Scotland C. Leman

CONTACT INFORMATION

Department of Statistics
Virginia Tech
401A Hutchenson Bldg
Blacksburg, VA 24061 USA

Cell: (540) 449-9860
Office: (540) 231-5441
E-mail: leman@vt.edu
WWW: www.apps.stat.vt.edu/leman

Statistician, with expert level knowledge in Bayesian methodology and stochastic based inference.

RESEARCH INTERESTS

Bayesian statistics, visual analytics, large scale stochastic simulation, statistical genetics, molecular evolution, epidemiology , coalescence theory, branching processes, time series analysis, hierarchical modeling, data augmentation, Markov chain Monte Carlo methods, mixing theory, importance sampling methodology

EDUCATION

Duke University, Durham, North Carolina USA

Ph.D., Statistics, April 2007

- Dissertation Title: "On Evolutionary Theory, Inference, And Simulation: A Genealogical Perspective "

M.S., Statistics, January 2005

Stanford University, Palo Alto, California USA

M.S., Scientific Computing and Computational Mathematics, May 2003

University of California, at Davis, Davis, California USA

B.S., Mathematics, May, 2001 (Cum Laude, Honors)

Virginia Tech, Blacksburg, Virginia USA
Assistant Professor **2007-2013**
Tenure track position in The Department of Statistics.

CURRENT PROFESSIONAL POSITION

Virginia Tech, Blacksburg, Virginia USA
Associate Professor **2013-2015**
The Department of Statistics (Tenured)
Courses taught:

- Introduction to Data Analytics and Visualization (under graduate) (2014),
- Statistical Inference: Stat 4114 (graduate) (2014, 2015),
- Bayesian statistics: Stat 5444 (graduate) (2007, 2008, 2009,2010, 2011,2012,2013,2014),
- Bayesian simulation: Stat 5304(graduate) (2008, 2009, 2010, 2011,2012,2015).
- Advanced Data Analytics: Stat 5226(graduate) (2012).
- Stochastic Modeling and Inference: Stat 6224(graduate) (2013).

REFEREED PUBLICATIONS

Hoegh A, Ferreira MAR and Leman SC; Spatiotemporal Model Fusion: Multiscale Modelling of Civil Unrest. *Journal of the Royal Statistical Society: Series C* (2016). doi:10.1111/rssc.12138 Hoegh A, Carzolio M, Crandell I, Hu X, Roberts L, Song Y, Leman SC; Nearest-neighbor matchup effects: accounting for team matchups for predicting March Madness. *Journal of Quantitative Analysis in Sports* (2015), 11(1), 29-37.

Leman S, House L, and Hoegh A; Developing a New Interdisciplinary Computational Analytics Undergraduate Program: A Qualitative-Quantitative-Qualitative Approach. *The American Statistician* (2015), 59(4):397-408.

House L, Leman SC, and Han C; Bayesian Visual Analytics (BaVA). *Journal of Statistical Analysis and Data Mining* (2015), 8(2):1-13.

Han C, Leman SC, and House L; Covariance-Guided Mixture Probabilistic Principal Component Analysis (C-MPPCA). *Journal of Computational and Graphical Statistics* (2015), 24(1):66-83.

Hoegh A, Leman SC, Saraf P, and Ramakrishnan N; Bayesian Model Fusion for Forecasting Civil Unrest. *Technometrics* (2015), 57(3), 332-340.

Awe OO, Crandell I, Adepoju AA, and Leman SC; A Time Varying Parameter State-Space Model for Analyzing Money Supply Economic Growth Nexus. *Journal of Statistical and Econometric Methods* (2015), 4(1), pp.73-95.

Ramakrishnan N, Lu C, Marathe M, Marathe A, Vullikanti A, Eubank S, Leman SC, Roan M, Brownstein J, Summers K; Model-Based Forecasting of Significant Societal Events. *Intelligent Systems*, IEEE (2015), 30:86-90

Hoegh A, Leman SC; A Spatio-Temporal Model for Assessing Winter Damage Risk to East Coast Vineyards. *Journal of Applied Statistics* (2015), 42(4):834-845.

