

A Contribution to Validating of Simulation Models in Production and Logistics

Von der Fakultät für Ingenieurwissenschaften, Abteilung Maschinenbau und
Verfahrenstechnik der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades

eines

Doktors der Ingenieurwissenschaften

Dr.-Ing.

genehmigte Dissertation von

M. Sc. Nan Liu
aus
Xi'an, China

Tag der mündlichen Prüfung: 23.Juli.2015

Promotionskommission:

Vorsitzender: Prof. Dr.-Ing. Andrés Kecskeméthy

1. Berichterstatter: Prof. Dr.-Ing. Bernd Noche

2. Berichterstatter: Prof. Dr.-Ing. Uwe Bracht

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Technische Logistik der Universität Duisburg-Essen und als Mitarbeiter bei SDZ GmbH in Dortmund.

Mein besonderer Dank geht an meinen Doktorvater, Herr Prof.Dr.-Ing. Bernd Noche, der mir die größte Freiheit und Unterstützung gegeben hat. Herrn Prof. Dr.-Ing. Uwe Bracht von Technische Universität Clausthal möchte ich mich für die Übernahme des Korreferats sehr herzlich bedanken. Herrn Prof. Dr.-Ing. Andrés Kecskeméthy, Leiter des Lehrstuhles für Mechanik und Robotik, danke ich für die Übernahme des Vorsitzes der Prüfungskommission.

Der größte Dank geht jedoch an meine Frau, meine Eltern, und meine Großmutter, die mir immer die Kraft gegeben haben, meine Träume zu verwirklichen. Euch möchte ich diese Arbeit widmen.

Düsseldorf, im August 2015

Nan Liu

Abstract: in this dissertation a new way of thinking in the validation of simulation model methods in production and logistics systems is presented. The focus is on finding the most important parameters for the validation of simulation models, the relationships between these parameters in the form of correlations which influence the complexity of the models, the required availability of data for these parameters, and the uncertainty which is largely dependent on data availability. Further, a short assessment is made of sequential bifurcation as a method for model validation in the context of finding important parameters, reducing complexity, and exploring uncertainties.

TABLE OF CONTENTS

Chapter 1 the problem of simulation model validation	11
1.1 A short history of simulation in production and logistics	11
1.2 Definition of ‘simulation’ and its use in production and logistics	11
1.3 Model validation in production and logistics systems	17
1.4 The problem of validation	18
Chapter 2 the problem of simulation model validation paradigms	28
2.1 Model validation paradigms	28
2.2 The model validation paradigm of Osman Balci	28
2.3 The model validation paradigms of Shannon, Law and Kelton, and Banks.....	33
2.4 The model validation paradigm by the Arbeitsgemeinschaft Simulation (ASIM)	34
2.5 The model validation paradigm of Rabe et al.	35
2.6 The model validation paradigm by Sargent	37
2.7 Model validation paradigms in production and logistics	40
Chapter 3 Complexity, uncertainty and data availability in the model validation of production and logistics systems.....	44
3.1 The three aspects of discrete-event simulation in model validation	44
3.1.1 Three aspects of model validation	46
3.1.2 Model complexity.....	46
3.1.3 Model uncertainty.....	46
3.1.4 Data availability	47
3.2 System complexity, and complexity in logistics and production systems.....	47
3.2.1 Complexity and system complexity.....	47
3.2.2 The measurement of system complexity	50
3.2.3 Complexity in a simulation model: correlation	51
3.3 Uncertainty in simulation models	60
3.3.1 Types of uncertainty.....	60
3.3.2 Examples of model uncertainty.....	61
3.3.3 The coefficient of variation	69
3.4 Data in model validation	72
3.5 The three factors of simulation model validation in production and logistics.....	74
Chapter 4 Current techniques of model validation and a new paradigm including sequential bifurcation	75
4.1 Current techniques of model validation	75
4.1.1 Current validation techniques by Sargent	75
4.1.2 Current validation techniques by Balci	82

4.2 The sequential bifurcation method for simulation model validation in production and logistics	97
Chapter 5 Conclusion	129
Bibliography	131

LIST OF FIGURES

FIGURE 1-1: WILLIAM SEALY GOSSET’S PAPER WITH MANUAL SIMULATION UNDER THE PSEUDONYM “STUDENT”	12
FIGURE 1-2: VARIOUS TYPES OF PRODUCTION PROCESSES	14
FIGURE 1-3: DISCRETE-EVENT SIMULATION CHARACTERISTICS	19
FIGURE 1-4: A SIMPLE M/M/1 MODEL IN DOSIMIS-3	22
FIGURE 2-1: THE LIFE CYCLE OF A SIMULATION STUDY BY OSMAN BALCI	29
FIGURE 2-2: A HIGH-LEVEL PROCEDURE FOR PROBLEM FORMULATION	31
FIGURE 2-3: A CLASSIFICATION OF SYSTEM CHARACTERISTICS BY SHANNON	32
FIGURE 2-4: ITERATIVE CONCEPTUAL MODEL DEVELOPMENT BY PACE	32
FIGURE 2-5: COMMUNICATIVE MODEL FORMS BY NANCE	33
FIGURE 2-6: SIMULATION PROCEDURE PARADIGMS BY SHANNON, LAW AND KELTON, BANKS ET AL	34
FIGURE 2-7: SIMULATION STUDY PROCEDURE BY ASIM	35
FIGURE 2-8: SIMULATION PROCEDURE PROPOSED BY RABE, SPIECKERMANN AND WENZEL	36
FIGURE 2-9: MODELLING PROCESS BY SARGENT	37
FIGURE 2-10: THE ITERATIVE PROCESS OF MODEL VALIDATION BY BANKS	38
FIGURE 2-11: MODEL CREDIBILITY AND COST	39
FIGURE 3-1: COMPLEXITY, UNCERTAINTY AND LOW DATA AVAILABILITY LEAD TO UNCERTAINTY IN VALIDATION AND PREDICTION	45
FIGURE 3-2: THREE CLASSES OF SYSTEM BY WEAVER	48
FIGURE 3-3: COMPLEXITY IN LOGISTICS AND PRODUCTION SYSTEMS	49
FIGURE 3-4: A SMALL PRODUCTION SYSTEM	55
FIGURE 3-5: SIMULATION MODEL OF THE SMALL PRODUCTION SYSTEM	55
FIGURE 3-6: TYPE II CORRELATION BY CORRELOGRAM	57
FIGURE 3-7: TYPE III CORRELATION BY CROSS-CORRELATION PLOT	58
FIGURE 3-8: TYPE III CORRELATIONS BY SCATTERPLOT	59
FIGURE 3-9: SIMULATION MODEL OF A SMALL PRODUCTION SYSTEM	60
FIGURE 3-10: ONE-ZIPPY SYSTEM (1), TWO-KLUNKIES SYSTEM (2), AND TWO-OLDIES SYSTEM (3)	62
FIGURE 3-11: COMPARING CONSTANT SYSTEM THROUGHPUT TIME AND ZIPPY SYSTEM THROUGHPUT TIME WITH POISSON ARRIVAL AND EXPONENTIAL SERVICE TIME	63
FIGURE 3-12: COMPARING CONSTANT SYSTEM WAITING TIME AND KLUNKY SYSTEM WAITING TIME WITH EXPONENTIAL ARRIVAL TIME AND EXPONENTIAL SERVICE TIME	64
FIGURE 3-13: COMPARING CONSTANT SYSTEM WAITING TIME AND OLDY SYSTEM WAITING TIME WITH EXPONENTIAL ARRIVAL TIME AND EXPONENTIAL SERVICE TIME	65
FIGURE 3-14: COMPARING CONSTANT SYSTEM WAITING TIME AND OLDY SYSTEM WAITING TIME WITH EXPONENTIAL ARRIVAL TIME AND EXPONENTIAL SERVICE TIME AND NO BUFFER	66
FIGURE 3-15: COMPARING ZIPPY SYSTEM THROUGHPUT WITH DIFFERENT CONFIGURATIONS	67
FIGURE 3-16: COMPARING KLUNKY SYSTEM THROUGHPUT WITH DIFFERENT CONFIGURATIONS	68
FIGURE 3-17: COMPARING OLDY SYSTEM THROUGHPUT WITH DIFFERENT CONFIGURATIONS	68
FIGURE 3-18: THROUGHPUT TIMES OF MODEL 1	71
FIGURE 3-19: THROUGHPUT TIMES OF MODEL 2	71
FIGURE 4-1: TURING TEST	81
FIGURE 4-2: MODEL VALIDATION PARADIGM WITH SEQUENTIAL BIFURCATION	102
FIGURE 4-3: A REALISTIC PRODUCTION AND TRANSPORTATION SYSTEM WITH THE USE OF SEQUENTIAL BIFURCATION ...	115

