Big Data in Environmental Science

Program

May 11-15, 2015

The University of British Columbia
Earth Sciences Building (ESB)
2207 Main Mall, Vancouver
Workshop on Big Data in Environmental Science

May 11 to 15, 2015 at the University of British Columbia, PIMS

Earth and Ocean Sciences Building

The timings indicated below are not intended to be too rigid; we will tolerate small deviations from the time table to make sure that discussion gets followed up on. The event ends Friday at lunch and we have scheduled a free afternoon on Wednesday for speakers to have some time to get out in Vancouver. We want, however, to draw your attention to a public lecture by Gavin Shaddick on spatio-temporal methods in environmental epidemiology to be given Wednesday afternoon at 4 PM in the same building as the one in which our conference is being held. A reception is to follow. Details concerning the location and title will be provided at the workshop.

Acknowledgments:

We have been funded by the Canadian Statistical Sciences Institute (CANSSI), by the Pacific Institute for the Mathematical Sciences (PIMS), by the University of British Columbia Department of Statistics, by STATMOS, by Simon Fraser University, and by the Department of Statistics and Actuarial Science at Simon Fraser. We are grateful to all of them for this. We are grateful to PIMS staff, and particularly to Ruth Situma, Ian Allison and Clare Kiernan of the UBC PIMS site, for a great deal of organizational help.

Organizing Committee:

Richard Lockhart, Simon Fraser University;

Charmaine Dean, Western University;

Peter Guttorp, University of Washington;

Will Kleiber, Colorado;

Bo Li, University of Illinois;

Steve Sain, the Climate Corp;

James V. Zidek, University of British Columbia.
Getting Started

WiFi Get connected: Select the "ubcvisitor" wireless network on your wireless device. Open up a web browser, and you will be directed to the login page.

Frequently Asked Questions

Q: Where do I check in on the first day?

Check-in and Package pick up can be done in the Atrium

Q: Where are the sessions?

All workshop sessions in ESB Room 2012 of the Earth Sciences Building at UBC

Q: Will the program change?

Program changes and updates will be announced at each session.

Q: When should I wear my badge?

Please wear your name badges at all times on site so that PIMS Staff recognize you as a guest.

Q: Where can I go for help on site?

If you need assistance or have a question during the conference, please feel free to talk to one of the organizers

Q: Where can I get refreshments and meals?

For snack or quick meals, please view the list of UBC eateries attached at the end of the program

Q: Where can I get directions for campus and the building?

You will find a copy of the building floor on page 3 and a campus map at the end of the program
Monday 11th May, 2015

**Speaker abstracts begin on page 8**

8:15am  
Registration and Check-in (ESB Atrium)

8:15am - 8:50am  
Morning coffee, pastries and fruit (ESB 2012 Lobby)

8:50am - 9:00am  
**Welcome Message:**
  
  - **Prof. Richard Lockhart:** Simon Fraser University, Organizing Committee Chair
  - **Prof. Martin Barlow:** Interim Director Pacific Institute for the Mathematical Sciences

9:00 am - 9:45am  
Andrew Trites, University of British Columbia

  *Tentative title: A biologist’s perspective on biologging and big data*

9:45 am - 10:30 am  
Jim Zidek & Seagle Liu: University of British Columbia

  *Interpolating the high dimensional track record of the fur seal: fusing a physical model with data*

10:30 am - 11:00 am  
Coffee (ESB 2012 Lobby)

11:00 am - 11:45 am  
Mevin Hooten, Colorado State University, via Skype

  *Fringe benefits: The hidden utility of constraints in telemetry studies*

11:45 am - 12:00 pm  
Discussion

12:00 pm - 1:30 pm  
Lunch (No-host; please see list of UBC eateries located at the end of the program and also online at [http://www.food.ubc.ca/](http://www.food.ubc.ca/))

1:30 pm - 2:20 pm  
Finn Lindgren, University of Bath

  *Stochastic partial differential equations and numerical methods for large scale spatial statistics*

2:30 pm - 3:00 pm  
Coffee (ESB 2012 Lobby)

3:00 pm - 3:45 pm  
Timothy Johnson, University of Michigan

  *Analysis of Point Pattern Imaging Data using Log Gaussian Cox Processes with Spatially Varying Coefficients*

