods.

Though the first published article on nonparametric kernel smoothing appeared in 1956, statisticians have historically encountered what Dr. Racine refers to as “three major obstacles” preventing widespread adoption of these techniques. These obstacles were (1) limitations arising from computational demands, (2) lack of open software and (3) the inability to handle the presence of both continuous and categorical data.

The first, perhaps most significant, barrier preventing the widespread use of nonparametric smoothing techniques was the absence of sufficient computing power. “The burden arises”, Dr. Racine explains, “from the fact that data-driven methods of ‘bandwidth’ selection are necessary, which by their very nature involve run times that increase exponentially with the amount of data at hand”. Only within the last decade or so has processing power increased to the point where statisticians can effectively and efficiently analyze the large data sets for which nonparametric methods are ideally suited. High performance computing clusters, such as SHARCNET, have had an enormous, theoretical and practical, impact on research in the field. Dr. Racine says normal analysis that might “take a normal desktop computer several weeks to a month to calculate”, can be run on a SHARCNET cluster to generate a solution in a matter of hours.

Furthermore, Dr. Racine has developed a library written in the ‘C’ programming language that has recently been adapted to work in combination with ‘R’—a language and environment for statistical computing and graphics popular among statisticians. For the past two decades Dr. Racine has extended his C library, now underlying the R library ‘np’, and has made it available via the Comprehensive R Archive Network (CRAN). An extension, ‘npRmpi’, has been tailored to take advantage of SHARCNET’s high performance computers that provides users with a set of high level R functions that can be used either for nonparametric data analysis or to further development of kernel methods. With the combination of np, npRmpi, and SHARCNET’s extensive resources, a major obstacle limiting nonparametric methods has been overcome.

Figures 1 and 2 illustrate the potential benefits from using nonparametric versus parametric methods, when appropriate. In this example, Dr. Racine applied both parametric and nonparametric estimation methods to a real-world data set based on “the Old Faithful Geyser” in Yellowstone National Park. The figures plot

the estimated joint probability distributions using each approach where the parametric model is taken to be joint normal, which is inappropriate. The comparison shows clearly, as Dr. Racine states, that “the underlying bimodal nature of the joint PDF is readily revealed by the kernel estimator but is obscured by the incorrectly specified parametric model”. These graphs exemplify how nonparametric methods avoid “imposing potentially false structure such as symmetry, unimodality, monotonically decreasing away from the mode and so on”, assumptions that could invalidate parametric inference.

“SHARCNET has become a key component in my research agenda. SHARCNET has helped shape the development of new methods via extensive finite-sample simulations that would otherwise be impossible.” -Jeffrey Racine

A second, more important example involves assessing the efficacy of right heart catheterization (RHC) in critically ill patients. RHC is a medical procedure that measures the cardiac function of the heart by inserting a catheter into the right ventricle to measure the pressure, oxygen content and also reveal cardiac abnormalities. Previous studies on this process have concluded that RHC in critically ill patients increases the risk of death, which is counter to conventional wisdom. However, these studies relied on rigid parametric models which may not have been appropriate. Since the majority of data was categorical, Dr. Racine used the same data sets of previous studies but instead applied nonparametric kernel techniques recently developed by him and his co-authors. Interestingly, when using the nonparametric estimator Dr. Racine discovered that the nonparametric method not only delivered improved predictions, but more importantly, he found that there doesn’t appear to be any increased risk of death, revealing that “the misspecified parametric model appears to provide an unreasonable estimate of the risk of death”.

The Old Faithful Geyser and the assessment of right-heart catheterization both highlight the potential benefits of using nonparametric smoothing methods, where appropriate. The use of the R software and libraries and the cluster computing resources offered by SHARCNET combined have significantly advanced nonparametric methods. Dr. Racine’s research has primarily focused on extending kernels methods to effectively handle categorical data which is frequently encountered in most if not all fields of study, and the techniques developed have widespread potential applications in statistics, medicine, and the social sciences.

Tech Bytes

BY DAVID MCCAUGHAN, HPTC CONSULTANT, SHARCNET

User Certification Q&A

What happens to existing users of the system?

All current users will automatically receive User I certification; if you require User II certification, please contact one of SHARCNET's technical staff who will be able to assist you in upgrading your access (heavy users familiar to SHARCNET may be proactively bumped to User II).

Does a PI have to have a certain level of certification in order to sponsor accounts with that level of access?