LIST OF TABLES

TABLE 4-1: MODEL VALIDATION TECHNIQUES.....	83
TABLE 4-2: SORTED SEQUENCE OF THE VARIABLES OF MODEL 1	105
TABLE 4-3: EXPERIMENT 1 - TWO SCENARIOS OF INPUT VARIABLES OF MODEL 1.....	106
TABLE 4-4: RESULTS OF EXPERIMENT 1 OF MODEL 1.....	106
TABLE 4-5: FIRST BIFURCATION OF VARIABLES OF MODEL 1.....	107
TABLE 4-6: EXPERIMENT 2 – SCENARIO 1 AND SCENARIO 2 AFTER FIRST VARIABLES BIFURCATION OF MODEL 1.....	107
TABLE 4-7: RESULTS OF EXPERIMENT 2 OF MODEL 1.....	107
TABLE 4-8: EXPERIMENT 3 – PART 1 - MAIN EFFECT INVESTIGATION OF MODEL 1.....	108
TABLE 4-9: EXPERIMENT 3 – PART 2 - MAIN EFFECT INVESTIGATION OF MODEL 1.....	108
TABLE 4-10: EXPERIMENT 3 – SCENARIO 1 AND SCENARIO 2 AFTER SECOND VARIABLES BIFURCATION OF MODEL 1	108
TABLE 4-11: RESULTS OF EXPERIMENT 3 OF MODEL 1.....	108
TABLE 4-12: EXPERIMENT 4 – VARIABLES OF FIRST PART OF EXPERIMENT 2 AND 3 FIXED TO LOW VALUES.....	109
TABLE 4-13: EXPERIMENT 4 - THIRD BIFURCATION OF VARIABLES IN THE SECOND PART OF EXPERIMENT 3.....	109
TABLE 4-14: SCENARIO 1 AND SCENARIO 2 AFTER THIRD VARIABLES BIFURCATION OF MODEL 1.....	109
TABLE 4-15: RESULTS OF EXPERIMENT 4 OF MODEL 1.....	110
TABLE 4-16: EXPERIMENT 5 - VARIABLES OF FIRST PART OF EXPERIMENT 2, 3 AND 4 FIXED TO LOW VALUES.....	110
TABLE 4-17: EXPERIMENT 5 – FOURTH BIFURCATION OF VARIABLES IN THE SECOND PART OF EXPERIMENT 4.....	110
TABLE 4-18: SCENARIO 1 AND SCENARIO 2 AFTER FOURTH VARIABLES BIFURCATION OF MODEL 1.....	110
TABLE 4-19: RESULTS OF EXPERIMENT 5 OF MODEL 1.....	111
TABLE 4-20: IMPORTANT INPUT VARIABLES IN THE SECOND PART OF EXPERIMENT 2 OF MODEL 1.....	111
TABLE 4-21: EXPERIMENT 6 - VARIABLES OF FIRST PART OF EXPERIMENT 2, 3, 4 AND 5 FIXED TO LOW VALUES.....	111
TABLE 4-22: EXPERIMENT 6 - FIFTH BIFURCATION OF VARIABLES IN THE FIRST PART OF EXPERIMENT 2.....	111
TABLE 4-23: SCENARIO 1 AND SCENARIO 2 AFTER FIFTH VARIABLES BIFURCATION OF MODEL 1	112
TABLE 4-24: RESULTS OF EXPERIMENT 6 OF MODEL 1.....	112
TABLE 4-25: EXPERIMENT 7 - VARIABLES OF FIRST PART OF EXPERIMENT 2, 3, 4, 5 AND 6 FIXED TO LOW VALUES	112
TABLE 4-26: EXPERIMENT 7 - SIXTH BIFURCATION OF VARIABLES IN THE SECOND PART OF EXPERIMENT 6.....	112
TABLE 4-27: SCENARIO 1 AND SCENARIO 2 AFTER SIXTH VARIABLES BIFURCATION OF MODEL 1	113
TABLE 4-28: RESULTS OF EXPERIMENT 7 OF MODEL 1.....	113
TABLE 4-29: FORTY TWO INPUT VARIABLES IN MODEL 2.....	116
TABLE 4-30: SORTED SEQUENCE OF TWENTY-NINE INPUT VARIABLES AFFECT THE SYSTEM THROUGHPUT IN MODEL 2....	116
TABLE 4-31: EXPERIMENT 1 - TWO SCENARIOS OF INPUT VARIABLES OF MODEL 2.....	117
TABLE 4-32: RESULTS OF EXPERIMENT 1 OF MODEL 2.....	117
TABLE 4-33: FIRST BIFURCATION OF VARIABLES OF MODEL 2.....	117
TABLE 4-34: EXPERIMENT 2 – SCENARIO 1 AND SCENARIO 2 AFTER FIRST VARIABLES BIFURCATION OF MODEL 2	118
TABLE 4-35: RESULTS OF EXPERIMENT 2 OF MODEL 2.....	118
TABLE 4-36: EXPERIMENT 3 – PART 1 - MAIN EFFECT INVESTIGATION OF MODEL 2.....	119
TABLE 4-37: EXPERIMENT 3 – PART 2 - MAIN EFFECT INVESTIGATION OF MODEL 2.....	119
TABLE 4-38: EXPERIMENT 3 – SCENARIO 1 AND SCENARIO 2 AFTER SECOND VARIABLES BIFURCATION OF MODEL 1	120
TABLE 4-39: RESULTS OF EXPERIMENT 3 OF MODEL 1.....	120
TABLE 4-40: EXPERIMENT 4 – VARIABLES OF FIRST PART OF EXPERIMENT 2 AND 3 FIXED TO LOW VALUES.....	120
TABLE 4-41: EXPERIMENT 4 - THIRD BIFURCATION OF VARIABLES IN THE SECOND PART OF EXPERIMENT 3.....	120
TABLE 4-42: SCENARIO 1 AND SCENARIO 2 AFTER THIRD VARIABLES BIFURCATION OF MODEL 2.....	121
TABLE 4-43: RESULTS OF EXPERIMENT 4 OF MODEL 2.....	121
TABLE 4-44: EXPERIMENT 5 - VARIABLES OF FIRST PART OF EXPERIMENT 2, 3 AND 4 FIXED TO LOW VALUES.....	122
TABLE 4-45: EXPERIMENT 5 – FOURTH BIFURCATION OF VARIABLES IN THE SECOND PART OF EXPERIMENT 4.....	122
TABLE 4-46: SCENARIO 1 AND SCENARIO 2 AFTER FOURTH VARIABLES BIFURCATION OF MODEL 2.....	123
TABLE 4-47: RESULTS OF EXPERIMENT 5 OF MODEL 2.....	123
TABLE 4-48: IMPORTANT INPUT VARIABLES IN THE SECOND PART OF EXPERIMENT 2 OF MODEL 2.....	123
TABLE 4-49: EXPERIMENT 6 - FIFTH BIFURCATION OF VARIABLES IN THE FIRST PART OF EXPERIMENT 2.....	123
TABLE 4-50: SCENARIO 1 AND SCENARIO 2 AFTER FIFTH VARIABLES BIFURCATION OF MODEL 2	124
TABLE 4-51: RESULTS OF EXPERIMENT 6 OF MODEL 2.....	124
TABLE 4-52: EXPERIMENT 7 - VARIABLES OF FIRST PART OF EXPERIMENT 2, 3, 4, 5 AND 6 FIXED TO LOW VALUES.....	124

TABLE 4-53: EXPERIMENT 7 - SIXTH BIFURCATION OF VARIABLES IN THE SECOND PART OF EXPERIMENT 6.....	125
TABLE 4-54: SCENARIO 1 AND SCENARIO 2 AFTER SIXTH VARIABLES BIFURCATION OF MODEL 2	125
TABLE 4-55: RESULTS OF EXPERIMENT 7 OF MODEL 2.....	125
TABLE 4-56: EXPERIMENT 8 - MAIN EFFECT VARIABLES.....	126
TABLE 4-57: EXPERIMENT 7 - SEVENTH BIFURCATION OF VARIABLES IN THE SECOND PART OF EXPERIMENT 2	126
TABLE 4-58: SCENARIO 1 AND SCENARIO 2 AFTER SEVENTH VARIABLES BIFURCATION OF MODEL 2.....	126
TABLE 4-59: RESULTS OF EXPERIMENT 8 OF MODEL 2.....	126

LIST OF ACRONYM

AGV	Automated Guided Vehicle
ANOVA	Analysis of Variance
AS/RS	Automated Storage And Retrieval System
BOM	Bill of Material
DES	Discrete-Event Simulation
DOD	Department of Defense
ENIAC	Electronic Numerical Integrator and Computer
GPSS	General Purpose Simulation System
GSP	General Simulation Program
IID	Identical and Independent Data
KPI	Key Performance Indicator
MANOVA	Multivariate Analysis Of Variance
MTBF	Mean Time Between Failure
MTTR	Mean Time To Repair
SKU	Stock Keeping Unit
SME	Subject-Matter Experts
SRM	Storage-Retrieval Machine
TIC	Theil's Inequality Coefficient
WIP	Work In Process

Introduction

As simulation practitioners, we would like to be able to build a simulation model and have a high level of confidence that the results from the model can be reliably applied to real situations. We would also like to be able to judge the validity of models built by others. In much the same way, decision makers using the results from models in important matters need the assurance that those results can be trusted. The validation procedure plays a decisive role in any simulation study in logistics and production systems as people attempt to increase productivity and enhance efficiency, and indeed gain competitiveness by using simulation techniques.

