SEQUENTIAL MONTE CARLO METHODS
Final Report

Program Leaders: Arnaud Doucet and Simon Godsill

1 Program and its Objectives:

This aim of this 12 month SAMSI program was to develop new approaches to scientific/statistical computing using innovative sequential Monte Carlo (SMC) methods. The program addressed fundamental challenges in developing effective sequential and adaptive simulation methods for computations underlying inference and decision analysis. The research blended conceptual innovation in new and emerging methods with evaluation in substantial applied contexts drawn from areas such as control, communications and robotics engineering, financial and macro-economics, among others. Researchers from statistics, computer science, information engineering and applied mathematics were involved, and the program promoted the opportunity for both methodological and theoretical research. The interdisciplinary aspects of the program were substantial, as was the attractiveness for students and postdocs.

2 Background

Monte Carlo (MC) methods are central to modern numerical modelling and computation in complex systems. Markov chain Monte Carlo (MCMC) methods provide enormous scope for realistic statistical modelling and have attracted much attention from disciplinary scientists as well as research statisticians. Many scientific problems are not, however, naturally posed in a form accessible to evaluation via MCMC, and many are inaccessible to such methods in any practical sense. For example, for real-time, fast data processing problems that inherently involve sequential analysis, MCMC methods are often not obviously appropriate at all due to their inherent "batch" nature. The recent emergence of sequential MC concepts and techniques has led to a swift uptake of basic forms of sequential methods across several areas, including communications engineering and signal processing, robotics, computer vision and financial time series. This adoption by practitioners reflects the need for new methods and the early successes and attractiveness of SMC methods. In such, probability distributions of interest are approximated by large clouds of random samples that evolve as data is processed using a combination of sequential importance sampling and resampling ideas. Variants of particle filtering, sequential importance sampling, sequential and adaptive Metropolis MC and stochastic search, and others have emerged and are becoming popular for solving variants of "filtering" problems; i.e. sequentially revising sequences of probability distributions for complex state-space models. Useful entree material and examples SMC methods can be found at the following SMC preprint site. Many problems and existing simulation methods can be formulated for analysis via SMC: sequential and batch Bayesian inference, computation of p-values, inference in contingency tables, rare event probabilities, optimization, counting the number of objects with a certain property for combinatorial structures, computation of eigenvalues and eigenmeasures of positive operators, PDE's admitting a Feynman-Kac representation and so on. This research area is poised to explode, as witnessed by this major growth in adoption of the methods.

The SAMSI SMC program aims were to:

- Address methodological and theoretical problems of SMC methods, including synthesis of concepts underlying variants of SMC that have proven apparently successful across multiple fields, and the development of methodological and theoretical advances.
- Develop the methodological research – with broad opportunities for test-bed examples, methods evaluation and refinement of generic approaches – in the contexts of a number of important applied problems (e.g. data assimilation, inference for large state spaces, finance, tracking, continuous time models).

The program was an opportunity for exchange between communities, helping, we hope, to shape the future of stochastic computation and sequential methods, involving statisticians, computer scientists and engineers as core participants as well as others working collaboratively in a range of applied fields.

3 Core Group

A core group of researchers have been based at SAMSI, complemented by external participants in the various working groups, which held weekly meetings via Webex connections to SAMSI.

3.1 Local faculty

- Mark Huber (Duke)
- Mike West (Duke)
- Nilay Argon (UNC)

3.2 Senior researchers (at SAMSI for significant periods of time in Fall, 2008):

- Susie Bayarri (University Valencia)
- Jaya Bishwal (University North Carolina Charlotte)
- Carlos Carvalho (University of Chicago)
- Arnaud Doucet (University British Columbia)
- Pena Edsel (University South Carolina)

- Liu Fei (University Missouri)
- Marco Ferrante (University Pavia)
- Nathan Green (DSTL)
- Hedibert Lopes (University of Chicago)
- Raquel Prado (University Santa Cruz)
- Sylvain Rubenthaler (University Nice)
- Yoshida Ryo (Institute Statistical Mathematics)
- Jochen Voss (University Warwick)

3.3 Researchers (Spring and Summer, 2009)

- Daniel Clark (University Herriott Watt)
- Mark Coates (University Mc Gill)
- Paul Fearnhead (University of Lancaster)
- Andrew Thomas (University St Andrews)
- James Lynch (University South Carolina)
- Ernest Fokoue (Kettering University US)

3.4 Postdoctoral fellows

- Ioanna Manolopoulou
- Das Sourish

3.5 Postdoctoral associates

- Armagan Artin (Duke)
- Christian Macaro (Duke)
- Julien Cornebise (University Paris VI)
- Gentry White (NSCU)
- Elizabeth Shamseldi (Duke)
- Bin Liu (Duke)

3.6 Graduate students

- Melanie Bain (Duke)
- Luke Bornn (University British Columbia)
- Deidra Coleman (NCSU)
- Ana Corberan (University of Valencia)
- Thomas Flury (Oxford University)
- Roman Holenstein (University British Columbia)
- Chunlin Ji (Duke)
- Olasunkanmi Obanubi (Imperial College)
- Gareth Peters (University New South Wales)
- Francesca Petralia (Duke)
- Sarah Schott (Duke)
- Minghui Shi (Duke)
- Baqun Zhang (NCSU)

4 Program Organization

4.1 Opening workshop

The Opening Workshop was held during September 7-10, 2008 at SAMSI, organized by Arnaud Doucet (British Columbia), Simon Godsill (Cambridge) and Mike West (Duke University). This highly successful event engaged significant parts of the statistical, engineering and mathematics community, as well as others in econometrics and sciences, and included themed sessions from all of the main program topics (working groups).

Tutorial talks were given by four world leaders in the various areas of SMC: Pierre del Moral (Bordeaux), Paul Fearnhead (Lancaster), Hedibert Lopes (Chicago) and Jun Liu (Harvard). These were pitched at various levels, allowing useful participation by attendees starting in the area as much as those already expert in one or more topics. Themed conference sessions were arranged to have a good balance between senior invited talks, new researcher talks and panel discussion. Most sessions stimulated very active discussion.

At a break-out session on the final afternoon, Working Group leaders were allocated and a broad declaration of interest was obtained from all workshop participants for their subsequent participation in the program.

4.2 Undergraduate workshop

The SAMSI Two-Day Undergraduate Workshop was held from October 31 - November 1, 2008. There were nine technical talks given by Jaya Bishwal, Jochen Voss, Gentry White, Nathan Green, Christian Macaro, Sourish Das, Julien Cornebise, Ioana Maolopoulou and Francesca Petralia, covering many aspects of SMC from basic methodology to applications in finance and defence. There was be also an interactive R session.

4.3 Fall SAMSI course on sequential Monte Carlo

This course provided an introduction to sequential Monte Carlo methodology, theory and applications. It was attended by approximately 40 people. Topics covered include: introduction to SMC, advanced SMC methods, SMC methods for parameter estimation in general state-space models, SMC methods as alternative to MCMC methods. The main instructor was Arnaud Doucet and two 'invited' instructors gave some lectures: Christophe Andrieu (Bristol) and Alexander Chorin (Berkeley).

4.4 Mid-term workshop

A mid-term workshop was organised on 19-20 Feb 2009 at the SAMSI Institute. This had participants from most of the working groups, including the leaders of the Continuous Time (Fearnhead - Lancaster), Tracking (Godsill - Cambridge), Big Data (West - Duke), Parameter Learning (Lopes - Chicago) and Model Assessment (Carvalho - Chicago) working groups. These leaders gave overviews of progress in the different working groups and other participants gave research updates on SAMSI related work. Three of the 15 talks were delivered successfully by Webex from remote locations. A particular focus of talks and discussion was the Tracking working group, which assembled many of its participants at the workshop.

4.5 Adaptive Design Workshop

An Adaptive Design, Computer Modeling and SMC workshop organized by Jim Berger and Suzie Bayarri was held from April 15-17, 2009; see http://www.samsi.info/workshop/smc-adaptive-design-sequential-monte-carlo-and-computer-modeling-workshop-april-15-17-2009 for details. There were 13 speakers.

4.6 Transition workshop

The transition workshop was held at SAMSI from 9-10 Nov. 2009; see http://legacy.samsi.info/workshops/2 transition200911.shtml for details. There were 20 talks and all the working groups were represented.

4.7 Working groups

The working groups met weekly throughout the program. The topics were

- Tracking and Large-Scale Dynamical Systems

- Theory
- Population Monte Carlo
- Continuous Time
- Parameter
- Model Assessment
- Big Data

4.7.1 Tracking and Large-scale dynamical systems

The working group leaders were Simon Godsill (whole year) and Nathan Green (Fall 08).

