Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

SIMULATION EXPERIMENTS: BETTER DATA, NOT JUST BIG DATA

Susan M. Sanchez

Naval Postgraduate School
Operations Research Department
1411 Cunningham Rd.
Monterey, CA 93943-5219, USA

ABSTRACT

Data mining tools have been around for several decades, but the term "big data" has only recently captured widespread attention. Numerous success stories have been promulgated as organizations have sifted through massive volumes of data to find interesting patterns that are, in turn, transformed into actionable information. Yet a key drawback to the big data paradigm is that it relies on observational data—limiting the types of insights that can be gained. The simulation world is different. A "data farming" metaphor captures the notion of purposeful data generation from simulation models. Large-scale designed experiments let us grow the simulation output efficiently and effectively. We can explore massive input spaces, uncover interesting features of complex simulation response surfaces, and explicitly identify cause-and-effect relationships. With this new mindset, we can achieve quantum leaps in the breadth, depth, and timeliness of the insights yielded by simulation models.

1 INTRODUCTION

There is no universally accepted definition for the term "big data." Many have argued that we have always had big data, whenever we had more available than we could analyze in a timely fashion with existing tools. Running a single multiple regression in the 1940's was very difficult, since matrix multiplication and inversion for even a small problem (say, a handful of independent variables and a few hundred observations) is a non-trivial task when it must be done by hand. From that perspective, big data can be viewed as any data set that pushes against the limits of currently available technology.

What can be done with big data? Most people immediately think of data mining, a term that is ubiquitous in the literature. The concept of data farming is less well-known, but has been used in the defense community over the past decade. At the Simulation Experiments & Efficient Designs (SEED) Center for Data Farming at the Naval Postgraduate School, we describe the differences between these metaphors as follows (Sanchez 2010, Sanchez 2014).

Miners seek valuable nuggets of ore buried in the earth, but have no control over what's out there or how hard it is to extract the nuggets from their surroundings. As they take samples from the earth, they gather more information about the underlying geology. Similarly, data miners seek to uncover valuable nuggets of information buried within massive amounts of data. They may look at "all" the "big data" available at a particular point in time, but they typically have no control over what data are out there, nor how hard it will be to separate the useful information from the rest of the data. Data mining techniques use statistical and graphical measures to try to identify interesting correlations or clusters in the data set.

Farmers cultivate the land to maximize their yield. They manipulate the environment to their advantage, by using irrigation, pest control, crop rotation, fertilizer, and more. Small-scale designed experiments can help them determine whether these treatments are effective. Similarly, data farmers manipulate simulation models to advantage—but using large-scale designed experimentation. This allows them to learn more about the simulation model's behavior in a structured way. In this fashion, they "grow" data from their

models, but in a manner that facilitates identifying the useful information. For large-scale simulation experiments, this often results in big data sets. Yet, although the data sets are big, they are far smaller than what would be needed to gain insights if the results were observational (i.e., obtained using *ad hoc* or randomly generated combinations of factor settings). The data sets are also better, in the sense they let us identify root cause-and-effect relationships between the simulation model input factors and the simulation output.

2 REVISITING CAUSATION AND CORRELATION

The simplest way of establishing cause-and-effect is via an experiment. Let's suppose that we're looking at a stochastic system, so our observation of Y comes with some error. An experiment involving a single factor X involves changing X and observing whether or not there is a change in the response Y. We're trying to uncover the ground truth about the relationship between X and Y over a range of interest, as shown in (1).

$$Y = g(X) + \varepsilon(X) \tag{1}$$

After collecting data, we will end up estimating this via some metamodels, as shown in (2). The subscripts on $\hat{g}$ and $\hat{\varepsilon}$ show these models were estimated from data obtained using design D.

$$\hat{Y} = \hat{g}_D(X) + \hat{\varepsilon}_D(X) \tag{2}$$

Three important concepts in the design of experiments (DOE) are: *control*, *randomization*, and *replication*. For real-world experiments, you exercise control over the situation by deciding which values of X are of interest. You also decide how to control for everything else that isn't of interest, perhaps by holding it constant or using a control group for comparison purposes. Randomization is used to guard against hidden or uncontrollable sources of bias. For example, if you are measuring the miles per gallon of different vehicles by using a single driver, randomizing the order in which they are driven will remove any systematic bias due to fatigue. With replication you collect multiple observations to assess the magnitude of the variability associated with Y, so you can construct confidence intervals or conduct significance tests.

In simulation experiments, we use these concepts in different ways. In the simulation world, the analyst has total control. Potential factors in simulation experiments include the *input parameters* or *distributional parameters* of a simulation model, whether they are controllable in the real world or not. The analyst also has control over the random number seeds and streams. This means that, unlike physical experiments where uncontrollable noise occurs, the results from a simulation experiment are perfectly repeatable, and randomization is not needed to guard against hidden or uncontrollable sources of bias. Replication means you get multiple *experimental units* (runs or batches) to get a sense of the magnitude of the variability associated with Y.

