

MULTIPLE FIDELITY SIMULATION OPTIMIZATION OF HOSPITAL PERFORMANCE UNDER HIGH CONSEQUENCE EVENT SCENARIOS

Jason R Schenk

Industrial, Welding, and Systems Engineering,
The Ohio State University, 210 Baker Systems,
1971 Neil Avenue, Columbus, Ohio 43210

Ning Zheng

Industrial, Welding, and Systems Engineering,
The Ohio State University, 210 Baker Systems,
1971 Neil Avenue, Columbus, Ohio 43210

Deng Huang

Scientific Forming Technologies Corporation, 5038
Reed Road, Columbus, Ohio 43220

Theodore T. Allen

Industrial, Welding, and Systems Engineering,
The Ohio State University, 210 Baker Systems,
1971 Neil Avenue, Columbus, Ohio 43210

ABSTRACT

In optimizing systems, experimental models are often available with different levels of cost and different levels of “fidelity” or trustworthiness, a fact that can be exploited. For example, a highly detailed model might be made for a few possible configurations, supplemented by a large number of rough models that are less expensive to construct. The purpose of this paper is to illustrate the application of a recently proposed Multiple Fidelity Sequential Kriging Optimization (MFSKO) method to derive the optimal resource allocation for disaster preparedness of a hospital. The system is evaluated via discrete event simulations of two sophistication levels. The MFSKO method integrates multiple fidelity data in search for the global optima with less total evaluation cost. Kriging meta-models are generated as by-products of the optimization.

1 INTRODUCTION

Drills, training exercises, and disaster preparedness help to ensure that a hospital can mobilize its resources effectively and quickly during a disaster. Preparations are costly, however, as evidenced by the extensive efforts of agencies from the Defense Department to local Emergency Management Agencies to maximize training transfer while minimizing associated costs (Bowers 2003 and Brady 2003). They require managerial effort and sap clinical staff and resources from regular operations and patient care. The question becomes “How much is enough?” when it comes to preparation. Experimental data, in this case derived from discrete event simulation, is also expensive. Creating a detailed model for each disaster plan quickly becomes prohibitively expensive. This paper shows that low-fidelity and inexpensive models can supplement high-fidelity models to reach a global optimum disaster preparation plan.

For the purpose of analysis, greater preparedness can be equated with increased resource response. As staff becomes more educated on disaster response, and the hospital organization becomes more efficient and structured in its disaster plan, more staff will respond more quickly during an event. The staff that does respond will act more proficiently, further increasing the effective capacity of the system. Similarly, organizational readiness will increase the number of beds that can be made available quickly throughout the hospital, not just in the Emergency Department (ED).

A non-linear black-box global optimization technique, the Multiple Fidelity Sequential Kriging Optimization (MFSKO) method, was developed recently (Huang et al. 2005). The method, which is an extension of the Efficient Global Optimization (EGO) (Jones et al. 1998), aims to solve expensive black-box problems in areas such as large-scale circuit board design and manufacturing process improvement. A uniqueness of the MFSKO method is that it can fully integrate simulations at *different* levels of sophistication into the optimization. Less sophisticated and therefore less expensive runs can be used in conjunction with higher cost runs, resulting in significant reduction in total evaluation cost. Also, there is an additional benefit that global meta-models of the simulations are created as by-products of the optimization.

In this research, MFSKO method is applied to solve the hospital disaster preparedness optimization problem. The goal is to find the optimal resource allocation, considering both the costs of training, drills, and other preparations, and the performance under high-consequence event scenario. In addition, we examine the appropriateness of the applying the MFSKO approach to discrete event simulation.

