

References

- [1] Leonardo S. Bastos and Anthony O’Hagan. Diagnostics for Gaussian Process Emulators. *Technometrics*, 51(4):425–438, nov 2009.

Abstract: Mathematical models, usually implemented in computer programs known as simulators, are widely used in all areas of science and technology to represent complex real-world phenomena. Simulators are often so complex that they take appreciable amounts of computer time or other resources to run. In this context, a methodology has been developed based on building a statistical representation of the simulator, known as an emulator. The principal approach to building emulators uses Gaussian processes. This work presents some diagnostics to validate and assess the adequacy of a Gaussian process emulator as surrogate for the simulator. These diagnostics are based on comparisons between simulator outputs and Gaussian process emulator outputs for some test data, known as validation data, defined by a sample of simulator runs not used to build the emulator. Our diagnostics take care to account for correlation between the validation data. To illustrate a validation procedure, we apply these diagnostics to two different...

Notes: r2018-03-20 Gaussian process emulators, their advantages and disadvantages, are discussed along with diagnostics to assess successful emulation. The paper discusses both numerical and graphical diagnostics. There is helpful discussion about how to identify what specific problems might be causing poor diagnostic results. Some of the techniques discussed in the paper include the use of standardized prediction errors (from a validation set), the Mahalanobis distance between the emulator and simulator outputs, and variance decompositions. Early in the paper there is a very helpful bit of background/foundation of Gaussian process emulation, including guidance on how to proceed when the observation variance is unknown and must be estimated (and the resulting Student process distribution).

- [2] M. J. Bayarri, J. O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R. J. Parthasarathy, R. Paulo, J. Sacks, and D. Walsh. Computer Model Validation with Functional Output. *The Annals of Statistics*, 35:1874–1906, 2007.

Abstract: A key question in evaluation of computer models is Does the computer model adequately represent reality? A six-step process for computer model validation is set out in Bayarri et al. [Technometrics 49 (2007) 138-154] (and briefly summarized below), based on comparison of computer model runs with field data of the process being modeled. The methodology is particularly suited to treating the major issues associated with the validation process: quantifying multiple sources of error and uncertainty in computer models; combining multiple sources of information; and being able to adapt to different, but related scenarios. Two complications that frequently arise in practice are the need to deal with highly irregular functional data and the need to acknowledge and incorporate uncertainty in the inputs. We develop methodology to deal with both complications. A key part of the approach utilizes a wavelet representation of the functional data, applies a hierarchical version of the scalar validation methodology to the wavelet coefficients, and transforms back, to ultimately compare computer model output with field output. The generality of the methodology is only limited by the capability of a combination of computational tools and the appropriateness of decompositions of the sort (wavelets) employed here. The methods and analyses we present are illustrated with a test bed dynamic stress analysis for a particular engineering system.

Notes: r2017-07-12; 2017-02-03 Six steps for their technique: 1. define problem 2. establish evaluation criteria 3. design experiments 4. approximate computer model output 5. analyze the combo of field and computer run data 6. feed back to revise the model, perform additional experiments, and so on. Current paper attempts to deal with 3 new problems: 1. irregular function output 2. uncertainty in inputs 3. prediction in altered/new settings The authors exhort the use of "tolerance bounds" as measures of model accuracy. One reason is that tolerance bounds account for model bias. See Bayarri et al [4] for details. Overview: sec. 2

describes example problem. Sec. 3 formulates the statistical problem and assumptions. Sec. 4 has methods of analysis. Sec. 5, results.

- [3] Maria J Bayarri, James O Berger, Rui Paulo, Jerry Sacks, John A Cafeo, James Cavendish, Chin-Hsu Lin, and Jian Tu. A Framework for Validation of Computer Models. *Technometrics*, 49(2):138–154, may 2007.

Abstract: We present a framework that enables computer model evaluation oriented toward answering the question: Does the computer model adequately represent reality? The proposed validation framework is a six-step procedure based on Bayesian and likelihood methodology. The Bayesian methodology is particularly well suited to treating the major issues associated with the validation process: quantifying multiple sources of error and uncertainty in computer models, combining multiple sources of information, and updating validation assessments as new information is acquired. Moreover, it allows inferential statements to be made about predictive error associated with model predictions in untested situations. The framework is implemented in a test bed example of resistance spot welding, to provide context for each of the six steps in the proposed validation process.

Notes: r2018-01-10; 2017-07-11 The authors advance the method of using "modularization" which Berger and Bayarri focus on more thoroughly in their 2009 paper with Liu. The paper includes useful citations on dealing with high-dimensional output data (p. 141). The authors note (p.145) that a non-constant mean function may be called for if you want extrapolation from your GP model. There is some discussion of the trade-offs between using MLEs for hyperparameters or using a full Bayesian analysis; the authors tentatively weigh in for the former option. The authors assert (p. 150) that they have shown that, contrary to common opinion, data can be used simultaneously for calibration/tuning and for model validation, by "incorporating the posterior distribution of the tuning parameters in the overall assessment of uncertainties."

- [4] Marc H. Bornstein and Helen G. Bornstein. The pace of life. *Nature*, 259(5544):557–559, feb 1976.

Abstract: The pace of life

- [5] D. Andrew Brown and Sez Atamturktur. Nonparametric Functional Calibration of Computer Models. *Statistica Sinica*, 28:721–742, feb 2018.

Abstract: Standard methods in computer model calibration treat the calibration parameters as constant throughout the domain of control inputs. In many applications, systematic variation may cause the best values for the calibration parameters to change between different settings. When not accounted for in the code, this variation can make the computer model inadequate. In this article, we propose a framework for modeling the calibration parameters as functions of the control inputs to account for a computer model's incomplete system representation in this regard while simultaneously allowing for possible constraints imposed by prior expert opinion. We demonstrate how inappropriate modeling assumptions can mislead a researcher into thinking a calibrated model is in need of an empirical discrepancy term when it is only needed to allow for a functional dependence of the calibration parameters on the inputs. We apply our approach to plastic deformation of a visco-plastic self-consistent material in which the critical resolved shear stress is known to vary with temperature.

Notes: r2018-01-25; 2017-05-11 The notion of state-aware calibration is introduced and discussed. There are several useful tips for computation, especially on pages 6 and 7.

- [6] Jenni Brynjarsdóttir and Anthony O'Hagan. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11), 2014.