There are certainly times when uncovering correlation is very useful. As Hogan (2014) points out, "Epidemiological studies demonstrated more than a half century ago a strong correlation between smoking and cancer. We still don't understand exactly *how* smoking causes cancer. The discovery of the correlation nonetheless led to anti-smoking campaigns, which have arguably done more to reduce cancer rates over the past few decades than all our advances in testing and treatment." Yet there are also many more times when identifying correlation is not nearly good enough. It's well-known that trolling through mountains of data can yield spurious correlations: an excellent discussion of how this affects published research results appears in The Economist (2013). If big data sets reveal a correlation between two variables X and Y, the ground truth can be one of four basic situations: (i) changes in X cause changes in Y; (ii) changes in Y cause changes in X; (iii) changes in X and Y are both the result of changes in other, potentially unknown or unobservable factors; or (iv) this is a spurious correlation. One drawback of big observational data is that we have no real way of testing these results without moving to a different environment. Common sense can help eliminate some spurious correlations (NPR 2014, Vigen 2014), but one of the tenets of big data is that it is not repeatable. Correlations that appear strong at one point in time and weak at another might mean that the underlying system has changed.

In their recent book, Mayer-Schönberger and Cukier (2013) claim that "the need for sampling is an artifact of a period of information scarcity" and that big data "leads society to abandon its time-honored preference for causality, and in many instances tap the benefits of correlation." But if the goal of analysis is to yield better outcomes via controlling or influencing the inputs, what could be more important than causality?

3 CHARACTERIZING BIGNESS

Why are so many either extolling or lamenting big data's focus on correlation rather than causation? I suggest it is because their view of big data is observational. This is not the case for those using simulation. Using the data farming metaphor, we grow the data for analysis, rather than mine existing data. However, just as a different mindset is needed for dealing with big observational data, so a different mindset is needed for generating and dealing with big simulation data.

3.1 The 3 (or more) V's of Big Data

Big data is typically characterized by 3 V's — volume, velocity, and variety (Laney 2001). These have all increased at an astonishing rate during the last decade, with much of the data being generated, captured from, or stored on the internet. Volume refers to the amount of data, and this is enormous. As of June 2014, IBM states that "90% of the data in the world today has been created in the last two years alone" (IBM 2015), but that is already out of date. Some have advocated a look at more V's (including veracity, validity, and volatility, viability, value, and victory), but others argue that these are derived from context-specific analytics while the original 3 V's capture the "intrinsic, definitional big data properties essence of big data" (Grimes 2013). Regardless, the nature of the V's determine the analysis tools that can be used, and the decisions that can result.

The 3 (or more) V's have a different flavor in simulation than in other big data situations. Velocity and volume are partially controlled by the analyst, who determines how to run the simulation (e.g., on a single core or on a high-performance computing cluster), how much data to output (e.g., aggregate statistics at the end-of-run, batch statistics, or full time-series output) for each performance measure, and the number of performance measures to study. Generated output can have a variety of types, but the variety does not include many of the problems that we find with observational data (e.g., incompatible data formats, inconsistent data semantics). In the near future I anticipate that many simulators may make use of big data tools for tying their models to real-time or near real-time data sets for model input (Elmegreen, Sanchez, and Szalay 2014). With regard to some of the "wanna V's," verification and validation have long been cornerstones of effective simulation practice. A structured V&V process means the simulation output should not suffer from lack of veracity. Of course, the question of validation—whether the simulation model's behavior is sufficiently close to that of the real-world problem of interest for decision-making purposes—is always with us.

3.2 The 3 F's of Data Farming

Moving beyond the 3 V's of observational big data, in the arena of large-scale simulation experiments we can focus on the 3 F's of inferential big data: *factors, features*, and *flexibility*. All these should be "big" when viewed with a simulation mindset.

Factors refers to a broad view of the inputs (or functions of inputs) that, if varied, affect the simulation and should be manipulated to increase our understanding of the simulation responses. A "big factor" view includes many aspects. Clearly, a large number of factors may be of interest: in fact, one can argue that if we didn't feel an input was important, then we would not have included it in our simulation model. But the big factor view also means that factors may vary over wide ranges, rather than limited ranges. They may be of different types—qualitative, discrete, or continuous—rather than homogenous. Factors can be further broken down into decision factors that can be controlled in the real-world settings; noise factors

that are difficult or impossible to control in reality, but can be controlled during the simulation experiment; and artificial (simulation-specific) factors—such as run length, warm- up period, batch size, or random number seed—that may not have an analogy in the real world, but can influence the way we conduct our simulation experiments and the results we obtain.