Regular participants included: Mark Briers (QinetiQ, UK), Daniel Clark (Heriot-Watt University UK), Avishi Carmi (Cambridge University UK), Julien Cornebise (SAMSI post-doc), Ernest Fokoue (Kettering University US), Simon Godsill (Cambridge University - program leader), Nathan Green (DSTL UK, long-term Fall 08 visitor) Chunlin Ji (Duke University - SAMSI Grad. student), Sze Kim Pang (Cambridge University UK), Gareth Peters (Univ. New South Wales, SAMSI Fall 08 visitor), Viktor Rozgic (University of Southern California), Francois Septier (Cambridge University, UK), Joshua Vogelstein (Johns Hopkins University), Gentry White (N.C. State University - SAMSI Post-doc), Namrata Vaswani (Iowa State University)

Working group - organisation This working group has focussed on methodology for problems in high-dimensional tracking, with applications in computer vision, tracking, meteorology, biological imaging, etc. Standard particle filters do not perform satisfactorily in this scenario and hence we are pushing the methodology further by development of novel approaches. These include elements of Markov chain Monte Carlo-based filters, genetic algorithm approaches and SMC samplers. The goals of the working group are be to produce papers on various topics involving multiple participants from the group, leading to future collaborative projects across a number of disciplines. The theme was organised into sub-groups addressing the following areas:

Subgroup 1: Multiple target tracking (lead Francois Septier (Cambridge)):

Participants include Simon Godsill, Francois Septier, Chunlin Ji, Mark Briers, Viktor Rozgic, Ernest Fokoue, Daniel Clark

The aim of this sub-group was to research new methodologies for high-dimensional and structured tracking problems using SMC. We have generated standard datasets, test scenarios and data simulation code for multiple target tracking with random birth and death of objects and various sensor characteristics based on point process models or pixellated image data. A number of methodologies have been tested on this scenario, including novel MCMC-based particle filters, resample-move filters, SMC samplers, variational Bayes, etc. New smoothing methods for random finite set models have also been developed by Dan Clark. In addition we have done work in detection and tracking of dynamic group objects, using a virtual-leader formulation and adaptations of our previous SDE-based models, and with dynamic graph structures, tracked over time with SMC. The main interest here is to track formations of objects which have a common pattern of motion. The movement of the whole group is of interest rather than tracking each object separately. Groups of targets can be considered as formations of entities whose number varies over time because targets can enter a scene, or disappear at random times. The groups can split, merge, to be relatively near to each other or move largely independently on each other. However, it is typical for group formations to maintain some patterns of movement. Monte Carlo approaches combined with evolving random graphs is proposed for group object state and structure estimation were developed and published by the Cambridge and Lancaster groups.

The Various participants have benefitted from the collaborations as can be seen from the large number of high quality publications arising from this working group.

1. S. K. Pang, S. Godsill, J. Li, F. Septier, and S. Hill. "Sequential Inference for Dynamically Evolving Groups of Objects" a Chapter in "Bayesian Time Series Models" edited by D. Barber, A.T. Cemgil and S. Chiappa, Cambridge University Press, 2010 ISBN : 9780521196765.

2. F. Septier, S. K. Pang, A. Carmi, and S. Godsill. "On MCMC-Based Particle Methods for Bayesian Filtering : Application to Multitarget Tracking" in IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2009), Aruba, Dutch Antilles, Dec. 2009

3. F. Septier, A. Carmi, and S. Godsill. "Tracking of Multiple Contaminant Clouds" in Proc. Int. Conf. on Information Fusion (FUSION 2009), Seattle, USA, Jul. 2009 Winner of the third best regular paper award.

4. A. Carmi, F. Septier, and S. Godsill. "The Gaussian Mixture MCMC Particle Algorithm for Dynamic Cluster Tracking" in Proc. Int. Conf. on Information Fusion (FUSION 2009), Seattle, USA, Jul. 2009 [BIBTEX]

5. F. Septier, A. Carmi, S. K. Pang, and S. Godsill. "Multiple Object Tracking Using Evolutionary and Hybrid MCMC-Based Particle Algorithms" in 15th IFAC Symposium on System Identification, (SYSID 2009), Saint-Malo, France, Jul. 2009 Invited paper.

6. . F. Septier, S. K. Pang, S. Godsill, and A. Carmi. "Tracking of Coordinated Groups using Marginalised MCMC-based Particle Algorithm" in IEEE Aerospace Conference, Big Sky, Montana, Mar. 2009

7. A. Carmi, S. Godsill, and F. Septier. "Evolutionary MCMC Particle Filtering for Target Cluster Tracking" in IEEE 13th DSP Workshop and the 5th SPE Workshop, Marco Island, Florida, Jan. 2009

8. Variational Mean Field Approach to Efficient Multitarget Tracking. E. Fokoue. JSM Proceedings 2009.

9. Sequential Monte Carlo Smoothing with Random Finite Set Observations. D. Clark and M. Briers. JSM Proceedings 2009

10. First-Moment Multi-Object Forward-Backward Smoothing Daniel E. Clark International Conference on Information Fusion, 2010

11. Extended object filtering using spatial independent cluster processes A Swain, D Clark International Conference on Information Fusion, 2010

12. Improved SMC implementation of the PHD filter B Ristic, D Clark, B N Vo, International Conference on Information Fusion, 2010

13. Performance evaluation of multi-target tracking using the OSPA metric B Ristic, B N Vo, D Clark International Conference on Information Fusion 2010

14. First-moment filters for spatial independent cluster processes. A. Swain and D. E. Clark. SPIE Defense, Security and Sensing Symposium, 2010.

15. Forward-Backward Sequential Monte Carlo Smoothing for Joint Target Detection and Tracking, D. Clark, B.-T. Vo and B.-N. Vo, IEEE Proc. 12th Annual Conf. Information Fusion, Seattle, Washington, 2009.

16. A. Gning, L. Mihaylova, S. Maskell, S. K. Pang, and S. Godsill, Group object structure and state estimation with evolving networks and monte carlo methods, IEEE Transactions on Signal Processing, 2010.

17. , Ground target group structure and state estimation with particle filtering, in Proc. of Interna- tional Conf. on Information Fusion, Germany, Cologne, 2008.

18. , Evolving networks for group object motion estimation, in Proc. of IET Seminar on Target Tracking and Data Fusion: Algorithms and Applications, Birmingham, UK, 2008, pp. 99106.

19. H. Bhaskar, L. Mihaylova, and S. Maskell, Population-based particle filters, in Proc. from the Insti- tution of Engineering and Technology (IET) Seminar on Target Tracking and Data Fusion: Algorithms and Applications, Birmingham, UK, 2008, pp. 3138.

20. H. Bhaskar, L. Mihaylova, S. Maskell, and S. Godsill, Evolving population Markov Chain Monte Carlo particle filtering, Journal paper (in preparation).

21. H. Bhaskar, L. Mihaylova, and A. Achim, Video foreground detection based on symmetric alpha-stable mixture models, IEEE Transactions on Systems and Circuits for Video Technology, 2010, in press.

22. D. Angelova and L. Mihaylova, Contour segmentation in 2d ultrasound medical images with particle filtering, Machine Vision and Applications Journal, 2010, in press.

23. L. Mihaylova, D. Angelova, D. Bull, and N. Canagarajah, Localisation of mobile nodes in wireless sensor networks with time correlated measurement noises, IEEE Transactions on Mobile Computing, 2010, in press.

24. L. Mihaylova and D. Angelova, Noise parameters estimation with Gibbs sampling for localisation of mobile nodes in wireless networks, in Proc. of the 13th International Conference on Information Fusion. ISIF, Edinburgh, UK, 2010.

25. A. Gning, L. Mihaylova, and F. Abdallah, Mixture of uniform probability density functions for nonlinear state estimation using interval analysis, in Proc. of 13th International Conference on Information Fusion. ISIF, Edinburgh, UK, 2010.

26. L. Mihaylova, A. Gning, V. Doychinov, and R. Boel, Parallelised gaussian mixture filtering for vehicular traffic flow estimation, in Informatik 2009, Lecture Notes in Informatics (LNI) - Proceedings Series of the Gesellschaft fur Informatik (GI), Eds. S. Fischer, E. Maehle, R. Reischuk, Vol. P-154. Germany, Luebeck, 2009, p. 2321 2333.

27. J.A. Pocock, S.L. Dance and A.S. Lawless (2010) State Estimation using the Particle Filter with Mode Tracking, submitted to Computers and Fluids

28. M. Hong, M. F. Bugallo, and P. M. Djuric, "Joint Model Selection and Parameter Estimation by Population Monte Carlo Simulation," to appear in Journal of Selected Topics in Signal Processing, 2009.

29. B. Shen, M. F. Bugallo, and P. M. Djuric, "Multiple marginalized population Monte Carlo," accepted in EUSIPCO 2010.

30. M. F. Bugallo and P. M. Djuric, "Target tracking by symbiotic particle filtering," Proceedings of the 2010 IEEE Aerospace Conference, Big Sky (Montana, USA), March 2010.

31. M. F. Bugallo, M. Hong, and P. M. Djuric, "Marginalized population Monte Carlo," Proceedings of the IEEE 34th International Conference on Acoustics, Speech and Signal Processing (ICASSP'2009), Taipei (Taiwan), April 2009.

32. P. M. Djuric and M. F. Bugallo, "Improved target tracking with particle filtering," Proceedings of the 2009 IEEE Aerospace Conference, Big Sky (Montana, USA), March 2009.

33. M. F. Bugallo and P. M. Djuric, "Complex systems and particle filtering," Proceedings of the Asilomar Conference on Signals, Systems, and Computers, Paci c Grove (California, USA), November 2008.

PhD Student: Viktor Rozgik Viktor Rozgic PhD Student, Advisor: Prof. Shrikanth Narayanan

Signal Analysis and Interpretation Laboratory Viterbi School of Engineering University of Southern California

Research and Impact to Thesis

I have found information about SAMSI opening workshop online and since I have been using Sequential Monte Carlo (SMC) methods in my work before I decided to attend it. I found the talks very interesting and I got involved in work of the Tracking Workgroup which was the good match for the topic of my thesis "Multimodal fusion for tracking and identification in Smart Environments". Collaboration with the grop members, talks I have heard in the workshops and weekly group meedings; and references and papers in progress shared over the group's webpage were very helpfull for my thesis work. Besides getting a much better perspective on the state-of-the-art Sequntial Monte Carlo Algorithms and understanding the current research directions I had a hands-on-experience in implementation and testing of the SMC algorithms on the synthetic multi-target tracking problem. For this oportunity I feel very gratefull to Prof. Simon Godsill and Dr. Francois Septier. I have managed to transplant and adapt part of this work to problems of audio-visual tracking, speaker segmentation and identitfication in meeting scenarios. Proposed work for the final part of my thesis includes work on multitarget tracking algorithms which is not focused only on the Meeting Monitoring environments and I hope that I am going to be able to continue collaboration with people I have meet during the workshop in the following period.