Abstract: Science-based simulation models are widely used to predict the behavior of complex physical systems. It is also common to use observations of the physical system to solve the inverse problem, i.e. to learn about the values of parameters within the model, a process often called calibration. The main goal of calibration is usually to improve the predictive performance of the simulator but the values of the parameters in the model may also be of intrinsic scientific interest in their own right. In order to make appropriate use of observations of the physical system it is important to recognise model discrepancy, the difference between reality and the simulator output. We illustrate through a simple example that an analysis that does not account for model discrepancy may lead to biased and over-confident parameter estimates and predictions. The challenge with incorporating model discrepancy in statistical inverse problems is the confounding with calibration parameters, which will only be resolved with meaningful priors. For our simple example, we model the model-discrepancy via a Gaussian Process and demonstrate that by accounting for model discrepancy our prediction within the range of data is correct. However, only with realistic priors on the model discrepancy do we uncover the true parameter values. Through theoretical arguments we show that these findings are typical of the general problem of learning about physical parameters and the underlying physical system using science-based mechanistic models.

Notes: r2018-01-31; 2017-09-08 Looks like in order to take the lesson of this paper to heart, one would have to be really thoughtful about the physical ground of the discrepancy function, in crafting as narrow constraints as possible. No black boxes here.

- [7] Siddhartha Chib and Edward Greenberg. Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49(4):327, nov 1995.

Abstract: We provide a detailed, introductory exposition of the Metropolis-Hastings algorithm, a powerful Markov chain method to simulate multivariate distributions. A simple, intuitive derivation of this method is given along with guidance on implementation. Also discussed are two applications of the algorithm, one for implementing acceptance-rejection sampling when a blanketing function is not available and the other for implementing the algorithm with block-at-a-time scans. In the latter situation, many different algorithms, including the Gibbs sampler, are shown to be special cases of the Metropolis-Hastings algorithm. The methods are illustrated with examples.

Notes: r2018-01 Useful tutorial on the Metropolis-Hastings algorithm which also very clearly lays out its theoretical underpinnings. Its relationship to the Accept/Reject method is described clearly.

- [8] Carla Currin, Toby Mitchell, Max Morris, and Don Ylvisaker. Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments. *Journal of the American Statistical Association*, 86(416):953–963, dec 1991.

Notes: r2018-03-09 The article discusses the use of Gaussian process emulators of computer code, with emphasis on design and on selection of a correlation function. Some interesting points about correlation functions are made, such as how the appropriate choice of correlation function yield prediction functions that are either linear or cubic splines in each dimension. The article includes some nice tables comparing the results using various correlation functions, compared also against some polynomial models fitted by least squares.

- [9] Andrew Gelman. Iterative and Non-Iterative Simulation Algorithms. *Computing Science and Statistics (Interface Proceedings)*, 24:433–438, 1992.

Abstract: The Gibbs sampler, Metropolis’ algorithm, and similar iterative simulation methods are related to rejection sampling and importance sampling, two methods which have been

traditionally thought of as non-iterative. We explore connections between importance sampling, iter-ative simulation, and importance-weighted resampling (SIR), and present new algorithms that combine aspects of importance sampling, Metropolis' algorithm, and the Gibbs sampler.

- [10] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis*. CRC Press, London, 3rd edition, 2013.

Abstract: Third edition. "Preface This book is intended to have three roles and to serve three associated audiences: an introductory text on Bayesian inference starting from first principles, a graduate text on effective current approaches to Bayesian modeling and computation in statistics and related fields, and a handbook of Bayesian methods in applied statistics for general users of and researchers in applied statistics. Although introductory in its early sections, the book is definitely not elementary in the sense of a first text in statistics. The mathematics used in our book is basic probability and statistics, elementary calculus, and linear algebra. A review of probability notation is given in Chapter 1 along with a more detailed list of topics assumed to have been studied. The practical orientation of the book means that the reader's previous experience in probability, statistics, and linear algebra should ideally have included strong computational components. To write an introductory text alone would leave many readers with only a taste of the conceptual elements but no guidance for venturing into genuine practical applications, beyond those where Bayesian methods agree essentially with standard non-Bayesian analyses. On the other hand, we feel it would be a mistake to present the advanced methods without first introducing the basic concepts from our data-analytic perspective. Furthermore, due to the nature of applied statistics, a text on current Bayesian methodology would be incomplete without a variety of worked examples drawn from real applications. To avoid cluttering the main narrative, there are bibliographic notes at the end of each chapter and references at the end of the book"— Part I: Fundamentals of Bayesian inference. Probability and inference – Single-parameter models – Introduction to multiparameter models – Asymptotics and connections to non-Bayesian approaches – Hierarchical models – Part II: Fundamentals of Bayesian data analysis. Model checking – Evaluating, comparing, and expanding models – Modeling accounting for data collection – Decision analysis – Part III: Advanced computation. Introduction to Bayesian computation – Basics of Markov chain simulation – Computationally efficient Markov chain simulation – Modal and distributional approximations – Part IV: Regression models. Introduction to regression models – Hierarchical linear models – Generalized linear models – Models for robust inference – Models for missing data – Part V: Nonlinear and nonparametric models. Parametric nonlinear models – Basis function models – Gaussian process models – Finite mixture models – Dirichlet process models – A. Standard probability distributions – B. Outline of proofs of limit theorems – Computation in R and Stan.

Notes: Comprehensive reference resource for Bayesian data analysis methodologies.

- [11] Stuart Geman and Donald Geman. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

Abstract: We make an analogy between images and statistical mechanics systems. Pixel gray levels and the presence and orientation of edges are viewed as states of atoms or molecules in a lattice-like physical system. The assignment of an energy function in the physical system determines its Gibbs distribution. Because of the Gibbs distribution, Markov random field (MRF) equivalence, this assignment also determines an MRF image model. The energy function is a more convenient and natural mechanism for embodying picture attributes than are the local characteristics of the MRF. For a range of degradation mechanisms, including blurring, nonlinear deformations, and multiplicative or additive noise, the posterior distribution is an MRF with a structure akin to the image model. By the analogy, the posterior distribution defines another (imaginary) physical system. Gradual temperature reduction in the physical system isolates low energy states (annealing), or what is the same thing, the most probable

states under the Gibbs distribution. The analogous operation under the posterior distribution yields the maximum a posteriori (MAP) estimate of the image given the degraded observations. The result is a highly parallel relaxation algorithm for MAP estimation. We establish convergence properties of the algorithm and we experiment with some simple pictures, for which good restorations are obtained at low signal-to-noise ratios.

- [12] Michael Goldstein and Jonathan Rougier. Reified Bayesian modelling and inference for physical systems. *Journal of Statistical Planning and Inference*, 139(3):1221–1239, mar 2009.

Abstract: We describe an approach, termed reified analysis, for linking the behaviour of mathematical models with inferences about the physical systems which the models represent. We describe the logical basis for the approach, based on coherent assessment of the implications of deficiencies in the mathematical model. We show how the statistical analysis may be carried out by specifying stochastic relationships between the model that we have, improved versions of the model that we might construct, and the system itself. We illustrate our approach with an example concerning the potential shutdown of the Thermohaline circulation in the Atlantic Ocean.