Features refers to the simulation responses. A "big feature" view includes many aspects. We may be interested in multiple responses, rather than limiting ourselves to a single one. For stochastic simulation models, our responses may have complex variance/covariance structures. It's worth mentioning that this characteristic, as pervasive and accepted as it is in simulation, is still the exception rather than the rule for traditional designed experiments. As with the factors, the responses may be of different types. We may be interested in short-term, transient behavior for one response, at the same time we're interested in the long-run quantiles of another response. We may also be searching for different types of features in the response surface landscapes. For example, in trade-off analyses, maxima and minima are often much less interesting than so-called "knees in the curve," where we find that further changes in one or more factors lead to diminishing (or increasing) returns. Other interesting features include thresholds where responses change suddenly, such as a model entity experiencing an abrupt shift from usually losing to usually winning; broad, flat regions that indicate we have a solution that is robust to uncertainties in the underlying factors over certain ranges; and Pareto sets containing those alternatives that cannot be completely dominated (in terms of desirable multivariate responses) by any other alternatives in this set.

Flexibility represents the fact that we may need to be able to answer many questions from our experiments, even if we don't know *a priori* all the questions that might be asked. A "big flexibility" view includes many aspects. Choosing a design that facilitates a broad variety of metamodeling, data mining, and graphical analysis tools is far more likely to yield success than picking a restrictive design that is only suitable for one particular approach or model. One way of gaining flexibility is to have a space-filling design: this term has typically been used for designs involving continuous-valued factors. The analogy for discrete-valued or qualitative factors is to have balanced or nearly-balanced designs. Note that in simulation, we can have additional flexibility built into how we store and retrieve our data. As long as we store the design points and random number seeds, we have the option of "growing" new data from our original experiment by rerunning our simulation and printing out additional input, if needed. For example, if we initially print out only end-of-run summaries and then find out that strange things have happened in a handful of the thousands of runs we've conducted, we can rerun that handful and inspect (or animate) the entire sequence of events to better understand what has transpired.

3.3 Fast Data Farming Environments

If inferential simulation data sets are already on hand, we can identify whether they are characterized by enough factors, features, and flexibility to address our questions. But if we are preparing to grow our data, then the timeliness of getting the results is also important.

First, let me state that efficient DOE is *absolutely required* for large-scale simulation experiments. In June 2008, a supercomputer called the "Roadrunner" was unveiled. It was assembled from components originally designed for the video game industry, it cost $133 million, and was capable of doing a petaflop (a quadrillion operations per second). The New York Times stated that "petaflop machines like Roadrunner have the potential to fundamentally alter science and engineering" by allowing researchers to "ask questions and receive answers virtually interactively" and "perform experiments that would previously have been impractical" (Markoff 2008). Five years later, China's "Tianhe 2" supercomputer took over as the new world leader, with over 33 petaflop capability. Yet let's take a closer look at the practicality of a brute-force approach. Suppose a simulation has 100 factors, each factor has two levels (low and high) of interest, and we decide to look at all combinations of these 100 factors. A single replication of this experiment would take over 40 million years on the Roadrunner, and over 935 millenia on the Tianhe-2, even if each simulation run consisted of a single machine instruction (Sanchez, Sanchez, and Wan 2014, Lucas et al. 2015). Efficient design of experiments can break this curse of dimensionality at a tiny fraction of the

hardware cost. Recent breakthroughs provide designs, such as those in Vieira Jr et al. (2013), that can be used to explore 100 or more factors, for models that take a more reasonable minute to run, in times ranging from a few hours on a single processor, to a few minutes or days on a computing cluster.

So, if you use designed experiments, is that enough to say that your turnaround time is fast? Not necessarily. For example, if changing the factor levels in your simulation model can only be accomplished through a GUI, then the bottleneck is often the analyst's time, rather than the computational time. Manually changing all the factor settings is a time-consuming and error-prone process. If you are facing a very short deadline, your only alternative may to use a small design and double-check or triple-check your input settings. But as we discuss in Section 4.4, setting up a data farming environment is extremely worthwhile. Automating the run generation process immediately expands your capability for growing data that doesn't suffer from input errors, and paves the way for running your model on a cluster.

Is a small design always preferable? For simulation experiments, the answer is a resounding 'No!' If your cluster has 1000 cores and your simulation takes a fixed amount of CPU time to complete, it will take no longer to conduct 1000 runs than to conduct a single run; if so, calling a design with 10 design points "faster" than one with 1000 designs points is both wrong and counterproductive! It is better to gather enough data, via larger designs and multiple replications, to be able to explore the simulation's performance without resorting to lots of simplifying assumptions. The bottom line is that a large-scale simulation is fast if you get the kind of data you need in time to act on it.

4 THEORY INTO PRACTICE

4.1 Setting Appropriate Goals

For over a decade, the SEED Center has advocated three basic goals of large-scale simulation experiments: (i) *developing a basic understanding* of a particular simulation model or system; (ii) *finding robust* decisions or policies; and (iii) *comparing the merits* of various decisions or policies (see, e.g., Sanchez and Lucas 2002; Kleijnen et al. 2005). We have stated that these require a different mindset than other goals that are often mentioned for computer experiments, namely: (i) constructing accurate *predictions*; (ii) *calibrating* the simulation to real-world data, or (iii) *optimizing* a system (see, e.g., Santner, Williams, and Notz 2003). Casting our results in a big data framework helps to clarify why we choose different goals. We believe that simulation is the proper choice for modeling complex problems. In these circumstances, it's unlikely that decision-makers will be interested only in the answer to a single, narrow question. Developing an understanding—including identifying important factors and interesting features—allows us to address a much richer range of questions. Similarly, we believe robust (resilient) alternatives found by seeking configurations that perform well over a variety of noise factor settings may be much more useful in practice than alternatives found by optimizing a single performance measure while implicitly holding noise factors constant. Finally, in trade-off analyses it may be much more useful to identify the existence of "knees in the curve," where we start to see increasing or decreasing rates of return, than to numerically predict exactly where these inflection points fall.