House L, Leman SC; Han C, Bayesian visual analytics: BaVA. *Statistical Analysis and Data Mining*: The ASA Data Science Journal (2015), 8: 1-13.

Leman SC, House L, Szarka J, Nelson H; Life on the bubble: Who?s in and who?s out of March Madness? *Journal of Quantitative Analysis in Sports* (2014), 10(3): 315-328

Bradel L, North C, House L, Leman SC; Multi-Model Semantic Interaction for Text Analytics. *Visual Analytics Science and Technology* (VAST) (2014), IEEE.

Nsoesie E, Leman SC; A Dirichlet Process Model for Prediction of Epidemic Curves, *BMC Infectious Diseases* (2014), 14:12.

Marakeby H, Badr E, Torkey H, Song Y, Leman SC, Monteil CL, Heath LS, Vinatzer BA; A System to Automatically Classify and Name Any Individual Genome-Sequenced Organism Independently of Current Biological Classification and Nomenclature. *PLoS ONE* (2014), 9(2): e89142.

Monteil CL, Studholme D, Leman SC, Moris C, Vinatzer B; Non-agricultural reservoirs contribute to emergence and evolution of Pseudomonas syringae crop pathogens. *New Phytologist* (2012),199(3): 800-11

Hu X, Bradel L, Maiti D, House L, North C, Leman SC; Semantics of Directly Manipulating Spatializations, *IEEE Transactions on Visualization and Computer Graphics* (2013), 19(12): 2052-2059.

Clarke CR, Chinchilla D, Hind SR, Taguchi F, Miki R, Ichinose Y, Martin GB, Leman SC, Felix G, Vinatzer B; Allelic Variation in FLS2-Dependent Detection of Two Distinct Epitopes of Pseudomonas Syringae Flagellin by Plants. *New Phytologist* (2013), 200(3): 847-60.

Leman SC, House L, Maiti D, Endert A, North C; Visual to Parametric Interaction. *PLoS ONE* (2012), 8(3): e50474. doi:10.1371/journal.pone.0050474

Leman SC, House L; Improving Mr. Myagi's Coaching Style: Teaching Data Analytics with Interactive Data Visualizations. *Chance* (2012), 25(2): 4-12

Velasco-Cruz C, Leman SC, Smith EP; Assessing the Risk of Rising Temperature on Brook Trout: A Spatial Dynamic Linear Risk Model. *Journal of Agricultural, Biological, and Environmental Statistics (JABES)* (2012), DOI: 10.1007/s13253-012-0088-8

Cai R, Yan S, Haijie L, Leman SC, Vinatzer BA; Reconstructing Host Range Evolution of Bacterial Plant Pathogens Using Pseudomonas Syringae pv. Tomato and Its Close Relatives as a Model. *Infection, Genetics and Evolution* (2011), 11(7): 1738-51

Cai R, Lewis J, Yan S, Liu H, Clarke SR, Campanile F, Almeida SF, Studholme DJ, Lindeberg M, Schneider D, Zaccardelli M, Setubal JC, Morales-Lizcano NP, Bernal A, Coaker G, Baker C, Bender CL, Leman SC, Vinatzer BA; The Plant pathogen Pseudomonas Syringae pv. Tomato is Genetically Monomorphic and Under Strong Selection to Evade Tomato Immunity, *Public Library of Science (PLoS) Pathogens* (2011), 7(8):e1002130.

Endert A, Fox S, Maiti D, Leman SC, North C; The Semantics of Clustering: Analysis of User-Generated Spatializations of Text Documents. *Advanced Visual Interfaces 2012 (AVI2012)* (2011), 978-1-4503-1287-5

Endert A, Han C, Maiti D, House L, Leman SC, North C; Observation-level Interaction with Statistical Models for Visual Analytics. *IEEE Visual Analytics Science and Technology (VAST)* (2011), 978-1-4673-0015-5

Sharakhova MV, Xia A, Leman SC, Sharakhov IV; Arm-Specific Dynamics of Chromosome Evolution in Malaria Mosquitoes, *BioMed Central (BMC) Evolutionary Biology* (2011) 11:91.

