This annotated reading list is not intended to be comprehensive. Rather it gives a few key papers for the main topics of the course. Further articles will be suggested for project work.

Design and Analysis of Computer Experiments

D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, pp. 455–492, 1998. Derek recommends you read the first part of the paper as a tutorial on the intuition of using a Gaussian process to represent a deterministic function.

J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments (with discussion)," *Statistical Science*, vol. 4, pp. 409–435, 1989. Treating a function as arising from a stochastic process or Gaussian process was already known in statistics, geostatistics (kriging), and optimization. But this paper introduced that formulation for deterministic computer experiments and made it feasible for applications with many input variables.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker, "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments," *Journal of the American Statistical Association*, vol. 86, no. 416, pp. 953–963, 1991. Toby Mitchell, Max Morris and Don Ylvisaker were pioneers in bringing computer experiments to the attention of statisticians. Mitchell and Morris were based at Oak Ridge National Laboratory, where simulations were run on the supercomputers of the day.

T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. New York: Springer, 2003. A broad coverage of the literature on design and analysis of computer experiments up to the time of publication.

C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press, 2006. A comprehensive account of Gaussian processes from a computer-science perspective. Its coverage is broader than computer experiments, including classification for instance.

Design of Computer Experiments

L. Pronzato and W. G. Müller, "Design of computer experiments: Space filling and beyond," *Statistics and Computing*, vol. 22, pp. 681–701, 2012. This paper provides an overview of many of the space filling methods for design of computer experiments.

M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979. These authors introduced Latin hypercubes sampling (design) for computer experiments. The original purpose was different but Latin hypercubes are now widely used for analysis via Gaussian processes.

J. Sacks, S. B. Schiller, and W. J. Welch, "Designs for computer experiments," *Technometrics*, vol. 31, pp. 41–47, 1989. This paper predates the Sacks, Welch, Mitchell, and Wynn *Statistical Science* paper by a few months and had the main ideas of using a stochastic process for analysis. But its main contribution is now viewed as introducing model-based design of computer experiments, specifically to minimize integrated mean squared error of prediction.

M. E. Johnson, L. M. Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, vol. 26, no. 2, pp. 131–148, 1990. This paper introduced "space-filling" designs via two distance-based criteria. Excellent intuition is provided via analogies to everyday problems. The maximin criterion is widely used in applications, often to choose within the class of Latin hypercube designs.

B. Tang, "Orthogonal array-based latin hypercubes," 1993. An orthogonal array is a fractional factorial with some balance properties. SFU's Boxin Tang used these designs from traditional design of physical experiments to improve the space-filling properties of Latin hypercubes in low-dimensional projections, typically all subsets of 2 or 3 inputs).

K. T. Fang, D. K. J. Lin, P. Winker, and Y. Zhang, "Uniform design: Theory and application," *Technometrics*, vol. 42, no. 3, pp. 237–248, 2000. The authors briefly review the vast literature on uniform designs and illustrate use of these designs with a computer code of a launching system.

J. L. Loeppky, J. Sacks, and W. J. Welch, "Choosing the sample size of a computer experiment: A practical guide," *Technometrics*, vol. 51, pp. 366–376, 2009. The authors argue that the accuracy of a GP emulator is affected by two summaries of the correlation sensitivity parameters and that $n = 10d$ runs will often be enough for moderate accuracy or diagnose that accuracy cannot be achieved without a much larger sample size.

Sensitivity Analysis and Visualization

M. Schonlau and W. J. Welch, "Screening the input variables to a computer model via analysis of variance and visualization," in *Screening Methods for Experimenta-*

tion in Industry, Drug Discovery and Genetics (A. M. Dean and S. M. Lewis, eds.), pp. 308–327, New York: Springer-Verlag, 2006. Sensitivity analysis and visualization are achieved together via analysis of variance applied to the prediction function. The methods had been used in papers for more than 10 years by Sacks, Welch, and coworkers and were part of Matt Schonlau's thesis. But this review lays out all the mathematics and how to do the computations.

J. E. Oakley and A. O'Hagan, "Probabilistic sensitivity analysis of complex models: a Bayesian approach," *Journal of the Royal Statistical Society B*, vol. 66, pp. 751–769, 2004. A fully Bayesian treatment of the above with some review of other definitions of sensitivity by Sobol, Saltelli, etc.

Sequential Design

D. Bingham, P. Ranjan, and W. J. Welch, "Design of computer experiments for optimization, estimation of function contours, and related objectives," in *Statistics in Action: A Canadian Outlook* (J. F. Lawless, ed.), pp. 109–124, Boca Raton, Florida: CRC Press, 2014. An overview of sequential design, largely based on the next two papers, that appeared as a chapter in a book by the Statistical Society of Canada to celebrate the International Year of Statistics. The chapter is available at `http://www.ssc.ca/webfm_send/1322`

D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, pp. 455–492, 1998. The paper defines an expected improvement (EI) criterion to choose the next function evaluation when searching for a global optimum. The method has been widely adopted in the optimization literature, engineering design, etc.

P. Ranjan, D. Bingham, and G. Michailidis, "Sequential experiment design for contour estimation from complex computer codes," *Technometrics*, vol. 50, no. 4, pp. 527–541, 2008. The authors develop another EI criterion, this time for mapping out where $y(\mathbf{x})$ equals some pre-specified critical value. Many papers by other researchers have followed this work, to find quantiles, percentiles, etc. of an output distribution.

Calibration and Validation (Assessment)

M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models (with discussion)," *Journal of the Royal Statistical Society B*, vol. 63, pp. 425–464, 2001. This seminal paper uses jointly models computer-model and physical data via GP models for calibration of unknown parameters in the presence of systematic discrepancy between the two types of data.

D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne, "Combining field data and computer simulations for calibration and prediction," *SIAM Journal on Scientific Computing*, vol. 26, no. 2, pp. 448–466, 2004. The authors give a readable account of how to conduct a Bayesian analysis of of the Kennedy-O'Hagan formulation, along with two detailed engineering examples.

D. Higdon, J. Gattiker, B. Williams, and M. Rightley, "Computer model calibration using high-dimensional output," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 570–583, 2008. Again the objective is calibration of unknown parameters in the presence of discrepancy between the computer-model runs and physical data. The authors tackle multivariate data, which are reduced in dimensionality via principal components. The principal component weights are then modelled by GPs.

M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu, "A framework for validation of computer models," *Technometrics*, vol. 49, no. 2, pp. 138–154, 2007. These authors point out that, as all computer models are wrong to some extent, "validation" of a computer code against physical data amounts to an assessment of the magnitude of the discrepancy.

W. Kleiber, S. Sain, M. J. Heaton, M. Wiltberger, C. S. Reese, and D. Bingham, "Parameter tuning for a multi-fidelity dynamical model of the magnetosphere," *Annals of Applied Statistics*, vol. 7, no. 3, pp. 1286–1310, 2013. The paper extends calibration (or tuning) to multivariate output from a space-time field. The authors also allow several versions of the computer model and use sequential design to improve calibration.