Special Proceedings: ISBN #: 978-81-950383-1-2
24th Annual Conference, 23-27 February 2022; pp 57-68

Computer Simulation: Some Views on Model Development, Experimentation, and Robust Design

K. C. James
Department of Statistics, *Cochin University of Science and Technology*, *Kerala*, *India*

Received: 07 June 2022; Revised: 11 July 2022; Accepted: 13 July 2022

Abstract

A computer simulation is a computation that emulates the behaviour of some real or conceptual systems over time. We conduct experimentation with such models to understand the behaviour of a system. Simulation models are widely used in modern scientific research, education, industry and manufacturing, and public policy matters. These models tend to be extremely complex, often with many factors and sources of uncertainty. The complexity reflected in the system simulation models is characterized by the presence of entity elements that are dynamically created, asynchronous interactions between the entities, the use of shared resources, and connectivity between the entities. Conceptual modelling is a very relevant task in simulation modelling, but is often neglected by analysts. Simulation itself does not serve as an optimization technique. Computer experiment design principles differ from physical experiment design principles, and the three concepts of blocking, replication, and randomization are inessential or irrelevant to computer experiment design. In this paper, a few ideas on how to develop discrete-event simulation models and perform designed experiments are discussed, which helps in better solutions for the analysts. The focus is on some recent developments in the field of simulations which include ideas of visual analytics, data farming, knowledge discovery, and robust design.

Key words: Computer simulation; Conceptual modelling; Visual analytics; Knowledge discovery; Robust design.

1. Introduction

Simulation involves building a model that mimics the behavior of a system, experimenting with the model to create observations of these behaviours, and attempting to comprehend, summarise, and/or generalise these behaviours. Simulation also entails testing and comparing various designs, as well as validating, explaining, and supporting simulation outcomes and research recommendations in many cases. Simulations can also be classified based on how they are implemented. The implementation methodologies for continuous system simulation, Monte Carlo simulation, discrete-event simulation (DES), hybrid simulation, and agent-based simulation are all different.

Simulation has several advantages. Many integrated operations systems are subject to both variability and complexity (combinatorial and dynamic). Because it is difficult to anticipate the performance of systems that are subject to any one of variability, interconnectedness, or complexity, predicting the performance of operations systems that are potentially exposed to all three is extremely difficult, if not impossible.

Corresponding Author : K C James
Email: jamesmech@cusat.ac.in

Simulation models, on the other hand, can explicitly depict a system's unpredictability, interconnection, and complexity. As a result, a simulation can be used to anticipate system performance, evaluate various system designs, and assess the impact of different designs and policies on system performance.

1. Simulation allows researchers to investigate and experiment with the internal interactions of a complex system or a subsystem within one.
2. Informational, organisational, and environmental changes can be simulated, and the impact on the model's behavior can be determined.
3. Because simulation resembles what happens in an actual system or what is perceived for a system in the design stage, it appeals to clients instinctively.

A simulation's output data should be identical to the outputs that may be recorded from the real system. Furthermore, theoretically solvable models can be used to create a simulation model of a system that does not rely on dubious assumptions (such as the same statistical distribution for every random variable). Simulation is frequently the technique of choice in problem-solving for these and other reasons. Simulation models, unlike optimization models, are "run" rather than "solved." The model is run and the simulated behaviour is evaluated given a specific set of input and model variables.

Computer simulation is applied in a large number of industrial systems that include

- Manufacturing systems
- Public systems: health care, military, natural resources
- Transportation systems
- Construction systems
- Restaurant and entertainment systems
- Business process reengineering/management
- Food processing
- Computer system performance

In a recent attempt, Discrete event simulation (DES) is used even to help livestock farmers, by simulating potential growth strategies and observing the impact in relation to existing farm processes (Gittins *et al*., 2020). To know more about a wide variety of application areas of simulations, readers may refer to any Winter Simulation Conference proceedings of recent years.

In the following sections, some considerations required for effective simulation modelling are discussed. In section 2 few ideas of conceptual modelling, a largely forgotten area by many modellers are presented. Section 3 includes a few thoughts on simulation experimentations and some recent developments in this area. In section 4 few ideas of robust design relevant for simulations are discussed.