Sharakhova MV, George P, Brusentsova IV, Leman SC, Bailey JA, Smith CD, Sharakhov IV; Genome mapping and characterization of the Anopheles gambiae heterochromatin, *BMC Genomics*, (2010), 11:459

Xia A, Sharakhova MV, Leman SC, Tu Z, Bailey JA, Smith CD, Sharakhov IV; Genome Landscape and Evolutionary Plasticity of Chromosomes in Malaria Mosquitoes, *PLoS One* (2010), 5(5): e10592

Almeida NF, Yan S, Cai R, Clarke CR, Morris CE, Schaad NW, Lacy GH, Jones JB, Castillo JA, Bull CT, Leman SC, Guttman DS, Setubal JC, Vinatzer BA; A Multilocus sequence typing & analysis database and website for plant-associated and plant-pathogenic microorganisms. *Phytopathology*, (2010), 100(3), pp. 208-215, 2010

Leman S, Levy F, Walker E; Modeling The Spread Of Infectious Disease Using Genetic Information Within A Marked Branching Process. *Statistics in Medicine*, (2009), 28(29): 3581 - 3717

Leman SC, Chen Y, Lavine M; The Multiset Sampler. *Journal of the American Statistical Association (JASA)* (2009), 104(487): 1029-1041

Levy F, Leman SC, Walker E; Nosocomial Transmission Clusters and Risk Factors in Moraxella Catarrhalis. *Epidemiology and Infection* (2009), 137(4):581-90.

Leman SC, Uyenoyama MK, Chen Y, Lavine M; The Evolutionary Forest Algorithm. *Bioinformatics*, (2007), 23(15):1962-1968,

Leman SC, Chen Y, Stajich JE, Noor, MAF, Uyenoyama MK; Likelihoods from summary statistics: A Recent Divergence Between Species. *Genetics*, (2005), 171(3):1419-1436.

Publications in Press

Han C, House L, Leman SC; Expert-Guided Generative Topographical Model?ing with Visual to Parametric Interaction. *PLoS ONE*, To Appear.

Chen X, House L, Self JZ, Leman SC, Robertson JE, Fry J, North C; Be the Data: An Exploratory Study of Embodied Experience for Learning Data Analytics. *American Educational Research Association* (AERA) Annual Meeting, To Appear.

Chen X, Self J, House L, North C; Be the Data: A New Approach to Immersive Analytics. In IEEE

Virtual Reality, *Immersive Analytics*, To Appear.

Self J, Hu X, House L, Leman S, North C; Designing Interactive Algorithms with Visual Analytics. *Human Centered Machine Learning Workshop at the Conference on Human Factors in Computing Systems* (CHI16), To Appear.

Technical Reports

Leman SC, House L, Han C, Bayesian Visual Analytics (BAVA), FODAVA Technical Report (#FODAVA-10-02).

Current funded research

UrbComp: Data Science for Modeling, Understanding, and Advancing Urban Populations, National Science Foundation (NSF), NRT-DTSE, 1545362.
Award Amount, $2,999,128.00
PIs: Naren Ramakrishnan, Leanna House, Layne Watson, Mark Embree, Scotland Leman

Human Machine Collaboration for Data Discovery, General Dynamics
Award Amount, $250,000
PIs: Chris North, Scotland Leman, Nicholas Polys, Naren Ramakrishnan

Data Analytics for Large Sensor Systems, Office of Naval Research (ONR).
Award Amount, $260,752.00
PIs: Eric Smith, William Devenport, Leanna House, Scotland Leman, William Alexander

Usable Big-Data Analytics via Multi-scale Visual To Parametric Interaction, National Science Foundation (NSF), III, 1447416 .
Award Amount, $998,914.00
PIs: Chris North, Scotland Leman, Leanna House, Yong Cao