4.2 Portfolio of Potential Designs for Large-scale Simulation Experiments

How do we run simulation experiments that meet the 3 F's of factors, features, flexibility in a way that is sufficiently fast? First and foremost, we use a designed experiment capable of meeting these requirements given our computing resources and the time frame available for making the decision. Sanchez and Wan (2012) created a "consumer report" chart that provides guidance to those interested in conducting large-scale simulation experiments. An updated version of this chart is kept on the SEED Center web pages, along with software and spreadsheets that can be freely downloaded (SEED Center for Data Farming 2015). The chart characterizes designs in terms of their factors, features, and flexibility; gives notes with additional guidance; provides citations for the source papers; and highlights designs that we've found to be good starting points.

Note that the field of DOE has been around for a long time. Many of the classic experimental designs could be used in simulation studies—but don't actually meet the needs of our "big data" view. The context for real-world experiments can be much more constrained than for simulations in terms of costs, number of factors, time required, ability to replicate, ability to automate, etc., so a framework specifically oriented toward simulation experiments is beneficial.

4.3 Analysis Approaches

Clearly, the design you choose will impact the types of analyses you can conduct. One common type of metamodel involves polynomial models. We find the main effects model too restrictive, but a good starting point for a single response Y is often a model that can include second- order effects, such as quadratic terms and two-way interactions. For some designs, such as the CCDs of Sanchez and Sanchez (2005), it is possible to simultaneously estimate all first-order and second-order terms. Also, MacCalman, Vieira Jr, and Lucas (2015) constructed second-order Latin hypercubes for a modest number of factors ($k \leq 12$).

Alternatively, space-filling designs based on orthogonal or nearly-orthogonal Latin hypercubes are much more efficient, growing as $O(k)$ or $O(k^2)$, rather than the $O(2^k)$ required to simultaneously estimate all 2^k potential effects. A variety of polynomial metamodels can be fit, from first-order models, models with higher-order terms involving a subset of factors, up to a ridiculously large-order polynomial involving a single factor. Stepwise regression or some other automated method can help determine the subset of terms that are most important. Note that in a big data world, many statistically significant terms may, nonetheless, be eliminated from the metamodel because they are not deemed to be of practical importance, i.e, dominated by other terms with much larger effects.

A second approach we find to be quite useful is the data mining approach of partition trees, also called classification and regression trees (CART). Partition trees employ a binning and averaging process to successively split a large group of heterogeneous data into two smaller groups of data, where each leaf in a split is individually more homogeneous, while the difference between the leaves is large. These are useful for model debugging, as well as the final analysis. For example, a partition tree identified, using only two splits from 27 potential factors, conditions for which a model of mineclearing operations was not working.

Two other metamodeling approaches of interest are kriging and stochastic kriging. These metamodeling approaches are quite flexible in terms of the model form; the small-multiples plots in Figure 1 show kriging fits for two responses against two factors for three different levels of a third factor. Kriging has been heavily used for deterministic computer experiments, and also adapted for stochastic simulation experiments (see, e.g., van Beers and Kleijnen 2003; Ankenman, Nelson, and Staum 2010). However, the analysis approach is computationally quite intensive, so kriging model-fitting is typically conducted for experiments involving relatively small numbers of factors and and design points. Adaptive sequential designs are often used, so that each new design point is chosen to reduce the uncertainty associate with model predictions at untested points as much as possible (Kleijnen and van Beers 2004). The community of those designing and analyzing deterministic computer experiments has thus focused on features and a one-observation-at-a-time view of feedback, although some recent work in batch sequential designs (Loeppky et al. 2010; Duan et al. 2015), may provide additional flexibility for other metamodeling and analysis approaches if the data become too large for kriging software to handle. Brantley et al. (2013) describe sequential methods for allocating a fixed computing budget across design points to efficiently identify terms in polynomial metamodels.

Once the most important factors and interactions have been identified, many other statistical and graphical approaches are also extremely useful at conveying the information to the decision maker. Figure 2, adapted from a presentation based on Marlow and Novak (2013) shows the results of a single replication from a simulation that tracks the state of each of 24 naval helicopters in a fleet over their 30 year life cycle. The colors represent ten different states of the helicopters—including embarked and ashore flight hours; scheduled, unscheduled, and deep maintenance—when a particular heuristic is used for scheduling flights and rotations.