2 DISCRETE EVENT SIMULATION OF EMERGENCE DEPARTMENT

The model used in this study simulates the operations of an Emergency Department (ED). The simulation was based on a real Emergency Department, with some minor modifications, using Rockwell Automation’s Arena® software, which is a stochastic discrete event simulation tool based in SIMAN. An ED model was used because during a disaster the hospital responds as if it were one large ED. The model was set to run for 10 simulated days of warm-up time under normal circumstances followed by a bolus of 500 patients arriving within one hour. This represents the amount of patients arriving to one of the local hospitals under a single high-rise bomb attack, similar to the Oklahoma City bombing. The model was designed to run for two more days and data was captured as the system recovered from the influx of patients. At the time of the disaster, additional physician and nursing staff were made available, as were additional beds. The amounts of these resources made available were the decision variables that the MFSKO manipulated and attempted to optimize. The rationale behind varying levels of additional resources is the assumption that additional training translates into additional responders and more efficient use of available resources. The preparation costs required to increase the additional resources responding during a disaster were set at \$10,000 per each additional physician, \$5,000 per each additional nurse, and \$10,000 per each additional bed that the system may make use of. These additional resources include actual additional responders as well as “effective” resources, created through increased preparedness of system. The output of the model used as a response variable was the average length of time until the patients were seen by a physician (door-to-doc time). This time was translated into a cost equaling \$1,000,000 per hour per patient, representing the increased morbidity and mortality resulting from delaying intervention. Each setting was replicated 10 times in order to minimize the Monte Carlo error associated with the stochastic nature of the model.

The original model was used as the “high-fidelity” model. It contained 88 nodes. Details include staff scheduling for nurses, registration staff, triage staff, and physicians, patient arrival patterns detailed by hour of day, procedure times, patient acuity, and others, taken from real-world observations. The data for the model was compiled over the period of one year, using interviews, staff meetings, and over 72-hours of direct observation. Every area of the ED is modeled, including a low acuity area called “Prompt Care”, an observation unit, triage, registration, ancillary testing, and the main ED. One of these areas is shown in Figure 1. The model also includes consultations and admissions processes. The model is able to react to changing variables much as the real system would. Details

such as utilization, wait times, process times, etc. are automatically captured for each setting of the system.

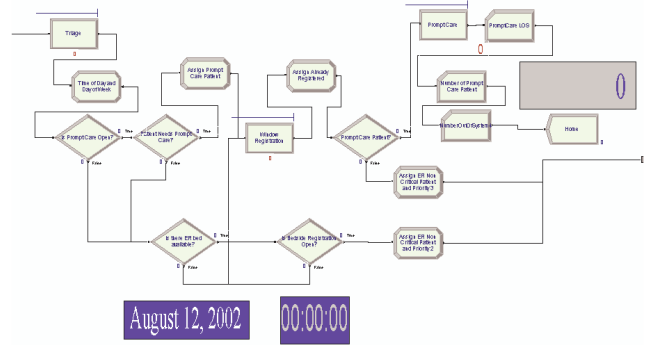


Figure 1 High Fidelity Model of ED Evaluation Process

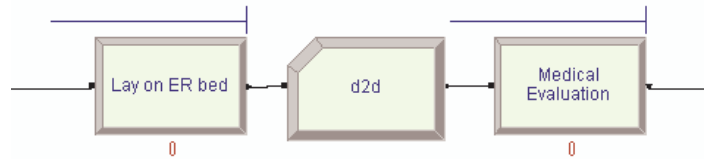


Figure 2 Low Fidelity Model of ED Evaluation Process

Stripping down the original model to 31 nodes created a “low-fidelity” model. Details such as data collection, registration, lab testing, etc. were aggregated or eliminated to reduce the size and increase the speed of this second model (see Figure 2).

3 MULTIPLE FIDELITY SEQUENTIAL KRIGING OPTIMIZATION

3.1 Overview

When cost-per-evaluation on a system of interest is high, one may utilize surrogate systems that provide cheaper but lower-fidelity information. For example, lab and pilot systems can be used to mimic production systems. In the context of discrete event simulation, as mentioned above, a simplified simulation model can be a cheaper surrogate for a more sophisticated model. We call these surrogate systems “lower-fidelity systems”, where the term “fidelity” relates to the extent to which a surrogate system can reproduce the input-output relationships of the system of interest. The system of interest, often a physical experiment or a high-resolution simulation, is called the “highest fidelity” system or the “real” system. Differences in cost for changing the model is especially apparent when testing different scenarios requires significant rework of the model struc-

ture. In this example, differing scenarios were admittedly simple to change in both the high-fidelity and low-fidelity models.

In the Multiple Fidelity Sequential Kriging Optimization (MFSKO) method, low-fidelity systems are exploited to reduce the total evaluation cost needed for the optimization. The method integrates all data on to build a kriging meta-model that provides a global prediction of the objective function and a measure of prediction uncertainty. The location and fidelity level of the evaluation are selected by maximizing a cost-related Expected Improvement (EI) function.