2. Conceptual Modelling, The Soft Operations Research Exercise

Simulations involve a number of steps as summarised in figure 1

1. A conceptual model: a description of the model that is to be developed

2. A computer model: the simulation model implemented on a computer
3. Improvements and/or understanding: derived from the results of the experimentation
4. An improvement in the real world: obtained from implementing the improvements and/or understanding gained

Although effective conceptual modeling is vital, it is also the most difficult and least understood stage in the modeling process (Law, 2015). Conceptual modelling is a very relevant task in simulation modelling, but is often neglected by analysts. The author believes that many of statistical modelling tasks also require good conceptual modelling exercises, but largely this step is ignored. It can be treated as a soft operations research exercise and a good conceptual model significantly enhances the accuracy and acceptability of the computer model. It minimizes the likelihood of incomplete, unclear, inconsistent, and wrong requirements. It helps build the credibility of the model and forms the basis for model verification and guides model validation. It helps experimentation by expressing the modeling objectives, and model inputs and outputs.

Conceptual modelling consists of the following sub-activities (Robinson, 2011):

- Develop an understanding of the problem situation
- Determine the modelling objectives
- Design the conceptual model: inputs, outputs, and model content

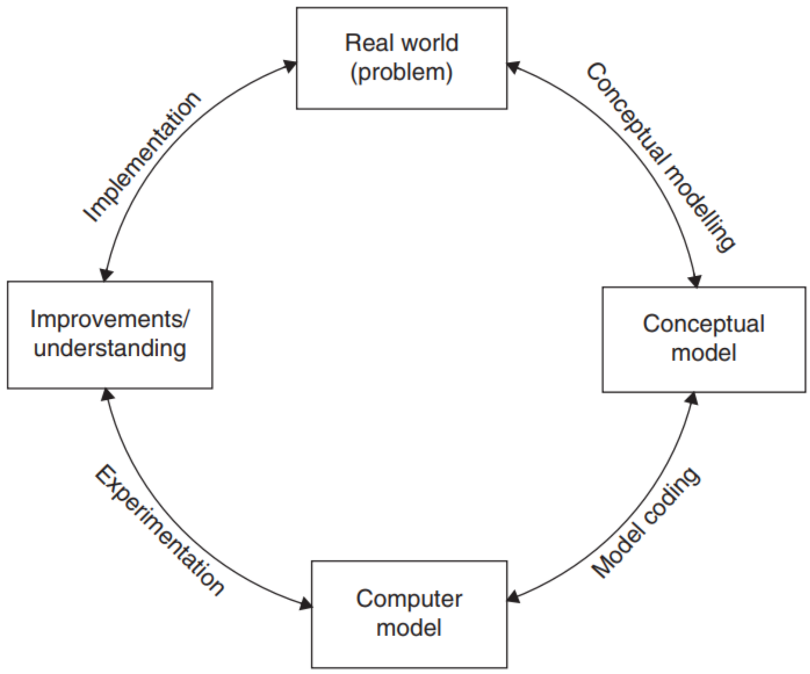

Figure 1: Simulation model development process (Source: Robinson, 2014)

For effective simulation, during conceptual modellers consideration should be given to

- Subject matter experts
- Organizing and structuring knowledge
- Adoption of “soft” OR approaches (Rosenhead and Mingers, 2004)
- Dimensions for determining the performance of a conceptual
- Identifying, adapting, and developing modeling frameworks

- Model simplification methods
- Model representation methods
- Use of software engineering techniques

A model should be created for a specific reason, and its validity should be determined in relation to that purpose. A constructed model should typically be a parsimonious model, which means it is as simple as feasible while yet accomplishing its goal. Furthermore, a model's accuracy (also known as model fidelity) should normally be limited to what is required to meet the model's function or purpose. If the goal of a model is to answer a range of questions, the model's validity must be assessed separately for each question. Soft operations research/problem structuring approaches have been used by OR practitioners for many years. When problem structuring approaches are used in combination with analytical approaches such as computer simulation, it is sensible to regard the two approaches as complementary(Pidd, 2007).

Model developers and users, decision-makers who use information derived from model results, and persons who are affected by model-based decisions are all interested in whether a model is valid. Strict verification and validations of conceptual models and computer models are essential to develop confidence in the customer's mind.