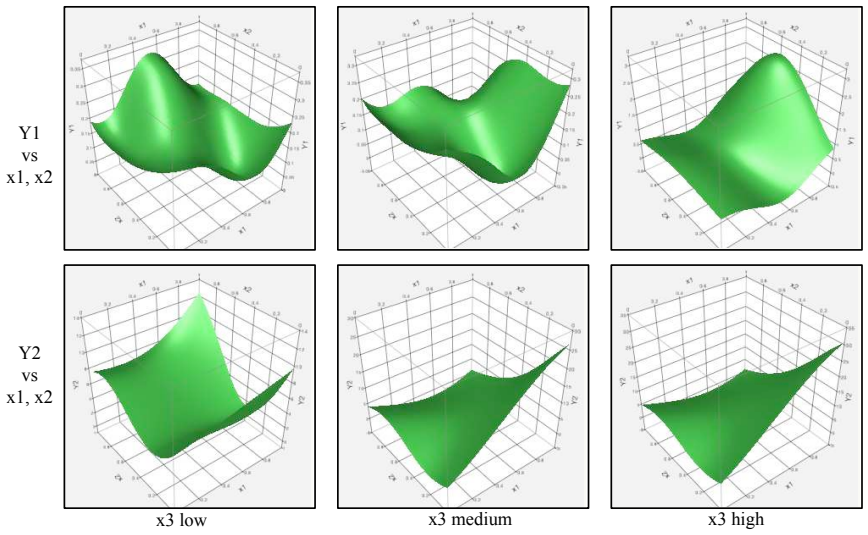
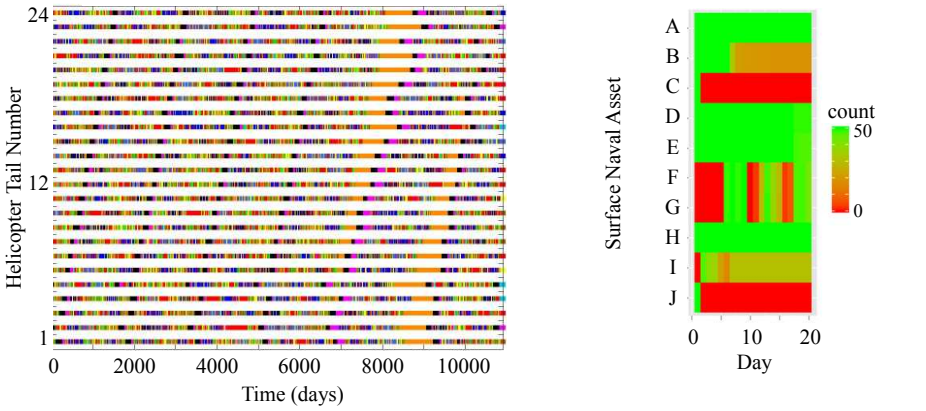


Figure 1: Kriging metamodels for two different responses, as a function of three factors.



(a) Helicopter Flight Status Over Time

(b) Inventory Level > 50% Capacity

Figure 2: Two graphs revealing several dynamic outputs for a single design point: (a) States of each of 24 naval helicopters over a 30 year life cycle for a single replication. (b) Likelihood of a variety of resources dropping below critical levels, over time, based on 50 replications.

4.4 Automated Implementation

Even using good designs, the data farming approach includes automating the process of data collection whenever possible. Once you have chosen the design (or design algorithm), you should use a computing script to automatically run the experiments, allocate individual runs to distributed computing assets, and consolidate the output in a form suitable for analysis. This requires some programming expertise and may take a bit more time initially, but the payoff is worthwhile. Software that facilitates this automation for simulations with XML input files is available at the SEED Center web pages.

Another issue to consider when running large-scale experiments is the number of replications. For run- based experiments with long individual run times, it makes sense to iterate through replications of the entire design, rather than iterate through a large number of replications of design point 1, then design

point 2, etc. This makes it more likely that you will have useful information if you stop the experiment early, and it allows you to halt once you've seen enough.

4.5 Testing the Results: Another Reason Simulators Have Better Data

In large-scale simulation experiments, if we identify that some factor X belongs in a metamodel for some response Y, the ground truth can be one of four different situations: (i) changes in X (either individually, or in conjunction with other factors) cause changes in Y; (ii) changes in X are partially confounded with changes in high-order interactions of other factors; (iii) changes in X are fully confounded with changes in high-order interactions of other factors; or (iv) we have a false positive effect. Yet, in the simulation world, it is possible to explicitly test the results and determine whether or not our effects are truly important. Our choice of design determines the degree of confounding we are willing to accept *a priori*. This, in conjunction with model confirmation runs or secondary experiments, lets us test to see if the metamodel is sufficiently accurate at previously-untested regions in the factor space. If so, we can feel comfortable using the metamodels, in place of running additional simulations, for quick-turn analysis. If not, we can use new information and additional designs to enhance the metamodels until their performance is satisfactory.