Be The Data, Institute for Creativity, Arts, and Technology (Virginia Tech).
Award Amount, $30,000.00
PIs: Leanna House, Scotland Leman, Chris North

Early Model Based Event Detection Using Surrogates, Intelligence Advanced Research Projects Activity (IARPA), OSI (4th year extension).
Award Amount, $3,000,000.00
PIs: Naren Ramakrishnan, Scotland Leman, Tien Chang Lu, Madhav Marathe, Michael Roan

Early Model Based Event Detection Using Surrogates, Intelligence Advanced Research Projects Activity (IARPA), OSI.
Award Amount, $14,494,495.00
PIs: Naren Ramakrishnan, Scotland Leman, Tien Chang Lu, Madhav Marathe, Michael Roan

Critical Thinking with Data Visualization, National Science Foundation (NSF), TUES, 1141096.
Award Amount $199,997
CoPIs: Chris North, Leanna House

Previous funded research

Data Analytics for Large Acoustic Array Data, Office of Naval Research (ONR).
Award Amount, $44,163.00
PIs: Eric Smith, William Devenport, Leanna House, Scotland Leman, William Alexander

First Year Experiences for Statistics Majors, Virginia Tech Quality Enhancement Plan (QEP)
Award Amount: $32,000 CoPIs: Jane Robertson, Leanna House, Chris Franck, Eric Vance, Eric Smith

Bayesian Analysis in Visual Analytics, National Science Foundation (NSF), CCF, 0937071.
Award Amount: $499,307 (PI)
CoPIs: Chris North, Leanna House

User-Guided Spatialization for Visualizing NSF Award Portfolios. NSF 10222504.
Award Amount: $24,900 (PI)
CoPIs: Chris North, Leanna House

Recent Presentations

Multi-scale Visual to Parametric Interaction, Paris, France, VAST 2014.

Visual To Parametric Interaction, Montreal, QC, Joint Statistics Meetings, Aug. 2013.

Semantics of Directly Manipulating Spatializations, Altalnta, GA, Visual Analytics Science and Technology, Oct. 2013.

Challenges and Advances in High Dimensional and High Complexity Monte Carlo (Banff, Alberta, Canada): Particle Filtered Tempering: De-Correlating the Ladder, Scotland Leman (Mar. 2012).

Foundations of Data and Visual Analytics Workshop (Atlanta, GA): Bayesian Visual Analytics, (Dec. 2011).

Case Studies in Bayesian Statistics and Machine Learning (Pittsburgh, PA): MultiSet Model Selection, Presentation (Nov. 2011).

Joint Statistical Meetings Invited Speaker, Session on Sports Statistics (Miami Beach, FL.): March Madness: Who's In and Who's Out (2011).

International Society for Bayesian Analysis presenter, Bayesian Visual Analytics (BaVA), Valencia, Spain, June, 2010.

Joint Statistical Meetings Session Organizer and Contributor (Washington D.C.): Bayesian Visual Analytics (2010).

Joint Statistical Meetings Session Organizer and Contributor (Washington D.C.): Bayesian Philosophies and Practicalities (2009).

International Society For Bayesian Analysis (Hamilton Island, AU): The Multiset Sampler (2008).

Bayesian Case Studies, Carnegie Mellon, Pittsburgh, PA: Speciation in Coalescence Models (2008)

Statistical and Applied Mathematical Sciences Institute (Durham, NC): The Multiset Auxiliary Variable Sampler (2008).

National Dynamics Simulation Science Laboratory (Blacksburg, VA): Modeling The Spread Of Infectious Disease Using Genetic Information Within A Marked Branching Process (2008).

International Society For Bayesian Analysis (Hamilton Island, Australia): The Random Forest Algorithm (2006).

Joint Statistical Meetings (2004-2007)

Invited Session Organizer

Roundtable discussion leader at the Joint Statistical Meetings: "Don't Forget: Quantitative Statistics Courses are Very Qualitative.", Seattle, WA, August, 2015.