Subgroup 2: Biological Cell tracking (lead Chunlin Ji):

Participants include: Chunlin Ji, Simon Godsill, Daniel Clark.

This group interfaces also with the multiple target tracking sub-group, being concerned with video imaging data involving fluorescently labelled multiple cells, which move around, grow, divide, etc. They have investigated a number of approaches including point process based methods, including PHD filters, and also pixel-based likelihood functions. Datasets have already been provided by Duke researchers. Two papers are in preparation:

- C. Ji & M. West (2009) Bayesian Nonparametric Modelling for Time-varying Spatial Point Processes

Chunlin Ji (SAMSI RA) is attached to the Tracking (Godsill) working group and participates actively in the Big data group with West on spatial dynamic modelling for biological cell tracking problems. Ji is developing SMC methods in the context of new classes of models. This research has grown out of existing work of Ji & West in static problems, now extended with new dynamic models that will form an additional part of Ji's PhD thesis research, and one initial paper is in draft at the time of this report (see manuscripts section). Ji has led discussions on this work at several Tracking working group and Big data group meetings, gave a talk at the February 2009 mid-program workshop, and will present this work at the 7^{th} Workshop on Bayesian Nonparametrics in Turin, Italy, in June 2009, and at the 2009 Joint Statistical Meetings in Washington DC 2009.

- C. Ji & M. West (2009) Bayesian Nonparametric Modelling for Time-varying Spatial Point Processes

A grant application has gone in from Dan Clark to the UK's BBSRC Tools and Resources panel (New Investigator program) on cell tracking work

(SAMSI RA) was attached to the Tracking (Godsill) working group.

Since Feb. 2009, I participated this SMC program in SAMSI as a postdoc (or research scholar). I have three papers, which are related to this program. Now I list them in the following.

1) Bin Liu, Merlise Clyde, Tom Loredo and Jim Berger. Adaptive Annealed Importance Sampling Method for Calculating Marginal Likelihood with Application to Bayesian Exoplanets Detection. We're revising the manuscript and will submit it soon.

2) Bin Liu, Chunlin Ji, Yangyang Zhang and Chengpeng Hao. Multi-target Tracking in Clutter with Sequential Monte Carlo Methods (first drafft prior to coming to SAMSI). Accepted by IET Radar, Sonar Navigation.

3) Bin Liu, Chunlin Ji, Yangyang Zhang, and Chengpeng Hao. Blending Sensor Scheduling Strategy with Particle Filter to Track a Smart Target, Wireless Sensor Networks.

All the above works are supported by Profs. Merlise Clyde and Jim Berger's Astro fund, namely, National Science Foundation of the USA under Grant No. 0507481.

Subgroup 3: Covert chemical release (lead Nathan Green):

Participants include: Nathan Green, Francois Septier, Avishy Carmi, Simon Godsill, Mark Briers, Gareth Peters.

This topic involves plume tracking and source term estimation, exploring contour tracking, cloud tracking, ABC methods, SMC samplers and other novel techniques. The subgroup has produced models and simulation code for the source term estimation problem, in which the task is to estimate the location of a covert chemical release through sequential monitoring of the pattern of the resulting chemical plume. DSTL have agreed in principle to provide LIDAR data for this problem, and this is still being negotiated at the time of this report. A number of advances have been made in the area.

For the pure cloud tracking problem (without source term estimation), we have studied the problem of sequential inference about complex evolving cloud structures from LIDAR data, presently all simulated from models. A dynamic Gaussian mixture approximation with unknown number of components is used for the cloud intensity. Some very successful results were obtained from very ambiguous thresholded data, which have impressed specialists at QinetiQ UK Ltd. and DSTL UK Ltd. A paper is submitted to the 2009 Fusion conference, and a further paper is in preparation for the SYSID conference:

- Tracking of Multiple Contaminant Clouds. Francois Septier, Avishy Carmi, Simon Godsill. Fusion 2009 (submitted).

- Multiple Object Tracking Using Evolutionary and Hybrid MCMC-Based Particle Algorithms. F. Septier, A. Carmi, S. K. Pang and S. J. Godsill. SYSID 2009.

In addition to this, work on sequential source term estimation has been undertaken using a new trans-dimensional ABC algorithm that is able for the first time to detect multiple unknown covert releases. A survey paper has been submitted already and a paper on the STEM application is also under preparation:

- S.A. Sisson, G. W. Peters, Y. Fan, and M. Briers, Likelihood-free samplers, Journal Submission, Dec 2008.

- G. W. Peters, M. Briers and ... Trans-dimensional ABC for source term estimation, In preparation.

This work has also led to an invite to the ABC workshop in Paris, June 2009. The work has attracted serious attention from the UK's DSTL defence organisation and is very likely to lead to new grant funding in the near future.

A final sub-topic in this area concerns emulation-based methods for approximation of complex source term simulation problems. A paper is in preparation:

- Emulation Based Priors for Source Term Estimation. Gentry White and Nathan Green, in preparation.

This looks into using an emulation based approach to construction priors for use in a sequential Monte Carlo model for source term estimation. This emulation based approach allows for the construction of priors based on prior information from both computer models as well as field data. These priors offer advantages over existing priors in that they avoid degeneracy in the SMC simulation. This work draws on work from the previous SAMSI program on Development, Assessment and Utilization of Complex Computer Models, including work from the Engineering Methodology working group and the paper "Mechanism-Based Emulation of Dynamic Simulation Models: Concept and Application in Hydrology" (Reichert et. al 2009). currently under submission.

Subgroup 4: Neuron tracking (lead Joshua Vogelstein):

This topic involves tracking of multiple neuronal activity measured in living brains and involves learning of sparse connectivity matrices in continuous-time spiking environments. To include continuous time spike modelling, inference for multiple (sparsely connected) neurons, parameter estimation, image models.

The work progressed very pleasingly and the following papers were written:

- Spike Inference from Calcium Imaging Using Sequential Monte Carlo Methods, Joshua T. Vogelstein, Brendon O. Watson, Adam M. Packer, Rafael Yuste, Bruno Jedynak and Liam Paninski, Biophysical Journal, Volume 97, Issue 2, 636-655, 22 July 2009.
- A BAYESIAN APPROACH FOR INFERRING NEURONAL CONNECTIVITY FROM CALCIUM FLUORESCENT IMAGING DATA, Yuriy Mishchencko, Joshua T. Vogelstein, and Liam Paninski, to appear in the Annals of Applied Statistics

Other works are in progress, in particular, modifying the original smc work to utilize Rao-Blackwellization. They also plan on improving the parameter estimation in the work, potentially by replacing the EM step with a fully Bayesian sampling step.

Regular participants include: Mark Briers (QinetiQ, UK), Daniel Clark (Heriot-Watt University UK), Avishi Carmi (Cambridge University UK), Julien Cornebise (SAMSI post-doc), Ernest Fokoue (Kettering University US), Simon Godsill (Cambridge University - program leader), Nathan Green (DSTL UK, long-term Fall 08 visitor) Chunlin Ji (Duke University - SAMSI Grad. student), Sze Kim Pang (Cambridge University UK), Gareth Peters (Univ. New South Wales, SAMSI Fall 08 visitor), Viktor Rozgic (University of Southern California), Francois Septier (Cambridge University, UK), Joshua Vogelstein (Johns Hopkins University), Gentry White (N.C. State University - SAMSI Post-doc), Namrata Vaswani (Iowa State University)

4.7.2 Theory

The Theory working group was led by Mark Huber (Duke). The goal was to develop and analyze algorithms arising in SMC and MCMC applications. The plan of attack was to examine several techniques from both fields, and attempt to answer questions such as: 1) when can a method from one field be used in the other, and 2) is it possible to prove something about the running time of these methods as algorithms? This second question typically reduces to questions about rate of convergence, or the variance of estimators.

Participants include Petar Djuric (Stony Brook), Jan Hannig (UNC), Jim Lynch (U. South Carolina), Jonathan Mattingly (Duke), Edsel Pena (U. South Carolina), Gareth Peters (SAMSI), Giovanni Petris (Arkansas), Clyde Schoolfield (Florida), Sarah Schott (Duke), Namrata Vaswani (Iowa State), Anand Vidyashankar (Cornell).

The theory working group has been exploring relationships between Markov chain approximations and SMC methods, with an eye towards provably good methodologies. Unfortunately, most of the existing work relies on having rapidly mixing Markov chains, at which point the use of SMC is not necessary. However, proper use of Monte Carlo samples remains a difficult issue. Therefore, we have started concentrating on a very general methodology called the "Product Estimator" for moving from samples to approximate integration. This method is very versatile and does not rely on having bounded variance of the random variables used in the Monte Carlo algorithm. Current analysis, however, relies on using a medians-of-averages approach. An approach using pure averages should converge more quickly, but this makes the analysis more difficult, requiring study of products of binomial random variables. The eventual goal is a tighter bound on the tails of these distributions, moving from what is now a constant of 16 to a value of 2.

Graduate student: Sarah Schott.

Since our working group lacked a postdoc, Sarah was organizing the meetings and keeping our web page. On the research side, she worked on the product estimator problem described above, beginning with simulation studies and currently working to extend large deviations inequalities for binomials from sums to products.

