Bayesian Analysis (2010) **5**, Number 4, pp. 677–682

Comment on Article by Vernon et al.

Pritam Ranjan*

The article is very well written and the methodology used here is quite practical. I would like to congratulate the authors for a thoughtful uncertainty analysis of the Galform model. The paper addresses an inverse problem where the objective is to find the set of all 17-dimensional input values of the computer simulator Galform that leads to a pre-specified output (in this case the measured observational data on the number of galaxies in the universe (Norberg et al. (2002))). A few of the major challenges in this problem are (a) the input space of the computer simulator is high dimensional, (b) the simulator runs are expensive, and (c) the simulator outputs are functional. The authors use the idea of history matching to address some of these problems (Craig et al. (1997) present a nice review on history matching). However, the paper also raises a few interesting questions and should provide opportunities for further research.

Most of my comments are not necessarily specific to this case study and apply generally to the history matching problems for expensive computer simulators with functional outputs. Section 1 discusses my concerns on the simplification of the simulator output data structure while solving the history matching problem. Though the idea of history matching is in the literatures for the last three decades, it is not commonly used in the area of computer experiments (Section 2 suggests a strategy for making it more popular). Finally, some modeling and design aspects of the problem that can use further attention are discussed in Section 3.

1 Simplification of the data structure

The Galform model takes a 17-dimensional input to run (Table 1 of the paper describes the input space), and produces several outputs related to various physical characteristics of the simulated galaxies. Two types of simulator outputs (the (log) number of galaxies per unit volume with respect to the luminosity functions bj and K) were used in the case study. Both of these outputs are functional in nature, and the history matching problem requires matching these luminosity curves with the observed data (Norberg et al. (2002)). In the spirit of Craig et al. (1997), the authors simplify the history matching problem by choosing a set of seven scalar values (three bj and four K luminosity values) from the two functional output curves, and use the Gaussian process modeling approach to construct seven different scalar valued emulators for the (log) number of galaxies per unit volume at these seven luminosity values. That is, the original history matching problem with the simulator producing multiple functional outputs is simplified to finding common solutions of seven inverse problems for simulators with scalar responses and 17-dimensional inputs.

Such a simplification is very appealing and can be useful in practice, but it is not

*Department of Mathematics and Statistics, Acadia University, Wolfville, NS, Canada, mailto:pritam.ranjan@acadiau.ca

 DOI:10.1214/10-BA524B

obvious how to choose a small representative set of points from the response curves for an arbitrary function valued simulator. I believe that the methodology in the paper can be strengthened by including a formal technique for selecting a set of points that is sufficient enough to capture the characteristics of the functional responses. On the other hand, we can take a step back and ask the bigger question "Is it necessary to simplify the data structure of the simulator response?". In an attempt to answer this question, I tried the following approach for couple examples (i.e., similar history matching problems based on Teismann et al. (2009)) and it works reasonably well.

Let f be the computer simulator that takes a d-dimensional input $x = (x_1, \ldots, x_d) \in (0,1)^d$ and produces a functional output $f(x) = \{z_t(x), t \in (0,T)\}$. Suppose the aim is to estimate the set of input values in $(0,1)^d$ for which the evaluation of $f(x)$ gives an acceptable match to $\{z_t^0, t \in (0,T)\}$, a pre-specified observation from the underlying physical system. The key steps of an algorithm that can be used to solve the inverse problem are as follows:

1. For every $x \in (0,1)^d$ transform the simulator to $g(x) = \int_0^T (z_t(x) - z_t^0)^2 dt$ so that the output is now scalar. In practice, one can use $g(x) = \sum_{k=1}^M (z_{t_k}(x) - z_{t_k}^0)^2/M$ over a fine grid $\{t_k = T(k-1)/(M-1),\ k = 1, ..., M\}$. Now the objective is to find the zeros of $g(\cdot)$.

2. Emulate $\{g(x), x \in (0,1)^d\}$ using a Gaussian process model, and find the zeros of the emulator if there exists a perfect match to the history. If we are interested in all reasonable approximations of the history then one can find the minima of $g(\cdot)$. Alternatively, one can emulate $\{\log(g(x)), x \in (0,1)^d\}$ using treed Gaussian process models (Gramacy and Lee (2008)), and find the minima. Note that by taking the log transformation the process becomes non-stationary near the zeros.

For finding the zeros and minima of the emulators of $g(\cdot)$ or $\log(g(\cdot))$, one can follow the implausibility measure type approach as suggested in the paper. As as alternative, a sequential design approach (discussed in the next section) can be used to estimate the zeros and minima of $g(\cdot)$ or $\log(g(\cdot))$.

2 Popularity in computer experiments

I really liked the notion of an *implausibility measure*, particularly its ability of reducing the input space and finding all possible inverse solutions in this context. Such a measure can specifically be very useful for computer simulators with high dimensional input space. Though the idea is not new in environmental applications like the oil industries, it is not very popular in the computer experiments literature. I believe one can establish a relationship between such a multi-stage (or multi-wave) algorithm using the implausibility measure and a sequential design technique (based on the expected improvement (EI) criterion) commonly used in the computer experiments for estimating pre-specified features of an expensive computer simulator.

The idea of a merit based criterion for selecting follow-up design points in a computer experiment started with Mockus et al. (1978), but, the sequential designs based on the EI criteria received significant attention after Jones et al. (1998) and Schonlau et al. (1998) developed EI criteria for estimating the global minimum of an expensive simulator. In the last decade, a host of merit based criteria have been proposed to achieve a good approximation of a specific feature of interest in a small budget of simulator runs. Santner et al. (2003) and Fang et al. (2006) present nice reviews of such criteria. Recently, Ranjan et al. (2008) proposed an EI criterion for solving the inverse problem under the assumption that the simulator is deterministic and the response is scalar.