JSM organizer (invited session), The I.J. Good Memorial session: "Remembering a Genius", Vancouver, BC, August, 2010. Speakers: Jim Berger, Adrian Raftery, Steve Fienberg; Discussant: David Banks; Chairman: Golde Holtzman

JSM organizer, Bayesian Philosophies and Practicalities, Washington, D.C., August, 2009.

Previous Professional Positions

IBM, San Jose, California USA
Developer/Tester/Technical Author **2003**
Worked with a team to develop and test current software in the web mining field.

- Wrote two published books describing the use of developed software and tools. These books were the core sources of information for the client to follow while integrating the software into their systems. These provided detailed methodology and examples through mock company scenarios.
- Developed and tested software critical for end user consumption of the products.

University of Chicago, Medical Center, Chicago, Illinois USA
Researcher/Developer/Consultant **2001-2003**
Responsible for developing a software interface to access experimental data from a large relational database and transport and process it into a statistical analysis tool.

- Developed Graphical User Interface in Java to access data and port it to existing software.
- Redesigned portions of the database to aid in efficient queries. Ultimately enhanced query times

by a factor of 1000.
- Analyzed data and performed inference.
- Algorithmic development for use in simulation and inference procedures.

Institute of Theoretical Dynamics, Davis, California USA
Research Assistant/Developer/Analyst **1999-2001**
Worked closely with scientists and mathematicians. In charge of analyzing simulation routines and results. Performed fluid flow experiments critical to answering fundamental scientific questions.
- Simulation development for three dimensional Navier-Stokes equations.
- Three dimensional mesh reconstruction used for simulating physical bodies.
- Analysis of simulations and data.

Previous Teaching Experience

National Bioinformatics Network, Cape Town, South Africa
Instructor **September, 2006**
Taught graduate level students an intensive course on the topics of: generalized linear models, survival analysis, multi-dimensional scaling, cluster analysis, principal component analysis and data visualization.
- Class size of 25 students.
- Emphasis was placed on applications related to bioinformatics and genetics.
- A mixture of theory and software was used to demonstrate concepts.

The Duke School, Durham, North Carolina
Instructor **September, 2005 - May, 2006**
Taught an after school class on *algorithmic thought and computer programming* to middle school students.
- Class sizes of approximately 15 students.
- Object oriented computer programming in C++.
- Debugging emphasized using the integrated development environment Xcode.

Duke University, Durham, North Carolina
Teaching Assistant **August, 2003 - present**
Duties at various times have included office hours, consulting and leading weekly computer lab exercises.

Honors and Awards

Mu Sigma Rho, National Honor Society, 2008

Phi Kappa Phi, National Honor Society, 2000

Pi Mu Epsilon, National Mathematics Honor Society, 2000

Golden Key, National Honor Society, 2000

Research Training Program (RTP) graduate, UC Davis, 2000

Graduated Cum Laude, Honors in Mathematics, 2001

Mu Sigma Rho, Statistics Honor Society, 2007

Professional Affiliations

- International Society for Bayesian Analysis (ISBA).
- The American Statistical Association (ASA).