Impact on research

Initially Mark Hubert introduced the product estimator as a side algorithm, a participant in the working group asked the question about the tightness of the constant. This raised an interesting point, and as Sarah and Huber studied the problem further, it proved far more deep a question than at first realized. In addition to this research avenue, the participants have learned much about SMC methodology over the course of the program, and still hope to utilize some of these methods in improving perfect simulation algorithms (the focus of my research program.)

4.7.3 Population Monte Carlo

The population Monte Carlo working group was led by Arnaud Doucet and Julien Cornebise.

Following up the discussions at the kick-off workshop in September 2008, the Population Monte Carlo working group was created to demonstrate the potential of Sequential Monte Carlo (SMC) methods for general stochastic computation problems. Although most of the current work on SMC address on-line inference problems, the objectives of this group is to focus on the development of SMC and its variants to address problems where Markov chain Monte Carlo (MCMC) methods are traditionally used. Standard MCMC are typically inefficient for multimodal target distributions and the objectives of this group is to develop powerful particle alternatives.

We have worked on five specific topics plus a side-collaboration on a connected matter.

Subgroup 1: Adaptive SMC samplers and Population Monte Carlo algorithms

SMC samplers and Population Monte Carlo is a general methodology which can be used as an alternative to MCMC methods. However it requires specifying a cooling schedule and some proposal distributions. We have developed several methods to automatize the tuning of these key parameters. We proposed a new method which allows us to compute on-the-fly a relevant cooling schedule. The resulting algorithms have been used to solve Approximate Bayesian Computation problems (in junction with Subgroup 3), and to perform inference in stochastic volatility models. We have developed methodologies to design automatically the parameters of the proposal distributions, both for SMC Samplers and for Population Monte Carlo variants. We also proposed algorithms that use Rao-Blackellisation of the key steps of PMC-related widespread algorithms to achieve variance reduction of the final estimates.

1. (Subgroup 1 and 4) Adaptivity for ABC algorithms: the ABC-PMC scheme. Beaumont, M., Robert, C.P., Marin, J.-M. and Cornuet, J.M., Biometrika 96(4), 983-990 (2009)

2. On variance stabilisation by double Rao-Blackwellisation. Iacobucci, A., Marin, J.-M., and Robert, C.P., Computational Statistics and Data Analysis 54, 698-710. (2009)

3. Adaptive Sequential Monte Carlo algorithms, J. Cornebise, Ph.D. thesis, University Pierre et Marie Curie Paris 6 (2009)

4. Chain Ladder Method: Bayesian Bootstrap versus Classical Bootstrap. Gareth Peters, Mario Wuthrich, Pavel Shevchenko - Insurance: Mathematics and Economics, to appear (2010)

5. (Subgroup 1 and 4) An adaptive SMC method for approximate Bayesian computation. Pierre Del Moral, Arnaud Doucet and Ajay Jasra, submitted January 2009.

6. Inference in Levy-driven stochastic volatility models. Ajay Jasra, Dave Stephens and Arnaud Doucet, Scandinavian Journal of Statistics, 2010.

7. A vanilla Rao–Blackwellisation of Metropolis-Hastings algorithms. Douc, R. and Robert, C.P, Annals of Statististics, 2010.

8. Adaptive Multiple Importance Sampling, Cornuet, J.M., Marin, J.-M., Mira, A. and Robert, C.P., available as arXiv:0907.1254, submitted (2009)

9. An adaptive sequential Monte Carlo sampler. Paul Fearnhead and Benjamin M Taylor, submitted May 2010.

10. (Subgroup 1 and 4) Auxiliary SMC samplers with applications to Partial Rejection Control and Approximate Bayesian Computation. J. Cornebise, G. Peters, O. Ratmann in preparation.

11. (Subgroup 1 and 3) Adaptation strategies for Particle Markov Chain Monte Carlo. J. Cornebise, G. Peters, in preparation.

Subgroup 2: SMC samplers for Normalizing Constant Calculations

It is possible to use SMC to compute normalizing constants of high-dimensional distributions. In physics this strategy is known as Jarzinsky's equality. Recently an alternative method known as nested sampling has appeared in the literature. This method enjoys several advantages compared to standard techniques. However, it remains inefficient when applied to multimodal distributions. We are currently studying an adaptive SMC version of nested sampling. Our preliminary results indicated that this new method outperforms significantly the original nested sampling algorithm in complex scenarios. We are currently investigating the theoretical properties of the resulting estimate.

1. Particle nested sampling, Arnaud Doucet and Christian P. Robert, in preparation.

Subgroup 3: Particle Markov chain Monte Carlo
Particle MCMC is a new class of methods which allows us to use SMC proposals within MCMC algorithms (Andrieu, Doucet & Holenstein, 2010). There are several open questions to address such as selecting the optimal trade-off between the number of MCMC iterations/number of particles or how to select adaptively the number of particles as a function of the current parameter value so as to ensure that the variance of the marginal likelihood is below a given threshold. We are currently studying theoretically the performance of these algorithms so as to identify the optimal tradeoff; our study relies on new sharp convergence results for SMC estimates of normalizing constants. We have also proposed some extensions of Particle MCMC methods which allow us to solve optimization problems. These extensions rely on new combinatorial identities for SMC schemes.

1. Particle Markov chain Monte Carlo methods. C. Andrieu, A. Doucet and R. Holenstein, Journal of the Royal Statistical Society B (with discussion) (2010).
2. Comment on "Particle Markov Chain Monte Carlo". Julien Cornebise and Gareth Peters (SAMSI technical Report) merged version of two comments sent to JRSSB (2010)
3. Ecological non-linear state space model selection via adaptive particle Markov chain Monte Carlo (AdPMCMC). Gareth Peters, Keith Hayes, Geoff Hossack (2010)
4. Channel Tracking for Relay Networks via Adaptive Particle MCMC. Ido Nevat, Gareth Peters, Arnaud Doucet and Jinhong Yuan (2010)
5. (Subgroup 1 and 3) J. Cornebise, G. Peters, Adaptation strategies for Particle Markov Chain Monte Carlo, in preparation.

Subgroup 4: Likelihood-free inference and Approximate Bayesian Computation (ABC) Likelihood-free inference is a rising challenge in Statistics, for models where the likelihood of the observations is either intractable or very expensive to compute.Approximate Bayesian Approximations (ABC) algorithms, which replace the likelihood of the observations by a distance on a space of summary statistics between Monte-Carlo-based simulated observations and the data. Our subgroup developed the use of SMC Samplers as the cornerstone in the MC part of ABC algorithms in lieu of crude MC or of MCMC, using the cooling schedule of SMC samplers as a series of progressive thresholds on the distance of the summary statistics, thus fighting the depletion of the sample and increasing dramatically the computational efficiency. We also brought those ABC algorithms to numerous models and applied problems where classical likelihood-based methods are helpless, especially in financial engineering, telecommunication relay systems, and in model choice for spatially correlated data.

1. Design Efficiency for "likelihood free" Sequential Monte Carlo samplers. Gareth Peters, Scott Sisson, Yanan Fan (2008)

2. (Subgroup 1 and 4) Beaumont, M., Robert, C.P., Marin, J.-M. and Cornuet, J.M. Adaptivity for ABC algorithms: the ABC-PMC scheme. Biometrika 96(4), 983-990.

3. On Sequential Monte Carlo, Partial Rejection Control and Approximate Bayesian Computation. Gareth Peters, Yanan Fan, Scott Sisson (2009)

4. Likelihood-Free Bayesian Inference for Alpha-Stable Models. Gareth Peters, Scott Sisson, Yanan Fan (2009)

5. Likelihood-free methods for model choice in Gibbs random fields. Grelaud, A., Marin, J.-M., Robert, C.P., Rodolphe, F. and Tally, F., Bayesian Analysis, 3(2), 427-442 (2009)

6. Bayesian Symbol Detection for Relay Systems via Likelihood-free Inference. Gareth Peters, Ido Nevat, Scott Sisson, Yanan Fan and Jinhong Yuan. IEEE Transactions on Signal Processing, to appear (2009)

7. Likelihood-free Samplers. Scott Sisson, Gareth Peters, Yanan Fan, Mark Briers, to appear (2010)

8. (Subgroup 1 and 4) An adaptive SMC method for approximate Bayesian computation. Pierre Del Moral, Arnaud Doucet and Ajay Jasra, submitted January 2009.

9. Filtering with Approximate Bayesian Computation. J. Cornebise, G. Peters, in preparation

10. (Subgroup 1 and 4) Auxiliary SMC samplers with applications to Partial Rejection Control and Approximate Bayesian Computation. J. Cornebise, G. Peters, O. Ratmann in preparation

Subgroup 5: Applications of PMC to cosmology

Some of the participants of the SAMSI program were involved in an ongoing partnership with Institut dâ■Astrophysique de Paris, and therefore gave way to innovative applications of Population Monte-Carlo on cosmology-related problems. They applied the easy parallelization and higher computational efficiency of PMC to lead Bayesian inference of the parameters for data consisting of CMB anisotropies, supernovae of type Ia, and weak cosmological lensing, and provided a comparison of those results to those obtained using state-of-the-art Markov chain Monte Carlo (MCMC), reducing the computation time from fays to hours using PMC on a cluster of processors.