Notes: To read

- [13] Robert B Gramacy and Herbert K. H Lee. Bayesian Treed Gaussian Process Models With an Application to Computer Modeling. *Journal of the American Statistical Association*, 103(483):1119–1130, sep 2008.

Abstract: Motivated by a computer experiment for the design of a rocket booster, this article explores nonstationary modeling methodologies that couple stationary Gaussian processes with treed partitioning. Partitioning is a simple but effective method for dealing with nonstationarity. The methodological developments and statistical computing details that make this approach efficient are described in detail. In addition to providing an analysis of the rocket booster simulator, we show that our approach is effective in other arenas as well.

Notes: To read

- [14] Heikki Haario, Eero Saksman, and Johanna Tamminen. Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, 20(2):265–273, jun 2005.

Notes: To read

- [15] Gang Han, Thomas J. Santner, William I. Notz, and Donald L. Bartel. Prediction for Computer Experiments Having Quantitative and Qualitative Input Variables. *Technometrics*, 51(3):278–288, aug 2009.

Abstract: This article introduces a Bayesian methodology for the prediction for computer experiments having quantitative and qualitative inputs. The proposed model is a hierarchical Bayesian model with conditional Gaussian stochastic process components. For each of the qualitative inputs, our model assumes that the outputs corresponding to different levels of the qualitative input have similar functional behavior in the quantitative inputs. The predictive accuracy of this method is compared with the predictive accuracies of alternative proposals in examples. The method is illustrated in a biomechanical engineering application.

Notes: To read

- [16] Ofir Harari, Derek Bingham, Angela Dean, David Higdon, Corresponding Author, and Dave Higdon. Computer Experiments: Prediction Accuracy, Sample Size and Model Complexity Revisited. *Statistica Sinica*, 2017.

Abstract: We revisit the problem of determining the sample size for a Gaussian process emulator and provide a data analytic tool for exact sample size calculations that goes beyond the $n = 10d$ rule of thumb and is based on an IMSPE-related criterion. This allows us to tie sample size and prediction accuracy to the anticipated roughness of the simulated data, and to propose an experimental process for computer experiments, with extension to a robust scheme.

Notes: r2017-07-10 Who would have thought the $n=10d$ rule would turn out to be an oversimplification? Here's an IMPSE-based alternative, along with a good takedown of the $10d$ rule.

- [17] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, apr 1970.

Abstract: A generalization of the sampling method introduced by Metropolis et al. (1953) is presented along with an exposition of the relevant theory, techniques of application and methods and difficulties of assessing the error in Monte Carlo estimates. Examples of the methods, including the generation of random orthogonal matrices and potential applications of the methods to numerical problems arising in statistics, are discussed.

- [18] Matthew J. Heaton, Abhirup Datta, Andrew Finley, Reinhard Furrer, Rajarshi Guhaniyogi, Florian Gerber, Robert B. Gramacy, Dorit Hammerling, Matthias Katzfuss, Finn Lindgren, Douglas W. Nychka, Furong Sun, and Andrew Zammit-Mangion. Methods for Analyzing Large Spatial Data: A Review and Comparison. *unpublished*, oct 2017.

Abstract: The Gaussian process is an indispensable tool for spatial data analysts. The onset of the "big data" era, however, has lead to the traditional Gaussian process being computationally infeasible for modern spatial data. As such, various alternatives to the full Gaussian process that are more amenable to handling big spatial data have been proposed. These modern methods often exploit low rank structures and/or multi-core and multi-threaded computing environments to facilitate computation. This study provides, first, an introductory overview of several methods for analyzing large spatial data. Second, this study describes the results of a predictive competition among the described methods as implemented by different groups with strong expertise in the methodology. Specifically, each research group was provided with two training datasets (one simulated and one observed) along with a set of prediction locations. Each group then wrote their own implementation of their method to produce predictions at the given location and each which was subsequently run on a common computing environment. The methods were then compared in terms of various predictive diagnostics. Supplementary materials regarding implementation details of the methods and code are available for this article online.

Notes: r2018-01-09 Very nice overview of methods for analyzing large spatial data, both established and new innovations, and good comparison of them with code available. The article is not so helpful for understanding these methods, though the references here are a valuable collection toward that end.

- [19] Dave Higdon, James Gattiker, Brian Williams, and Maria Rightley. Computer Model Calibration Using High-Dimensional Output. *Journal of the American Statistical Association*, 103(482):570–583, jun 2008.

Abstract: This work focuses on combining observations from field experiments with detailed computer simulations of a physical process to carry out statistical inference. Of particular interest here is determining uncertainty in resulting predictions. This typically involves calibration of parameters in the computer simulator as well as accounting for inadequate physics in the simulator. The problem is complicated by the fact that simulation code is sufficiently demanding that only a limited number of simulations can be carried out. We consider applications in characterizing material properties for which the field data and the simulator output

are highly multivariate. For example, the experimental data and simulation output may be an image or may describe the shape of a physical object. We make use of the basic framework of Kennedy and O’Hagan. However, the size and multivariate nature of the data lead to computational challenges in implementing the framework. To overcome these challenges, we make use of basis repre...

Notes: r2017-03-30 The authors’ approach is based on Kennedy and O’Hagan 2001; latter use GP models. The current authors extend it to allow for highly multivariate output, keeping computational costs low enough for MCMC to be workable. High dimensionality for them still means less than 100. More than that is ”beyond the scope of the approach given here”. Although K&O’H’s approach was not ”fully Bayesian”, the current authors give it a ”fully Bayesian” description, which better matches their own extension.

- [20] Dave Higdon, Marc Kennedy, James C. Cavendish, John A. Cafo, and Robert D. Ryne. Combining Field Data and Computer Simulations for Calibration and Prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466, jan 2004.

Abstract: We develop a statistical approach for characterizing uncertainty in predictions that are made with the aid of a computer simulation model. Typically, the computer simulation code models a physical system and requires a set of inputs—some known and specified, others unknown. A limited amount of field data from the true physical system is available to inform us about the unknown inputs and also to inform us about the uncertainty that is associated with a simulation-based prediction. The approach given here allows for the following: uncertainty regarding model inputs (i.e., calibration); accounting for uncertainty due to limitations on the number of simulations that can be carried out; discrepancy between the simulation code and the actual physical system; uncertainty in the observation process that yields the actual field data on the true physical system. The resulting analysis yields predictions and their associated uncertainties while accounting for multiple sources of uncertainty. We use a Bayesian form...