The purpose of this dissertation is to provide a fundamental understanding of the process of validating simulation models in logistics and production systems along with an easy-to-apply method of validation. The focus is on creating a validation procedure that is easy in application and sufficiently efficient. This is achieved by defining an understanding of three aspects in logistics and production systems, which is vital for the validity of the simulation model.

The fundamental properties of discrete-event simulation in logistics and production systems are randomness, dynamism and discreteness.

The results of validation procedure must answer the question: valid or not? If the answer is yes, what kind of risks will the model user take? If no, what are the potential flaws residing in the model?

If variation is the chief problem in dealing with any logistics system, when one validates a simulation model variation is often reflected by a sudden increase or decrease of entities in the system or sub-system. Can such a sudden change indicate an abnormal phenomenon in the system, or is it caused merely by the inherent randomness of the system?

There is fundamentally no difference between the validation of a simulation model and the validation of any logical premise. The prediction power of a model is the most important of any simulation results. Since there is no perfect model, the best model is the real system itself.

In the following chapter, these questions will be discussed.

Chapter 1 the problem of simulation model validation

1.1 A short history of simulation in production and logistics

Production and logistics companies need to reduce time, cost and risk when they plan any investment in new production facilities, roll out new product lines or add new distribution centres to delivery networks or modify existing ones. It is important for decision-makers to be able to predict the consequences and quantify the risks by providing adequate information with sufficient confidence before any investment is made. Therefore, simulation technology has been used in industries to satisfy these challenges. In order for us to understand simulation in production and logistics systems, several things need to be cleared up first:

1. What is simulation?
2. What is a production and logistics system?
3. Why is simulation used in analysing a production and logistics system?
4. How do we measure performance in a production and logistics system for a simulation?

1.2 Definition of ‘simulation’ and its use in production and logistics

There are no universal definitions of simulation in the community. Followings are several most widespread ones.

Banks definition of simulation:

Simulation is the imitation of the operation of the real-world process or system over time. Simulation involves the generation of an artificial history of the system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system that is represented [BCNN05].

The definition given by VDI Guideline 3633 [VDI36] is as follows:

Simulation is the reproduction of a system with its dynamic processes in an experiment-capable model in order to gain knowledge which can be applied to the real system. In particular, the processes evolve over time. In a broader sense, the preparation, execution and evaluation of targeted experiments of a simulation model together is understood as simulation [NaWi10].

In order to gain better understanding of contemporary statues of simulation technique, it is a necessity to review its history. The inception of simulation in industrial processes can be traced back to the year 1908. The mathematician and chemist William Sealy Gosset, who

worked at the Arthur Guinness brewery, used the pseudonym “Student” and published a paper later known as the ‘Student’s t-distribution’. He used a rough form of manual simulation to verify his speculation about the exact form of the probability density function of the Student’s t-distribution.

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

By STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a “population” of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Figure 1-1: William Sealy Gosset’s paper with manual simulation under the pseudonym “Student” [Goss08]

The birth of the first computer, the ENIAC, and the application of the Monte Carlo method on the computer by Ulam, von Neumann, Metropolis and others in the mid-1940s marked a period of fast growth in simulation.

The first tool for systematically building the simulation of an industrial plant, including machines with the status of busy, failed, idle or unavailable, was the general-purpose simulator General Simulation Program (GSP) , developed by British professor Keith Douglas Tocher in the year 1960 [Toch63].

The more well-known General Purpose Simulation System (GPSS) [Gree72], developed by Geoffrey Gordon in the 1960s, was initially used to simulate urban traffic control systems, telephone call interception and switching, airline reservation processing, and steel mill operations.

From the late 1970s to the present, the publishing of many notable textbooks, e.g. by Emshoff and Sisson, Fishman, and Law and Kelton, Kiviat, Villanueva, as well as Markowitz’s development of SIMSCRIPT II.5, Pegden’s SIMAN (later evolved into Arena) [Pegd86], Sargent’s contributions to formal verification and validation – all brought simulation finally to numerous practitioners in industry.

In the meantime, production and logistics systems had also experienced rapid changes. The essential focus on throughput and the standardization of production and logistics systems of the 1950s is now widening and shifting to the customization of products, agility and flexibility of production, and the pre-eminence of the market. Therefore, the application of simulation as a decision-supporting tool in these areas becomes more and more important. When a new production facility is designed, a new distribution centre is to be built, an old factory is to be drastically modernized, or a new staff deployment is planned, simulation is used to evaluate the proposed designs and schedules. Managers need to find potential weaknesses and flaws in the system, eliminate them, and improve the system's efficiency and productivity before making capital investments, without disrupting the operation of the current system.

The advantages of simulation against other analysis tools are [LaMc97] [BCNN96]:

- **Avoidance of risks:** new policies, operating procedures, decision rules, information flows, organizational procedures, etc. can be explored without disrupting the ongoing operation of the real system
- **Cost saving:** new hardware design, physical layout, transportation systems, etc. can be tested without committing resources to their acquisition
- **Hypothesis testing:** hypotheses of how and why certain phenomena occur can be tested
- **Time efficiency:** time can be compressed or expanded, allowing for a speed-up or slow-down of the phenomena under investigation
- **Repeatability:** different system designs can be simulated in identical environments or the same system in different operating environments
- **Bottleneck analysis:** bottleneck analysis can be performed to indicate where work-in-process, information, materials and so on are being excessively delayed
- **Predictability:** "What if"-questions can be answered. This is particularly useful in the design of new systems

Simulation can be applied to different types of production processes [Wall03]:

1. **Project:** when a product is large, complex or produced as one-of-a-kind in a finite duration with a deadline; when all equipment and personnel are moved to where they are needed in a very flexible way, e.g. in large engineering contracts, construction, etc.
2. **Job shop:** when products are produced in low volume by a small number of skilled personnel assigned to the job, with temporary, intermediate storage capability, while jobs wait for subsequent processing, and highly adaptable equipment is arranged by function. In the job shop products are highly variable in their form, structure, material and/or processing requirements. Each job goes through different machines according to its own requirements, such as routing, material input and varying duration. Example: machine repairing store, etc.
3. **Batch:** when multiple products are produced in larger volumes than in the job shop but smaller than in line production. These products are produced in batches and the WIP goes through different machines as batches. This has a better flow of WIP than

the job shop, since many processes are repetitive. Examples of batch production are numerical control machines, flexible manufacturing systems and group technology.

4. **Production line:** when a product is made in large volumes and standardized, highly automated production processes are arranged by product flow in “takt” time. Assembly lines work with a fixed input in constant flow and produce their products with minor differences. Personnel perform the same operations at each production run in a standard way. Between machines there are normally conveying systems such as conveyors, AGVs (automated guided vehicle), power and free systems to transport intermediate products. As an example, automobile production is such a system.
5. **Continuous flow:** a product is produced in large volume with a high level of automation. Materials flow continuously through successive processes and are converted into one or more products. Personnel are usually not involved in these processes expect to monitor and maintain equipment. Examples of continuous flow production are chemical production, oil refineries, etc. Continuous flow production is not the focus of DES (discrete-event simulation).

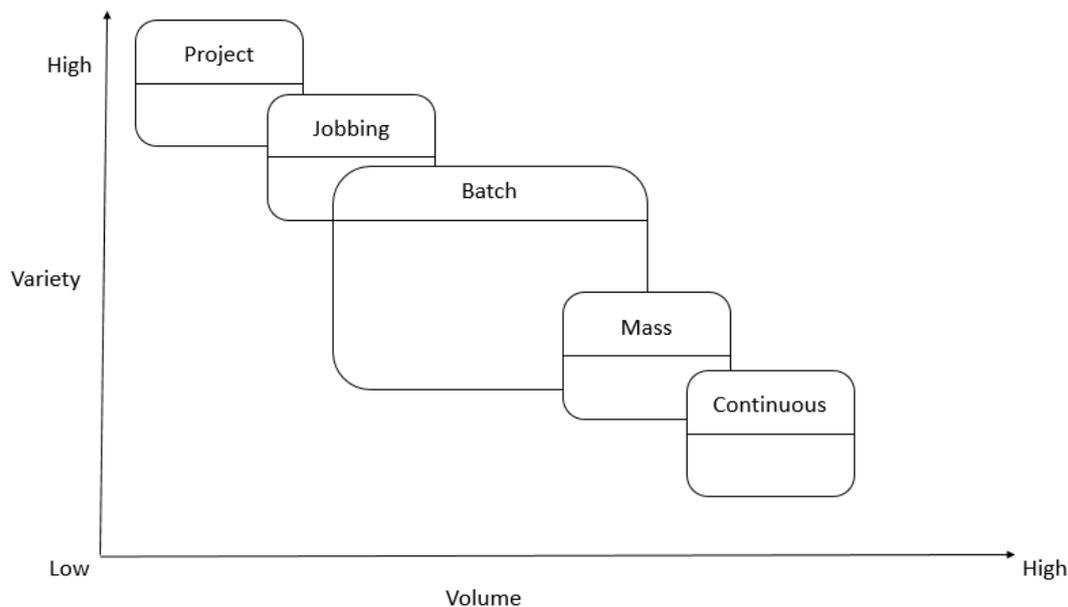


Figure 1-2: Various types of production processes [Wall03]