3. Experimentation and Knowledge Discovery

The focus of early experimental designs was mostly on physical experiments. Traditionally, simulation experts conduct experiments on the computer model for predetermined system specifications focusing on single model aspects and specific analysis questions. Modellers compare multiple system configurations and choose the one which presents the best system performance. Computer experiment design principles differ from physical experiment design principles, and the three concepts of blocking, replication, and randomization are inessential or irrelevant to computer experiment design. The "space-filling property" is commonly used in deterministic computer models based on partial differential equations to cover the experimental region with design points. Such analysis are nowadays finding value even in stochastic discrete event simulations.

Recent developments in big data analytics have also influenced experimentation and analysis of simulation. In the following subsections, some of these developments are presented. Subsection 3.2 discuss how big data concepts have a different flavor in simulations; 3.3 discusses data farming; 3.4 describes briefly various tools used in the knowledge discovery process.

3.1. Experimentation: traditional simulation vs. knowledge discovery

Traditionally, simulation studies make use of several runs on predetermined experimental scenarios. Recent developments in the "Knowledge Discovery" process as applied to simulations focus on the use of large sets of experimental data from simulations to find out hidden patterns for useful interpretations of the system. Table 1 below differentiates these approaches. Figure 2 shows a procedure for knowledge discovery (Feldkamp *et al*., 2015a) which make use of concepts of big data, data farming, data mining and visualisations.

Table 1: Traditional Simulation Vs Knowledge discovery

Traditional Simulation	Knowledge discovery
• Project goals formulated beforehand • Simulation study is carried out by comparing predetermined scenarios that the user already had in mind before. • The target function has to be set up beforehand for optimization. • Analyst usually takes an educated guess which input parameters (factors) might be influential on the project scope.	• Use a combination of data mining and visual analysis • Find hidden and potentially interesting • Knowledge generated outside of prior defined project scopes

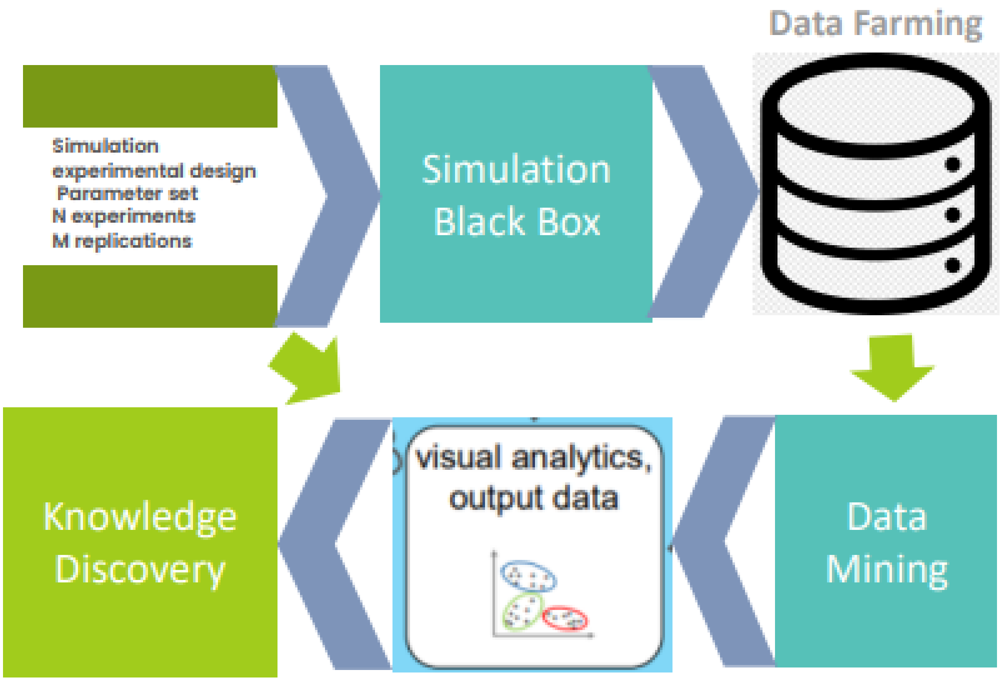

Figure 2: Knowledge discovery Process

3.2. Big data: the 3 (or more) V's have a different flavor in simulation

The term "Big Data" refers to a large amount of data that can't be stored or processed by conventional data storage or processing equipment. Big Data is generated on a massive scale, and it is being processed and analysed by many global corporations in order to unearth insights and enhance their businesses. Simulation experiments can generate a huge amount of data if the experiment considers a large number of factors and levels. This data can be considered as having the "V"(volume, velocity, veracity, etc) characteristics of conventional big data. However, we can find some subtle differences as shown in Table 2.