3:45 pm - 4:30 pm  
Marco Ferreira, Virginia Tech

  *Dynamic Multiscale Spatiotemporal Models for Gaussian and Poisson Processes*
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:15am - 8:50am</td>
<td>Morning coffee, pastries and fruit (ESB 2012 Lobby)</td>
</tr>
<tr>
<td>9:00am - 9:45am</td>
<td><strong>Soutir Bandyopadhyay</strong>, Lehigh University</td>
</tr>
<tr>
<td></td>
<td><em>Spatial Methods for Nonstationary Fields Using Compact Basis Functions</em></td>
</tr>
<tr>
<td>9:45 am - 10:30 am</td>
<td><strong>Hao Zhang</strong>, Purdue University</td>
</tr>
<tr>
<td></td>
<td><em>Modeling the Complexity of Data Structure in Environmental Sciences</em></td>
</tr>
<tr>
<td>10:30 am - 11:00 am</td>
<td>Coffee (ESB 2012 Lobby)</td>
</tr>
<tr>
<td>11:00 am - 11:45 am</td>
<td><strong>Jonathan Bradley</strong>, University of Missouri</td>
</tr>
<tr>
<td></td>
<td><em>Efficient Parameterizations for Multiscale Multivariate Spatio-Temporal Data</em></td>
</tr>
<tr>
<td>11:45 am - 12:00 pm</td>
<td>Discussion</td>
</tr>
<tr>
<td>12:00 pm - 1:30 pm</td>
<td>Lunch</td>
</tr>
<tr>
<td>1:30 pm - 2:20 pm</td>
<td><strong>Finn Lindgren</strong>, University of Bath</td>
</tr>
<tr>
<td></td>
<td><em>Towards realistic stochastic modelling of global temperatures</em></td>
</tr>
<tr>
<td>2:30 pm - 3:00 pm</td>
<td>Coffee (ESB 2012 Lobby)</td>
</tr>
<tr>
<td>3:00 pm - 3:45 pm</td>
<td><strong>Robert Lund</strong>, Clemson University</td>
</tr>
<tr>
<td></td>
<td><em>Changepoints and Associated Climate Controversies</em></td>
</tr>
<tr>
<td>3:45 pm - 4:30 pm</td>
<td><strong>Dorit Hammerling</strong>, NCAR</td>
</tr>
<tr>
<td></td>
<td><em>A new ensemble-based consistency test for the Community Earth System Model</em></td>
</tr>
<tr>
<td>4:45pm - 6:00pm</td>
<td>Reception and Networking Event (ESB Atrium)</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8:15am - 8:50am</td>
<td>Morning coffee, pastries and fruit (ESB 2012 Lobby)</td>
</tr>
</tbody>
</table>
| 9:00 am - 9:45am| Peter Craigmile, Ohio State University  
*Heteroscedastic asymmetric spatial processes (HASP)* |
| 9:45 am - 10:30 am| James Balamuta, University of Illinois  
*Powering Up the Computation for Big Data* |
| 10:30 am - 11:00 am| Coffee (ESB 2012 Lobby)                                                                            |
| 11:00 am - 11:45 am| Luke Bornn, Simon Fraser University  
*Lost Moments: The Effect of Pre-processing on Environmental Data* |
| 11:45 am - 12:30 pm| Will Kleiber, University of Colorado  
*Coherence for Random Fields* |
| 12:30pm - 12:45pm| Group Photo (Please assemble at ESB Atrium)                                                       |

**No scheduled talks on Wednesday afternoon, however participants are encouraged to attend the afternoon lecture by Prof. Gavin Shaddick hosted by the Peter Wall Institute for Advanced Studies.**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
</table>
| 4:00pm - 5:00pm | Gavin Shaddick, University of Bath & International Visiting Research Scholar at UBC’s Peter Wall Institute of Advanced Studies  
*Public Lecture in ESB 2012 with a reception to follow; Details to be announced* |
Thursday 14th May, 2015

8:15am- 8:50am  Morning coffee, pastries and fruit (ESB 2012 Lobby)

9:00 am -9:45am  Doug Nychka, NCAR
                      Extremes in regional climate: What to do with 8000 histograms

9:45am- 10:30am  Sudipto Banerjee, UCLA
                      On Gaussian Process Models for High-Dimensional Geostatistical Datasets

10:30am- 11:00am  Coffee (ESB 2012 Lobby)

11:00 am – 11:45pm  Matthias Katzfuss, Texas A & M
                      A multi-resolution approximation for big spatial data

12:15 pm - 1:45 pm  Lunch

1:30 pm - 2:15 pm  Bo Li, University of Illinois
                      Evaluating Climate Field Reconstructions in Reduced Dimension

2:15 pm - 3:00 pm  Bruno Sanso, University of California at Santa Cruz
                      Using MARS for functional computer model emulation and sensitivity analysis

3:00 pm - 3:30 pm  Coffee (ESB 2012 Lobby)

3:30 pm - 5:00 pm  Jennifer Bryan, UBC
                      Visualization with ggplot2; bring your laptop
Friday 15th May, 2015

8:15am- 8:50am  Morning coffee, pastries and fruit (ESB 2012 Lobby)

9:00 am - 9:45am  Mike Brauer, University of British Columbia

Tentative title: An environmental epidemiologist’s perspective on the role of big data in environmental health risk assessment

9:45 am - 10:30 am  Renjun Ma, University of New Brunswick

Spatiotemporal Analysis of Environmental Health Risk

10:30 am - 11:00 am  Coffee (ESB 2012 Lobby)

11:00 am - 11:45 am  Gavin Shaddick, University of Bath

Incorporating large scale exposure modeling into environmental epidemiological studies

11:45 am - 12:00 pm  Discussion and Final Remarks

Survey:
Please help PIMS to improve the quality of its events and plan for the future by filling out the survey located on the event webpage at http://www.pims.math.ca/scientific-event/150511-bdes.
Speaker Abstracts and Titles

James Balamuta, University of Illinois

Powering Up the Computation for Big Data
On top of the statistical methods that have been demonstrated to largely reduce the computation for large data sets, the efficient usage of computing resource can further speed up the computation. I will introduce several ways to improve the computing. The improvements will largely focus on effectively utilizing R's C/C++ API with Rcpp and by applying parallelization techniques afforded to R using explicit parallelism for OpenMP, MPI, and parallel as well as implicit parallelism through a parallelized BLAS. Finally, we illustrate these techniques using two spatial data examples.