The aspects of these production systems that can be investigated using simulation are:

1. Numbers and types of machines and their rates, capacities, breakdowns (MTBF[mean time between failure], MTTR[mean time to repair]), maintenance
2. Numbers, types, forms of material handling equipment
3. Control strategies for AGVs, power and free systems, forklifts, tucker trains
4. Locations and sizes of inventory buffers for supplies, spare parts, work-in-process (WIP), and finished goods and inventory policies
5. Product flow, routing and resources needed, bills of material (BOMs)
6. Evaluation of production schedules, e.g. made-to-stock, made-to-order
7. Evaluation of changes in production plans, e.g. product volume or mix
8. Evaluation of the effect of adding new equipment to an existing system
9. Personnel shift scheduling, job responsibilities and qualification

10. Throughput analysis, throughput time or flow time analysis, bottleneck analysis

Performance measures for manufacturing systems evaluated by simulation are [Noch98]:

1. Throughput under average and peak loads
2. Manufacturing lead time
3. Time in queues
4. Utilization of equipment, personnel
5. WIP size
6. Buffer size
7. Staffing requirements
8. Punctuality of delivery

A detailed classification of simulation applications in material handling systems, made by Bernd Noche [NoWe91], is as follows:

1. Warehouse, buffer, storage
 - a. Warehouse planning
 - b. Warehouse management
 - c. Location
 - d. Warehouse type
 - e. Warehouse equipment
 - f. Warehouse topology
 - g. Number of docks
 - h. Warehouse strategy
 - i. Warehouse pre-storage zones
 - j. Inventory management and investigation
 - k. Buffer dimensioning
 - l. Intermediate warehouses
 - m. Pallet storage
 - n. Small part storage
 - o. Oversize items
 - p. Automated storage and retrieval systems (AS/RS) with storage-retrieval machines (SRM)
2. Transportation, cargo handling, cargo preparation
 - a. Intra-plant transportation
 - b. Inter-plant transportation
 - c. Distribution
 - d. Loading areas
 - e. Transportation system types
 - f. Transportation system topology
 - g. Transportation means allocation
 - h. Facility control
 - i. Capacity calculation
 - j. Palletising, depalletizing
 - k. Loading and unloading of transportation means
 - l. Provision of material
 - m. Terminal locations
 - n. Buffer dimensioning

- o. Ramp design
- 3. Consolidation, sorting
 - a. Consolidation technique
 - b. Consolidation strategy
 - c. Sorting strategy
 - d. Generation of consolidation orders
 - e. Single/multi-stage consolidation

Performance measures for material handling systems evaluated by simulation are:

- 1. Warehouse, buffer, storage
 - a. Warehouse capacity, number of vehicles
 - b. Service degree, degree of fulfilment
 - c. Stack height, channel length
 - d. Types of vehicle
 - e. Lane guides
 - f. ABC zone sizes
 - g. Space requirements
 - h. Minimum picking quantity, minimum inventory level
 - i. Number of buffers
 - j. Intermediate warehouses, numbers of objects
- 2. Transportation, cargo handling, cargo preparation
 - a. Numbers of vehicles, load range
 - b. Maximum waiting time of orders
 - c. Fleet size
 - d. Staging area size
 - e. Maximum velocity
 - f. Block section length
 - g. Proportion of rush orders
 - h. Queue length
 - i. Entity mixing
 - j. Height/weight of loading unit
 - k. Numbers of transport auxiliaries
 - l. Trans-shipment site, coordination
 - m. Trailer capacity, buffer size
 - n. Numbers of ramps
- 3. Consolidation, sorting
 - a. Container sizes
 - b. Sequences of entity collection
 - c. Numbers of customers per consolidation
 - d. Numbers of consolidation stages, order splitting

Aspects of randomness in the manufacturing system are:

- 1. Arrival rates of raw material
- 2. Machine processing times
- 3. Machine downtimes
- 4. Machine reparation times
- 5. Loading/unloading times
- 6. Machine setup times

In the randomness above, arrival rates of raw materials, machine processing times, loading/unloading times, and machine setup times and repairing times are parameters that can be controlled or scheduled in most cases. An uncontrollable factor such as machine downtime has a major impact on the performance of an automated manufacturing system.

1.3 Model validation in production and logistics systems

Balci and Sargent [BaSa84], in 1984, identified sixteen terms relevant to model evaluation: acceptability, accuracy, analysis, assessment, calibration, certification, confidence, credibility, evaluation, performance, qualification, quality assurance, reliability, testing, validation and verification. Model verification and validation have been defined thus:

- Model verification refers to building the model right; and
- Model validation refers to building the right model.

And later, in 2002, Balci, Nance, Arthur and Ormsby [BNAO00] renewed the definitions of verification and validation in this way:

- Verification deals with transformational accuracy.
- Validation deals with behavioural or representational accuracy.

I would like to adopt the verification and validation definition from the 1998 AIAA Guide [AIAA98]:

- Verification is the process of determining that a model's implementation accurately represents the developer's conceptual description of the model and the solution to the model
- Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model

The validity of a production or logistics system is largely dependent on several indicators, including throughput, the utilization of equipment and machines, throughput time, the time parts spend in queues, punctuality of deliveries, stockout rate. With these indicators, suggestions about the number of machines, size of buffers, product volume and mix, labour requirements, etc. can be made for a planned system after experiments have been made on the model. So an invalidated model will be harmful to any experiments based on it. Furthermore, the results of the experiments will be misleading to the decision makers using it.

1.4 The problem of validation

“Is there any knowledge in the world which is so certain that no reasonable man could doubt it?” In the first chapter, “Appearance and reality” of Bertrand Russell’s essay “The problem of philosophy”, he asks the ultimate question, “that is really one of the most difficult that can be asked” [Russ12].

The problem of the validation of a simulation model also involves such fundamental questions:

1. What is a model?
2. Is there such a thing as a validated model?
3. Is there any model built for a system which is so certain that no reasonable expert (simulation expert, system engineer, project manager) could doubt it?
4. And if it exists, what kind of character does it have?

In order to discuss these questions, I would like to start with the question of what a model is. There is no universally agreed answer to this question. Human beings began building models when they drew the first pictures of objects on their walls.

For different disciplines, there are different definitions. For example, in mathematics, models are used to interpret theories; in physics, Rene Descartes initiated the idea that an understanding of the world for a scientist does not consist of “postulates representing his own beliefs but as useful models from which one could deduce consequences in agreement with observations” [HoBr01].

There are three necessary properties for a model, according to Stachowiak [Stac73]:

1. Mapping feature
A model is based on an original.
2. Reduction feature
A model only reflects a relevant selection of the original’s properties.
3. Pragmatic feature
A model needs to be usable in place of the original with respect to some purpose.

From these properties we can see that a model is established from the original, but is simplified, since only the features that are relevant to the purpose of the study are chosen and built into the model. It is not simply a copy of the original. For example, in order to investigate the effect on throughput time of adding another machine to the production, another production line is built and a new machine added to the line. To take another example: in a car crash test, an exact copy of the original car is built and an experiment is done; this is simply a car crash test, not a model. The effects can definitely be seen. However, these are exact copies of the originals: they cannot provide the same advantages of models, such as reducing costs, performing different experiments, investigating extreme situations, compressing time, repeating tests, etc. There are of course drawbacks to models – typically, their inaccuracy. But the most accurate ‘model’ can only be the original itself, because a model is always an

abstraction. It takes the essential attributes from the original system and imitates the interactions between them for the purposes of investigation.

In a discrete-event simulation model (DES model), some extra properties are present. Firstly, it is stochastic, which means that at least one parameter in the model is random. Compared to a deterministic model in which the outcomes are fixed under predefined interactions among the entities in a system, the input of the DES model consists of variables in the form of a probability distribution, and its output is random, e.g. the number of orders in a warehouse, the inter-arrival time of the customers in a bank, breakdowns of the machines in a factory, etc. Because of this randomness, the output can be regarded only as an estimation of the real properties of the model. Secondly, it is dynamic. Time plays a vital role in the model. Contrary to the static model, in which time plays no role, the properties of this model evolve over time, e.g. the stock level in a warehouse, the number of customers in a bank, the throughput of a production line, etc. Thirdly, it is discrete. Compared to a continuous model, in which the parameters change constantly over the time, the parameters of a discrete model change at certain points in time, although in some literature it is said that sometimes the boundary between a continuous model and a discrete model is not so clear and often they appear in some mixed form.