Table 2: Big data vs simulation data: The 3 (or more) V's have a different flavor in simulation

Big Data Characteristics	Big Data in Simulation
Analysts usually will not have any control over the data. Data may come from sources like information generated every second from social media, cell phones, cars, credit cards, M2M sensors, images, video, *etc*	Velocity and volume are partially controlled by the analyst. Analyst 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.
Data is not under the control of analysts and data may be Structured Data, Semi-Structured Data, or Unstructured Data. Most big data contain lots of missing data, errors, and incompatible data formats	The variety does not include many of the problems that we find with observational data (*e.g.*, incompatible data formats, inconsistent data semantics).
A large variety may be seen in big data	Simulations can have a variety of types of inputs and responses

3.3. Data farming

"A 'data farming' metaphor captures the notion of purposeful data generation from simulation models. Large-scale experiments let us grow the simulation output efficiently and effectively. We can use modern statistical and visual analytic methods to explore massive input spaces, uncover interesting features of complex simulation response surfaces, and explicitly identify cause-and-effect relationships"(Sanchez, 2018). Data can be grown in simulation experiments to extract many useful insights

- *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.*
- *they "grow" data from their models, but in a manner that facilitates identifying useful information.*
- *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.*

3.4. Visual analytics

Results of simulation experiments can be shown visually by means number of charts. Proper selection of charts can reveal interesting patterns. Visual Analytics is a key technique of a knowledge discovery process for discrete event simulations.

- *In traditional simulation studies, techniques such as animation of process flow, time plots, and graphs of selected outputs are used for visually representations*
- *In Visual Analytics, Data mining algorithms and visualization are used to build up knowledge and draw conclusions from it.*

- *This approach is advantageous because the human mind is able to identify patterns and relations in visual representations quickly.*

Feldkamp *et al.*(2015a, b) lists the following visual tools combined with data mining tools to extract patterns in simulation experiment data(see Table 3). Few case studies on use of visual analytics in simulations in various application areas are found in recent literature. Table 4 lists some of them and the types of visual analytics representations used in such case studies.

Table 3: Knowledge discovery tools used in simulations

Visualization tools	Data Mining methods
• Box plots • Histograms	• Measures of central tendency and variation • Distribution analysis
• Scatter matrix and plots • Parallel coordinate plots • Spider charts	• Multidimensional patterns • Linear regression • Logistic regression
• Flowcharts • Heatmaps • Network graphs	• Correlation tables • Association rules • Bayesian networks • Classification trees

Table 4: Visualization tools and data mining methods for knowledge discovery in some recent literature

No	Reference	Author	Tools used
1	Using Simulation as a Knowledge Discovery Tool in An Adversary C2 Network	Ntuen *et al*, 2009	A hierarchical cluster tree
2	Knowledge Discovery Based Simulation System in Construction	Emad E, 2011	Fuzzy Clustering
3	Knowledge Discovery In Simulation Data: A Case Study Of a Gold Mining Facility	Feldkamp *et al.*, 2016	Correlation matrix of input and output parameters, Matrix scatter plot of selected parameters, Clustering, Linear regression model Radarplots 3D Scatterplot
4	Interactive Visual Analysis of Large Simulation Ensembles	Matkovic *et al*, 2015	Scatterplot Histogram
5	Visual Analytics of Manufacturing Simulation Data.	Feldkamp *et al*, 2015a&b	Correlation matrix of input and output parameters Matrix scatter plot of selected parameters Clustering Linear regression model Radarplots 3D Scatterplot