In this paper, a minimal outline the MFSKO method is described. For further details, please refer to Huang (2005). The basic procedure of the optimization method is as follows:

Step 1: Build the initial kriging meta-model for the system of interest using multiple fidelity data.

Step 2: Use a cross validation to make sure the kriging prediction and the measure of uncertainty are satisfactory. Appropriate transformations such as the logarithm or the inverse may be applied to the objective function if needed.

Step 3: Find the location and fidelity level of the new evaluation that maximize the Expected Improvement (EI) function. If the maximal EI is sufficiently small, terminate the optimization scheme.

Step 4: Conduct an evaluation where the EI is maximized. Update the kriging meta-model with the new data point. Go to Step 3.

(As of convention, step 1 is also referred to as the “initial fit” stage, while Steps 3 and 4 are called the “infill” stage. The sequentially added evaluations are also called the “infill” points.)

3.2 Multiple Fidelity Kriging Meta-modeling

Formulations for multiple fidelity kriging models were first published by Kennedy et al. (2000). Suppose there are a total of m systems to draw evaluations from, including the real and the surrogates. Denote the output functions of these systems in increasing order of fidelity by $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})$, where \mathbf{x} is the input vector. To build kriging meta-models for multiple fidelity systems, an autoregressive model is assumed as follows:

$$f_l(\mathbf{x}) = f_{l-1}(\mathbf{x}) + \delta_l(\mathbf{x}), \quad (l = 2, 3, \dots, m) \quad (1)$$

$$\text{and } f_1(\mathbf{x}) = \delta_1(\mathbf{x}). \quad (2)$$

We use kriging to model the lowest-fidelity system, $\delta_1(\mathbf{x})$, as well as the difference between systems, $\delta_l(\mathbf{x})$ ($l = 2, 3, \dots, m$), i.e., we have

$$\delta_l(\mathbf{x}) = \mathbf{b}_l(\mathbf{x})^T \boldsymbol{\beta}_l + Z_l(\mathbf{x}) + \varepsilon_l \quad (l = 1, 2, \dots, m) \quad (3)$$

where \mathbf{b}_l and $\boldsymbol{\beta}_l$ are the basis functions and coefficients, respectively, of the linear model. Z_l is the systematic departure and ε_l is the random error.

The systematic departure from the linear model, Z_l , is modeled as a zero-mean stationary Gaussian stochastic process, with the covariance between two points $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{x}' = (x'_1, \dots, x'_d)$:

$$\text{cov}[\delta_l(\mathbf{x}), \delta_l(\mathbf{x}')] = \sigma_{Z,l}^2 \exp \left[- \sum_{j=1}^d \theta_{l,j} (x_j - x'_j)^2 \right] \quad (4)$$

where d is the dimension of input space, $\sigma_{Z,l}^2$ is the variance of the stochastic process, and $\theta_{l,j}$ is a “roughness” parameter associated with the dimension j .

We denote by Y_1, Y_2, \dots, Y_n the data drawn from an n -point design with point locations $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ and system indexes $\{l_1, l_2, \dots, l_n\}$, respectively. Note that $1 \leq l_1, l_2, \dots, l_n \leq m$. The data may contain random errors, which are assumed to be independent and identically distributed (IID). We denote by $\sigma_{\varepsilon,l}^2$ the variance of the random error associated with system l . To describe the kriging model predictor, introduce the following notation:

$$\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \dots, \hat{\boldsymbol{\beta}}_m]^T$$

$$\mathbf{h}_l(\mathbf{x}) = [\mathbf{b}_1(\mathbf{x})^T, \mathbf{b}_2(\mathbf{x})^T, \dots, \mathbf{b}_l(\mathbf{x})^T, 0, \dots, 0]^T$$

($l = 1, 2, \dots, m$)

