Description The purpose of these lectures is to introduce the main ideas and techniques underlying Bayesian approaches to the analysis of computer model experiments, in particular in situations where the main interest is in calibration and validation. There are usually multiple sources of error and uncertainty involved in these problems, and this makes Bayesian statistics particularly well-suited to produce sensible solutions in this context.

We will start by reviewing the main principles of the Bayesian approach to Statistics, including basic Markov chain Monte Carlo (MCMC) tools needed to obtain simulated samples from the posterior distribution of the unknown quantities involved in the postulated statistical model.

We then proceed by describing the use of Gaussian processes in producing emulators of computationally intensive computer models. This is central to everything that follows, because simulation-based inference will require the evaluation of the computer model a very large number of times within MCMC algorithms, and this is not feasible without a fast approximation to the output the computer model.

Next, we address the issue of calibration: combining computer model data and field data with the purpose of producing an estimate of the uncertain parameters appearing in the computer model. This involves specifying a statistical model that relates the output of the computer model and the real process it aims at reproducing, thereby accounting for model uncertainty.

The approach to validation that we espouse is based on producing model-based predictions in untested situations, and assessing the magnitude of the associated error.

Although the main ideas are very general, the practical implementation of the methodology is dependent on the specific details of each particular problem. We illustrate this by analyzing applications producing outputs of different complexities, including functional output.

Short bio Rui Paulo has been an Assistant professor at the Mathematics Department of ISEG, the school of Economics and Management of the Technical University of Lisbon, Portugal, since 2006. Prior to that, he was a Lecturer of Statistics at the Mathematics Department of the University of Bristol, UK, between 2004 and 2006. Paulo earned his PhD degree in Statistics and Decision Sciences from Duke University in 2002, with a dissertation which contained the use of novel objective priors for the parameters of Gaussian processes in the context of computer models. He subsequently spent two years as a post-doctoral fellow jointly at the National Institute of Statistical Sciences and at the Statistical and Applied Mathematical Sciences Institute, North Carolina, USA, doing research in the areas of stochastic computation and of statistical analysis of complex computer models. His research interests also include computational statistics, particularly stemming from Bayesian methods, objective Bayesian methodology and model selection. He has published a number of articles in the field of analysis of complex computer models and has actively participated in research conferences and projects in that area. He currently is the principal investigator of a three-year research grant involving two other colleagues, which is focused on several aspects of the analysis of complex computer models.

Tentative outline of lectures

1. Introduction: Bayesian statistics and MCMC methods
    - Bayes theorem as an inferential tool. Bayesian methodology, its advantages and practical difficulties.
    - Priors distributions: subjective priors, conjugate priors, objective priors.
    - Simulation-based inference. Gibbs sampling and the Metropolis-Hastings algorithm.
2. Emulation of computer models
    - Computer models: notation, different types of inputs.
    - Gaussian stochastic processes as priors for unknown functions.
    - Model data and posterior distribution of computer model given model data and the hyperparameters.
    - Fully Bayesian emulator: prior distributions for the hyperparameters, associated MCMC. Likelihood-based emulator, strategies to compute estimates of the hyperparameters.
3. Calibration and validation of computer models
    - Modeling the field data
    - Modeling the relationship between computer model output and reality: the bias function
    - Confounding issues, modularization techniques
    - Sampling strategies
    - Pure-model prediction and bias-corrected prediction, tolerance bounds
4. Extensions
    - Functional output: smooth functions
    - Functional output: irregular functions
    - Multivariate output

Tentative outline of practical sessions

1. Introduction
    - Bayesian analysis of a probit or linear regression model using conjugate priors and objective priors. Implement a Gibbs sampler and a Metropolis-Hastings algorithm to perform simulation based inference.

2. Model approximation

    - Implement the formulae for the conditional mean and covariance function
    - Compute maximum likelihood estimates of the hyperparameters
    - Implement an MCMC algorithm to produce a sample from the posterior distribution of the hyperparameters
    - Produce predictions of the output of the computer model at untested input configurations using the estimates only and using the posterior distribution

3. Calibration and validation

    - Implement an MCMC algorithm to produce a sample from the posterior distribution of all unknowns
    - Predict reality at untested input configurations using only the computer model (pure-model prediction) and using the computer model and the statistical model (bias-corrected prediction)