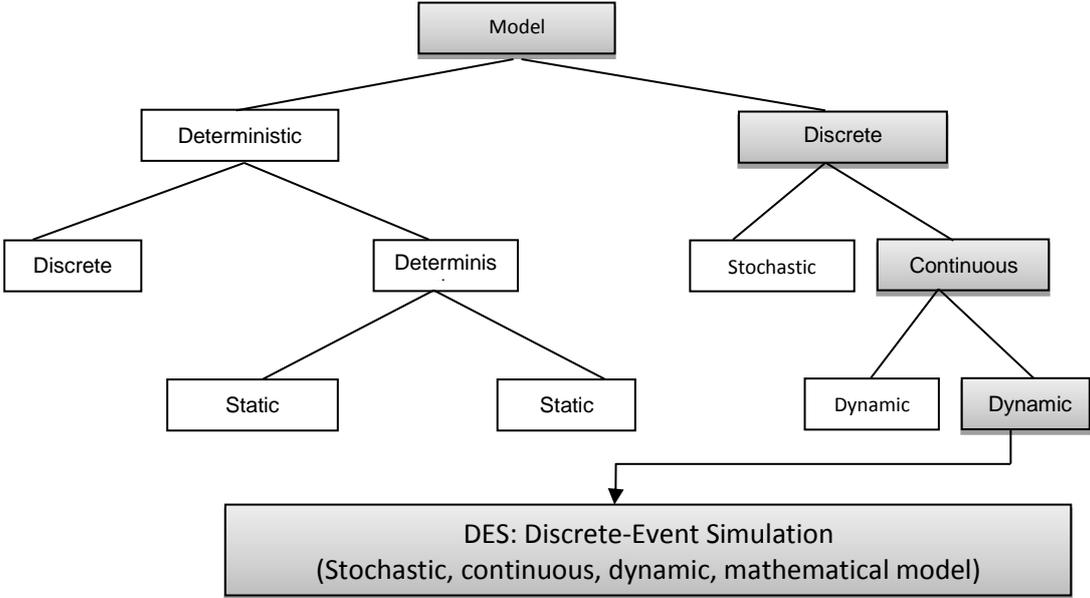


Figure 1-3: Discrete-event simulation characteristics [BCNN96]

Therefore, what we are discussing here is a stochastic, dynamic and discrete model which is based on the original system, intended to be usable with some purpose, and with reduced properties. It is based on mathematical relationships and the input data is probabilistic. Each part of the input data is obvious and meaningful. However, the interactions between them in the model are so complex that we can merely draw conclusions from them. Because of the complexity and the randomness in the model, we have to ask the question of whether we can trust the model that we build. Can we draw conclusions that are repeatable, reliable and predictable so that we can build things from it or are the results merely better than chance?

Consider the validation conclusion from a study of emergency vehicle base locations in the following example:

“Another relevant issue involved how much our model output could deviate from system output and still remain valid... Unfortunately, we were building a model, so some errors and approximations were unavoidable. Model validation here was thus effected by means of a group decision between the client and analysts; when both groups were satisfied, the model was simply considered valid” [GDCV90].

Is it a fact that the deviation in the output of the model from that of the real system results from the errors in the model, as the authors described? If so, how can we call the model a validated model? What is the criterion for a validated model? And how can we prove a model is validated? Is it based only on satisfaction or agreement between the “client and analyst”?

In a classic article on the validation of computer simulation models, Naylor and Finger pointed out that to validate any kind of model means to prove the model to be true, which means, firstly, that we need criteria to distinguish the errors from the truth. Then these criteria can be used for other models. But is there a set of criteria for validating a model? No. Because the criteria are exclusively sampling rules that rely on probability theory. The simulation model and its output represent the “essence of inductive reasoning” because they are generated by “a set of inductive inferences (behavioural assumptions or operating characteristics) about the behaviour of a given system”. The validation procedure is based on possibility, not certainty, in accordance with the assumptions of the model.

In order to be clear about these essential questions, we need a short review of the various positions on validation in simulation.

Naylor and Finger, in an essay published in 1967, described and discussed three positions on verification – rationalism, empiricism and positive economics – and then proposed a multi-stage validation methodology themselves.

The first position to be discussed was that of rationalism. The rationalist regards any model as merely a system of logical deductions from a series of synthetic premises of unquestionable truth, called, by Immanuel Kant, the synthetic a priori, which are “not themselves open to empirical verification or general appeal to objective experience” [BIRH62]. Kant argued that, “when once reason has learnt completely to understand its own power in respect of objects which can be presented to it in experience, it should easily be able to determine, with completeness and certainty, the extent and the limits of its attempted employment beyond the bounds of experience” [Kant71]. So Kant and his followers believe there is a foundation to the truth and we all have access to it through our cognitive apparatus. Experience does not teach us anything and we gain knowledge only by organizing our minds. When we build a simulation model, there are at least some postulates in the model that will be treated as “a priori” without being tested by experience. Another argument cited by Naylor and Finger [NaFi67] to support rationalism in economics is that “there are not postulates the existence of whose counterparts in reality admits of extensive dispute once their nature is fully realized. We do not need controlled experiments to establish their validity: they are so much the stuff of our everyday experience that they have only to be stated to be recognized as obvious.

Indeed, the danger is that they may be thought to be so obvious that nothing significant can be derived from their further examination. Yet, in fact, it is on postulates of this sort that the complicated theorems of advanced analysis ultimately depend” [Robb35]. So the problem is that there is a set of basic assumptions beneath a particular system. Naylor and Finger denied this position by applying the argument of Reichenbach: “scientific philosophy ... refuses to accept any knowledge of the physical world as absolutely certain. Neither the individual occurrences, nor the laws controlling them, can be stated with certainty. The principles of logic and mathematics represent the only domain in which certainty is attainable; but these principles are analytic and empty. Certainty is inseparable from emptiness; there is no synthetic a priori” [Reic51]. Also, if we look back into scientific history, non-Euclidean geometry defeated Euclidean geometry at the beginning of 19th century; the wave theory of light replaced the particle theory at the close of the 19th century; the theory of relativity overturned Newton’s laws of motion. All these make us sceptical enough about rationalism as a foundation for the truth. However, in the domain of the simulation of production and logistics systems there are indeed a number of theories that have been developed from operational research and which can be regarded as synthetic a priori. The most notable is Little’s law, which can be used as a “back of the envelope” calculation for the validity of a simulation model. Little’s law is simple, general and can basically be used for any kind of queuing system. And since it is a mathematical equation, it holds as long as both sides in the equation are equal. So we do not need to go out to the production floor in the factory and collect data to prove it, as we do in validating a simulation model.

There is one semiconductor factory example, used by John Little and Stephen C. Graves [LiSt08]:

Semiconductor devices are manufactured in extremely capital-intensive fabrication facilities. The manufacturing process entails starting with a silicon wafer and then building the electronic circuitry for multiple identical devices through hundreds of process steps. Suppose that the semiconductor factory starts 1000 wafers per day, on average; this is the input rate. The start rate has remained fairly stable over the past 9 months. We track the amount of work-in-process (WIP) inventory. The WIP varies between 40,000 and 50,000 wafers; the average WIP is 45,000 wafers.

Since we know the arrival rate $\lambda = 1000$ wafers/day, the WIP is the system queue length of 45,000 wafers. Using Little’s law:

$$L = \lambda \omega$$

So

$$\omega = L/\lambda = 45,000/1000 = 45 \text{ days}$$

We can build a simple model as follows:

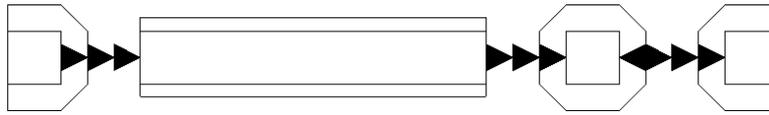


Figure 1-4: A simple M/M/1 model in Dosimis-3 [Dosi12]

With a fairly low level of detail, this model shows the arrival of the wafers with the arrival rate of 1000/month uniformly distributed; the WIP constitutes the wafers in the queue and the workstation; the wafers leave the model from the sink. No breakdowns are built into the model, which is known as an M/M/1 queue and has known transient and steady-state characteristics. We run this model for nine months and ten replications. If the throughput time of the system is around forty-five days, with a certain degree of confidence, then the model is valid for use. Of course, this is only a very simplified model. If it is built with more detail, such as more machines, more control logic, and we also consider dynamic factors such as downtime and reparation time of machines, it will be more difficult to use the present theories to validate the model. However, these theories can still be used for rough checking. For instance, we still use Little's law as our example. If the arrival rate of wafers were to increase; it is reasonable to expect that the throughput time would also tend to rise. With a large scale model with high complexity and input variables, it is more efficient to use sensitivity analysis to pick the most critical input variables for testing.