Media Coverage

- WSLS, Channel 11 News, Jan 14, 2015, "Former Virginia Tech professor portrayed in WWII film, The Imitation Game" by Dawn Jeffries;
  http://www.wsls.com/story/27849889/former-virginia-tech-professor-portrayed-in-wwii-film-the-imitation-game
- The Oklahoman, Feb 27, 2015, "NCAA Tournament: Evidence says name brand programs will get the selection committee's nod every time", by Barry Tramel;
  http://newsok.com/ncaa-tournament-evidence-says-name-brand-programs-will-get-the-selection-committees-nod-every-time/article/5397173
- The Daily Bruin, Mar 17, 2015, "Q&A: Statistics analyst talks prestige bias in March Madness", by Andrew Erickson;
  http://dailybruin.com/2015/03/17/qa-statistics-analyst-talks-prestige-bias-in-march-madness
- The American Statistical Association News, Feb 24, 2015, "A ?MARQUEE? BIAS CAN INFLUENCE WHICH ?BUBBLE? TEAMS GET INTO MARCH MADNESS", by Jeffrey Meyers
  http://www.amstat.org/newsroom/pressreleases/2015-MarqueeFactorHelpsTeamsGetintoMarchMadness
- Sports Are 80 Percent Mental, Mar 3, 2011, "Is There Bias In Selection Of March Madness Teams?" by Dan Peterson;
  http://blog.80percentmental.com/2011/03/is-there-bias-in-selection-of-march.html
- fromtheeditr, March 2, 2011, "The Bias Behind 'Bracketology'; A Study" by Dan Smith;
  http://fromtheeditr.blogspot.com/2011/03/bias-behind-bracketology-study.html
- WSLS, Channel 10, Roanoke, VA, March 2, 2011 "Stats professors: Virginia Tech up against odds for NCAA tourney bid" by Ken Heineck
- WDBJ, Channel 7, Roanoke, VA, March 1, 2011 "Calculating VT's odds of making the NCAA Tournament Could biases have affected VT getting a bid in 2010?"
- Virginia Tech News, March 1, 2011, "March Madness: Statisticians quantify entry biases," by Catherine Doss

  Also appeared:

  - MSN Fox Sports,
    msn.foxsports.com/collegebasketball/story/MARCH-MADNESS-STATISTICIANS-QUANTIFY-ENTRY-BIASES-69543943
  - Science Daily,
    www.sciencedaily.com/releases/2011/03/110301111259.htm
  - Science Blog,
    scienceblog.com/43241/march-madness-statisticians-quantify-entry-biases/
  - Science Newsline Medicine,
    www.sciencenewsline.com/medicine/2011030112000067.html
  - Red Orbit,
    www.redorbit.com/news/sports/2005304/march_madness_statisticians_quantify_entry_biases/
- Tech Talk Live, February 28, 2011, Mentioned by Coach Seth Greenberg during an interview with Bill Roth
- College of Science Magazine, Virginia Tech, September 13, 2010, "Statisticians help researchers see their data in a new way," by Catherine Doss
- College of Science Magazine Video, Virginia Tech;
  http://www.science.vt.edu/media/statistics-house-leman-video.html
- The Washington Post, March 17, 2010, "Hokies turn to statistic for answers to NCAA snub," by Mark Viera; http://voices.washingtonpost.com/hokies-journal/2010/03/hokies_turn_to_statistic_for_a.html#more

- Richmond Times Dispatch, Mar 16, 2010, "Virginia Tech stats professors doing NCAA tournament study for Greenberg," by Darryl Slater; http://www.mytimesdispatch.com/index.php/sports/comments/virginia_tech_stats_professors_doing_ncaa_tournament_study_for_greenberg/11450/

Graduate Students

Graduated Ph.D. Students: Ciro Velasco-Cruz (joint with Eric Smith, 2012),Chao Han (joint with Leanna House, 2012), Dipayan Maiti (2012), Lucas Roberts (2014), Xinran Hu (2014)

Ph.D. Advisor to: Yuhyun Song, Ian Crandell, Marcos Carzollio, Andy Hoegh, JT Fry, Matt Slifko

Advised Master Students: Ho Cho(2012), Han Li (2012), Elaine Nsoesie (2011), Hayley Nelson (2012)

Outside Departmental Ph.D. Advisor to: Elaine Nsoesie (NDSSL), Rongman Cai (Biology).

Technical Skills

- Software: Matlab, R, CodeWarrior, Office.
- Languages: C/C++, Java, Python, Perl, Fortran, Matlab, R, SQL, some use of Unix shell scripts.
- Technical Skills: Probability and statistics. Ability to develop and deploy Bayesian models.
- Operating Systems: Macintosh, *nix Flavors, Windows.