4.6 Reality, and Back Again

All the techniques we've described allow us to get better understanding of our simulation model. We've found them very useful throughout the simulation process: during model development and model verification, as well as for "production runs." Yet it's important to remember that our experiments are providing us with information about the simulation model's behavior, not necessarily the behavior of the real-world system they attempt to emulate. In cases where it is possible to get some real-world data for validation purposes, we can use our simulation experiments to suggest real-world test cases that may be most informative or interesting. In other situations, where we do not have (and may hope never to have) real-world data, the ability to perform large-scale designed experiments allows us to examine a plethora of "what-if" cases in an efficient and effective manner.

4.7 Finding Out More

For more on the philosophy and tactics of designing large-scale simulation experiments, examples of graphical methods that facilitate gaining insight into the simulation model's performance, and extensive literature surveys, we refer the reader to Kleijnen et al. (2005), Sanchez and Wan (2012), or Sanchez et al. (2012). Books that discuss DOE for computer or simulation models include Santner, Williams, and Notz (2003); Kleijnen (2007); and Law (2014). Note that their goals for analysts performing simulation experiments, and hence their design recommendations, may differ from those in this paper.

Software, spreadsheets, and other resources for a broad portfolio of designs, and a host of applications, are available via the SEED Center's website.

5 THOUGHTS FOR THE FUTURE

I believe several aspects of the intersection between big data, simulation, and decision making will be of increasing interest in the near future. Here are a few that I hope will resonate with our simulation community.

5.1 Future Simulation Clients

All too often, the elegance and rigor of having a closed-form or mathematically tractable solution have been touted as advantageous over a simulation modeling approach (Lucas et al. 2015). This ignores the introduction of "type III errors" (Mitroff and Featheringham 1974) that occur when we solve the wrong problem. This bias towards stylized analytical models becomes harder to justify in the face of readily

available big data. For example, when it's not necessary to task someone to go and collect a lot of information because that information is already available, it is harder to justify assuming i.i.d. exponential random variates. Increasingly, the lack of a closed-form solution is not an issue when our software is capable of computing results to a desired level of precision in a small amount of time; this is called *computational tractability* (Lucas et al. 2015). Climate change, economics, transportation, combat, and social dynamics are just a few of the areas where closed-form analytic models will not suffice—computational models are better at capturing the complexity of the underlying systems, and so are better choices for investigating these types of problems.

When stakeholders have complex problems and are studying complex systems, they are not likely to be interested in answers to simple questions. Just as "having" big data from the internet meant that companies found new and exciting things to do with it, we've seen that having big data from simulation experiments offers the opportunity for new and interesting ways of looking at the results. "How should I set up my transportation network?" and "What are the impacts of the affordable care act on health costs and health outcomes?" are much more complicated (and interesting) questions than "What is the expected time-in-system for a customer in an M/M/1 queue with no balking and unlimited buffering?"

The current fascination with big data has several secondary effects. The rapid evolution of data science means that a greater number of simulation and non-simulation professionals will be becoming more adept at scripting, modeling, graphical displays, and statistical analyses. Decision makers may, similarly, be less likely to shy away from using observational or model-driven data to inform their decisions. At the same time, enhancements to methods for rapidly creating, merging, searching, displaying, and analyzing data from large repositories may prove to be useful new tools for the simulation community. Comfort with computerized and computer-based decisions is also increasing in other ways. If we trust a well-written computer program to drive a car (Jaffe 2014), why not trust a well-written computer program for other types of decisions?

5.2 Future Simulation Methods

Most often, we see examples of simulation studies defined to address a specific question. I believe it is time to view simulation-based decision making as an ongoing process, not an end state. Why do we churn up the CPU cycles when we're in the midst of an analysis activity, and then let our computer sit idle for the rest of the time? One intriguing idea is to go back and forth between models of different types or different fidelities, as we seek to learn more about these systems: this is being done in some scientific computing communities, such as computational physics, and it may have interesting parallels in the discrete-event simulation community. Another approach, even if we begin with a specific question, is to generate more output (in a structured manner) so that we are prepared for the next round of questions from our model, and we have been able to identify interesting features in the response surface metamodels that might not have initially been apparent.

There has been a great deal of good work in our community on simulation optimization, ranking and selection, and response surface modeling. But these presuppose that the decision maker knows what questions to ask. I believe that more research is needed on multi-objective procedures, exploitation of parallel computing, adaptive methods, and the design and analysis of large-scale simulation experiments. At the same time, there appear to be new opportunities for some established research areas. Importance sampling becomes potentially more of interest as we pull in real-time data. Can we easily update the state of our simulation, identify branching opportunities, and move forward quickly in parallel? Regarding simulation optimization and other adaptive search techniques, it may be that we should be doing optimization on metamodels, rather than on the simulations themselves—and that we need automated ways of reoptimizing as these metamodels evolve over time.