In a second work, PMC was used to compare cosmological models in the context of dark energy and primordial perturbations: cosmology has spawned a multitude of different models, which Bayesian paradigm compares using directly the posterior probabilities of models, in favour of each of one of the given models. From a practical viewpoint, especially for high-dimensional parameter spaces, the calculation of the evi- dence is very challenging. While fast approximations exist, such as the Bayesian Information Criterium or variational Bayes they can fail dramatically for posterior distributions which are not well approximated by a multivariate Gaussian. They used PMC method to estimate the Bayesian evidence and assess the accuracy and reliability of this estimate, using the same set of sampled values used for parameter estimation and to to calculate the Bayesian evidence. Thus with the PMC method, model selection comes at the same computational cost as parameter estimation.

1. Estimation of cosmological parameters using adaptive importance sampling. Wraith, D., Kilbinger, M., Benabed, K., Cappe, O., Cardoso, J.-F., Fort, G., Prunet, S., Robert, C.P. Physical Review D, 80, 023502 (2009)

2. Kilbinger, M., Wraith, D., Robert, C.P. , Benabed, K., CappÃ©, O., Cardoso, J.-F., Fort, G., Prunet, S., Bouchet, F. Bayesian model comparison in cosmology with population Monte Carlo. Available as arXiv:0912.1614, submitted (2009)

Subgroup 6: Stability of Mixture Kalman Filters

A collaboration developed during the stay at SAMSI of two members of the PMC working group on a neighboring topic. This stay at the SAMSI allowed Sylbain Rubenthaler to work with Nicolas Chopin (CREST-ENSAE Paris) on a problem he had been considering for a long time : the stability of the so-called "mixture Kalman filters". They wrote an article on the stability of particle algorithms with a Feynman-Kac representation such that the potential function may be expressed as a recursive function which depends on the complete state trajectory. This includes the mixture Kalman filter. They studied the asymptotic stability of such particle algorithms as time goes to infinity. Sylvain Rubenthaler also started a ongoing collaboration with Amarjit Budhiraja (UNC Chapel Hill) on reflected diffusion, more precisely: the numerical approximation of the stationnary law of a relfected diffusion by means.

1. Stability of Feynman-Kac formulae with path-dependent potentials. Chopin, Nicolas and Del Moral, Pierre and Rubenthaler, Sylvain, to appear in Stochastic Processes and their Applications (2010)

4.7.4 Particle Learning

The particle learning working group was led by Hedibert Lopes.

Key points

- PL as a flexible framework sequential Bayesian computation, comparable and complementary to MCMC schemes;

- Empirical evidence of the superiority (in terms of MC error) of *resample-sample* filters (APF) over *sample-resample* filters (SIS).
- Hybrid filters that combines auxiliary particle filter (APF), Storvik's and PL for state and parameter learning;

Scientific papers **Accepted papers**

1. Particle learning and smoothing. *Statistical Science (to appear).* By Carlos Carvalho, Michael Johannes, Hedibert Lopes and Nicholas Polson.
2. Particle learning for sequential Bayesian computation (with discussion). *Bayesian Statistics 9 (to appear).* By Hedibert Lopes, Carlos Carvalho, Michael Johannes and Nicholas Polson.
3. Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting (to appear).* By Hedibert Lopes and Ruey Tsay.
4. Bayesian inference for stochastic volatility modeling. In Böcker, K., editor, *Rethinking Risk Measurement, Management and Reporting: Bayesian Analysis and Expert Elicitation (to appear).* By Hedibert Lopes and Nicholas Polson.
5. Bayesian computation in finance. In Chen, M.-H., Dey, D., Müller, P., Sun, D. and Ye, K., editors, *Frontiers of Statistical Decision Making and and Bayesian Analysis — In Honor of James O. Berger (to appear).* By Satadru Hore, Michael Johannes, Hedibert Lopes, Robert McCulloch and Nicholas Polson.
6. Extracting SP500 and NASDAQ volatility: The credit crisis of 2007-2008. In O'Hagan, T. and West, M., editors, *Handbook of Applied Bayesian Analysis.* By Hedibert Lopes and Nicholas Polson.

Submitted papers

7. Tracking flu epidemics using Google trends and particle learning. By Vanja Dukić, Hedibert Lopes and Nicholas Polson.
8. Sequential parameter learning and filtering in structured AR models. By Raquel Prado and Hedibert Lopes.
9. Particle learning for general mixtures. By Carlos Carvalho, Hedibert Lopes, Nicholas Polson and Matt Taddy.
10. Bayesian statistics with a smile: a resampling-sampling perspective. By Nicholas Polson, Hedibert Lopes and Carlos Carvalho.
11. Sequential parameter estimation in stochastic volatility models. By Maria Ríos and Hedibert Lopes.

Work in progress

12. Particle learning for generalized dynamic conditionally linear models. By Carlos Carvalho, Hedibert Lopes and Nicholas Polson. (75% finished).

13. Particle learning for fat-tailed distributions. By Hedibert Lopes and Nicholas Polson. (75% finished).

14. Sequential Monte Carlo estimation of DSGE models. By Hao Chen, Francesca Petralia and Hedibert Lopes. (75% finished).

15. Learning in a regime-switching macro-finance Nelson-Siegel model. By Bruno Lund and Hedibert Lopes. (75% finished).

16. Learning expected inflation from nominal and real yields. By Satadru Hore, Hedibert Lopes and Juha Seppälä. (50% finished).

17. Sequential Monte Carlo methods for long memory stochastic volatility models. By Christian Macaro and Hedibert Lopes. (50% finished).

Books Two books published by Working Group members during the 2008-2010 period contain chapters with detailed explanation, literature review, examples and comparisons of sequential Monte Carlo filters.

Raquel Prado and Mike West (2010) *Time Series: Modelling, Computation and Inference.* Boca Raton: Chapman & Hall/CRC Press.

Giovanni Petris, Sonia Petrone and Patrizia Campagnoli (2009) *Dynamic Linear Models with R.* New York: Springer.

PhD Students

Francesca Petralia and *Hao Chen*

Francesca (Department of Statistical Sciences, Duke University), Hao (Fuqua School of Business, Duke University) and I are finishing the paper *Sequential Monte Carlo estimation of dynamic stochastic general equilibrium models.* The paper's novelty is the proposal of a full SMC scheme to estimate non-linear DSGE models in an on-line fashion, which allows simultaneous filtering of states and parameters. We apply our filter on a neoclassical growth model, where structural parameters are sequentially estimated compared to currently used MCMC schemes. In addition, we illustrate sequential model checking/assessment via SMC by comparing the Smets and Wouters DSGE model for the Euro area with several Bayesian vector autoregressive (BVAR) models. Hao presented the paper during the 2009 JSM meetings in Washington, D.C. and during the 2010 Seminar on Bayesian Inference in Econometrics and Statistics (SBIES) in Austin, Texas.

Bruno Lund

December 2009 PhD Thesis on *Term structure models with non-affine dynamics and macro-variables*, Graduate School of Economics, Getulio Vargas Foundation, Rio de

Janeiro. I was Bruno's co-advisor along with Caio Almeida. Bruno spent the 2008-2009 academic year working under my supervision as a visiting PhD student at Chicago Booth. We have just finished the paper *Learning in a regime-switching macro-finance Nelson-Siegel model*, which study the US post-WW2 joint behavior of macro-variables and the yield-curve, which extends previous work by incorporating the possibility of regime changes in order to track NBER cycles. The paper contributes to the sequential Monte-Carlo estimation literature by providing an interesting context for the combination of our particle learning (PL) algorithm with Liu and West's (2001) filter.

Maria Paula Rios

Maria is working under my supervision in the Statistics and Econometrics group at the University of Chicago Booth School of Business. We have just finished the paper *Sequential parameter estimation in stochastic volatility models*, which combines Liu and West's (2001) with Storvik's (2001) particle filters to the class of Markov switching stochastic volatility (MSSV) models (Carvalho and Lopes, 2007). We show that Carvalho and Lopes' (2007) particle filter degenerates or has larger Monte Carlo error than our extension. Our filter takes advantage of recursive sufficient statistics that are sequentially tracked and whose behavior resembles that of a latent state with conditionally deterministic updates. The performance of our filter is also assessed when comparing its sequential estimation of SP500 volatilities to the VIX index, which is the CBOE volatility index.

Post-doctoral researchers Christian Macaro (DSS, Duke University) and I are working on the paper *Sequential Monte Carlo Methods for Long Memory Stochastic Volatility Models*, which proposes to implement an alternative representation based on the aggregation of first order autoregressive models. The resulting framework is particularly suitable for the implementation of SMC methods, since only the marginal posterior distribution of the particles is required. Comparisons with alternative truncation representation is also provided. An application emphasizes the benefits of this proposal when dealing with real time estimation of volatility of financial high frequency data.

4.7.5 Model Assessment and Adaptive Design

The working group leader was Carlos Carvalho.

Summary Goals and Outcomes Following up the discussions in the kick-off workshop (September 2008) the "Model Selection and Adaptive Design" (**MAAD**) working group was formed with the intent to enhance, explore and demonstrate the potential of particle based methods to address issues related to model uncertainty and sequential design/decision making. The group focuses on applications (listed below) where either the computation of model probabilities or the exploration of model spaces represent an enormous challenge that requires effective computation strategies. The central goal of our efforts is to make use of "state of the art" SMC techniques in trying to tackle these issues.

Since its formation, the group has met weekly at SAMSI for discussions of relevant issues and to report on progress made by many of the participants.