Soutir Bandyopadhyay, Lehigh University

Spatial Methods for Nonstationary Fields Using Compact Basis Functions
Kriging is a non-parametric regression method used in geostatistics for estimating curve and surfaces and it forms the core of most statistical methods for spatial data. In climate science these methods are extremely useful for estimating how climate varies over a geographic region when the observational data is sparse or the computer model runs are limited. A statistical challenge is to implement spatial methods for large sample sizes and also account for the heterogeneity in the physical fields, both common features of many geophysical problems. Equally important is to provide companion measures of uncertainty so that the estimated surfaces can be compared in an objective way and are suitable for decision making. Here we present a statistical method that expands the spatial field in a large number of basis functions of varying resolution but all with compact support. Parsimonious models for the precision matrix of the basis coefficients are able to approximate standard covariance models but also scale to large numbers of spatial locations.

Sudipto Banerjee, Professor and Chair, Department of Biostatistics, UCLA Fielding School of Public Health

On Gaussian Process Models for High-Dimensional Geostatistical Datasets
With the growing capabilities of Geographical Information Systems (GIS) and user-friendly software, statisticians today routinely encounter geographically referenced datasets containing observations from a large number of spatial locations. Over the last decade, hierarchical spatial process models have become widely deployed statistical tools for researchers to better understand the complex nature of spatial variability. However, fitting hierarchical spatial models often involves expensive matrix decompositions whose computational complexity increases in cubic order with the number of spatial locations. This renders such models infeasible for large spatial data sets. In this talk, I will present
two approaches for constructing well-defined spatial stochastic processes that accrue substantial computational savings. Both these processes can be used as “priors” for spatial random fields. The first approach constructs a low-rank process operating on a lower-dimensional subspace. The second approach constructs a Nearest-Neighbor Gaussian Process (NNGP) that can be exploited as a dimension-reducing prior embedded within a rich and flexible hierarchical modeling framework to deliver exact Bayesian inference. Both these approaches lead to Markov chain Monte Carlo algorithms with floating point operations (flops) that are linear in the number of spatial locations (per iteration). We compare these methods and demonstrate its use in inferring on the spatial distribution of forest biomass from the US Forest Inventory database spanning the continental US.

*Joint work with Abhirup Datta, Andrew O. Finley and Alan E. Gelfand.*

**Luke Bornn, Simon Fraser University**

Lost Moments: The Effect of Pre-processing on Environmental Data

Monitoring networks for environmental and meteorological data are typically vast and irregularly spaced with significant missingness across space and time. As a result, this data is often aggregated both spatially and temporally to ease analysis. In this talk, I will explore the biases introduced and information lost when using this aggregated data, and propose methods for mitigating these impacts.

**Jonathan Bradley, University of Missouri**

Efficient Parameterizations for Multiscale Multivariate Spatio-Temporal Data

Many data sources report related variables of interest that are also referenced over multiple geographic regions and time; however, there are relatively few general statistical methods that one can readily use that incorporate dependencies over different spatial locations, spatial scales, time points, and variables. Additionally, many multivariate spatio-temporal areal datasets are extremely high-dimensional, which leads to practical issues when formulating statistical models. We use the multivariate spatio-temporal mixed effects model (MSTM) in a fully Bayesian framework to analyze data of this type. Moreover, we introduce the use of Obled and Creutin eigenfunctions within this framework to allow for multivariate spatio-temporal data observed on different spatial scales. We illustrate our method through an empirically motivated simulation study based on Quarterly Workforce Indicators (QWI) published by the US Census Bureau’s Longitudinal Employer-Household Dynamics (LEHD) program. We also provide a demonstration of our approach using an environmental dataset.

**Mike Brauer, University of British Columbia**

Title: Coming soon.

Abstract: Coming soon.
Visualization with ggplot2

I will give a brief overview of some recent, exciting additions to the toolkit for statisticians and data analysts who work in R. Examples including the RStudio IDE and the dplyr package for data manipulation. Then we will go directly into a hands-on tutorial on ggplot2, a package written by Hadley Wickham, implementing the Grammar of Graphics from Leland Wilkinson.

Please bring your laptop! Make sure you are running a fairly recent version of R (R version 3.2.0 was released 2015-04-16). Install the ggplot2 and gapminder packages (run install.packages(c("ggplot2", "gapminder"))). If you haven't already, this might be a great time to install and try out the RStudio IDE (http://www.rstudio.com/products/rstudio/download/).

