

Continuing Education Course Proposal

2017 Joint Statistical Meetings

Baltimore, MD, USA

1. TITLE OF PRESENTATION: Applied Bayesian Computational Methods, with an Introduction to the NIMBLE Platform **VERY TENTATIVE title, feel free to change**
still thinking about how best to phrase the NIMBLE part or whether just to leave it out of the title

2. NAMES AND ADDRESSES OF PRESENTERS:

Abel Rodríguez
Applied Mathematics and Statistics
School of Engineering, MS: SOE2
University of California
Santa Cruz, CA 95064, USA
E-mail: abel@soe.ucsc.edu
Phone: +1 (831) 459-1047
Fax: +1 (831) 459-4829

Christopher Paciorek
Department of Statistics
367 Evans Hall
University of California
Berkeley, CA 94720, USA
E-mail: paciorek@stat.berkeley.edu
Phone: +1 (510) 842-6670
Fax: +1 (501) 642-7892

3. ABSTRACT:

[here's a first pass; not wedded to it](#)

This day-long short course will introduce participants to modern Bayesian computational methods, with a focus on the wide variety of new and old Markov chain Monte Carlo (MCMC) algorithms. The course will combine discussion of the methods and the principles underlying them with demonstration of the methods. The course will discuss computations for model fitting, model selection, model assessment and MCMC assessment. There will be a heavy emphasis on hands-on implementation of algorithms, primarily within the NIMBLE platform (r-nimble.org), an R-based system that allows for flexible use of a variety of MCMC approaches, as well as flexible programming of new algorithms that can be applied to hierarchical models specified in the BUGS language.

4. OUTLINE:

- Morning Session I: Overview of techniques and tools

Brief refresher on Bayesian statistics. Overview of computational techniques used in Bayesian statistics, including a discussion of relative advantages and disadvantages. Introduction to simulation-based techniques for Bayesian inference. Introduction to the NIMBLE platform, including the BUGS language for specifying models.

(Break)

- Morning Session II: Basic MCMC methods

Gibbs sampling. Example: Linear mixed-effect models in NIMBLE. General Metropolis-Hastings algorithms and Metropolis-within-Gibbs. Example: Conditionally autoregressive models in NIMBLE. Slice samplers. Example: Log-linear regression in NIMBLE. Evaluation of Markov chain Monte Carlo algorithms. Example: Log-linear regression in NIMBLE revisited.

are log-linear models particularly amenable to slice sampling - if so, sounds good to me!

perhaps reformat as bulleted list as follows for each of the 4 sessions (don't feel strongly and not sure of length limits on the document)?

- Gibbs sampling. Example: Linear mixed-effect models in NIMBLE.
- General Metropolis-Hastings algorithms and Metropolis-within-Gibbs. Example: Conditionally autoregressive models in NIMBLE.
- Slice samplers. Example: Log-linear regression in NIMBLE.
- Evaluation of Markov chain Monte Carlo algorithms. Example: Log-linear regression in NIMBLE revisited.

(Lunch)

- Afternoon Session I: Improving MCMC performance

Techniques for improving the performance of MCMC (blocking, reparameterization, Rao-Blackwellization, parameter expansion). Examples: variable selection in linear regression, linear and binary mixed-effect models revisited in NIMBLE. Latent variables and other tricks. Example: Binary regression models, robust regression, mixture models in NIMBLE. Reversible-jump MCMC algorithms. Example: variable selection in non-linear regression in NIMBLE. Introduction to Hamiltonian Monte-Carlo methods.

(Break)

- Afternoon Session II: Advanced MCMC methods

Hamiltonian Monte-Carlo methods. Examples: Gaussian process regression, circular regression in Stan. Estimating normalizing constants. Example: Comparing linear and non-linear regression models in NIMBLE. MCMC for intractable likelihoods. Example: Exponential Random Graph models [not sure if we can do this in NIMBLE as I'm not sure what's involved - worth some discussion].

perhaps work a survival analysis model in somewhere though not sure what tool it best accompanies - perhaps along with or instead of the CAR model example assuming we'll have some sort of random-walk/smooth process as the baseline hazard; or we could look at what's involved in accelerated failure time modeling - I'd need to investigate but am hoping it's doable with what's feasible in NIMBLE