Specific Goals and Areas of Focus At the current stage the group has identified four main areas of focus, as described below:

1. Particle Model Selection

Our goal is to develop a general class of particle methods to accommodate uncertainty in variable selection in high-dimensional settings. There is a rich Bayesian literature on variable selection and stochastic search methods for linear regression models, but very little work has been done for nonparametric models that allow the conditional distribution of a response to change flexibly with predictors. Our initial plan was to develop an efficient particle stochastic search (PSS) approach for high-dimensional variable selection in linear regression, while simultaneously developing a Particle Learning algorithm for posterior computation in probit stick-breaking processes (PSBPs). PSBPs are a recently proposed nonparametric Bayes modeling framework, which allow conditional distributions to change flexible with predictors. Due to the conjugacy of the PSBP after data augmentation, it should be possible to adapt the Particle Learning algorithm to include a PSS component. This will allow selection of variables having any impact on the conditional distribution of a response, while also accommodating responses having arbitrary scales (continuous, categorical, count, etc). An additional topic that the group will focus on is development of efficient particle methods for calculating Bayes factors for comparing non-nested models. The idea is to initially devote a similar number of particles to each model, but then through resampling as the algorithm progress, devote increasing numbers of particles to the better models in the list. This will allow accurate posterior computation and estimation of marginal likelihoods for good models, while not wasting computational effort on poor models.

We have made substantial progress in the above areas. Here are specifics of each project:

- **"Particle stochastic search for high-dimensional variable selection"** (Shi and Dunson) - we have continued to make progress in refining our particle stochastic search (PSS) algorithm and have compared the algorithm in a variety of settings to shotgun stochastic search (SSS). We also have results comparing to SSVS for simulated examples and a real data application taken from Hans et al. SSS paper. The paper with the above title is in final preparation stage and will be submitted within a few weeks. We will then move our focus to variable/model selection in nonparametric Bayes regression models, adapting PSS to allow for the inclusion of parameters/latent variables common to the different models in the particles. This will allow variable selection in PSBP mixture models and other interesting cases. We plan to apply this in dynamic mixture model settings as well.

- **"Bayesian distribution regression via augmented particle learning"** (Dunson and Das) - we have continued to make progress in developing and implementing an efficient sequential Monte Carlo algorithm for posterior computation and marginal likelihood estimation in a broad class of mixture models that allow the mixing weights to varying with time, space and predictors. This class of mixture models is referred to as probit stick-breaking mixtures and has the appealing property of facilitating efficient computation through a data augmentation strategy. In particular, for many useful special cases, one can obtain the marginal likelihood in closed form integrating out all of the parameters but conditioning on latent normal variables. Our proposed "augmented particle learning" (APL) algorithm proceeds by sequentially adding subjects in parallel to each of a large number of particles, sampling from the conditional posterior distributions of the latent variables as subjects are added and resampling appropriately. The method avoids the need for sequential importance sampling for updating of particles, instead relying on direct sampling, with marginalization used to improve efficiency. We have primarily code for count regression models which allow the conditional distribution of a count response to change flexibly with a predictor, and have already obtained good results for a mixture of Poisson case with no predictors. A manuscript with the initial results will be submitted within few weeks. An abstract follows:

To limit assumptions in modeling of conditional response distributions, hierarchical mixtures-of-experts models allow the mixing weights in a regression model to vary flexibly with predictors. Nonparametric Bayes methods can be used to incorporate infinitely-many components, allowing effective model dimension to increase with sample size. However, MCMC algorithms for posterior computation often encounter mixing problems due to multimodality of the posterior. Focusing on a broad class of probit stick-breaking process priors for conditional response distributions indexed by time, space or predictors, we propose an efficient augmented particle filter for posterior computation and approximation of marginal likelihoods. The algorithm sequentially updates random length latent normal vectors within each particle as subjects are added, avoiding truncation of the infinite collection of random measures. Through marginalization after data augmentation, the approach bypasses the need to update parameters, dramatically improving efficiency while avoiding degeneracies. The method can be applied broadly for continuous, count or categorical response variables. The methods are illustrated using simulated examples and an epidemiologic application.

Primary subgroup participants: David Dunson (Duke Univ.), Minghui Shi (PhD student, Duke Univ.), Sourish Das (SAMSI and Duke Univ.) and Artin Armagan (Duke Univ.).

2. Adaptive Design

For expensive data, as those arising from computer models, astronomy data, destructive experiments, etc, careful designs which contemplate how many data points will be obtained, where and when, is mandatory. These design problems, for these expensive experiments, have to, almost unavoidably, be sequential and adaptive so as to best use the very scarce and expensive information. Sequential decision problems (of which sequential designs are particular cases) involve "look ahead" computations for all possible future observations, which might be computationally challenging for complex models. We intend to explore SMC methods to help with these computations.

An initial paper is under way as detailed below:

- **"Adaptive sampling for Bayesian variable selection"** (Fei Liu, Fan Li and Dunson) - the problem is to sequentially select subjects based on their predictor values, with the response value obtained for the selected subjects and the objective being optimal performance in model selection. We have the methods details worked out and Fei Liu has implemented a couple of simple examples where she demonstrates substantial advantages relative to selecting the subjects in a random order. We have discussed strategies for proving improvements theoretically under the assumption that the number of subjects in the pool to draw from is large, so that we can avoid finite population sampling complications. Fan has found an interesting data example to motivate the approach and the paper should be completed in a month or two depending on Fei's time. Fei and David have discussed moving on to a "active transfer learning" problem in which there are multiple related regression models and one wants to borrow information in selection of models across the related models.

Primary subgroup participants: Susie Bayarri (Univ. of Valencia and SAMSI), Jim Berger (SAMSI and Duke Univ.), Merlise Clyde (Duke Univ.), Tom Loredo (Cornell Univ.), Ana Corberan (PhD student, Univ. of Valencia), Fei Liu (Univ. of Missouri) and Fan Li (Duke Univ.).

3. Sequential Model Monitoring

In this subgroup we focus on problems of sequential model reassessment and model space exploration as new observations become available. The examples we have been developing so far involve sequential posterior inference about graphical structures underlying the covariance matrix of innovations in dynamic linear models. These models have been applied in large scale sequential portfolio allocation where the graphs provide a regularization tool for the covariance matrix of assets. The development of sequential model selection procedures that address uncertainty about graphs while allowing for on-line updates is an open research area and one of key importance in further applications of DLMs in real forecasting problems. In our first attempt to solve this problem, we have been using particles systems as discrete approximations for the posterior distribution of models. Hao Wang has been coding some of the ideas discussed and the initial results are promising. We have made significant progress in this area and an first draft of a paper by Hao Wang, Craig Reeson and Carlos Carvalho is ready and should be submitted before the summer. It follows the title and abstract:

- **"Sequential Learning in Dynamic Graphical Models"** – *We propose a natural generalization of the dynamic matrix-variate graphical model (Carvalho and West 2007) to time varying graphs. The generalization uses the multi- process modelling idea to introduce sequential graphical model selection procedures that address uncertainty about graphs while allowing for efficient on-line updates. To develop an efficient Bayesian approach for sequentially searching high-dimensional graphical models, we describe a feature-inclusion particle stochastic search algorithm, or FIPSS. The FIPSS algorithm allows parallel exploration of the search space using estimates of edge inclusion probabilities. The model is illustrated using financial time series for predictive portfolio analysis.*

Primary subgroup participants: Carlos M. Carvalho (Univ. of Chicago and SAMSI), Hao Wang (PhD student, Duke Univ.) and Craig Reeson (Undergraduate student, Duke Univ.)

4. Dynamic Control

The main objective of this subgroup is to study problems that have a dynamic (sequential) decision making component as well as some uncertainty about the system parameters that would require Bayesian updating. We are particularly interested in problems that arise in health care settings where a decision maker (a doctor/nurse, emergency response officer, hospital management, etc.) will have to give decisions regarding the treatment options of patients or allocation of scarce resources to a group of patients. These decisions are dynamic in nature as the conditions of patients change with time. Such problems have been studied commonly in the Operations Research literature. However, almost all of the earlier studies assume that the decision maker has complete information about the states of patients and the system parameters. There are several situations where such an assumption of perfect information may not be realistic. For example, for rarely observed diseases or disasters that involve nuclear agents, there does not exist sufficient data to estimate parameters that are needed in solving the dynamic control problem. Our objective is to study such dynamic control problems where the decision maker will learn about the disease or the emergency event under consideration as the decisions are made sequentially. As an initial step, we will consider the following problem. Consider a system with several patients in need of care from a single resource (a doctor or an operating room). The patients are affected by the same disease or the same traumatic event but they could be in different stages of criticality. The stage that a patient is in may affect the cost of keeping that patient waiting, the service requirement of that patient, and also the success probability of the operation. The decision maker cannot observe the true states of patients but can observe certain signals that the patients send (for example, heart rate, blood pressure, etc.). Based on these signals, the decision maker decides which patient should be taken into service next with the objective of maximizing the total expected utility. As we mentioned earlier, the decision maker does not know exactly how the signals and the true states of patients relate and how the patients conditions degrade. When each patient is taken into service, we can observe the true condition of the patient and based on such information collected, we can update the unknown parameters related to the disease/condition. Then, with this updated information, we make the next decision to serve another patient.

Nilay Argon and Melanie Bain are currently using SMC methods in solving a dynamic control problem that arises in the aftermath of mass-casualty incidents. To be more specific, we consider a mass-casualty event (such as a plane crash or a terrorist bombing) that resulted in several casualties in need of care. Due to the massive number of casualties, the medical resources are overwhelmed and decision makers need to prioritize patients for service. Depending on their injuries, the patients could be in different stages of health. The stage that a patient is in may affect his/her probability of survival and also service requirement. The decision maker cannot observe the true states of patients but can observe certain signals that the patients send (for example, pulse, breathing rate, etc.). Based on these signals, the decision maker decides which patient should be taken into service dynamically with the objective of maximizing the total expected number of survivors. We initially assume that the decision maker knows how the signals and the true states of patients relate. We also assume that the patientsÕ conditions degrade according to a discrete time Markov chain with a known transition probability matrix.