Peter Craigmile, The Ohio State University

Heteroscedastic Asymmetric Spatial Processes (HASP)

The Gaussian stochastic process is commonly used for modeling time series and geostatistical data. The Gaussianity assumption, however, is known to be insufficient or inappropriate in many settings. In this talk, I discuss the development of specific non-Gaussian models to capture the asymmetry and heavy tails of many real-world datasets indexed in space and time. Introducing a general framework for constructing non-Gaussian spatial processes using transformations of a latent multivariate Gaussian process, we develop a heteroscedastic asymmetric spatial process (HASP) for capturing the non-Gaussian features of environmental or climatic data, such as the heavy tails and skewness. The conditions for this non-Gaussian spatial process to be well defined are discussed at length. The properties of the HASP, especially its marginal moments and covariance structure, are established along with a Markov chain Monte Carlo (MCMC) procedure for sampling from the posterior distribution. The HASP model is applied to the study of a US nitrogen dioxide concentration dataset, demonstrating the ability of HASP to capture asymmetry and heavy tails benefits its predictive performance.

This research project is joint with Jiangyong (Matt) Yin, Ph.D., Google Inc.

Marco Ferreira, Virginia Tech

Dynamic Multiscale Spatiotemporal Models for Gaussian and Poisson Processes

We discuss classes of dynamic multiscale models for Poisson and Gaussian spatiotemporal processes. Specifically, we use multiscale spatial factorizations to decompose the process at each time point into spatiotemporal multiscale coefficients. We then connect these spatiotemporal multiscale coefficients through time with state-space-like evolutions. In the case of Gaussian processes we use a Gaussian state-space evolution, whereas in the case of Poisson
processes we use a novel Dirichlet evolution. Further, we propose simulation-based full Bayesian posterior analysis. In particular, for both Gaussian and Poisson processes we develop filtering equations for updating information forward in time and smoothing equations for integration of information backward in time, and use these equations to develop forward filter backward samplers for the spatiotemporal multiscale coefficients. Because the multiscale coefficients are conditionally independent a posteriori, our full Bayesian posterior analysis is scalable, computationally efficient, and highly parallelizable. We present results on the spatial and spatiotemporal dependence structure. Finally, we illustrate the usefulness of our multiscale spatiotemporal methodology with two applications. The first application examines mortality ratios in the state of Missouri, and the second application considers tornado reports in the American Midwest.

Dorit Hammerling, IMAGE NCAR
A New Ensemble-Based Consistency Test for the Community Earth System Model
Climate simulations codes, such as the National Center for Atmospheric Research Community Earth System Model (CESM), are especially complex and continually evolving. Their on-going state of development and installation on new systems require frequent software verification in the form of quality assurance to both preserve the quality of the code and to instill confidence in the model in situations where bit-to-bit reproducibility is not feasible. To formalize and simplify this previously subjective and computationally-expensive verification process, we have developed a new tool for evaluating climate consistency. Because an ensemble of simulations allows us to gauge the natural variability of the models climate, our new tool uses an ensemble approach for consistency testing. In particular, an ensemble of CESM climate runs is created, from which we obtain a statistical distribution that can be used to determine whether a new climate run is statistically distinguishable from the original ensemble. The CESM Ensemble Consistency Test, referred to as CESM-ECT, is objective in nature and accessible to CESM developers and users. The tool has proven its utility in detecting errors in software and hardware environments and in providing rapid feedback to model developers. We will also discuss ongoing research to extend the current tool.

This is joint work with Allison Baker and Daniel Milroy.

Mevin Hooten, Colorado State University
Fringe Benefits: The Hidden Utility of Constraints in Telemetry Studies
Advances in animal telemetry data collection techniques have been a catalyst for the creation of statistical methodology for analyzing animal movement data. While the technology for data collection is improving dramatically over time, we are left with massive archives of historical animal telemetry data that are subject to measurement error (i.e., location uncertainty). This form of error arises as a combination of factors due to accuracy of the telemetry device and system, animal behavior, atmospheric interference, and landscape features. Furthermore, the measurement error
varies with both location and time and the information available about the accuracy is not easily incorporated into statistical models and is often in flux due to ongoing manufacturer findings. Thus, there is a need for 1.) approaches to better estimate the telemetry error distribution and 2.) improved methods to incorporate it into animal movement models. Using both simulations and real data, we describe how certain forms of auxiliary information can be surprisingly useful for learning about telemetry error. We show how hierarchical models, with appropriate constraints on the data and process components, can increase the utility of common forms of telemetry data for understanding animal space use and resource selection.