5. LEARNING OUTCOMES:

(a) Learning outcomes (performance objectives):

- Students will understand the relative advantages and disadvantages of different computational techniques used in Bayesian statistics.
- Students will be able to understand and implement modern simulation-based computational algorithms for inference in Bayesian models, including Gibbs samplers, random-walk Metropolis-Hastings sampler, Reversible Jump Markov chain Monte Carlo algorithms and Hamiltonian Monte Carlo algorithms, among others.
- Students will be able to evaluate the performance of simulation-based algorithms using standard concepts such as convergence, mixing, equivalent sample size, etc.
- Students will be able to use the BUGS language and the NIMBLE platform to implement simulation-based algorithms for Bayesian hierarchical models.

This course combines material that was previously presented by Rodríguez in a well-received short course at the World Meeting of the International Society for Bayesian Analysis (ISBA) in the summer of 2016 and by Paciorek in the forms of an ASA SBSS webinar, ISBA World Meeting summer 2016 short course, and International Statistical Ecology Conference summer 2016 short course, among other venues. The emphasis of the short course will be on application and implementation, with relatively minimal coverage of theoretical foundations.

(b) Content and instructional methods:

- The lectures will use slides specifically prepared for the course, which will include a number of references to papers on methodology and applications. The slides will be made available to participants.
- Chris, do you want to mention any specific book here? I am not really following anybody in particular, but it might be useful to mention something to anchor expectations in terms of complexity/level and provide back-up for our slides.
[draft] The material will be at the level of Gelman et al. (2013) Bayesian Data Analysis, Third Edition, although material is drawn from a variety of sources.
- The different computational strategies discussed in the slides will be illustrated using real-life data examples taken from the literature and, in some cases, from the presenters' own work. Examples will focus on widely used hierarchical models, including linear and non-linear multiple regression models, generalized linear models, spatial regression models, survival models, etc.
- Illustrations will focus on the NIMBLE platform which provides a variety of algorithms and is an extensible system for programming algorithms, but will use other software, such as Stan for Hamiltonian Monte Carlo, as needed. A significant component of the course will focus on introducing participants to the NIMBLE platform and ensuring that they are capable of extending its use to models beyond those covered in our illustrations.

- Interactivity will be encouraged. Students will be encouraged to bring their own laptops so that they can run examples along with the instructors and ask questions. We plan to use a dynamic and interactive approach in which both instructors alternate leading discussion and in which illustrations meld seamlessly with discussion on foundations and algorithmic structure.

6. PRESENTERS:

Abel Rodríguez obtained a Ph.D. in Statistics and Decision Sciences from Duke University in 2007 with a dissertation on Bayesian nonparametric modeling and inference using Dirichlet processes. He is currently a Professor of Statistics at the University of California, Santa Cruz and an Associate Editor for the *Annals of Applied Statistics* and *Bayesian Analysis*. His research areas include Bayesian nonparametric and semiparametric modeling, machine learning, time series analysis, spatial modeling, econometrics, and proteomics. He has been presenter and/or co-presenter of a number of shortcourses at the Joint Statistical Methods (2011, 2013, 2015) and the International Society for Bayesian Analysis World Conference (2014, 2016), as well as a number of other international venues including Brazil, Colombia and Mexico. He has published over 30 papers and has directed eight Ph.D. students.

Christopher Paciorek obtained a Ph.D. in Statistics from Carnegie Mellon University in 2003 with a dissertation on nonstationary Gaussian processes for regression and spatial modeling. He is currently an Associate Research Statistician, Lecturer, and Statistical Computing Consultant at the University of California, Berkeley. He is an Associate Editor for *Reproducibility for JASA*, an Associate Editor for *Advances in Statistical Climatology Meteorology and Oceanography*, and is on the board of statistical reviewers for *JAMA*. His research areas are in applied Bayesian statistics and spatial statistics with primary application to environmental and public health research including climate change, paleoecology, and global health. He is one of the core developers and co-PI of the NIMBLE project. He has published over 60 papers. He has presented workshops on a wide variety of computing topics to various audiences at UC Berkeley.

7. AUDIO-VISUAL EQUIPMENT: A second screen is requested (in addition to the standard equipment provided). Students are expected to use their laptops during the session, so enough power outlets would likely be helpful.