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{l_1}(\mathbf{x})^T \\ \mathbf{h}_{l_2}(\mathbf{x})^T \\ \dots \\ \mathbf{h}_{l_n}(\mathbf{x})^T \end{bmatrix}$$

$$\mathbf{V} = [\text{cov}(Y_i, Y_j)]_{1 \leq i, j \leq n}$$

$$= [\text{cov}(f_{l_i}(\mathbf{x}_i), f_{l_j}(\mathbf{x}_j))]_{1 \leq i, j \leq n} + [\sigma_{\varepsilon, l_i}^2 \delta_{ij}]_{1 \leq i, j \leq n}$$

$$\mathbf{t}_l(\mathbf{x}) = [\text{cov}(f_{l_1}(\mathbf{x}_1), f_l(\mathbf{x})), \dots, \text{cov}(f_{l_n}(\mathbf{x}_n), f_l(\mathbf{x}))]^T$$

$$\mathbf{y}^T = [Y_1, \dots, Y_n]$$

where T denotes the transpose, $\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$ for $i \neq j$. The best linear predictor (BLP) of $f_m(\mathbf{x})$ is:

$$\hat{f}_m(\mathbf{x}) = \mathbf{h}_m(\mathbf{x})^T \hat{\boldsymbol{\beta}} + \mathbf{t}_m(\mathbf{x})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}}) \quad (5)$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T \mathbf{V}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{V}^{-1} \mathbf{y}.$$

The hyper-parameters, include: $\sigma_{z,l}^2$, $\sigma_{\varepsilon,l}^2$, and $\theta_{l,i}$, for $l = 1, 2, \dots, m$, and $i = 1, 2, \dots, d$, are obtained by the Maximum Likelihood Estimations (MLE).

Kriging model also has a Bayesian interpretation, where the posterior distribution for the function of system l , $f_l^p(\mathbf{x})$, has a mean equal to the BLP predictor, $\hat{f}_l(\mathbf{x})$. And the posterior covariance is

$$\begin{aligned} \text{cov}[f_l^p(\mathbf{x}), f_l^p(\mathbf{x}')] &= \\ \text{cov}[f_l(\mathbf{x}), f_l(\mathbf{x}')] - [\mathbf{h}_l(\mathbf{x})^T, \mathbf{t}_l(\mathbf{x})^T] &\begin{bmatrix} \mathbf{0} & \mathbf{H}^T \\ \mathbf{H} & \mathbf{V} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_l(\mathbf{x}') \\ \mathbf{t}_l(\mathbf{x}') \end{bmatrix} \end{aligned} \quad (6)$$

3.3 The Expected Improvement Function

As described in 3.1, the Expected Improvement (EI) function is the criterion determining the location and fidelity level of the subsequent evaluation. The EI function takes the following form:

$$EI(\mathbf{x}, l) \equiv E[\max(\hat{f}_m(\mathbf{x}^*) - f_m^p(\mathbf{x}), 0)] \cdot \alpha_1(\mathbf{x}, l) \cdot \alpha_2(\mathbf{x}, l) \cdot \alpha_3(l) \quad (7)$$

where

$$\begin{aligned} \alpha_1(\mathbf{x}, l) &= \text{corr}[f_l^p(\mathbf{x}), f_m^p(\mathbf{x})] \\ \alpha_2(\mathbf{x}, l) &= \left(1 - \frac{\sigma_{\varepsilon,l}}{\sqrt{s_l^2(\mathbf{x}) + \sigma_{\varepsilon,l}^2}} \right) \\ \alpha_3(l) &= \frac{C_m}{C_l}, \end{aligned}$$

and

$$s_l^2(\mathbf{x}) = \text{cov}[f_l^p(\mathbf{x}), f_l^p(\mathbf{x})]$$

where C_m and C_l are the cost-per-evaluations on system m and l , respectively.

In addition, \mathbf{x}^* stands for the current ‘‘effective best solution’’, which is determined by maximizing a utility function, $u(\mathbf{x})$.

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}} [u(\mathbf{x})] \quad (8)$$

where

$$u(\mathbf{x}) = -\hat{f}_m(\mathbf{x}) - c s_m(\mathbf{x}).$$

The use of $u(\mathbf{x})$ implies a willingness to trade 1 unit of the predicted objective function for c unit of the standard deviation of prediction uncertainty.