No: certification is associated with individual user accounts; a supervisor who does not personally utilize SHARCNET systems is free to remain at User 0 without affecting the ability of their students to certify.

Are sponsored accounts automatically assigned the certification of the supervisor?

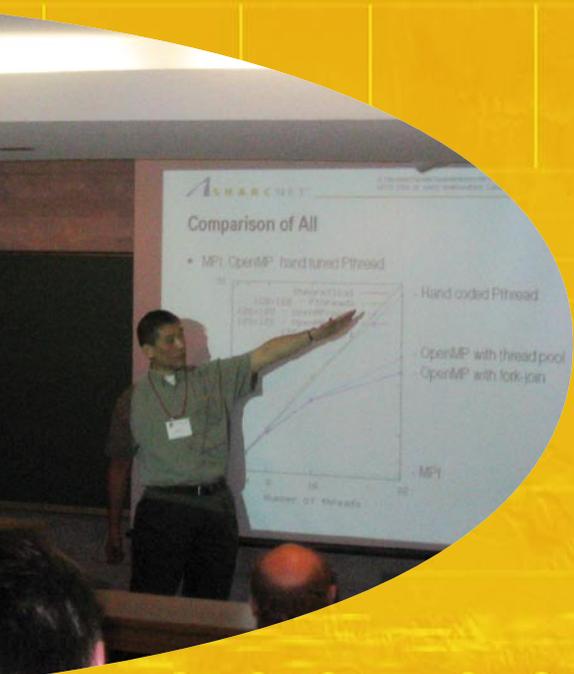
No: certification is associated with individual user accounts; a supervisor's experience with the system is treated independent of those she/he supervises.

Can I fast-track a new account to User I/II status?

Yes: if you need an account you sponsor upgraded quickly (or alternatively lack access to a local AccessGrid facility at which to attend a literacy seminar), please contact one of SHARCNET's technical staff and they will be able to assist you in upgrading an account.

I have more questions/concerns and would like more information.

Please direct inquiries to David McCaughan (dbm@sharcnet.ca) who will be happy to discuss the certification program and its implementation at length.



Baolai Ge,
HPTC Consultant

“The User Certification Program is designed to streamline and standardize the process of new (and ongoing) user education without placing undue hardships on users for basic system access.”

-David McCaughan

New User Certification Program

BY DAVID MCCAUGHAN, HPTC
CONSULTANT, SHARCNET

Beginning in September 2007, SHARCNET will be introducing a new **User Certification Program**. As SHARCNET's user community continues to expand, particularly into non-traditional fields, we need to ensure users are sufficiently literate to make the most appropriate use of available hardware and software resources. The User Certification program is designed to streamline and standardize the process of new (and ongoing) user education without placing undue hardships on users for basic system access.

There are two components to the certification program: a **qualification stream** aimed at encouraging informed, appropriate usage of SHARCNET hardware; and a **technical stream** providing a series of proficiency certificates as a means of formalizing the value-added aspect of more in-depth education within SHARCNET institutions. As qualification stream designations will affect basic system access, this will be the focus of this article: a future article will address the technical stream in greater depth.

Qualification Stream Certifications

Qualification certifications are the means by which we ensure users are aware of the basics of our systems and their software environments prior to allowing them less restricted access to production equipment. This minimizes the impact of a naïve user by limiting the resources they are permitted to access prior to demonstrating that they are aware of appropriate system policies and practices.

Initial levels of qualification will be required for new users in order to access any significant computing resources within SHARCNET. Higher levels of certification will enable even more extensive access to compute resources, and while certainly optional for anyone,

users demonstrating high levels of resource consumption may be indirectly upgraded through staff consultation.

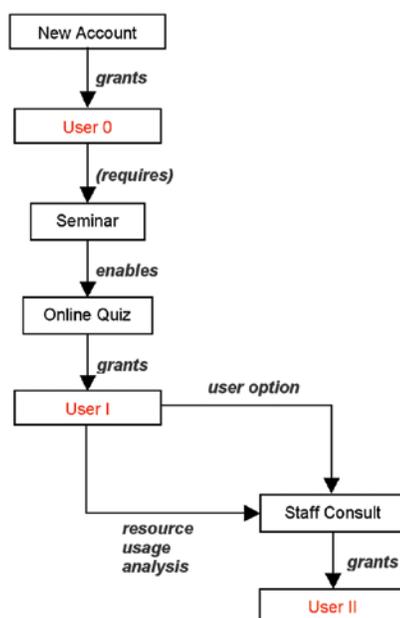


Figure 1 depicts the current design for the qualification certification process, including both certification levels (in red text) and the process by which they are attained.