Notes: r2018-01-15 Found this extremely helpful. There’s a lot of overlap with the 2007 flyer plate paper, but here it is, in my opinion, explained more clearly and directly. After reading this, the flyer plate paper makes more sense to me. Anyway, it is mostly helpful in just laying out clearly the basics of the kind of analysis they are doing in the 2007 paper, and the kind we are doing in NSF-DEMS.

- [21] V. Roshan Joseph and Huan Yan. Engineering-Driven Statistical Adjustment and Calibration. *Technometrics*, 57(2):257–267, apr 2015.

Abstract: Engineering model development involves several simplifying assumptions for the purpose of mathematical tractability, which are often not realistic in practice. This leads to discrepancies in the model predictions. A commonly used statistical approach to overcome this problem is to build a statistical model for the discrepancies between the engineering model and observed data. In contrast, an engineering approach would be to find the causes of discrepancy and fix the engineering model using first principles. However, the engineering approach is time consuming, whereas the statistical approach is fast. The drawback of the statistical approach is that it treats the engineering model as a black box and therefore, the statistically adjusted models lack physical interpretability. This article proposes a new framework for model calibration and statistical adjustment. It tries to open up the black box using simple main effects analysis and graphical plots and introduces statistical models inside the engineering m...

Notes: To read

- [22] Georgios Karagiannis, Bledar A. Konomi, and Guang Lin. On the Bayesian calibration of expensive computer models with input dependent parameters. *Spatial Statistics*, aug 2017.

Abstract: Computer models, aiming at simulating a complex real system, are often calibrated in the light of data to improve performance. Standard calibration methods assume that the optimal values of calibration parameters are invariant to the model inputs. In several real world applications where models involve complex parametrizations whose optimal values depend on the model inputs, such an assumption can be too restrictive and may lead to misleading results. We propose a fully Bayesian methodology that produces input dependent optimal values for the calibration parameters, as well as it characterizes the associated uncertainties via posterior distributions. Central to methodology is the idea of formulating the calibration parameter as a step function whose uncertain structure is modeled properly via a binary treed process. Our method is particularly suitable to address problems where the computer model requires the selection of a sub-model from a set of competing ones, but the choice of the best' sub-model may change with the input values. The method produces a selection probability for each sub-model given the input. We propose suitable reversible jump operations to facilitate the challenging computations. We assess the performance of our method against benchmark examples, and use it to analyze a real world application with a large-scale climate model.

Notes: 2018-01-30 Similar to state-aware, but this seems to focus on discrete input-dependent parameters.

- [23] Marc C. Kennedy, Clive W. Anderson, Stefano Conti, and Anthony O'Hagan. Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, 91(10-11):1301–1309, oct 2006.

Abstract: In this paper we present a number of recent applications in which an emulator of a computer code is created using a Gaussian process model. Tools are then applied to the emulator to perform sensitivity analysis and uncertainty analysis. Sensitivity analysis is used both as an aid to model improvement and as a guide to how much the output uncertainty might be reduced by learning about specific inputs. Uncertainty analysis allows us to reflect output uncertainty due to unknown input parameters, when the finished code is used for prediction. The computer codes themselves are currently being developed within the UK Centre for Terrestrial Carbon Dynamics.

- [24] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, aug 2001.

Notes: r2018-01-11; 2016-05-30 Very thorough and detailed development of the foundation of Bayesian calibration of computer models. Useful walkthrough of that foundation. 2018-04-16 The authors assert that theirs is the first attempt to model all the sources of uncertainty in calibration of computer models. This despite the fact that theirs is not a full Bayesian analysis; they use MLEs for some hyperparameters. The authors are explicit in considering calibration to be a matter of finding “best-fitting” values for calibration parameters; that is, calibration is for them hopelessly confounded with minimizing model inadequacy. (For the same reason, they even suggest performing model calibration on parameters whose true physical value is actually known.) In the examples considered by the authors, they integrate by crude quadrature (rather than MCMC).

- [25] F. Liu, M. J. Bayarri, and J. O. Berger. Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119–150, mar 2009.

Notes: r2018-01-09 “Modularizing”, ie keeping separate the distinct components of a Bayesian model, can be a useful way to fudge the analysis to keep poor components of the model from ruining the whole thing. The paper is particularly helpful in the treatment of discrepancy in a model. Useful references are given on pages 123 and 127; the latter gives four papers responsible for the “introduction” of Gaussian process response-surface methodology (GASP) for constructing emulators.

- [26] F. Liu, M.J. Bayarri, J.O. Berger, R. Paulo, and J. Sacks. A Bayesian analysis of the thermal challenge problem. *Computer Methods in Applied Mechanics and Engineering*, 197(29-32):2457–2466, may 2008.

Abstract: A major question for the application of computer models is Does the computer model adequately represent reality? Viewing the computer models as a potentially biased representation of reality, Bayarri et al. [M. Bayarri, J. Berger, R. Paulo, J. Sacks, J. Cafeo, J. Cavendish, C. Lin, J. Tu, A framework for validation of computer models, *Technometrics* 49 (2) (2007) 138154] develop the simulator assessment and validation engine (SAVE) method as a general framework for answering this question. In this paper, we apply the SAVE method to the challenge problem which involves a thermal computer model designed for certain devices. We develop a statement of confidence that the devices modeled can be applied in intended situations.

Notes: This paper is essentially a case study in applying the methodology they promulgate in their 2007 paper "A framework for validation of computer models". They call this model the SAVE method: simulator assessment and validation engine.

- [27] Jason L. Loeppky, Derek Bingham, and William J. Welch. Computer Model Calibration or Tuning in Practice. *Technometrics*, 2006.

Abstract: Computer models to simulate physical phenomena are now widely available in engineering and science. Before relying on a computer model, a natural first step is often to compare its output with physical or field data, to assess whether the computer model reliably represents the real world. Field data, when available, can also be used to calibrate or tune unknown parameters in the computer model. Calibration is particularly problematic in the presence of systematic discrepancies between the computer model and field observations. We introduce a likelihood alternative to previous Bayesian methodology for estimation of calibration or tuning parameters. In an important special case, we show that maximum likelihood estimation will asymptotically find values of the calibration parameter that give an unbiased computer model, if such a model exists. However, the calibration parameters are not necessarily estimable. We also show in some settings that calibration or tuning need to take into account the end-use prediction strategy. Depending on the strategy, the choice of the parameter values may be critical or unimportant.

Notes: r2018-04-06 Very clearly written. The authors show that MLE calibration is asymptotically unbiased – not in the sense that the calibration parameter estimate is unbiased, but in the sense that the calibration parameter estimate provides that the discrepancy function is 0; also this asymptotically the case in that it holds if you can draw indefinitely many times from the simulation.

- [28] M. D. McKay, R. J. Beckman, and W. J. Conover. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, may 1979.