**Timothy Johnson, University of Michigan**

**Analysis of Point Pattern Imaging Data using Log Gaussian Cox Processes with Spatially Varying Coefficients**

Log Gaussian Cox Processes (LGCP) are used extensively to model point pattern data. In these models, the log intensity function is modeled semi-parametrically as a linear combination of spatially varying covariates with scalar coefficients plus a Gaussian process that models the random spatial variation. Almost exclusively, the point pattern data are a single realization from the driving point process. In contrast, our motivating data are lesion locations from a cohort of Multiple Sclerosis patients with patient specific covariates measuring disease severity. Patient specific covariates enter the model as a linear combination with spatially varying coefficients. Our goal is to correlate disease severity with lesion location within the brain. Estimation of the LGCP intensity function is typically performed in the Bayesian framework using the Metropolis adjusted Langevin algorithm (MALA) and, more recently, Riemannian manifold Hamiltonian Monte Carlo (RMHMC). Due to the extremely large size of our problem | 3D data (64 x 64 x 64) on 240 subjects | we show that MALA performs poorly in terms of posterior sampling and that RMHMC is computationally intractable. As a compromise between these two extremes, we show that posterior estimation via Hamiltonian Monte Carlo performs exceptionally well in terms of speed of convergence and Markov chain mixing properties. To speed up posterior estimation, critical parts of the HMC algorithm are ported to a GPU and run in parallel. Further, we compare the statistical efficiency of the fully Bayesian approach with two deterministic approximations: variation Bayes and integrated nested-Laplacian approximation (INLA) in a simple LGCP model.

**Matthias Katzfuji, Texas A & M**

**A Multi-Resolution Approximation for Big Spatial Data**

Automated sensing instruments on satellites and aircrafts have enabled the collection of big spatial data over large and inhomogeneous spatial domains. If these kinds of datasets can be efficiently exploited, they can provide new insights on a wide variety of issues. However, traditional spatial statistical techniques such as kriging are not computationally feasible for big datasets. We propose a multi-resolution approximation (M-RA) of Gaussian processes observed at irregular (i.e., non-gridded) locations in space. The M-RA process is specified as a linear combination of
basis functions at multiple levels of spatial resolution, which can capture inhomogeneous spatial structure from very fine to very large scales. The basis functions are chosen to optimally approximate a given covariance function, and no restrictions on the covariance function are necessary. All computations involving the M-RA, including fully Bayesian parameter inference and prediction, are highly scalable for massive datasets. Crucially, the inference algorithms can also be parallelized to take full advantage of distributed computing environments.

**Will Kleiber, University of Colorado**

**Coherence for Random Fields**

Multivariate spatial field data are increasingly common and modeling typically relies on building cross-covariance functions to describe cross-process relationships. An alternative viewpoint is to model the matrix of spectral measures. We develop the notions of coherence, phase and gain for multidimensional stationary processes. Coherence, as a function of frequency, can be seen to be a measure of linear relationship between two spatial processes at that frequency band. We use the coherence function to give natural interpretations to cross-covariance parameters of the Matérn class, where the smoothness indexes dependence at low frequencies while the range parameter can imply dependence at low or high frequencies. Estimation follows from smoothed multivariate periodogram matrices. We illustrate the estimation and interpretation of these functions on two datasets, forecast and reanalysis sea level pressure and geopotential heights over the equatorial region. Examining these functions lends insight that would otherwise be difficult to detect using standard cross-covariance formulations.

**Bo Li, University of Illinois**

**Evaluating Climate Field Reconstructions in Reduced Dimension**

The large scale climate field reconstructions (CFRs) of the Common Era target hemispheric or global patterns of temperature change. They provide important spatial information that in turn can be used to infer dynamical insights about the causes of past climate variability and change. An important tool for assessing the performance of CFRs is the pseudoproxy experiments, which are controlled and systematic experiments based on millennium-length forced transient simulations with fully coupled general circulation models. All CFRs, modeled climate and real climate, are high dimension and correlated random processes. Evaluating the difference between CFRs as well as their performance relative to their target climate fields presents statistical challenges. We propose to assess the first and second moment structures and the trend of climate fields in a reduced dimension via functional data analysis approach. The evaluation is through a sequence of formal hypothesis tests that are constructed based on the functional principal components or dominant modes of variation inherent in the data. Our method well respects the properties of the climate fields and is robust by being nonparametric.
Finn Lindgren, University of Bath.

Stochastic Partial Differential Equations and Numerical Methods for Large Scale Spatial Statistics

Towards Realistic Stochastic Modeling of Global Temperatures

Abstracts: TBD

Robert Lund, Clemson University

Changepoints and Associated Climate Controversies

This talk overviews changepoint issues in climatology. Changepoints (inhomogeneities) are ubiquitous features in climatic time series, arising, for example, when stations relocate or instrumentation is changed. Changepoints confound many inference problems and are very important data features. Here, we show why changepoint information is essential in making accurate trend conclusions. Examples are given where inferences are questionable when changepoints are ignored. The talk will delve into two recent contentious climate issues: 1) the recent increase in Atlantic Basin hurricanes; and 2) the “warming hole” (lack of warming) seen in the Eastern United States.