$\alpha_1(\mathbf{x}, l)$ serves as a factor to discount the EI when an evaluation from a surrogate system is used. When $l = m$, $\alpha_1(\mathbf{x}, l) = 1$. The term $\alpha_2(\mathbf{x}, l)$ is meaningful only when outputs of system l contain random errors. It accounts for the diminishing return of additional replicates as the prediction becomes more accurate. At last, $\alpha_3(l)$ represents an adjustment to the sampling strategy according to the evaluation costs. A point on lower-fidelity system will be favored if everything else is equal.

As mentioned in Section 3.1, Step 3, the location and fidelity level of the next evaluation, \mathbf{x}_{n+1} and l_{n+1} , are selected by maximizing EI, i. e.:

$$(\mathbf{x}_{n+1}, l_{n+1}) = \arg \max_{\mathbf{x}, l} EI(\mathbf{x}, l). \quad (9)$$

The optimization scheme stops when the maximal EI is sufficiently small. In this study, we use a stopping criterion that is 0.1% of the active span of the responses.

3.4 The Initial Fit Design

In the initial fit stage, the kriging model is generated with evaluations at locations from a preset experiment design. Design problems for kriging models have been studied in a wealth of literature in the area of the Design and Analysis of Computer Experiments (DACE). For summaries on this area, see Santner et al. (2003).

For this study, the 26-point initial-fit design used is shown in Figure 3. Twenty of them are on the low-fidelity simulation, and six of them are on the high-fidelity simulation. The low-fidelity points form a Latin Hypercube design with maximal minimum distance between points (Stein 1987); and the high-fidelity points are a subset of the low-fidelity points, and form another Latin Hypercube design.

4 RESULTS AND DISCUSSIONS

For this analysis, the ratio of additional nurses to doctors responding was fixed at 1:1. Therefore the only settings are the number of beds and the number of staff. This al-

lows us to show the results in two-dimensional contour plots. It took the MFSKO method 50 low-fidelity evaluations and 16 high-fidelity evaluations to reach the stopping criterion. The optimal setting for the test problem is 19 additional doctors, 19 additional nurses, and 108 additional beds, and the global minimum of the cost function is 301.9k. Figure 3 shows all evaluated points and history of the infill points. Note that, by maximizing the Expected Improvement (EI) function, the search pattern displays a balance between local and global search. Many points concentrate near the global optimal setting, while other scattered around to reduce uncertainties about the unexplored areas. In addition, for the infill points, most of the “exploring” (global search) are done with low-fidelity evaluations, while the high-fidelity evaluations are mostly utilized to “refine” the optimal setting (local search). This pattern seems to agree with our intuitions. At last, the best solution was found by point 62, but the algorithm continues to evaluate four more points before the Expected Improvement (EI) in the entire domain is sufficiently small.

As mentioned previously, an addition benefit of the MFSKO method is that kriging meta-models are generated at the end of the optimization scheme. Figure 4 shows the contour maps of the kriging models for both the low-fidelity and high-fidelity simulations. In general, the two produce similar trends. However, the low-fidelity model seems to slightly overestimate the cost. In addition, the high fidelity model appears to be bumpier and have multiple local optima. This illustrates an additional potential benefit of using a multi-level approach: the high-fidelity model gives the meta-model detail and accuracy, while the low-fidelity model gives smoothness and added points at reduced cost.

The kriging model also gives the prediction uncertainty in terms of Mean Square Errors. Figure 5 displays the prediction uncertainty on the high-fidelity simulation. Note that the uncertainty is smaller in areas near the optima, where more evaluations have been allocated.