Abstract: Two types of sampling plans are examined as alternatives to simple random sampling in Monte Carlo studies. These plans are shown to be improvements over simple random sampling with respect to variance for a class of estimators which includes the sample mean and the empirical distribution function.

Notes: Latin hypercube sampling is introduced and discussed.

- [29] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, jun 1953.

Abstract: A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the twodimensional rigidsphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a fourterm virial coefficient expansion.

- [30] T Mori and K Tanaka. Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metallurgica*, 21(5):571–574, may 1973.

Abstract: Having noted an important role of image stress in work hardening of dispersion hardened materials, (1,3) the present paper discusses a method of calculating the average internal stress in the matrix of a material containing inclusions with transformation strain. It is shown that the average stress in the matrix is uniform throughout the material and independent of the position of the domain where the average treatment is carried out. It is also shown that the actual stress in the matrix is the average stress plus the locally fluctuating stress, the average of which vanishes in the matrix. Average elastic energy is also considered by taking into account the effects of the interaction among the inclusions and of the presence of the free boundary.

Notes: The Mori-Tanaka model is introduced and discussed.

- [31] Thomas Muehlenstaedt, Jana Fruth, and Olivier Roustant. Computer experiments with functional inputs and scalar outputs by a norm-based approach. *Statistics and Computing*, 27(4):1083–1097, jul 2017.

Notes: r2017-07-10 The authors propose an appropriate distance metric and demonstrate its use in creating LHC designs for experiments with functional input. I wonder how much its utility depends on the degree of knowledge we have about the shape of the functional input.

- [32] Radford M Neal. Regression and Classification Using Gaussian Process Priors. In J M Bernardo, J O Berger, A P Dawid, and A F M Smith, editors, *Bayesian Statistics 6*, volume 6, pages 475–501. Oxford University Press, New York, 1998.

Abstract: Gaussian processes are a natural way of specifying prior distributions over functions of one or more input variables. When such a function defines the mean response in a regression model with Gaussian errors, inference can be done using matrix computations, which are feasible for datasets of up to about a thousand cases. The covariance function of the Gaussian process can be given a hierarchical prior, which allows the model to discover high-level properties of the data, such as which inputs are relevant to predicting the response. Inference for these covariance hyperparameters can be done using Markov chain sampling. Classification models can be defined using Gaussian processes for underlying latent values, which can also be sampled within the Markov chain. Gaussian processes are in my view the simplest and most obvious way of defining flexible Bayesian regression and classification models, but despite some past usage, they appear to have been rather neglected as a general-purpose technique. This may be partly due to a confusion between the properties of the function being modeled and the properties of the best predictor for this unknown function. In this paper, I hope to persuade you that Gaussian processes are a fruitful way of defining prior distributions for flexible regression and classification models in which the regression or class probability functions are not limited to simple parametric forms. The basic idea goes back many years in a regression context, but is nevertheless not widely appreciated. The use of general Gaussian process models for classification is more recent, and to my knowledge the work presented here is the first that implements an exact Bayesian approach. One attraction of Gaussian processes is the variety of covariance functions one can choose from, which lead to functions with different degrees of smoothness, or different sorts of additive structure. I will describe some of these possibilities, while also noting the limitations of Gaussian processes.

I then discuss computations for Gaussian process models, starting with the basic matrix operations involved. For classification models, one must integrate over the "latent values" underlying each case. I show how this can be done using Markov chain Monte Carlo methods. In a full-fledged Bayesian treatment, one must also integrate over the posterior distribution for the hyperparameters of the covariance function, which can also be done using Markov chain sampling. I show how this all works for a synthetic three-way classification problem. The methods I describe are implemented in software that is available from my web page, at <http://www.cs.utoronto.ca/radford/>. Gaussian processes and related methods have been used in various contexts for many years. Despite this past usage, and despite the fundamental simplicity of the idea, Gaussian process models appear to have been little appreciated by most Bayesians. I speculate that this could be partly due to a confusion between the properties one expects of the true function being modeled and those of the best predictor for this unknown function. Clarity in this respect is particularly important when thinking of flexible models such as those based on Gaussian processes.

Notes: r2018-04-05 The paper is a pretty basic introduction to the use of Gaussian processes for regression and classification. There's a nice example near the beginning building the concept of GP regression from a case of linear regression. Correspondingly, there is some discussion of covariate matrices producing linear estimates.

- [33] Jeremy E. Oakley and Anthony O'Hagan. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769, aug 2004.

Notes: To read

- [34] A. O'Hagan. Curve Fitting and Optimal Design for Prediction, 1978.

Abstract: The optimal design problem is tackled in the framework of a new model and new objectives. A regression model is proposed in which the regression function is permitted to take any form over the space X of independent variables. The design objective is based on fitting a simplified function for prediction. The approach is Bayesian throughout. The new designs are more robust than conventional ones. They also avoid the need to limit artificially design points to a predetermined subset of X . New solutions are also offered for the problems of smoothing, curve fitting and the selection of regressor variables.

Notes: r2018-03-16; 2017-04-20 O'Hagan walks the reader through the use of Gaussian process models for smoothing, curve fitting, and prediction. Much time is spent on design; it is the focus of the work. However the fundamental mathematics of GP models is laid out probably as clearly and thoroughly as it is possible to do. This includes the formulae for multivariate GP models, as well as GP models with uninformative prior means.

- [35] Anthony O'Hagan. Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering and System Safety*, 91:1290–1300, 2006.

Abstract: The Bayesian approach to quantifying, analysing and reducing uncertainty in the application of complex process models is attracting increasing attention amongst users of such models. The range and power of the Bayesian methods is growing and there is already a sizeable literature on these methods. However, most of it is in specialist statistical journals. The purpose of this tutorial is to introduce the more general reader to the Bayesian approach.

Notes: r2018-01-11 Helpful tutorial, though very entry-level. But I found his discussion of sensitivity analysis helpful, and also his list of ongoing areas of research related to Bayesian analysis of computer codes.

- [36] Rui Paulo, Gonzalo García-Donato, and Jesús Palomo. Calibration of computer models with multivariate output. *Computational Statistics and Data Analysis*, 56:3959–3974, 2012.

Abstract: The problem of calibrating computer models that produce multivariate output is considered, with a particular emphasis on the situation where the model is computationally demanding. The proposed methodology builds on Gaussian process-based response-surface approximations to each of the components of the output of the computer model to produce an emulator of the multivariate output. This emulator is then combined in a statistical model involving field observations, which is then used to produce calibration strategies for the parameters of the computer model. The results of applying this methodology to a simulated example and to a real application are presented.

Notes: To read

- [37] Peter Z G Qian, C F Jeff, and Wu H Milton Stewart. Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments. *Technometrics*, 50(2):192–204, 2008.