6	Improving Navy Recruiting with Data Farming	Hogarth *et al*, 2016	Scatter plots Regression models Partition trees
7	A data farming analysis of a simulation of Armstrong's stochastic salvo model	Kesler *et al*, 2019	Pairwise scatter plot Partition tree

4. Robust Design

Robust design is a system optimization and enhancement approach based on the idea that a system shouldn't be judged solely on its average performance. A "good" system must be somewhat insensitive to uncontrollable causes of variation in the system's environment, in addition to demonstrating acceptable mean performance. The purpose of robust design is to help people make better decisions

- it focuses the decision-making process on factors that are controllable in practice;
- it identifies levels and consistency of performance based on those controllable factors;
- robust configurations are more likely to yield better engineering implementations;
- those real-world implementations have in many cases achieved greater reliability and performance at a lower cost.

In the simulation context, robust design can be viewed from different perspectives as depicted in Table 5. This table shows a comparison of experiments on real systems vs computer simulation. Please note that sometimes analogous/physical models/prototypes are easier to experiment with and may draw better results. Because of the expense, effort, and dangers involved in making and observing changes in a real system, one view is that simulation is largely used as a surrogate for a real system; another view is that robust design is an inherent element of the simulation process.

Table 5: Robust design - comparison of experiment with a real system vs simulation

Robust Design: Experiment with a real system	**Robust Design: Experiment with a Simulation model**
Conduct experiments on the real system	Simulation is largely used as a surrogate for a real system
Expense, effort, and dangers involved in making and observing changes in a real system are considerable	Initial development of models involves lots of time and effort to develop a valid and credible model. Lots of calibration efforts are required to fine-tune the model
Changes in the system for experimentation are often difficult and risky	Changes to the model and experimentation are relatively easy.
A large number of inputs and factor levels may not be physically possible always	Large number of input and factor levels can be studied with ease
Running/completion of experiments can take long time and effort. Replication is difficult.	Total time to perform an experiment is significantly less. Replication means we get multiple experimental units (runs or batches) to gain a sense of the magnitude of the variability associated with response and replications are easy in a simulation model.

Randomization is used to guard against hidden or uncontrollable sources of bias.	Results from simulation experiments are perfectly repeatable, and randomization is not needed to guard against hidden or uncontrollable sources of bias.
Homogeneous (*i.e.*, constant) variance is commonly assumed for physical experiments	Heterogeneous (*i.e.*, non-constant) variance is pervasive in stochastic simulation. Consequently, we should not view response variability as merely a nuisance for estimating means or other output statistics, but as an important characteristic of the simulation's behavior.
Experiments on real system help system optimization and improvement process that springs from the view that a system should not be evaluated based on mean performance alone	Robust design can be seen as a process of simulation optimization, where the "best" answer is not overly sensitive to small changes in the system inputs. Kleijnen (2017) calls this "robust optimization." If robust configurations are identified, then the actual results are more likely to conform to the anticipated results after implementation.

Accuracy and Precision: The predicted value of the distribution of outcomes in relation to some desired aim is referred to as accuracy. For example, if our goal is to determine an object's genuine weight, a scale that produces readings with a distribution that has the true weight as its anticipated value is considered an accurate scale, even if individual readings differ by a significant amount. The dispersion of the outcome distribution is referred to as precision. It is considered to be an accurate scale if the measurements are firmly grouped. Figure 3 illustrates many possible combinations of accuracy and precision for somebody shooting at a target. Subplot (a) has shots with low precision because most points are spread from their center of mass. Subplot (b) has high accuracy and precision—the center of mass is on-target, without much the spread. Subplot (c) is precise but not accurate. Many other possibilities are also there.