We first formulated the above problem as a partially observable Markov decision process (POMDP). The POMDP we obtained could have a very large belief state depending on the number of patients involved and also the number of health stages that we define. Hence, we will need to use an approximate method to solve this problem. We have thus far considered two approaches from the literature. One is by Thrun (2000), where particle filtering is used to reduce the size of the belief space, and the other is by Luo, Fu, and Marcus (2008), which is based on projecting the high-dimensional belief space to a low-dimensional family of parametrized distributions. We are currently implementing ThrunÕs approach.
Primary subgroup participants: Nilay Tanik Argon (UNC), Abel Rodriguez (Univ. of California, SC), Melanie Bain (PhD student, UNC) and Kai Wang (PhD student, Duke Univ.)

Papers

- Dynamic Financial Index Models: Modeling Conditional Dependencies via Graphs. By by Hao Wang, Craig Reeson and Carlos M. Carvalho. http://ftp.stat.duke.edu/Wor 21.pdf (submitted to Bayesian Analysis)

- Dynamic Stock Selection: A Structured Factor Model Framework By C. Carvalho et al., Invited paper, 2010 Valencia Meeting on Bayesian Statistics.

4.7.6 Continuous Time

The working group leaders are Jashya Bishwal, Paul Fearnhed and Jochen Voss. During the initial formative phase, the working group in continuous time models, parameter estimation and finance started a review of the relevant literature. Several articles were presented during the group meetings. Topics covered include the following:
- filtering discrete observations of a continuous time signal
- exact algorithms
- filtering and parameter estimation using a windowed SMC method (in discrete time)
- change point problems for continuous time processes
- filtering for a CIR process given Poisson observations

The working group focussed on four areas. Firstly, SMC for hierarchical branching process models, with application for qPCR analysis. Secondly, methods for survival data, where the underlying hazard depends on an unobserved stochastic process. The application for this work is to modelling, analysis and prediction of mortgage default rates. Thirdly, they have looked at inference for diffusion models via "least-action": defining and calculating a best path for the unobserved diffusion, and constructing Laplace approximations to the transition density of the diffusion. Applications here include models in systems biology. Finally, we looked at new inference methods for α-stable Lévy processes.

The research has led to a grant proposal to the UK's EPSRC, which will develop on the least action filtering and α-stable models by Rogers, Godsill and West. This latter area led to several grant proposals and a funded project from Fraunhoffer Institute in Germany (contact Ralf Korn) in which a PhD student, Tatjana Lemke, funded by Fraunhoffer, is based in Cambridge with Simon Godsill and working on Monte Carlo inference for α-stable Levy processes. There is a further grant on Laplace approximations, with a focus on their use for diffusion models, being prepared by Fearnhead, Tawn and Sherlock.

- 'Subprime Mortgage Default' James B. Kau ,Donald C. Keenan, Constantine Lyubimov, V. Carlos Slawson (submitted 2010)
- Anand Vidyashankar (In prep).
- Simulation and Inference for Stochastic Kinetic Models via limiting Gaussian Processes. Paul Fearnhead, Vasilieos Giagos and Chris Sherlock (In prep).
- Inference for autoregulatory genetic networks using diffusion approximations. Vasilieos Giagos, PhD Thesis, Lancaster University (In prep; submission by September 2010)
- ENHANCED POISSON SUM REPRESENTATION FOR ALPHA-STABLE PROCESSES, Tatjana Lemke and Simon Godsil, Accepted for Proc. IEEE ICASSP 2011.

4.7.7 Big Data & Distributed Computing

Summary Emerging from discussions at the Sept. 2008 opening workshop, the *Big Data & Distributed Computing (BD&DC)* working group was defined by research challenges and themes cutting across numerous applications of stochastic computation and sequential methods: scaling of models and analysis methods for increasingly large data sets and in problems with increasing large spaces of underlying parameters and latent variables. Exploiting multicore, cluster and parallel hardware promotes a need for basic research and innovations in the development of computational algorithms and also of model specification and structuring. The opportunity to make progress in these areas, linked to specific motivating applications, led to the formation of this focused working group. BD&DC subgroups were involved in specific research projects under the general goals, with a series of interconnections with some of the other working groups involving cross-cutting projects.

Goals and Areas of Focus *Exploration, evaluation, development and application of effective computational methods for model fitting and model assessment in problems involving large data sets and high-dimensional latent process parameters (the latter are examples of* *"big missing data"): sequential Monte Carlo methods, stochastic model search, sequential importance sampling and also annealing/optimization methods.* Specific research subgroups are as follows.

BD&DC.1 Distributed computing for SMC *(Manolopoulou, Mukherjee, West, Yoshida).*

Sequential Monte Carlo methods for model learning, estimation and comparison using distributed computing on clusters. This is relevant to several of the specific modelling, methodology and application areas. Activities include study of theory and methods of implementing a variety of SMC methods on clusters, using some of the specific model contexts of interest to this working group for development. Strategies for parallelised and cluster-based computation were explored for problems involving large data sets and high-dimensional latent processes, i.e., large missing data sets, the latter focused on state-space modelling of long time series. This research interacted with researchers in the Tracking working group.

BD&DC.2 SMC in dynamic non-linear models in systems biology (*Ji, Mukherjee, West).*

SMC and distributed computing in mechanistic, nonlinear state-space models, with motivating applications in dynamic stochastic models arising in studies of cellular communication in biological networks in systems biology. This area involves model development and use of customised sequential Monte Carlo methods, and so interacted substantially with some of the other program working groups. Specific characteristics of motivating problems were (a) state-space models with many uncertain parameters that are observed over time, and for which sequential learning is either desirable or necessary; (b) very high-dimensional latent processes. In systems biology problems, models are developed mathematically on very fine, discrete time scales, but actual data is observed at much cruder time scales, so that the fine time scale states become missing data in very high dimensions.

This research area also intersected with studies involving stochastic computation for model fitting in complex computer model emulation, related to the program activities at the intersection of research in computation and computer modelling design and development. This aspect of BD&DC research was represented in talks at the SAMSI workshop on *Adaptive Design, Sequential Monte Carlo and Computer Modeling* in April 2009 as well as in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

BD&DC.3 SMC for inference on "rare events" with very large data sets *(Manolopoulou, West).*

Sequential analysis and decision-guided sample selection and learning about rare events in mixture modelling with very large data sets. Motivating applications come from problems of inference on characteristics of rare sub-populations of biological cells in studies using flow cytometry technology in immunology, vaccine design and other areas. In such studies, a single experiment can easily generate hundreds of millions of observations in, typically 10-20 dimensions, representing marker proteins on the surface of cells. Random sampling to fit models, such as mixture models for classification and discrimination of sub-populations, is standard, though model fitting becomes challenged by sample size and so sequential methods are inherently interesting. Moreover, a specific focus on generating maximal information of rare sub-populations leads to statistical design and biased sampling strategies that are inherently sequential and for which simulation-based methods need to be developed. The interest in and role for distributed computation is evident.

BD&DC.4 SMC in nonparametric analyses *(Das, Dunson, Li, Liu).*

Sequential methods in nonparametric statistical regression and density estimation models, with motivating applications in problems in epidemiology and public health, and in studies of huge data sets in e-commerce and internet traffic research, among others. This study introduced new classes of SMC algorithms for posterior computation and marginal likelihood estimation in a flexible class of mixture models, which allow mixture weights to vary with time, space and predictors. The proposed augmented particle learning (APL) algorithm has had excellent performance in simulation experiments for a variety of data types, avoiding degeneracies common to SMC algorithms through use of marginalization after data augmentation. The algorithm has major advantages over MCMC algorithms in avoiding mixing problems that plague MCMC for mixture models, while also allowing marginal likelihood estimation, which allows testing of competing nonparametric models and parametric vs nonparametric models.

Application areas are numerous. Part of this research generated components of a successful NIH proposal, that proposed the further development of the APL algorithm in applications in statistical genetics and gene-environment interactions studies. This involves models that allow quantitative traits to vary flexibly with high-dimensional single nucleotide polymorphisms and environmental factors.

BD&DC.5 Adaptive sequential sampling for Bayesian variable selection *(Li, Liu, Dunson).*

Intersections of interests with the working group on sequential Monte Carlo in model assessment focussed on adaptive strategies for the variable selection problems and efficient Sequential Monte Carlo methods for the evaluation of designs in massive data sets. The problem is to sequentially select subjects based on their predictor values, with the response value obtained for the selected subjects and the the objective being optimal performance in model selection. The ideas were developed and methods detailed; implementations in a couple of demonstrate substantial advantages relative to selecting the variables in a random order. An interesting study now underway applies this to problems with censored data.

This aspect of BD&DC research was represented at the SAMSI workshop on Adaptive Design, Sequential Monte Carlo and Computer Modeling in April 2009, as well as in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009.

BD&DC.6 Sequential model search in very large model spaces *(Dunson, Shi, West, Yoshida).*

Stochastic and related deterministic/annealing based methods of search over very large, discrete model spaces such as arising in sparse multivariate factor models with many response variables, and in regression model uncertainty (linear and nonlinear) with many candidate covariates. Advances in computational methods included innovations in annealed entropy methods and Bayesian shotgun stochastic search methods. Some specific motivating applications came from genomics and public health contexts. One key development involved including model indices within the particles of SMC methods leading to an efficient algorithm for massive dimensional variable selection (Dunson and Shi), another involved a synthesis of entropy annealing based global optimization with stochastic search for very large-scale sparse statistical models (Yoshida and West).