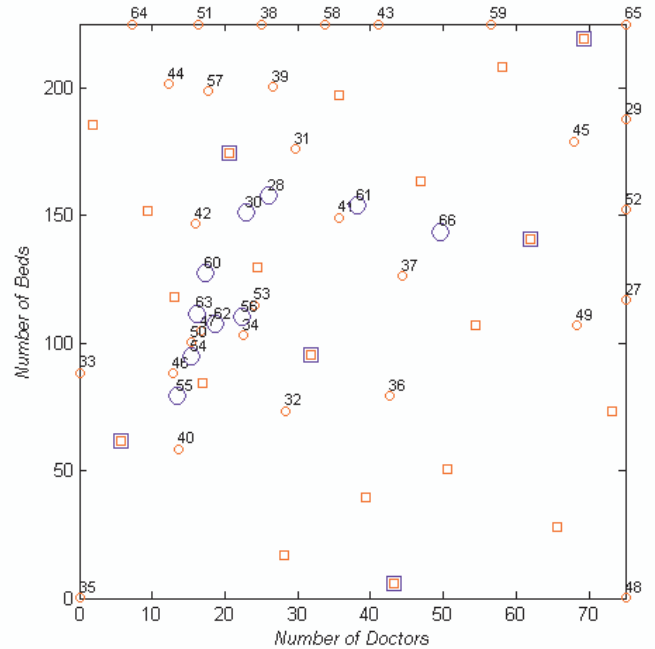
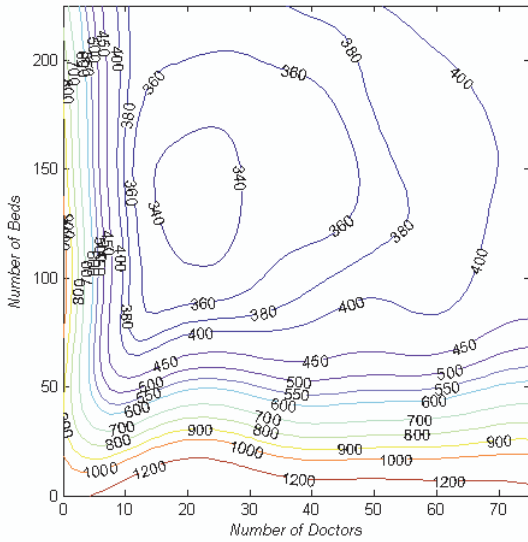
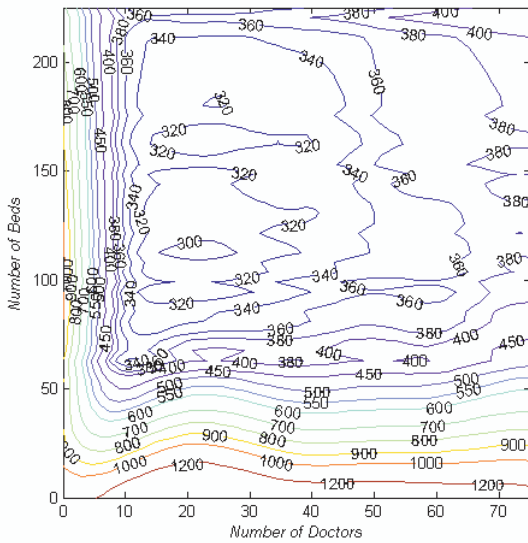


Figure 3: Evaluation points
 (□: low-fidelity initial-fit design points;
 □: high-fidelity initial-fit design points;
 ○: low-fidelity infill points;
 ○: high-fidelity infill points.

The numbers indicate the sequence of the infill points.)



a)



b)

Figure 4. Meta-models generated by MFSKO a) Low-fidelity simulation b) High-fidelity simulation

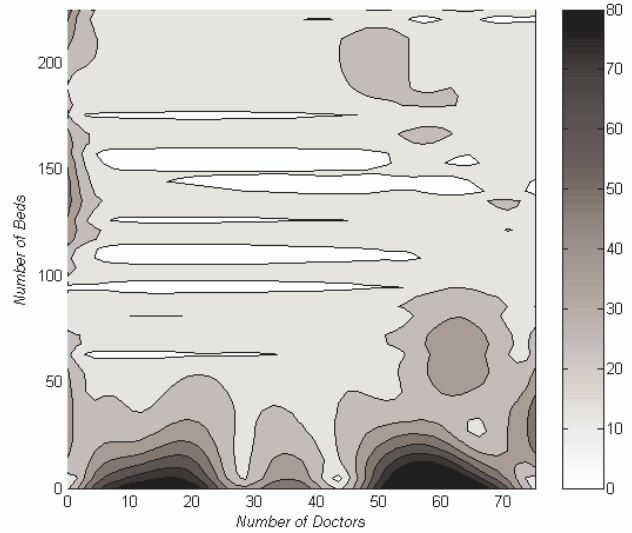


Figure 5. Kriging prediction uncertainty (Mean Squared Errors) of the high-fidelity simulation

5 CONCLUSIONS

In this research, a hospital disaster preparedness optimization problem was solved using the Multiple Fidelity Sequential Kriging Optimization (MFSKO) method. The area under study was a hospital Emergency Department (ED), since a hospital responds as a large ED during a disaster. Considering both the costs under regular operations and the performance under high-consequence event scenario, the optimal setting is 19 doctors, 19 nurses, with 108 beds.