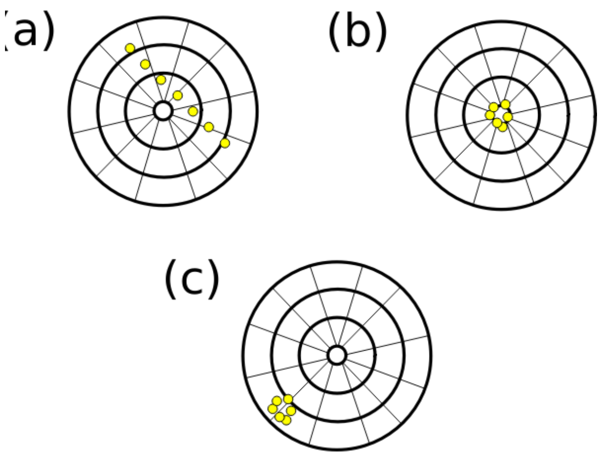

Figure 3: accuracy vs precision

Sanchez and Sanchez (2020) illustrate robustness with a nice example to make it clear that robustness is not solely determined by either the accuracy or the precision of the outcomes relative to the target. Consider response distributions for alternate configurations as depicted in Figure 4. Because the mean response for A is perfectly on target, it is the most accurate. If we were looking for the most accurate system, A would be the best option. System B, on the other hand, is a close second in terms of mean and precision—due to its smaller variance, it is far more likely to produce results close to. Based on this example, we could even argue that the means do not justify the ends in terms of robustness. Both C and D have mean performance above the target value, but when accuracy is taken into account, option D is significantly more likely to be farther from the target. The conclusion here is that robustness is not solely determined by either the accuracy or the precision of the outcomes relative to the target. Tradeoffs may be necessary.

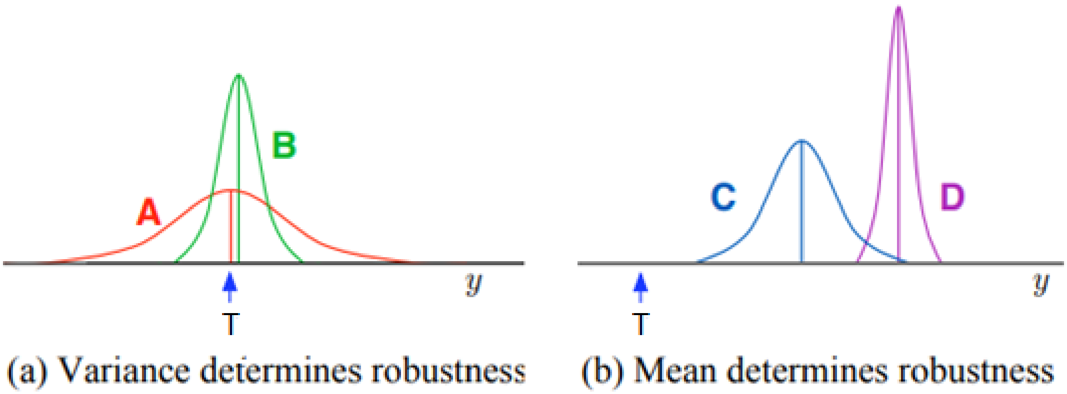

Figure 4: comparison of the robustness of systems A, B, C, and D

4.1. Loss function

Robust design analysis makes use of loss functions described by Taguchi. The quality loss function estimate costs associated when the product or process characteristics are shifted from the target value. Such functions help to assess the degree of risk associated with having outcomes that deviate significantly from the specified target. Risk should be non-negative, so loss functions are monotonically non-decreasing as the magnitude of deviations from the target increases. However, a loss function can be asymmetric about the target.

One such loss function commonly used is the quadratic loss function as given below.

$$L(y) = k(y - T)^2$$

where y is the observed outcome based on input x, T is the target value, and k is a scaling constant that is often used to adjust the loss to cost. When k is set to 1, it is referred to as scaled loss.

We get configurations with low loss when responses are consistently (as measured by variance) close to the target. We accept a tradeoff for small expected deviations from the target with sufficient improvement in consistency of the outcomes, or vice-versa and the variance is non-homogeneous.

4.2. Robust analysis and optimization with simulation metamodels

Computer simulation models of proposed or existing real systems are frequently used to make design decisions. Because it is impractical to build several prototype versions of the real system, or because the cost or other constraints prevent experimenting with the real system, analysts use the simulation model as a surrogate. As these models can be fairly complicated, simpler approximations are frequently created; models of the model, or metamodels. (Kleijnen, 2017). Simulation metamodels can take many forms like multiple regression, partition trees, and forests, or kriging.