5.3 Future Simulation Software

Data capture is being automated at an incredible rate from real-world systems, ranging from satellite imagery, to web site navigation, to social network analysis, to engine systems. In the future, I see a growth in automating linkages between real-world data and simulation modeling environments. This increases the potential for using simulation as a real-time decision support and control system, as it has in recent biopharmeceutical applications (Johnston et al. 2008). Major simulation packages may adopt an "app" approach, and use common data exchange protocols and interface protocols to link simulation models with external data sets. If so, the same protocols might also allow the practitioner to easily find or create suitable apps for analysis (e.g., simulation experiments, simulation optimization, ranking & selection, importance sampling) as well as for big data visualization. I expect an increase in the use of adaptive, automated analysis methods.

Simulation software developers should start taking more advantage of cloud computing, coupled with the ability to run models remotely via a web interface. Software developers might consider whether there's an analog to a subscription service for running simulation models, rather than licensing software for individual machines or users. At the server side, intelligent resource allocation ("automated data farming") can take advantage of parallel processor capabilities in stand-alone clusters or in clouds.

The use of smarter computational agents is an area that is ripe for improvement. The medical field has a few applications where intelligent software agents search through large data sets and find correlations. These have led to theories (e.g., on environmental or genealogical links to certain diseases later in life) that can then be examined more thoroughly and tested by medical researchers. Can we do the same with simulation? One way is to construct intelligent agents to search through model-driven data sets, identifying important factors and interesting features in the responses. Another way is to embed some of this capability into our models themselves; for example, rather than relying on calls to random variate generators with fixed parameters, we might allow intelligent agents within our simulation model to access near-real-time big data and assess whether or not these distributional models still appear to be valid.

5.4 But Wait, There's More!

I have listed a few changes that I feel are on the horizon, but if there's one thing that the past few decades have taught us, it's that we never truly know what the future will hold. When the internet got started, we viewed e-mail as a faster alternative to letters, and word-processing as a potential way of cutting down on waste paper—in other words, incremental change instead of revolutionary change. Similarly, when the web got started, we did not envision how this connectedness would change our society. So whatever the future holds, the simulation community should be poised to identify, respond to, and ideally blaze a trail that leverages emerging technologies.

Our simulation community has an important role to play. We have been interested in many of these ideas before they captured the public's attention. Because we have been wrestling with them for years, we already have a rich literature of effective ways to deal with complex problems. If we take steps to push this work out to broader communities, we will help those unfamiliar with the current state-of-the art in simulation avoid reinventing the wheel—or worse, repeating the mistakes of the past. More importantly, we will help jump-start the process of improving decisions that may affect our businesses, our lives, and our planet.

As we look to the future, our simulation community has an important role to play. A data farming approach—where we design experiments that handle big sets of factors, big sets of features, yet are flexible and fast—allows simulation researchers and practitioners to stake out the area of inferential big data. We have the chance to become recognized as the gold standard for model-based decision making within the big data analytics community, and we should seize this opportunity.

ACKNOWLEDGMENTS

This work was supported in part by the USMC Expeditionary Energy Office, and the ONR/NPS CRUSER initiative. Portions are adapted from Sanchez and Wan (2012) and Elmegreen, Sanchez, and Szalay (2014).

REFERENCES

Ankenman, B. E., B. L. Nelson, and J. C. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58:371–382.

Brantley, M. W., L. H. Lee, C.-H. Chen, and A. Chen. 2013. "Efficient Simulation Budget Allocation with Regression". *IIE Transactions* 45:291–308.

Duan, W., B. E. Ankenman, S. M. Sanchez, and P. J. Sanchez. 2015. "Sliced Full Factorial-Based Latin Hypercube Designs as a Framework for a Batch Sequential Design Algorithm". *Technometrics,* forthcoming.

Elmegreen, B. E., S. M. Sanchez, and A. S. Szalay. 2014. "The Future of Computerized Decision Making". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 943–949. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers, Inc.

Grimes, S. 2013, August 7. "Big Data: Avoid 'Wanna V' Confusion". *Information Week*.

Hogan, J. 2014, 9 June. "So Far, Big Data is Small Potatoes". Scientific American Blog Network. Available via http://blogs.scientificamerican.com/cross-check/2014/06/09/so-far-big-data-is-small-potatoes/ [accessed 2 July 2015].

IBM 2015. "What is Big Data?" Available via http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html [accessed 2 July 2015].

Jaffe, E. 2014, 28 April. "The First Look at How Google's Self-Driving Car Handles City Streets". *The Atlantic Citylab*.

Johnston, L., L. Schruben, A. Yang, and D. Zhang. 2008. "Establishing the Credibility of a Biotech Simulation Model". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 822–826. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers, Inc.

Kleijnen, J. P. C. 2007. *Design and Analysis of Simulation Experiments*. New York: Springer.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa. 2005. "A User's Guide to the Brave New World of Designing Simulation Experiments". *INFORMS Journal on Computing* 17 (3): 263–289.

Kleijnen, J. P. C., and W. C. M. van Beers. 2004. "Application-Driven Sequential Designs for Simulation Experiments: Kriging Metamodeling". *Journal of the Operational Research Society* 55:876–883.

Laney, D. 2001, 6 February. "3D Data Management: Controlling Data Volume, Velocity, and Variety". In *Application Delivery Strategies*, Number 949. META Group Inc.