Abstract: Standard practice when analyzing data from different types of experiments is to treat data from each type separately. By borrowing strength across multiple sources, an integrated analysis can produce better results. Careful adjustments must be made to incorporate the systematic differences among various experiments. Toward this end, some Bayesian hierarchical Gaussian process models are proposed. The heterogeneity among different sources is accounted for by performing flexible location and scale adjustments. The approach tends to produce prediction closer to that from the high-accuracy experiment. The Bayesian computations are aided by the use of Markov chain Monte Carlo and sample average approximation algorithms. The proposed method is illustrated with two examples, one with detailed and approximate finite elements simulations for mechanical material design and the other with physical and computer experiments for modeling a food processor.

Notes: r2017-08-22 Clearly stated model for integrating high and low data. Nice conditional distributions are given. The authors explore many tweaks to make the model suit various situations.

- [38] Peter Z. G Qian, Huaqing Wu, and C. F. Jeff Wu. Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors. *Technometrics*, 50(3):383–396, aug 2008.

Abstract: Modeling experiments with qualitative and quantitative factors is an important issue in computer modeling. We propose a framework for building Gaussian process models that incorporate both types of factors. The key to the development of these new models is an approach for constructing correlation functions with qualitative and quantitative factors. An iterative estimation procedure is developed for the proposed models. Modern optimization techniques are used in the estimation to ensure the validity of the constructed correlation functions. The proposed method is illustrated with an example involving a known function and a real example for modeling the thermal distribution of a data center.

Notes: r2017-07-13 The authors describe and provide theoretical grounding for a framework that allows one to use qualitative input to determine which model to apply to the quantitative input.

- [39] Brian J Reich, Eric Kalendra, Curtis B Storlie, Howard D Bondell, and Montserrat Fuentes. Variable selection for high dimensional Bayesian density estimation: application to human exposure simulation. *Journal of the Royal Statistical Society*, 61(1):47–66, 2011.

Abstract: Numerous studies have linked ambient air pollution and adverse health outcomes. Many studies of this nature relate outdoor pollution levels measured at a few monitoring stations with health outcomes. Recently, computational methods have been developed to model the distribution of personal exposures, rather than ambient concentration, and then relate the exposure distribution to the health outcome. Although these methods show great promise, they are limited by the computational demands of the exposure model. We propose a method

to alleviate these computational burdens with the eventual goal of implementing a national study of the health effects of air pollution exposure. Our approach is to develop a statistical emulator for the exposure model, i.e. we use Bayesian density estimation to predict the conditional exposure distribution as a function of several variables, such as temperature, human activity and physical characteristics of the pollutant. This poses a challenging statistical problem because there are many predictors of the exposure distribution and density estimation is notoriously difficult in high dimensions. To overcome this challenge, we use stochastic search variable selection to identify a subset of the variables that have more than just additive effects on the mean of the exposure distribution. We apply our method to emulate an ozone exposure model in Philadelphia.

- [40] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–423, 1989.

Abstract: Many scientific phenomena are now investigated by complex computer models or codes. A computer experiment is a number of runs of the code with various inputs. A feature of many computer experiments is that the output is deterministic - rerunning the code with the same inputs gives identical observations. Often, the codes are computationally expensive to run, and a common objective of an experiment is to fit a cheaper predictor of the output to the data. Our approach is to model the deterministic output as the realization of a stochastic process, thereby providing a statistical basis for designing experiments (choosing the inputs) for efficient prediction. With this model, estimates of uncertainty of predictions are also available. Recent work in this area is reviewed, a number of applications are discussed, and we demonstrate our methodology with an example.

Notes: r2018-03-13 Foundational with respect to much of the work done in computer model calibration. The case is made that despite their deterministic nature, computer models are a proper subject of statistical study, for which we may perform uncertainty analysis. The authors walk through the process of building a Gaussian process emulator of computer code for this purpose. Various covariance functions are discussed. The authors fit covariance function parameters through MLE. Roughly half the paper is devoted to questions of design.

- [41] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*. Springer, New York, 2003.

Abstract: The computer has become an increasingly popular tool for exploring the relationship between a measured response and factors thought to affect the response. In many cases, the basis of a computer model is a mathematical theory that implicitly relates the response to the factors. A computer model becomes possible given suitable numerical methods for accurately solving the mathematical system and appropriate computer hardware and software to implement the numerical methods. For example, in many engineering applications, the relationship is described by a dynamical system and the numerical method is a finite element code. The resulting computer "simulator" can generate the response corresponding to any given set of values of the factors. This allows one to use the code to conduct a "computer experiment" to explore the relationship between the response and the factors. In some cases, computer experimentation is feasible when a properly designed physical experiment (the gold standard for establishing cause and effect) is impossible; the number of input variables may be too large to consider performing a physical experiment, or power studies may show it is economically prohibitive to run an experiment on the scale required to answer a given research question. This book describes methods for designing and analyzing experiments that are conducted using a computer code rather than a physical experiment. It discusses how to select the values of the factors at which to run the code (the design of the computer experiment) in light of the research objectives of the experimenter. It also provides techniques for analyzing the resulting data so as to achieve these research goals. It illustrates these methods with code that is available to the reader at the companion web site for the book. Thomas Santner has been a professor in the Department of Statistics at The Ohio State University since 1990.

At Ohio State, he has served as department Chair and Director of the department's Statistical Consulting Service. Previously, he was a professor in the School of Operations Research and Industrial Engineering at Cornell University. He is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics, and is an elected ordinary member of the International Statistical Institute. He visited Ludwig Maximilians Universitt in Munich, Germany on a Fulbright Scholarship in 1996-97. Brian Williams has been an Associate Statistician at the RAND Corporation since 2000. His research interests include experimental design, computer experiments, Bayesian inference, spatial statistics and statistical computing. He holds a Ph. D. in statistics from The Ohio State University. William Notz is a professor in the Department of Statistics at The Ohio State University. At Ohio State, he has served as acting department chair, associate dean of the College of Mathematical and Physical Sciences, and as director of the department's Statistical Consulting Service. He has also served as Editor of the journal *Technometrics* and is a Fellow of the American Statistical Association. Physical Experiments and Computer Experiments – Basic Elements of Computer Experiments – Analyzing Output from Computer Experiments-Predicting Output from Training Data – Space Filling Designs for Computer Experiments – Criteria Based Designs for Computer Experiments – Other Issues.