Renjun Ma, University of New Brunswick

Spatiotemporal Analysis of Environmental Health Risk

Big data with complex spatiotemporal structures are common in environmental studies. In order to account for such spatiotemporal structures, spatially and temporally correlated random effects are often incorporated into generalized linear models for such data. The estimation of these models often poses theoretical and computational challenges. We propose an orthodox best linear unbiased predictor (BLUP) approach to these models. Our approach is illustrated with application to Ohio lung cancer data where the annual lung cancer deaths for 88 counties were obtained from 1968-1988. With estimated spatial and temporal random effects, we will also discuss the identification of high/low risk areas, spatial clustering as well as temporal trend. An adaptation of this approach to analyzing spatially correlated survival data will also be illustrated with application to American Cancer Society study of air pollution and mortality where over half million of subjects were followed every half month since 1982.

Doug Nychka, NCAR

Extremes in Regional Climate: What to do with 8000 Histograms

As attention shifts from broad global summaries of climate change to more specific regional impacts there is a need for the data sciences to quantify the uncertainty in regional predictions. A regional climate model (RCM) is a large and complex computer code based on physics that simulates the detailed flow of the atmosphere in a particular region from the large scale information of a global climate model. Part of the value of these simulations is to explore the
potential extremes in weather that are due to natural variation and also to climate change. Here we present an application that combines log-spline density estimates to discern tail behavior in a distribution with spatial methods for large data sets (LatticeKrig). This is applied to estimate return levels for daily precipitation from a subset of the North American Regional Climate Change and Assessment Program. Here the regional models comprise about 8000 grid locations over North America and so pose challenges for the statistical analysis of functional data.

Bruno Sanso, University of California at Santa Cruz

Using MARS for Functional Computer Model Emulation and Sensitivity Analysis

Sophisticated computer programs that produce realistic simulations of, for example, physical or biological systems, often require large amounts of computer power and take long time to produce results. Such programs usually depend on inputs that are subject to uncertainty. Due to the computational burden of running the code for each set of inputs, a full exploration of the impact of input uncertainty on the model output may be unfeasible. A popular alternative is to develop a surrogate model, based on statistical methods, that provides a fast to compute approximation to the computer model output. This is referred to as a statistical emulator. The most popular emulators are based on Gaussian processes (GP). GPs are very flexible but become computationally unfeasible when the number of inputs is very large. We implement, as an alternative, a Bayesian version of a multivariate adaptive regression splines model (B-MARS). We focus on a case study consisting of a deformation experiment for the protecting plate of an X-ray imager. We are interested in finding how sensitive the deformation in the plate is to configuration variables like plate width and spacer width, using simulations that output the profile of the deformed plate. We compare the results and the Sobol sensitivity indexes for functional emulators based on GPs, dynamic trees and B-MARS. We explore the potential for scalability of the B-MARS approach.

In collaboration with Devin Francom, PhD student at UCSC.

Gavin Shaddick, University of Bath

Incorporating Large Scale Exposure Modeling into Environmental Epidemiological Studies

Abstract: TBD

Andrew Trites, University of British Columbia

Abstract: TBD
Hao Zhang, Purdue University

Modeling the Complexity of Data Structure in Environmental Sciences

Data in environmental sciences are not only huge in size, but may be of different spatial and temporal scales. This coupled with the spatio-temporal correlation makes the analysis of such data a challenging task. Statistical models should capture the complex structure that exists in the data. Otherwise, contradictory and misleading results may occur and this is well documented in literature that I will review. As models become complex, so should be the statistical theory that provides a basis for comparing models and methods. I will illustrate this point through an example in ecology.

James V. Zidek and Seagle Liu, University of British Columbia

Interpolating the High Dimensional Track Record of The Fur Seal: Fusing a Physical Model With Data

Many biologging studies deploy biologgers equipped with magnetometers and accelerometers to record animal movements at infra-second frequencies, thus allowing their tracks to be to be reconstructed at high-resolution by dead reckoning (DR). But stochastic factors limit the accuracy of the DR paths. So a conventional (but ad hoc) method was developed, which uses irregularly observed GPS points and simply shifts the DR paths to pass through them. While appealing simple, the conventional method lacks the stochastic foundation that enable quantitative uncertainty statements about the true path to be made. The Bayesian melding (BM) approach provides such a foundation for melding model (the DR path) with data (the GPS measurements). However that approach is computational intensive at the best of times and here the challenges are daunting, due the high dimensional data records. Our implementation of the BM uses a Brownian Bridge process to combine the fine-resolution (but seriously biased) DR path and the sparse (but precise) GPS measurements. But several key approximations and a conditional independence property of the Brownian Bridge process were needed to make it work. A cross-validatory assessment of the method will be described and show that the BM works pretty well when applied to data obtained from northern fur seals (Callorhinus ursinus) foraging in the Bering Sea. The GPS corrected high-resolution path also revealed that the total distance traveled by the fur seals was much greater than that calculated by simply joining the dots (linear interpolation of the GPS observations)! The method has been published in a CRAN package.