It is our primary goal to avoid making the qualification process onerous or burdensome on the user community. Emphasis is on demonstrating understanding, and these levels of certification are primarily for internal use, filtering access to resources proportional to the user's ability to demonstrate they are able to make use of the resources appropriately.

The Certification Process

Once an account is approved, new users are automatically granted User 0 with restricted system access. New users will be required to attend an introductory literacy seminar (which is already offered bi-weekly over Access Grid) in order to enable a short online quiz on awareness of policies and usage, completion of which grants User I certification immediately. User I is considered the "standard" level

of system access, designed to satisfy the majority of current usage patterns. A further certification level, User II, exists to enable essentially unrestricted access to SHARCNET systems, which will require consultation with a SHARCNET staff member to assess the user's appreciation of more advanced issues such as network latency, parallel granularity and other issues central to maximizing efficiency of parallel applications (note this does not imply any development issues). Users can elect to seek this level of certification at their discretion, however a user may be contacted based on resource usage analysis by SHARCNET staff, with the end result being an upgrade in qualification to better support their work. Table 1 provides a summary of system access by certification level.

Certification	System Access
User 0 (new account)	8 CPUs total (across SHARCNET) 24 hours (wall clock) run-time
User I	256 CPUs total (across SHARCNET) 168 hours (1 week wall clock) run-time
User II	Unrestricted access

Table 1: System Access

This system will go into effect in mid-September, at which point all infrastructure for the provision of certification-based access will be in place; however, there will be a grace period of several months during which time system access restrictions will not be enforced in order to provide a smooth transition for all involved. Users will be notified by e-mail when a job they run would exceed their level of system access once the system is fully in force.

Shared Hierarchical Academic Research Computing Network

Board of Directors

Ted Hewitt (Chair), The University of Western Ontario
Hugh Couchman (Scientific Director), SHARCNET
Cindy Munro (Secretary), SHARCNET
Alan Wildeman, University of Guelph
Allison Sekuler, McMaster University
Paul Maxim, Wilfrid Laurier University
Keith Taylor, University of Windsor
Greg Weiler, Fanshawe College
Robert Eagle, Sheridan College
George Dixon, University of Waterloo
Michael Owen, Brock University
Bill Smith, University of Ontario Institute of Technology
Michael Siu, York University
Rui Wang, Lakehead University
Patrice Sawyer, Laurentian University
Jim Parker, Trent University
Sara Diamond, Ontario College of Art & Design
John McCormick, Perimeter Institute
Deborah Stacey (Researcher Representative), University of Guelph
Greg Chappell, HP Canada
Andrew Bjerring, CANARIE

Site Leaders

Thomas Wolf, Brock University
Garth Santor, Fanshawe College
Stephen Tullis/James Wadsley, McMaster University
Ed Sykes, Sheridan College
David Calvert, University of Guelph
Dhavid Aruliah, University of Ontario Institute of Technology
Hans De Sterk, University of Waterloo
Mark Daley, The University of Western Ontario
Pierre Boulos, University of Windsor
Ilias Kotsireas, Wilfrid Laurier University
Huaxiong Huang, York University
Sharon Hollinsworth, Lakehead University
Kalpdrum Passi, Laurentian University
Bill Atkinson, Trent University
John McCormick, Perimeter Institute

SHARCNET Funding

Canada Foundation for Innovation (CFI)
Ontario Innovation Trust (OIT)
Ontario Research Fund (ORF)

Private Sector Partners

Bell Canada • www.bell.ca
Hewlett Packard • www.hp.com
Nortel Networks • www.nortel.com
Optical Regional Advanced Network of Ontario (ORANO) • www.orion.on.ca
Platform Computing • www.platform.com
Quadrics Supercomputing World
• www.quadrics.com
Silicon Graphics • www.sgi.com

Academic Partner Institutions

The University of Western Ontario
University of Guelph
McMaster University
Wilfrid Laurier University
University of Windsor
Fanshawe College
Sheridan College
University of Waterloo
Brock University
University of Ontario Institute
of Technology
York University
Lakehead University
Laurentian University
Trent University
Perimeter Institute
Ontario College of Art & Design