The second position is empiricism, which is the opposite of rationalism. This takes empirical experience as the form of knowledge. Insofar as knowledge is regarded as a posteriori, it depends upon sense experience [Stan08]. "They insist that sense observation is the primary source and the ultimate judge of knowledge and that it is self-deception to believe the human mind to have direct access to any kind of truth other than that of empty logical relations" [Reic51]. People who believe in empiricism want to take the particular part and reduce it to something they can perceive if they validate a model. This particular part is called empirical foundation – which Bertrand Russell coined as the "ultimate furniture of the world". The empiricists accept nothing that is not observable and require all assumptions in the model to be empirically tested through observable or descriptive data [FeCa06].

Naylor and Finger rejected this position, using Blaug's [Blau78] suggestion of a middle field between extreme a priorism and empiricism, known as positive economics. However, I do believe that in some certain phase of the model validation procedure, the modeller does in some very high degree rely on a posteriori data. The example used by Kelton and Law [LaKe00] in their book, "Simulation Modelling and Analysis" to validate simulation models uses an SME's opinions to validate simulation results. The example is shown below:

At a structured walk-through for a transportation system, a significant percentage of the assumptions given to us by our corporate sponsor were found to be wrong by the SMEs present. As a result, various people were assigned responsibilities to collect information on different parts of the

system. The collected information was used to update the assumptions documented, and a second walk-through was successfully performed.

Another example is the suggestion made by Banks, Carson, Nelson and Nicol [BCNN05] in their book, “Discrete-event System Simulation” – the so-called ‘face validity’. It proposes that the first goal is to build a model that can be accepted “on its face” by SMEs. The authors mention further that “potential users and knowledgeable persons can also evaluate model output for reasonableness and can aid in identifying model discrepancies. ... Another advantage of user involvement is the increase in the model’s perceived validity, or credibility, without which a manager would not be willing to trust simulation results as a basis for decision making.”

These suggestions show the importance of a user’s empirical experience in validating a model in various validation phases. Hence, it is fair to say that this position, if not taken to its extreme form, should be adopted in certain phases of model validation.

However, identifying and employing high quality SMEs is known to be very difficult [Glas98] [Pace98]. There is always bias in human judgment and imperfection in SME knowledge, which will frequently arise subconsciously and lead to unfavourable decisions.

The third position discussed by Naylor and Finger is that compromise between rationalism and empiricism just mentioned – positive economics. People who adopt this position include Milton Friedman, who argues that the validity of a model depends on its predictive power rather than the assumptions on which it is based. The “positive economics” of Milton Friedman [Frie53] is actually a type of instrumentalism. In the view of the instrumentalists, the general propositions of a simulation model are reduced to the role of convenient arrangements (instruments) which we use to order our observations. Friedman makes statement such as:

...theory has no substantive content...its function is so serve as a filling system for organizing empirical material ... only factual evidence can show whether (a theory) is “right” or “wrong”, or better, tentatively “accepted” as valid or “rejected” ... the only relevant test of the validity of a hypothesis is comparison of its predictions with experience.

... truly important and significant hypotheses will be found to have “assumptions” that are wildly inaccurate descriptive representations of reality, and in general, the more significant the theory, the more unrealistic the assumptions (in this sense) [Frie53].

Despite the insufficiency inhabiting this philosophical position, Naylor and Finger admit that it can indeed endorse various structures the modeller might put into simulation models.

Ultimately, a large number of the systems we simulate have an unavoidable conventionalist element. For example: in a production system there are a certain number of finished goods which are defective and need to be reworked or scrapped. We do not know the degree of the deficiency or which parts are defective, but we wish to know the throughput of the production system in the long run. We could aggregate them to some percentage of the whole finished set. It depends entirely on an assumption or a set of assumptions how we implement this with regard to simplicity, convenience, prediction ability or other instrumentalist-like criteria. Another example is the Response Surface for objective learning validity in a simulation. This is “a set of statistical procedures used to develop an empirical model of the relationships between the input and output variables of a system when the inner dynamics of the system are unknown” [Carv91].

After Naylor and Finger [NaFi67] had discussed these three philosophical positions, they proposed a fourth approach to validation. They called it ‘multi-stage verification’. It is a three-stage procedure combining the three positions mentioned above: rationalism, empiricism, and instrumentalism. They argue that these three positions are all necessary for the validation procedure.

Now we will offer a review of this important procedure on which many proposed model validation methods rely.

The first stage is the formulation of a set of postulates or hypotheses which describe the behaviour of the target system. This set of postulates or hypotheses is not arbitrarily chosen. They are based on rationalism. In other words, they are Kant’s “synthetic a priori” [Kant71].

Like the scientist, the scientific philosopher can do nothing but look for his best deposits. But this is what he can do; and he is willing to do it with the perseverance, the self-criticism, and the readiness for new attempts which are indispensable for scientific work. If error is corrected whenever it is recognized as such, the path of error is the path of truth [Reic51].

The meanings of the postulates or hypotheses are not clearly defined by Naylor and Finger [NaFi67]. They say merely that they “would not object to” the “general knowledge” of the system or knowledge from “similar” systems. They argue further that the purpose of the first stage is to select a limited number of postulates, such as component specifications, variables, functional relationships, “on essentially a priori grounds”, as tentative hypotheses about the behaviour of the system to be simulated.

So, what would constitute a set of postulates on a priori grounds as hypotheses about the behaviour of the production and logistics system to be simulated?

Banks suggested it would be the high face validity of the model, which results from collaboration between simulation modeller and model user or knowledgeable persons. During this process the deficiencies of the model can be detected, and, because of the user

involvement, the enhanced credibility will lead to a willingness of the model user to accept the simulation results. In other words, what Banks proposed as the set of postulates are the experiences of the SMEs, since they have knowledge of the system and will check the hypotheses about its behaviour.

There are ultimately many postulates that we can rely on when we simulate a production and logistics system. For example:

- The theoretical downtime of machines can be given by the equipment suppliers.
- The technical details of each piece of equipment can all be acquired from the equipment suppliers.
- Peak periods of customer visits in a business day can be worked out by consulting the manager of the system.

The second stage is the attempt to “verify” the postulates by using the available statistical tests and/or SME opinions.

For production and logistics systems this includes two kinds of postulate, according to Banks [Bank98]: structural postulates and data postulates.

Structural postulates are the control logics in production and logistics systems, such as

Kanban systems, drumbeat systems (where all activities are forced to operate with a common cycle time), conwip systems (where the amount of work-in-progress on a group of machines is controlled by kanbans rather than that individual buffers are controlled), and period batch control systems (where operations are grouped so that they have similar overall cycle times).

The data postulates involve collecting the reliable data and related statistical analyses of it.

For example, the simulation model of a manufacturing line consists of different kinds of machines. The data postulates are:

- Product data such as routing, bill of material (BOM)
- Production order, number and deadline
- Labour break regulation
- Labour shift
- Processing times of the machines
- Setup times of the machines
- Speed of conveying equipment such as accumulating conveyors, AGVs, etc.
- Outage of machines due to random breakdowns and maintenance
- Mean time to repair (MTTR) and mean time between failures (MTBF) of machines.

This data needs to be collected from the production manager and line managers if the historical data is to be used to make a trace-driven simulation. If the data needs to be analysed, for example, cleaned, aggregated, grouped, etc., and then based on the processed data to identify the appropriate theoretical distributions and estimate their parameters, then these distributions need subsequently to be verified by the SMEs empirically and theoretically by means of a goodness-of-fit test.

The third stage is the procedure of testing a model's predictability regarding system behaviour in reality from the perspective of the intended uses of the model.

Predictability has further been divided into two categories: historical validation and validation by forecasting.

Historical validation in the context of production and logistics simulation models is the quantitative comparison of simulation outcomes with the historical outcomes of an existing system – a retrospective prediction. The outcome of this is that it can be decided whether or not the model's results match those of the real system on some level. If they do not match, then either the model or the experiment or both should be revised.

Validation by forecasting is concerned with prospective predictions. This is actually the purpose of building the simulation model. We might be interested in predicting either the behaviour of the system, such as detecting deadlocks or livelocks in the conveyor systems, bottleneck allocation in the manufacturing systems, reliability analysis, control strategies of automated guided vehicles (AGV), numbers and types of machines for a certain production line, product processing routing problems, etc.; or predicting the behaviour of the system under different or endogenous influences or a combination of these, such as location and size of inventory buffer optimization under different business policies, system throughput under different shift schedules, etc. The problem with validation by forecasting is that no historical data is available, since the real system or part of it does not yet exist. And even if the system does exist, when the modeller wants to predict the behaviour of the system in different environments or under extreme conditions, there will again be no data from the real system.

The central idea in this methodology is that of a model's ability, and accuracy, to predict system behaviour from the perspective of the intended uses. In order to achieve this, in the first two stages any assumptions should be judged with empirical and statistical tests. And "any hypotheses which can be rejected on a priori grounds should be so rejected", since it is much more efficient to correct mistakes in the early stages than in the finished model. Furthermore, as Kleijnen claimed, a model's validity is determined by its assumptions [KlGr92].

This methodology has been applied by Banks [Bank98] etc. as follows:

1. Build a model that has high face validity
2. Validate model assumptions
3. Compare model input-output transformations to corresponding input-output transformations in the real system.