Co-authors: Brian Bataille, and Andrew Trites
Recreational Activities
at the University of British Columbia

Aquatic Centre
6121 University Boulevard
(604) 822-4522  www.aquatics.ubc.ca
The UBC Aquatic Centre features a 50-metre indoor pool, seasonal 55-yard outdoor pool, whirlpool, fitness/weight room, sauna/steam rooms, seasonal patio area and diving boards from one to ten meters.

Summer Hours: Please call for swim times, lessons, etc.

Beaty Biodiversity Museum
2212 Main Mall
(604) 827-4955  www.beatymuseum.ubc.ca
A new public museum dedicated to enhancing the public’s understanding and appreciation of biodiversity. It is home to over 20,000 fossils from all over the world, including the largest blue whale exhibit in Canada.

Summer Hours: Wed- Sun: 11:00am-5:00pm

Belkin Art Gallery
1825 Main Mall (beside Fredric Wood Theatre)
(604) 822-2759  www.belkin.ubc.ca
The Morris and Helen Belkin Art Gallery’s mandate is to research, exhibit, collect, publish, educate and develop programs in the field of contemporary art and in contemporary approaches to the practice of art history and criticism.

Summer Hours: Tues-Fri: 10:00am-5:00pm
Sat-Sun: 12:00pm-5:00pm

Botanical Garden
6804 Marine Drive
(604) 822-9666  www.ubcbotanicalgarden.org
Established in 1916, the UBC Botanical Garden has an outstanding collection of temperate plants displayed according to their geographic areas. Exhibits of regional plants include the Native Garden and Alpine Garden.

Summer Hours: Daily 9:00am-5:00pm

Tennis Courts
2525 West Mall & 6010 Thunderbird Boulevard
(604) 822-2505
All guests staying at the University of British Columbia are welcome to use the tennis courts located at Place Vanier and Totem Park Residences. There are additional courts at the UBC Coast Club located at 6160 Thunderbird Blvd. Please call for information on reservations, fees and special packages.

Please Note: Rates and Schedules are subject to change.
Museum of Anthropology
6393 NW Marine Drive
(604) 822-5087 www.moa.ubc.ca
The Museum of Anthropology is one of North America’s premier museums. School programs focusing on the Northwest Coast First Nations are available. All programs encourage discussion, observation and hands-on experience with touchable objects to learn about people and cultures. School programs must be arranged in advance.

Summer Hours: Daily 10:00am-5:00pm
Tues: 10:00am-9:00pm

Nitobe Memorial Garden
1903 Lower Mall
(604) 822-9666 www.nitobe.org
Considered to be the best traditional, authentic Japanese Tea and Stroll garden in North America and among the top five Japanese gardens outside Japan, the Nitobe Garden includes a rare authentic Tea Garden with a ceremonial Tea House. The exquisite work of art was created out of two-and-a-half acres (one hectare) of pristine forest by landscape architects and gardeners recommended by the government of Japan.

Summer Hours: Daily 10:00am-5:00pm

Pacific Spirit Regional Park
Park Office 4915 West 16th Avenue
(604) 224-5739
The Pacific Spirit Regional Park encompasses 763 hectares of forest and foreshore surrounding UBC, and boasts 35 kilometres of walking trails. Experience a variety of landscapes, from estuary marshes, rock and cobble beaches, wooded ravines, ancient bog and upland forests. Regional Park Interpreters offer customized group programs on themes ranging between edible plants, birds, and bog ecology.

Student Recreation Centre
6000 Student Union Boulevard
(604) 822-6000 www.rec.ubc.ca or www.birdcoop.ubc.ca
The SRC is one of Canada’s premier University fitness facilities. It includes 1,800 square-feet of gym space, a full service fitness and weight room, a 2,300 square-foot dance studio, and a 1,600 square-foot traditional martial arts dojo.

Summer Hours: Mon-Thurs: 6:30am-9:00pm
Friday: 630am-7:00pm
Saturday: 10:00am-6:00pm
12:00pm-6:00pm

University Golf Course
5185 University Boulevard
(604) 224-1818 www.universitygolf.com
Designed to satisfy players of every level, the course features low-mowed rough and few hazards of water to carry over. Still, it does present challenges even for the experienced golfer. Greens on Par 3’s are well protected by sand and require stealth accuracy. Move back to the championship tees and put a little more distance between you and the pins. 18 holes, Par 72.

Summer Hours: First tee time: 6am
Last tee time: 8pm

Please Note: Rates and Schedules are subject to change.
Campus Dining
at the University of British Columbia

From world-class catering to casual dining, coffee shops and internationally-inspired food outlets, UBC offers a delicious assortment of food services solutions. Here is an overview of food service providers certain to deliver a satisfying campus dining experience.