BD&DC.7 Novel modeling for tracking *(Ji, Manolopoulou, West).*

Intersections of interests with the working group on sequential Monte Carlo in compute-intensive tracking problems generated novel model and computational methods development for tracking problems with prototype applications in monitoring and tracking many cells in systems biology experimental data. Data arising from motivating applications include studies in computational immunology driven by experiments in vaccine design, where the motion of multiple different cell types is monitored by measured fluorescent intensities of cell surface marker proteins. Research here involved novel Bayesian dynamic, non-parametric models for inhomogeneous spatial intensity functions and sequential Monte Carlo methods development for model fitting.

BD&DC.8 SMC in very large inverse problems *(Mukherjee, West).*

Emerging from BD&DC discussions in spring 2009, a new project emerged involving a subgroup of this working group, involving approaches to computer model-data synthesis and inversion in atmospheric chemistry (CO) monitoring and data synthesis. With a focus on short time-scale inference on improved understanding of the impact of earth surface fires (tropical forest fires, savannah fires, etc) on variations atmospheric CO, a core challenge is integration of *massive* amounts of high-resolution data from new satellites launched in 2009 with predictions from deterministic biophysical simulation models. Sequential Bayesian methods development were initiated as a spin-off from this BD&DC discussion, and have led to a new, free-standing collaboration with atmospheric chemists as well as defining a core thesis topic for Mukherjee. Two papers are in draft (May 2010). This new collaboration with disciplinary computer modelers is an excellent example of a *really* big, and cluster compute-intense BD&DC problem.

Participants This working group involves local faculty participants, SAMSI postdoctoral fellows, SAMSI visitors, SAMSI and non-SAMSI graduate students, and represents various areas of statistics and computational science. Several junior and female researchers are involved, including some who were quite new to the general area of Sequential Monte Carlo, as well as the specific areas of this working group, prior to the program. See Table 1 for the list of primary and active participants, as well as additional participants who either had some engagement in formative discussions, or who had occasional interactions in BD&DC meetings, or who were short-term SAMSI visitors.

Student Involvement

Name (gender)	Position	Affiliation	Dept/Discipline
(A)			
Carlos Carvalho (m)	Assistant Professor	Chicago	Statistics & Econ.
Sourish Das (m)	Postdoc	SAMSI	Statistics
David Dunson (m)	Professor	Duke	Statistical Science
Chunlin Ji (m)	Graduate RA	Duke & SAMSI	Statistical Science
Fan Li (f)	Assistant Professor	Duke	Statistical Science
Fei Liu (f)	Assistant Professor	Missouri	Statistics
Ioanna Manolopoulou (f)	Postdoc	SAMSI	Statistics
Chiranjit Mukherjee (m)	PhD student	Duke	Statistical Science
Minghui Shi (f)	Graduate RA	Duke & SAMSI	Statistical Science
Ryo Yoshida (m)	Assistant Professor	ISM Tokyo	Statistics
Hao Wang (m)	Graduate student	Duke	Statistical Science
‡Mike West (m)	Professor	Duke	Statistical Science
(B)			
Ernest Fokoue (m)	Assistant Professor	Kettering	Mathematics
Amadou Gning (m)	Postdoc	Lancaster	Communication Systems
Steve Koutsourelakis (m)	Assistant Professor	Cornell	Engineering
Lyudmila Mihaylova (f)	Lecturer	Lancaster	Communication Systems
Mario Morales (m)	Consultant	Emetricz	Engineering/Statistics
Deb Roy (m)	Assistant Professor	Penn State	Statistics
Andrew Thomas	Lecturer	St. Andrews University	Ecology & Statistics
Joshua Vogelstein (m)	PhD student	Johns Hopkins	Neuroscience

Table 1: *(A)* Primary and local participants (‡ *Working Group leader*); *(B)* Additional researchers (initial participants, collaborators and/or short-term visitors in the BD&DC group) {tab:BigData1}

Chunlin Ji (SAMSI RA) was primarily attached to the Tracking (Godsill) working group but also participated actively in the BD&DC group with West on spatial dynamic modelling for biological cell tracking problems and nonlinear modeling. Ji developed SMC methods in the context of new classes of models. This research grew out of existing work of Ji & West in static problems, now extended with new dynamic models that will formed additional parts of Ji's PhD thesis research, and led to expanded follow-on research in the area with SMC postdoc Ioanna Manolopolou mentioned below. Ji led discussions on this work at several BD&DC meetings, gave a talk at the February 2009 mid-program workshop, presented this work at the 7^{th} Workshop on Bayesian Nonparametrics in Turin, Italy, in June 2009 and at the 2009 Joint Statistical Meetings in Washington DC 2009. Ji defended his PhD in December 2009 and moved to a postdoctoral position with Jun Liu in Statistics at Harvard.

Chiranjit Mukherjee (Duke graduate student) actively participated in the BD&DC group (though was not officially supported by the program). Mukherjee developed studies of

SMC methods for model fitting and comparison in nonlinear dynamical models arising from systems biology (and other applications). These studies involve very long time series but for which most of the underlying states are unobserved, and his work has explored, evaluated and developed novel approaches to SMC using distributed computation. In March 2009, Mukherjee presented and passed his PhD preliminary exam based on this work, has one paper published from this work, led several discussions on the topic at the BD&DC meetings, presented a poster at the February 2009 mid-program workshop, presented this work in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009 and at several Duke workshops. Mukhjeree then became involved in the new, emerging project initiated in a BD&DC meeting on large-scale data integration for inverse inference and data synethesis with computer models in atmospheric chemistry. He will present this work at the ISBA World Meeting in Spain, June 2010. These two topics from his BD&DC interactions form the basis of his ongoing thesis research; he will defend in May 2011.

Francesca Petralia (SAMSI RA) was attached to the Particle Learning (Lopes) working group but also participated actively in the BD&DC group. Petralia presented a talk on her work with SMC in econometric models in the Particle Learning working group (Lopes, leader) in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009. Petralia is currently finalizing a paper, with Lopes, on this SMC project.

Minghui Shi (50% SAMSI RA) worked on sequential model search methodology for large, discrete model spaces, typified by "large p" regression model uncertainty. With Dunson, Shi developed novel extensions of shotgun stochastic search that incorporate new ideas from SMC. Shi passed her PhD preliminary exam on this topic in April 2009, and the topic underlies her emerging thesis area. Shi led discussions on this work at BD&DC meetings, presented a poster at the February 2009 mid-program workshop, presented this work in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009, and will present extensions at the ISBA World Meeting in Spain, June 2010.

Hao Wang (Duke graduate student) actively participated in the BD&DC group as well as other working groups, (though was not officially supported by the program). Wang worked, in part, on SMC methods for dynamic graphical models with Carvalho and West, led discussions on the topic at the BD&DC meetings, presented a poster at the February 2009 mid-program workshop, and presented this work at the SAMSI workshop on *Adaptive Design, Sequential Monte Carlo and Computer Modeling* in April 2009 as well as in a SAMSI Topic Contributed Session at the 2009 Joint Statistical Meetings in Washington DC 2009. With Carvalho, Wang wrote a paper on SMC in dynamic graphical models (currently under review) and the work formed a component of his PhD thesis. Wang successfully defended his PhD in spring 2009, and will move to a tenure-track faculty position, Assistant Professor of Statistics, University of South Carolina, in summer 2010.

Other Activities

- BD&DC short-term visitor Andrew Thomas (St. Andrews University) has, as a result of discussions and interactions in the BD&DC group, initiated development of new software for SMC based on the OpenBUGS software, and this is expected to be developed and set up as Open Source software.

Manuscripts Manuscripts linked to research in BD&DC and acknowledging SAMSI/NFS support:

1. D.B. Dunson & S. Das (2009) Bayesian distribution regression via augmented particle learning. In preparation.

2. C. Ji & M. West (2009) Spatial dynamic mixture modelling for unobserved point processes and tracking problems. Under revision.

3. F. Liu, D. B. Dunson & F. Zou (2010) High-dimensional variable selection in meta analysis for censored data, *Biometrics*, to appear.

4. F. Liu, F. Li & D. B. Dunson (2010) Adaptive design for variable selection in normal linear models, in preparation.

5. F. Liu & M. West (2009) A dynamic modelling strategy for Bayesian computer model emulation, *Bayesian Analysis*, **4**(2), 393-412.

6. I. Manolopolou, C. Chan & M. West (2009). Selection sampling from large data sets for targeted inference in mixture modeling, *Bayesian Analysis*, to appear.

7. I. Manolopolou, X. Wang, C. Ji, H.E. Lynch, S. Stewart, G.D. Sempowski, S.M. Alam, M. West and T.B. Kepler (2009) Statistical analysis of immunofluorescent histology, submitted for publication.

8. C. Mukherjee & M. West (2009) Sequential Monte Carlo in model comparison: Example in cellular dynamics in systems biology, *JSM Proceedings, Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association*, 1274-1287.

9. M. Shi & D. Dunson (2009) Bayesian variable selection via particle stochastic search, submitted for publication.

10. H. Wang, C. Reeson & C.M. Carvalho (2009) Dynamic financial index models: Modeling conditional dependencies via graphs, invited revision resubmitted.

11. R. Yoshida & M. West (2009) Bayesian learning in sparse graphical factor models via annealed entropy, *Journal of Machine Learning Research*, to appear.

5 Conclusions and expected outcomes

To conclude, the program has been successfull. It has generated much activity across a wide range of topics. We had well-structured Working Groups, holding weekly meetings by Webex. The results of these collaborative research efforts have started appearing in leading statistical journals.The collaborations enabled by the program no doubt lead to major grant applications in the area of SMC and its scientific applications.