Quadratic loss function may be used to fit metamodels of loss directly or fit separate metamodels for the mean and the variability. We often find it convenient to fit the standard deviation as our measure of variability, since it is on the same scale as the mean, but other options such as variance or log(variance) are possible. Fitting separate metamodels help identify which factors, interactions, or non-linear terms are the key causal drivers of average performance and variability. Numerical examples are presented in Sanchez and Sanchez (2020).

5. Conclusion

In this paper we discuss a number of issues related to simulations: how to develop discrete-event simulation models and perform designed experiments to identify significant variables, thus helping simulation optimization searches for optimal solutions. The focus is on some recent developments in the field of simulations which include ideas of visual analytics, data farming, knowledge discovery, and robust design.

Concepts in robust design and analysis in the context of simulation demonstrate how robustness often changes our perspective when contrasted with simulation optimization approaches. Robust solutions can be designed to yield consistently good performance even in the face of uncertainty and uncontrollable factors by incorporating those aspects of the system into the problem formulation.

The process of conceptual modelling is sometimes neglected by analysts and obviously, this can impact the credibility of the model. The author feels that there should be more research connecting the field of statistical modelling with soft operations research and soft systems methodologies. Research on data science tools such as visualisations and data mining applications is sparse in simulation literature and is an open area that requires more research inputs.

References

Emad, E. (2011). *Knowledge Discovery Based Simulation System in Construction, Unpublished Ph.D. Thesis*, Concordia University.

Feldkamp, N., Bergmann, S. and Strassburger, S. (2015a). Knowledge discovery in manufacturing simulations. *In Proceedings of the 2015 ACM SIGSIM PADS Conference*, 3–12.

Feldkamp, N., Bergmann, S. and Strassburger, S. (2015b). Visual analytics of manufacturing simulation data. *In Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 779–790. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Feldkamp, N., Bergmann, S. and Strassburger, S. (2016). Knowledge Discovery In Simulation Data: A Case Study Of a Gold Mining Facility. *In Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 779–790. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gittins, P., McElwee, G. and Tipi, N. (2020). Discrete event simulation in livestock management. *Journal of Rural Studies*, Volume 78, August 2020, Pages 387-398.

Hogarth A. R., Lucas T. W., and McLemore C.S. (2016). Improving navy recruiting with data farming, Proceedings of the 2016 Winter Simulation Conference

Kesler, G., Lucas, T. W. and Sanchez, P. J. (2019). A data farming analysis of a simulation of Armstrong's stochastic salvo model, *Proceedings of the 2019 Winter Simulation Conference*

Kleijnen, J. P. C. (2017). Design and analysis of simulation experiments: a tutorial. *In Advances in Modeling and Simulation*, edited by A. Tolk, J. Fowler, G. Shao, and E. Yucesan, 135–158. Cham, Switzerland: Springer International Publishing AG.

Law, A. M. (2015). *Simulation Modeling and Analysis*, 5th edition. McGraw Hill, Boston.

Matkovic, K., Gracanin, D., Jelović, M. and Hauser, H. (2015). Interactive visual analysis of large simulation ensembles, *Proceedings of the 2015 Winter Simulation Conference*.

Ntuen, C. A., Alabi1, O. A., Seong, Y. and Park, E. H. (2009). Using Simulation as a Knowledge Discovery Tool in An Adversary C2 Network, *Proceeding of 14th International Command & Control Research and Technology Symposium C2 and Agility*, Washington, DC, June 15-17

Pidd, M. (2007). Making sure you tackle the right problem: linking hard and soft methods in simulation practice, *Proceedings of the 2007 Winter Simulation Conference*, S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

Robinson, S. (2011). Conceptual modeling for simulation. *In Encyclopedia of Operations Research and Management Science*, Edited by J.J. Cochran, New York: Wiley.

Robinson, S. (2014). *Simulation: The Practice of Model Development and Use*, Palgrave Macmillan, 2014

Rosenhead, J. and Mingers, J. (2004). Problem structuring methods in action. *European Journal of Operational Research*, **152**, 530 – 554

Sanchez, S M. (2018). Data farming: better data, not just big data, *Proceedings of the 2018 Winter Simulation Conference*, M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

Sanchez, S. M. and Sanchez, P. J. (2020). Robustness revisited: simulation optimization viewed through a different lens, *Proceedings of the 2020 Winter Simulation Conference* K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.