Notes: r2017-07 Ch. 1: Gives some examples of computer models, and discusses the relevant differences between computer and physical experiments. For example, replication is ordinarily not an issue for computer experiments, as it is in physical experiments. Ch. 2: Includes a basic introduction to Gaussian process modeling. Ch. 3: Discusses basics of making predictions using Gaussian process models: BLUPs, MSPE, etc. Ch. 4: Continuation of chapter 3. Focuses on predictive distributions, rather than point estimates. Ch. 5: Discusses latin hypercube sampling, stratified sampling, uniform designs, and other space-filling designs. Ch. 6: Discusses entropy-based designs, MSPE-based designs, and other optimizing designs. Ch. 7: Discusses various methods of sensitivity analysis of computer models.

- [42] Terrance Savitsky, Marina Vannucci, and Naijun Sha. Variable Selection for Nonparametric Gaussian Process Priors: Models and Computational Strategies. *Statistical Science*, 26(1):130–149, 2011.

Abstract: This paper presents a unified treatment of Gaussian process models that extends to data from the exponential dispersion family and to survival data. Our specific interest is in the analysis of data sets with predictors that have an a priori unknown form of possibly nonlinear associations to the response. The modeling approach we describe incorporates Gaussian processes in a generalized linear model framework to obtain a class of nonparametric regression models where the covariance matrix depends on the predictors. We consider, in particular, continuous, categorical and count responses. We also look into models that account for survival outcomes. We explore alternative covariance formulations for the Gaussian process prior and demonstrate the flexibility of the construction. Next, we focus on the important problem of selecting variables from the set of possible predictors and describe a general framework that employs mixture priors. We compare alternative MCMC strategies for posterior inference and achieve a computationally efficient and practical approach. We demonstrate performances on simulated and benchmark data sets.

Notes: r2018-03-22 This paper provides a thorough and helpful discussion of Gaussian process regression before coming to focus specifically on the problem of variable selection in a Gaussian process regression. This is valuable because it allows one to undertake variable selection when one has little notion of what might be the parametric form of the relationship of the potential covariates to the response. The selection is accomplished via spike-and-slab mixture priors. The authors discuss two distinct MCMC schemas for carrying out the variable selection, comparing them with each other and with other methods.

- [43] Lifeng Shang and Antoni B Chan. On Approximate Inference for Generalized Gaussian Process Models. *Technical Report -City University of Hong Kong*, 2013.

Abstract: A generalized Gaussian process model (GGPM) is a unifying framework that encompasses many existing Gaussian process (GP) models, such as GP regression, classification, and counting. In the GGPM framework, the observation likelihood of the GP model is itself parameterized using the exponential family distribution (EFD). In this paper, we consider efficient algorithms for approximate inference on GGPMs using the general form of the EFD. A particular GP model and its associated inference algorithms can then be formed by changing the parameters of the EFD, thus greatly simplifying its creation for task-specific output domains. We demonstrate the efficacy of this framework by creating several new GP models for regressing to non-negative reals and to real intervals. We also consider a closed-form Taylor approximation for efficient inference on GGPMs, and elaborate on its connections with other model-specific heuristic closed-form approximations. Finally, we present a comprehensive set of experiments to compare approximate inference algorithms on a wide variety of GGPMs.

Notes: 2018-03-28 It's not clear to me what sort of value the work in this paper has. It's nice to provide an umbrella generalization of existing methods, but what does that get us, exactly? The paper is both very big and very difficult to read and I very much lament the authors' decision not to include illustrative examples of simple cases of GP regression/classification seen through the lens of their umbrella methodology.

- [44] Agnieszka Sorokowska, Piotr Sorokowski, Peter Hilpert, Katarzyna Cantarero, Tomasz Frackowiak, Khodabakhsh Ahmadi, Ahmad M. Alghraibeh, Richmond Aryeetey, Anna Bertoni, Karim Bettache, Sheyla Blumen, Marta Błażejewska, Tiago Bortolini, Marina Butovskaya, Felipe Nalon Castro, Hakan Cetinkaya, Diana Cunha, Daniel David, Oana A. David, Fahd A. Dileym, Alejandra del Carmen Domínguez Espinosa, Silvia Donato, Daria Dronova, Seda Dural, Jitka Fialová, Maryanne Fisher, Evrim Gulbetekin, Aslhan Hamamcolu Akkaya, Ivana Hromatko, Raffaella Iafrate, Mariana Iesyp, Bawo James, Jelena Jaranovic, Feng Jiang, Charles Obadiah Kimamo, Grete Kjellvik, Frat Koç, Amos Laar, Fívia de Araújo Lopes, Guillermo Macbeth, Nicole M. Marcano, Rocio Martinez, Norbert Mesko, Natalya Molodovskaya, Khadijeh Moradi, Zahrasadat Motahari, Alexandra Mühlhauser, Jean Carlos Natividade, Joseph Ntayi, Elisabeth Oberzaucher, Oluyinka Ojedokun, Mohd Sofian Bin Omar-Fauzee, Ike E. Onyishi, Anna Paluszak, Alda Portugal, Eugenia Razumiejczyk, Anu Realo, Ana Paula Relvas, Maria Rivas, Muhammad Rizwan, Svjetlana Salkičević, Ivan Sarmány-Schuller, Susanne Schmehl, Oksana Senyk, Charlotte Sinding, Eftychia Stamkou, Stanislava Stoyanova, Denisa Šukolová, Nina Sutresna, Meri Tadinac, Andero Teras, Edna Lúcia Tinoco Ponciano, Ritu Tripathi, Nachiketa Tripathi, Mamta Tripathi, Olja Uhryn, Maria Emília Yamamoto, Gyesook Yoo, and John D. Pierce. Preferred Interpersonal Distances: A Global Comparison. *Journal of Cross-Cultural Psychology*, 48(4):577–592, may 2017.

Abstract: Human spatial behavior has been the focus of hundreds of previous research studies. However, the conclusions and generalizability of previous studies on interpersonal distance preferences were limited by some important methodological and sampling issues. The objective of the present study was to compare preferred interpersonal distances across the world and to overcome the problems observed in previous studies. We present an extensive analysis of interpersonal distances over a large data set ($N = 8,943$ participants from 42 countries). We attempted to relate the preferred social, personal, and intimate distances observed in each country to a set of individual characteristics of the participants, and some attributes of their cultures. Our study indicates that individual characteristics (age and gender) influence interpersonal space preferences and that some variation in results can be explained by temperature in a given region. We also present objective values of preferred interpersonal distances in differe...

- [45] Curtis B. Storlie, William A. Lane, Emily M. Ryan, James R. Gattiker, and David M. Higdon. Calibration of Computational Models With Categorical Parameters and Correlated Outputs via Bayesian Smoothing Spline ANOVA. *Journal of the American Statistical Association*, 110(509):68–82, jan 2015.