These EI criteria are simply the expected gain of the suitably defined loss functions tailored towards estimating pre-specified features of the simulator. I believe that a definition of the implausibility measure in the spirit of the EI criteria can make the methodology in the paper more popular in the computer experiments literature.

3 Modeling and design issues

At the end of Section 7.1 of the paper, it is mentioned that the non-implausible regions of the input space turned out to be connected (after every wave) in the application considered here, and thus only one emulator per wave was constructed for the i-th simulator ($1 \leq i \leq 7$). The authors used the Gaussian process modeling approach (Sacks et al. (1989)) for building the emulators of these seven scalar valued simulators for each wave. My concern is "whether or not these non-implausible regions are convex?" Even if all the regions are convex for this application, it may not be true for other applications. If the input space is non-convex, the standard Gaussian process modeling approach should not be used and one might want to consider geodesic distance based models (e.g., Pratola (2006)). One can avoid this problem by constructing the emulators at the i-th wave ($i \geq 1$) on the entire input space by using all $\sum_{k=1}^{i} n_k$ design points, where n_k is the number of design points chosen for the k-th wave (i.e., $n_1 = 993, n_2 = 1414, n_3 = 1620$ and $n_4 = 2011$). One small drawback of this suggestion is that the computational time in fitting the Gaussian process model increases with the wave. This is because of the increment in the size of the datasets used for building emulators at the i-th wave. Nonetheless, efficient techniques like covariance tapering (Furrer et al. (2006)) can be used to reduce the computational time.

One can also use more updated techniques in the computer experiments literature for estimating the active input variables (e.g., Schonlau and Welch (2006), Linkletter et al. (2006)) and the mean function g_{ij} (e.g., Joseph et al. (2008)).

The discussion on the selection of design points for different waves of the analysis (Sections 4.1, 4.2 and 7.1 of the paper) is very thorough and helpful in understanding the methodology. For a practitioner trying to implement the methodology to another application, it would be beneficial to know the justification (or some *ad-hoc* thumb rule) for choosing the size of the Latin hypercube designs (i.e., why 1000 points at the first wave, 9500 points at the second wave, and so on?).

The total number of simulator runs used for estimating the inverse solution of the

Galform model appears to be quite large. In general, the sufficient number of simulator runs depends on several factors like the complexity of the simulator, the dimensionality of the input space and the number of distinct inverse solutions. But, I suspect that the approach outlined in Section 1 of this discussion along with the EI criterion for finding the minima or zeros can be less expensive (i.e., requires fewer simulator runs) as compared to that in the methodology used in this paper.

Although my comments appear to be somewhat critical, I really liked the paper. The authors have done an excellent job of analyzing such an interesting dataset and writing a very comprehensive article. I believe, the paper makes a very good contribution in the field and provides opportunities for further research.

References

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1997). "Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments." In *Case Studies in Bayesian Statistics*, volume 3, 36–93. New York: Springer-Verlag. 677

Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC. 679

Furrer, R., Genton, M. G., and Nychka, D. (2006). "Covariance tapering for interpolation of large spatial datasets." *Journal of Computational and Graphical Statistics*, 15: 502–523. 679

Gramacy, R. B. and Lee, H. K. H. (2008). "Bayesian treed Gaussian process models with an application to computer modeling." *Journal of the American Statistical Association*, 103(483): 1119–1130. 678

Jones, D., Schonlau, M., and Welch, W. (1998). "Efficient Global Optimization of Expensive Black-Box Functions." *Journal of Global Optimization*, 13: 455–492. 679

Joseph, V. R., Hung, Y., and Sudjianto, A. (2008). "Blind Kriging: A New Method for Developing Metamodels." *ASME Journal of Mechanical Design*, 130(3): 031102–1–8. 679

Linkletter, C., Bingham, D., Hengartner, N., Higdon, D., and KQ, Y. (2006). "Variable Selection for Gaussian Process Models in Computer Experiments." *Technometrics*, (48): 478–490. 679

Mockus, J., Tiesis, V., and Zilinskas, A. (1978). "The application of Bayesian methods for seeking the extremum." In *Dixon L, Szego G (eds) Towards Global Optimisation 2*, 117–129. Elsevier. 679

Norberg, P., Cole, S., and et al (2002). "The 2dF Galaxy Redshift Survey: The b_J-band galaxy luminosity function and survey selection function." *Monthly notices of the Royal Astronomical Society*, 336(3): 907–934. 677

Pratola, M. (2006). *Design on Non-Convex Regions: Optimal Experiments for Spatial Process Prediction*. Simon Fraser University. 679

Ranjan, P., Bingham, D., and Michailidis, G. (2008). "Sequential experiment design for contour estimation from complex computer codes." *Technometrics*, 50(4): 527–541. 679

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments." *Statistical Science*, (4): 409–435. 679

Santner, T. J., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag. 679

Schonlau, M., Jones, D., and Welch, W. (1998). "Global versus local search in constrained optimization of computer models." *New Developments and Applications in Experimental Design, Institute of Mathematical Statistics Lecture Notes, Hayward, California*, 34: 11–25. 679

Schonlau, M. and Welch, W. (2006). "Screening the input variables to a computer model via analysis of variance and visualization." In *Screening Methods for Experimentation in Industry, Drug Discovery, and Genetics. (Editors Dean, A. and Lewis, S.)*, 308–327. New York: Springer-Verlag. 679

Teismann, H., Karsten, R., Hammond, R., Hardman, J., and Franklin, J. (2009). "On the Possibility of Counter-Productive Intervention: The Population Mean for Blowflies Models Can Be an Increasing Function of the Death Rate." *Journal of Biological Systems*, 47: 739–757. 678