UBC Food Services
www.food.ubc.ca

Serving only locally-roasted fair trade organic shade-grown coffee at all UBC Food Services non-franchise locations

Wescadia Catering
Conference and special event catering
www.catering.ubc.ca

Sage Bistro at University Centre
Casual fine dining available for breakfast, lunch and special events
www.sage.ubc.ca

The Point Grill at Marine Drive Residence
New upscale casual dining restaurant open for brunch, lunch, and dinner. Open M-F

Triple O’s at David Lam Research Centre
Casual dining in a family-friendly environment. Open daily

Residence Dining
Totem Park and Place Vanier Cafeterias
For information about group meal plans, please call 604-822-6204 or email rene.atkinson@ubc.ca

Pacific Spirit Place Cafeteria at the SUB
Student Union Building, 6138 Student Union Blvd
Pacific Spirit Place is open weekdays for breakfast and lunch. For information about group meal plans, please call 604-822-9310 or email fred.cheng@ubc.ca

Bakeshop A&W
Pasta Bar Koya Japan
Salad Bar Manchu Wok
Pizza Pizza Subway

Proudly Brewing Starbucks Coffee
Starbucks Coffee at Student Union Building
The Barn at Main Mall
Starbucks Coffee at Fred Kaiser
Steamies Café at the Bookstore
Pond Café at Ponderosa Centre

More Great Locations…

Niche Café at Beaty Biodiversity Museum
Caffé Perugia at Life Sciences Centre
Café MOA at Museum of Anthropology
Ike’s Café at Irving K. Barber Learning Centre
Tim Horton’s at Forest Sciences Centre

For guests, visitors, or groups visiting the UBC Campus, the UBC Food Services gift card is the easiest way for you and your group to dine at any of our locations.
Food Outlets
at the Student Union Building (SUB)

The SUB features a variety of food outlets all under one roof and conveniently located at the heart of campus. Get a delicious bagel or muffin to go, grab a slice of pizza at Pie R Squared, pick up some freshly made sushi or sit and enjoy a juicy beef burger at Pit Pub. The SUB has something for everyone!

Concourse and Sub-Level

Blue Chip Cookies
Proudly serving organic, fair trade coffees, cappuccinos and lattés. All our cookies and fabulous baked goods are made in-house and baked fresh daily.

Bernoulli’s Bagels
Montreal-style bagels, sandwiches, and bagel melts using high-quality ingredients and freshly squeezed vegetable or citrus juice!

The Delly
Fresh sandwiches made to order. A wide selection of salads, wraps, curries, soups and pasta made daily.

The Honour Roll
Maki rolls, nigiri, sushi, donburi rice bowls and bento boxes are made fresh throughout the day. Ask about party platters and catering.

The Pit Burger Bar
Charbroiled hamburger specials, veggie burgers, hot wings, beer-battered fish & chips and more!

The Pit Pub
Satellite big-screen sports, six high-definition TV’s, great drink prices, and a great atmosphere!

The Moon Noodle House
Great wonton soup, daily specials, fresh steamed veggies, combos and hot & sour soup.

The Patio BBQ
On the south side of the SUB, Monday to Friday (weather permitting) offering grilled 1/4 pound burgers, veggie burgers, smokies and drinks.

The Pendulum Restaurant
Delicious grilled sandwiches and paninis, and lots of vegetarian and vegan dishes!

Pie R Squared
Great house-made pizza slices, great prices, cold drinks. Now offering soft-serve ice cream and doughnuts.

www.catering.ubc.ca

NEED CATERING? For catered events or meals on the go, Wescadia Catering offers a multitude of menu ideas to meet a range of dietary needs. We pride ourselves on our knowledgeable, friendly staff, professional service and quality ingredients.
University Boulevard
Restaurants and Food Outlets

University Boulevard boasts a vibrant neighbourhood feel, and features dozens of places to enjoy a sit-down meal, people-watch over coffee, or grab a quick bite on the run. Visitors will feel right at home choosing from internationally-recognized franchises and unique offerings from local entrepreneurs.

The Boulevard Coffee Roasting Co.
at David Strang, 5870 University Blvd.
theboulevard.ca

Mahony & Sons Public House
at David Strang, 5990 University Blvd.
www.mahonyandsons.com

The Well Café
at Regent College, 5800 University Blvd.

The Well Café

University Village
5700 Block, University Blvd.

Blenz Coffee Shop
Booster Juice Juice & Snack Bar
Mio Japan Japanese Fast Food
McDonald’s Breakfast – Late-Night Fast Food
Pearl Fever Tea House & Snack Bar
Pita Pit Lunch – Late-Night Take-Out & Delivery

University Village

One More Sushi Japanese Dining
Only U Café Deli & Diner
Starbuck’s Coffee Shop
University Pizza Take-Out & Delivery
Vera’s Burger Shack Diner
Village Restaurant Chinese Dining

International Food Fair
University Marketplace, Lower Level

A-1 Vietnamese Food Pho & Noodle House
Curry Point East Indian
Donair Town Persian, Mediterranean, Catering
Leona Mediterranean Food Lebanese

International Food Fair

Malaysian Cuisine Malaysian, Thai
Osaka Sushi Japanese
Timpo Mongolian BBQ Stir-Fry
Yi Kou Xiang Chinese

Also Recommended...

Westward Ho! Public House & Grill Room at the University Golf Club
www.universitygolf.com/dine