Abstract: It has become commonplace to use complex computer models to predict outcomes in regions where data do not exist. Typically these models need to be calibrated and validated

using some experimental data, which often consists of multiple correlated outcomes. In addition, some of the model parameters may be categorical in nature, such as a pointer variable to alternate models (or submodels) for some of the physics of the system. Here, we present a general approach for calibration in such situations where an emulator of the computationally demanding models and a discrepancy term from the model to reality are represented within a Bayesian smoothing spline (BSS) ANOVA framework. The BSS-ANOVA framework has several advantages over the traditional Gaussian process, including ease of handling categorical inputs and correlated outputs, and improved computational efficiency. Finally, this framework is then applied to the problem that motivated its design; a calibration of a computational fluid dynamics (CFD) model of...

Notes: r2017-07-13 This work sees itself as an advance on the work of Qian et al 2008 and Zhou et al 2011 in that it focuses specifically on calibration (rather than sensitivity analysis or uncertainty analysis). The authors support the use of "Bayesian Smoothing Spline ANOVA", which may offer some advantages in computational efficiency. But the authors also, for comparison, extend the approach of Higdon 2008 for use with qualitative input.

- [46] Curtis B. Storlie, Brian J. Reich, Jon C. Helton, Laura P. Swiler, and Cedric J. Sallaberry. Analysis of computationally demanding models with continuous and categorical inputs. *Reliability Engineering & System Safety*, 113:30–41, may 2013.

Notes: To read

- [47] Ambler Thompson and Barry N. Taylor. Guide for the Use of the International System of Units (SI). *Special Publication (NIST SP) - 811*, 2008.
- [48] Peter A Thompson and Eric W Marchant. A Computer Model for the Evacuation of Large Building Populations. *Fire Safety Journal*, 24:131–148, 1995.

Abstract: The computer model 'SIMULEX' is designed to simulate the escape movement of thousands of individual people through large, geometri-cally complex building spaces. The model is intended for use both as a research and design tool to analyse the evacuation of large populations through a wide range of building environments. The computer program assigns a variety of attributes to each individual in the building population. These attributes include a co-ordinate position, angle of orientation, and a walking speed for each person. Specific algorithms that facilitate the simulation of escape movement include distance mapping, wayfinding, overtaking, route deviation, and adjustments to individual speeds due to the proximity of crowd members. These algorithms contribute to a computer package that displays the building plan and the position and progress of individual building occupants as they walk towards, and through the exits.

- [49] Rui Tuo. Adjustments to Computer Models via Projected Kernel Calibration. *unpublished*, may 2017.

Abstract: Identification of model parameters in computer simulations is an important topic in computer experiments. We propose a new method, called the projected kernel calibration method, to estimate these model parameters. The proposed method is proven to be asymptotic normal and semi-parametric efficient. As a frequentist method, the proposed method is as efficient as the L_2 calibration method proposed by Tuo and Wu [Ann. Statist. 43 (2015) 2331-2352]. On the other hand, the proposed method has a natural Bayesian version, which the L_2 method does not have. This Bayesian version allows users to calculate the credible region of the calibration parameters without using a large sample approximation. We also show that, the inconsistency problem of the calibration method proposed by Kennedy and O'Hagan [J. R. Stat. Soc. Ser. B. Stat. Methodol. 63 (2001) 425-464] can be rectified by a simple modification of the kernel matrix.

Notes: To read

- [50] Brian Williams, Dave Higdon, Jim Gattiker, Leslie Moore, Michael McKay, and Sallie Keller-McNulty. Combining experimental data and computer simulations, with an application to flyer plate experiments. *Bayesian Analysis*, 1(4):765–792, dec 2006.

Abstract: A flyer plate experiment involves forcing a plane shock wave through stationary test samples of material and measuring the free surface velocity of the target as a function of time. These experiments are conducted to learn about the behavior of materials subjected to high strain rate environments. Computer simulations of flyer plate experiments are conducted with a two-dimensional hydro- dynamic code developed under the Advanced Strategic Computing (ASC) program at Los Alamos National Laboratory. This code incorporates physical models that contain parameters having uncertain values. The objectives of the analyses presented in this paper are to assess the sensitivity of free surface velocity to variations in the uncertain inputs, to constrain the values of these inputs to be consistent with experiment, and to predict free surface velocity based on the constrained inputs. We implement a Bayesian approach that combines detailed physics simulations with experimental data for the desired statistical inference (Kennedy and O’Hagan 2001; Higdon, Kennedy, Cavendish, Cafeo, and Ryne 2004).

Notes: r2018-01-12; 2017-11-01 Section 2.3 was especially helpful for building a GP emulator of a computationally expensive emulator. The rest of the paper deals with related topics such as using the emulator for sensitivity analysis. I just wonder if the sort of approach they use here is going to be appropriate for us, given how many observations we have. They use only 20 simulation observations, we use 504 times three simulator outputs.

- [51] Raymond K. W. Wong, Curtis B. Storlie, and Thomas C. M. Lee. A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):635–648, mar 2017.

Notes: To read

- [52] Yulei Zhang and William I. Notz. Computer Experiments with Qualitative and Quantitative Variables: A Review and Reexamination. *Quality Engineering*, 27(1):2–13, jan 2015.

Abstract: ABSTRACTIn this article, we review and reexamine approaches to modeling computer experiments with qualitative and quantitative input variables. For those not familiar with models for computer experiments, we begin by showing, in a simple setting, that a standard model for computer experiments can be viewed as a generalization of regression models. We then review models that include both qualitative and quantitative variables and present some alternative parameterizations. Two are based on indicator functions and allow one to use standard quantitative inputs-only models. Another parameterization provides additional insight into possible underlying factorial structure. Finally, we use two examples to illustrate the benefits of these alternative models

Notes: Very helpful overview focusing on the work of Han et al 2009, Qian et al 2008, and Zhou et al 2011.

- [53] Qiang Zhou, Peter Z. G. Qian, and Shiyu Zhou. A Simple Approach to Emulation for Computer Models With Qualitative and Quantitative Factors. *Technometrics*, 53(3):266–273, aug 2011.

Abstract: We propose a flexible yet computationally efficient approach for building Gaussian process models for computer experiments with both qualitative and quantitative factors. This approach uses the hypersphere parameterization to model the correlations of the qualitative factors, thus avoiding the need of directly solving optimization problems with positive definite constraints. The effectiveness of the proposed method is successfully illustrated by several examples.

Notes: r2017-07-14 The computational efficiency of the technique described in Qian et al 2008 is vastly improved here. However, this version still requires $m(m-1)/2$ parameters when the qualitative factor has m levels. For a single factor of 1000 materials, that is almost 500,000 variables; for two factors of 20 and 50 levels resp., it is 1415 parameters. Restrictions of the covariance matrix may allow reduction in the number of parameters without too far mitigating the utility of the framework.