One major drawback of this methodology is its ability to validate a model without a real system – validation by forecasting, since no data is available to compare the outcomes of the model and the real system. The model ought to be sitting on a mass of historical data in order to be validated. Moreover, there is no indication of the criteria for each stage.

In the following chapters I will discuss the procedure of simulation model validation in the following areas:

1. Paradigms
2. System complexity
3. System uncertainty
4. Data availability
5. Subjectivity: expert opinions
6. Objectivity: statistical tests
7. Sequential bifurcation

The aim in this dissertation is to propose a model validation process for identifying the most important input variables of simulation models in the production and logistics industry, and to reduce system complexity and uncertainty and lower the requirements for data availability in model validation.

Chapter 2 the problem of simulation model validation paradigms

2.1 Model validation paradigms

The purpose of simulation is not the building of a simulation model. A simulation study involves many processes. However, a great deal of research has been conducted into the building of more true-to-life simulation models. Even in practice, most of the time employed in simulation projects is used for model building. Since many simulation studies are used to analyse system design and performance before the real projects begin, model builders are often overwhelmed by the huge amount of data (or sometimes very little data) available, very high system complexity and tight project schedules. If a project is not carefully planned, the outcomes of the simulation will not only be large cost overruns, prolonged delays in completion and user dissatisfaction with the simulation results. There will also be setbacks in more important projects planned for after the simulation studies which are dependent on the results of the study to build production lines, choose locations and add distribution centres into sales networks, or improve the efficiency of current systems. The problem, or even challenge, here is how to conduct the project properly. Evaluating the acceptability and accuracy of simulation outcomes is never performed after the simulation model has been built and experimental results from the model have been obtained. The model validation process takes up the largest part of a simulation study. It must be conducted at each phase throughout the study. It must be carefully planned and executed. It should address the particular circumstances of the study and take into account factors such as overall validation requirements and objectives, validation agreement between model builder and model user, permitted time, personnel resources, etc. One of the key factors of success in a simulation study is the following of a well-structured approach and comprehensive validation paradigm. Numerous paradigms are available in the simulation community for the validation procedure. In this chapter I will discuss several of these paradigms.

2.2 The model validation paradigm of Osman Balci

Various simulation study paradigms have been proposed for the model builder to follow [Bank98]. Osman Balci proposed a paradigm with ten major processes (see Figure 2-1 below):

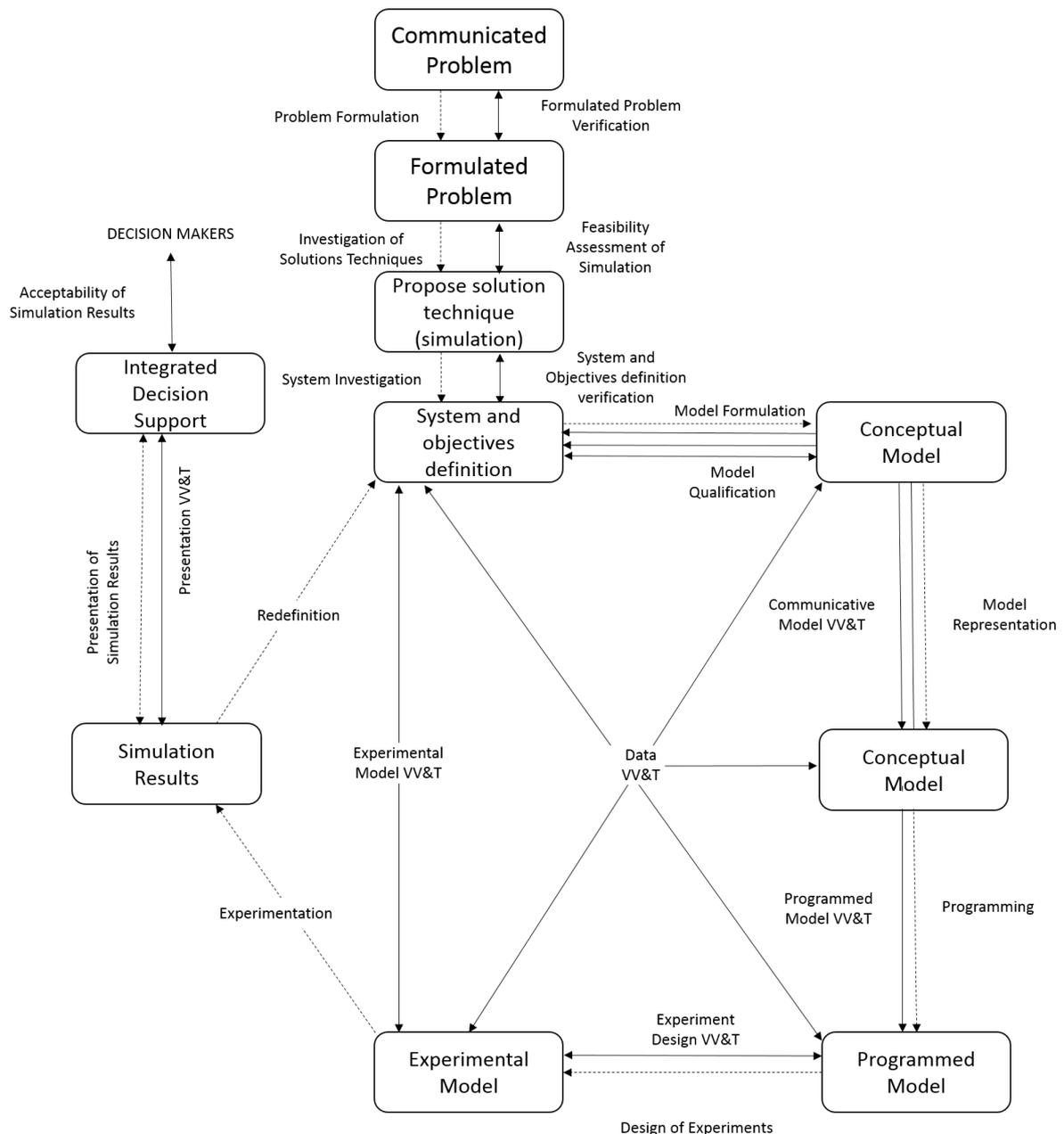


Figure 2-1: The life cycle of a simulation study by Osman Balci [Balc94]

Any scientific research begins with the formulation of a problem to be solved. A problem correctly formulated is half solved. Accuracy in the formulation of the problem greatly affects the acceptability and credibility of any simulation results. Albert Einstein [EiIn38] thought that the correct formulation of a problem is in some degree more important than its solution. Inadequate problem definition and lack of user participation in problem formulation will lead one to committing the ‘Type III error’ of solving the wrong problem. Balci and Nance [BaNa85] conducted an extensive survey of the literature and identified several approaches to problem formulation and thirty-eight indicators for assessing the accuracy of problem formulation at the beginning of a simulation study. The first four approaches, thought insufficient by Woolley and Pidd [WoPi81], are as follows:

1. The checklist approach
The problem formulation here might be: do this, then that, then..., and so on. The model builder is guided through an array of questions from which the required information will lead to the cause of the problem.
2. The definition approach
This problem formulation might be something like: what are the decision parameters? The model builder identifies and obtains the variables required to build a model advised by decision makers and model users.
3. The science approach
This problem formulation asks the question: what is really going on here? The model builder needs to collect quantitative data, observing the problem domain in order to gain insight into the real system, and define the “actual” problem.
4. The people approach
This problem formulation asks the question: what is everyone saying and why? Various people have different perceptions of the same problem. This approach requires the model builder to communicate the problem’s definition which is acceptable to both the model user and the model builder.

Woolley and Pidd [WoPi81] suggest the ‘exploration approach’ as a combination of the definition, science and people approaches above. This approach is a continual cycle: question → answer → reflect → question →... The answers from the model user or people who have knowledge about the system will inspire the model builder to reflect and ask further questions until the problem is sufficiently structured.

Balci and Nance also propose a guide to the high-level problem formulation procedure, with a better structure and more details, as in the flow chart seen here in Figure 2-2 which aids decision making [BaNa85].

Solutions to problems are divided into two categories:

1. Prescriptive solutions, which give the model user a value judgment on the “goodness” or “badness” of actions being taken on a problem. Problems that require a prescriptive solution might for example be such as this: Should a centralized or decentralized warehousing strategy be employed? Where the distribution centre or centres should be built? Which SKU should be stored in which distribution centre and which transportation strategy should be employed to reach the highest productivity with the least total transportation costs? What action should be taken to minimize stock-out situations and maximize the state of readiness of distribution centres?
2. Descriptive solutions give the model user no value judgment on the “goodness” or “badness” of actions being taken. Problems that require descriptive solutions might be as follows: How high will the utilization of the packaging machine in the warehouse be next month? What would be the effect of a 10% increase in customer orders on the service grade of the warehouses? What are the most significant parameters that affect the overall performance of the distribution network?