Law, A. M. 2014. *Simulation Modeling and Analysis*. 5^{th} ed. New York: McGraw-Hill.

Loeppky, J. L., L. M. Moore, and B. J. Williams. 2010. "Batch Sequential Designs for Computer Experiments". *Journal of Statistical Planning and Inference* 140 (6): 1452–1464.

Lucas, T. W., W. D. Kelton, P. J. Sánchez, S. M. Sanchez, and B. L. Anderson. 2015. "Changing the Paradigm: Simulation, Now a Method of First Resort". *Naval Research Logistics* 62 (4): 293–303.

MacCalman, A. D., H. Vieira Jr, and T. W. Lucas. 2015. "Second Order Nearly Orthogonal Modified Latin Hypercubes for Exploring Stochastic Simulations". *Journal of Simulation,* forthcoming.

Markoff, J. 2008, 9 June. "Military Supercomputer Sets Record". New York Times.

Marlow, D., and A. Novak. 2013. "Using Discrete-Event Simulation to Predict the Size of a Fleet of Naval Combat Helicopters". In *MODSIM2013, 20th International Congress on Modelling and Simulation*, 2506–2512. Modelling and Simulation Society of Australia and New Zealand.

Mayer-Schönberger, V., and K. Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. 1^{st} ed. New York: Houghton Mifflin Harcourt Publishing Company.

Mitroff, I. I., and T. R. Featheringham. 1974, November. "On Systemic Problem Solving and the Error of the Third Kind". *Behavioral Science* 19 (6): 383–393.

NPR. 2014, 17 May. "Drawing Phony Connections With Mismatched Metrics". Interview, Weekend Edition, NPR News. Available via http://www.npr.org/2014/05/17/313343183/drawing-phoney-connections-with-mismatched-metrics [accessed 2 July 2015].

Sanchez, P. J. 2010, April. "Data Farming". Presentation, CENIC Conference, Corporation for Education Network Initiatives in California. Available via http://harvest.nps.edu/papers/DataFarming.mp4 [accessed 2 July 2015].

Sanchez, S. M. 2014. "Simulation Experiments: Better Data, Not Just Big Data". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. D. Diallo, O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers, Inc.

Sanchez, S. M., and T. W. Lucas. 2002. "Exploring the World of Agent-Based Simulation: Simple Models, Complex Analyses". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yucësan, C. Chen, J. L. Snowdon, and J. Charnes, 116–126. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sanchez, S. M., T. W. Lucas, P. J. Sanchez, C. J. Nannini, and H. Wan. 2012. "Designs for Large-Scale Simulation Experiments, with Applications to Defense and Homeland Security". In *Design and Analysis of Experiments: Special Designs and Applications* (1st ed.), edited by K. Hinkelmann, Volume 3, Chapter 12, 413–441. New York: John Wiley & Sons.

Sanchez, S. M., and P. J. Sanchez. 2005. "Very Large Fractional Factorial and Central Composite Designs". *ACM Transactions on Modeling and Computer Simulation* 15 (4): 362–377.

Sanchez, S. M., P. J. Sanchez, and H. Wan. 2014. "Simulation Experiments: Better Insights by Design". In *Proceedings of the 2014 Summer Simulation Conference*. San Diego, California: The Society for Modeling & Simulation International.

Sanchez, S. M., and H. Wan. 2012. "Work Smarter, Not Harder: A Tutorial on Designing and Conducting Simulation Experiments". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Urmacher, 1929–1943. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers, Inc.

Santner, T. J., B. J. Williams, and W. I. Notz. 2003. *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.

SEED Center for Data Farming. 2015. "SEED Center". Available via http://harvest.nps.edu [accessed 2 July 2015].

The Economist. 2013, 19 October. *Unreliable Research: Trouble at the Lab*. The Economist.

van Beers, W. C. M., and J. P. C. Kleijnen. 2003. "Kriging for Interpolation in Random Simulation". *Journal of the Operational Research Society* 54:255–262.

Vieira Jr, H., S. M. Sanchez, K. H. K. Kienitz, and M. C. N. Belderrain. 2013. "Efficient, Nearly Orthogonal-and-Balanced, Mixed Designs: An Effective Way to Conduct Trade-off Analyses via Simulation". *Journal of Simulation* 7 (4): 264–275.

Vigen, T. 2014. "Spurious Correlations". Available via http://www.tylervigen.com [accessed 2 July 2015].

AUTHOR BIOGRAPHY

SUSAN M. SANCHEZ is a Professor in Operations Research at the Naval Postgraduate School, and Co-Director of the Simulation Experiments & Efficient Designs (SEED) Center for Data Farming. She also holds a joint appointment in the Graduate School of Business & Public Policy. She has a B.S. in Industrial & Operations Engineering from the University of Michigan, and a Ph.D. in Operations Research from Cornell. She has long been active in within the simulation community, including the WSC Board of Directors. Her web page is http://faculty.nps.edu/smsanche/, and her email is ssanchez@nps.edu.