The MFSKO method is a convenient and robust global optimization tool for this problem. Simulations of two levels of sophistications were fully integrated in algorithm, which presumably reduced the total evaluation costs. In addition, meta-models of the simulations are generated as by-products of the optimization, which help visualize and interpret of the problem.

One limitation of the present design is that the algorithm uses a continuous decision space. Further generations of the MFSKO method can be developed to handle discrete data.

This method could be easily augmented to include more complex variables, such as changes to the actual disaster response procedures under different disasters, and the use of new technology. This would require that the actual structure of the model be changed for each set of inputs, increasing the cost savings of this approach. Secondly, integrated distributed simulations, such as the one described by Jain and McLean (2003), could replace the ED model used in this analysis. Future work is needed to apply MFSKO to categorical or non-sequential data, such as alternative disaster management structures.

REFERENCES

- Brady, T. (2003), Emergency Mangement: Capability Analysis of Critical Incident Response. Winter Simulation Conference 1863-1867.
- Bowers, F. and Prochnow, D. (2003), JTLS-JCATS Federation Support of Emergency Response Training. Winter Simulation Conference 1052-1060.
- Huang, D. and T. Allen (2005), Design and Analysis of Variable Fidelity Experimentation Applied to Engine Valve Heat Treatment Process Design, The Journal of the Royal Statistical Society: Series C, 54, 2, 1-21.
- Huang, D. (2005), Experimental Planning and Sequential Kriging Optimization Using Variable Fidelity Data, Ph.D. Dissertation, The Ohio State University.
- Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A. (2005), Sequential Kriging Optimization Using Multiple Fidelity Evaluations, submitted to the Structural and Multidisciplinary Optimization.
- Jain, S. and McLean, C. (2003), A Framework For Modeling And Simulation For Emergency Response. Winter Simulation Conference, 1068-1076.
- Jones, D., Schonlau, M. and Welch, W. (1998), Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization 13, 455-492.
- Kennedy, M. C. and O'Hagan, A. (2000) Predicting the output of a complex computer code when fast approximation are available. Biometrika, 87, 1, 1-13.
- Santner T.J., Williams, B.J. and Notz, W.I. (2003), The Design and Analysis of Computer Experiments, Springer, New York.
- Stein, M. (1987), Large Sample Properties of Simulation Using Latin Hypercube Sampling. Technometrics 29, 143-151.

AUTHOR BIOGRAPHIES

THEODORE T. ALLEN is an associate professor in the Department of Industrial, Welding, and Systems Engineering at the Ohio State University (OSU). His research interests include six sigma, design of experiments, and simulation optimization with focus on applications in health care and manufacturing. He is an associated editor of the Journal of Manufacturing Systems, a senior member of ASQ, and a member of the INFORMS-Simulation Society. His e-mail address is allen.515@osu.edu and his Web address is <http://www-iwse.eng.ohio-state.edu/ISEFaculty/allen/allen.htm>.

DENG HUANG, Ph.D. is currently a principal research scientist at the Scientific Forming Technologies Corporation, Columbus, Ohio. His research interest centers on optimization and statistics methods applied to manufacturing

process designs. He also develops and supports the material microstructure models in the Finite Element software DEFORM™. Dr. Huang has published papers on global optimization, design of experiments, and computer modeling of material processing. He received his Ph.D. degree from OSU. His email address is dhuang@deform.com.

JASON SCHENK is an Ph.D. student in the Department of Industrial, Welding & Systems Engineering at OSU. He received his M.S. from the University of Iowa with his thesis involving the simulation of the University of Iowa hospitals and clinics. His research interests include simulation and hospital and military systems. His e-mail address is schenk.32@osu.edu.

NING ZHENG is a Ph.D. student in the Department of Industrial, Welding, and Systems Engineering at the Ohio State University. He has a M.S. degree from OSU and a B.S. degree in Mechanical Engineering from Tsinghua University, China. His research interests include applications of statistical and simulation methods to system design and to product design, and manufacturing. His email address is zheng.481@osu.edu.