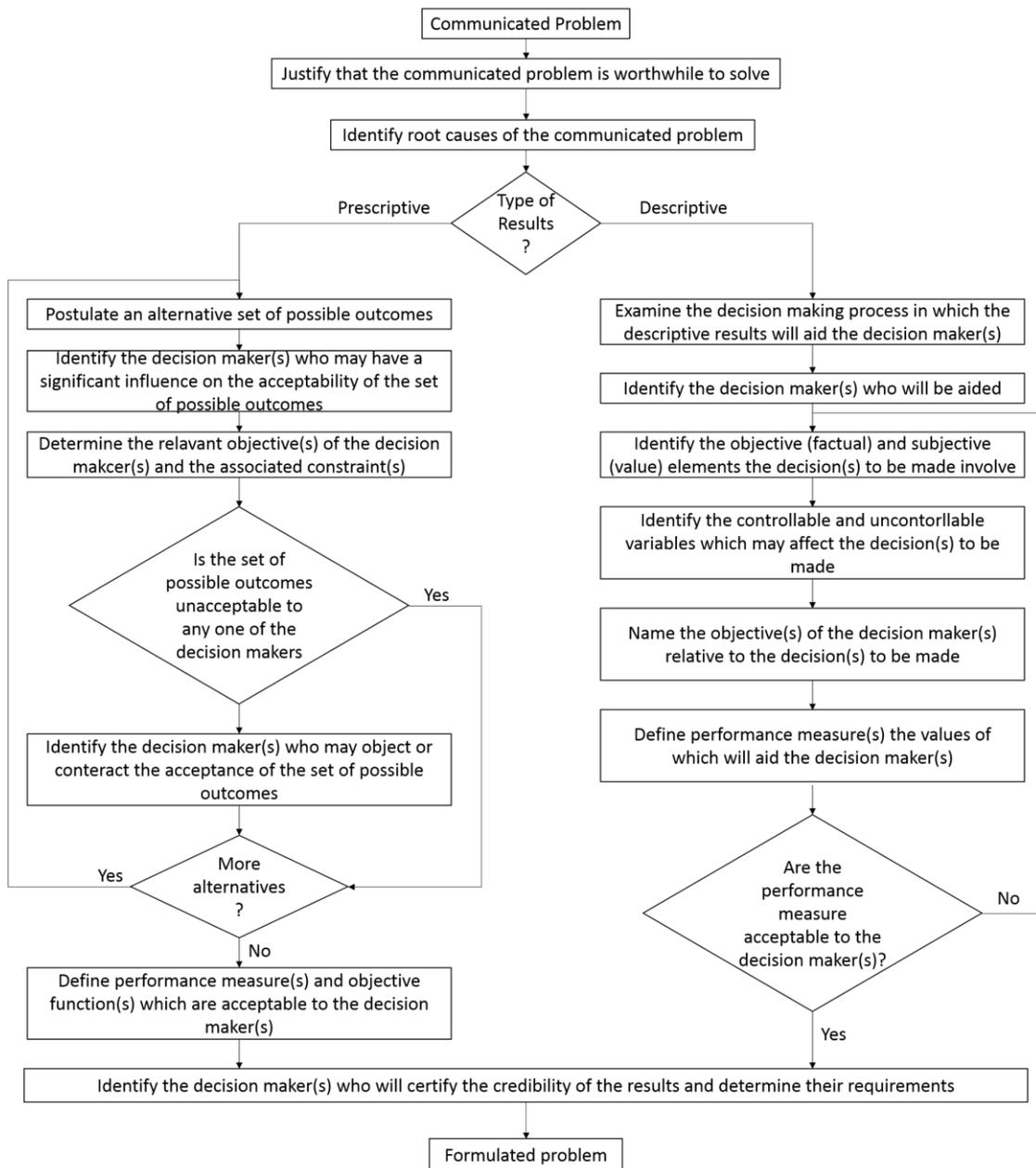


Figure 2-2: A high-level procedure for problem formulation [BaNa85]

The next step after problem formulation and the investigation of solution techniques is in this case certainly the simulation technique, as we discussed previously.

In the next phase, the system investigation, the characteristics of the study objectives that are identified during the problem formulation should be examined. Shannon identified the system characteristics as:

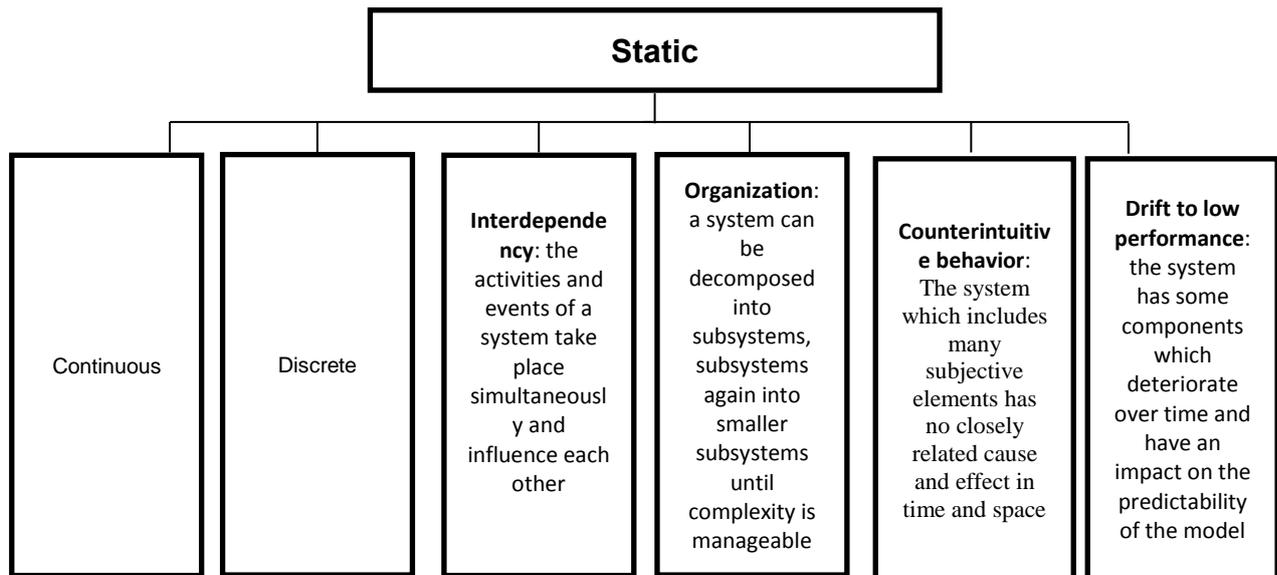


Figure 2-3: A classification of system characteristics by Shannon [Shan48]

During model formulation, from the concept a model is made. A concept model is an abstraction or simplification of the real or proposed system [Zeig76]. A more exact definition, given by Robinson [Robi04], is: “The conceptual model is a non-software-specific description of the simulation model that is to be developed, describing the objectives, inputs, outputs, content, assumptions, and simplification of the model.” Pace [Pace99] thinks a conceptual model should consist of assumptions, algorithms, characteristics, relationships and data. A conceptual model should be judged by its completeness, consistency, coherence and correctness. He suggests an iterative conceptual model development, as in Figure 2-4, below:

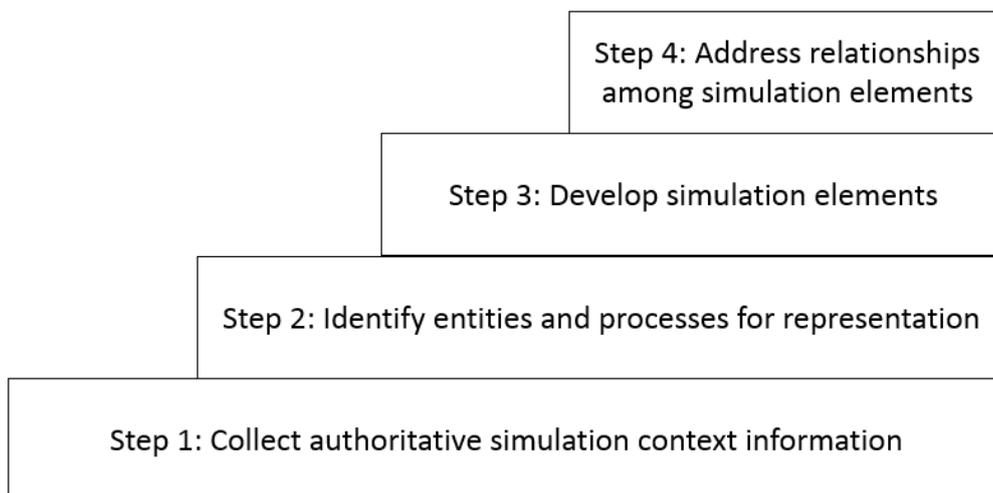


Figure 2-4: Iterative conceptual model development by Pace [Pace99]

When the conceptual model is built, it should be translated into a communicative model. Nance [Nanc94] defines a communicative model as “a model representation which can be communicated to other humans, can be judged or compared against the system and the study objectives by more than one human”. A communicative model can take the following forms, as in Figure 2-5. Different forms can be used for people with varying backgrounds.

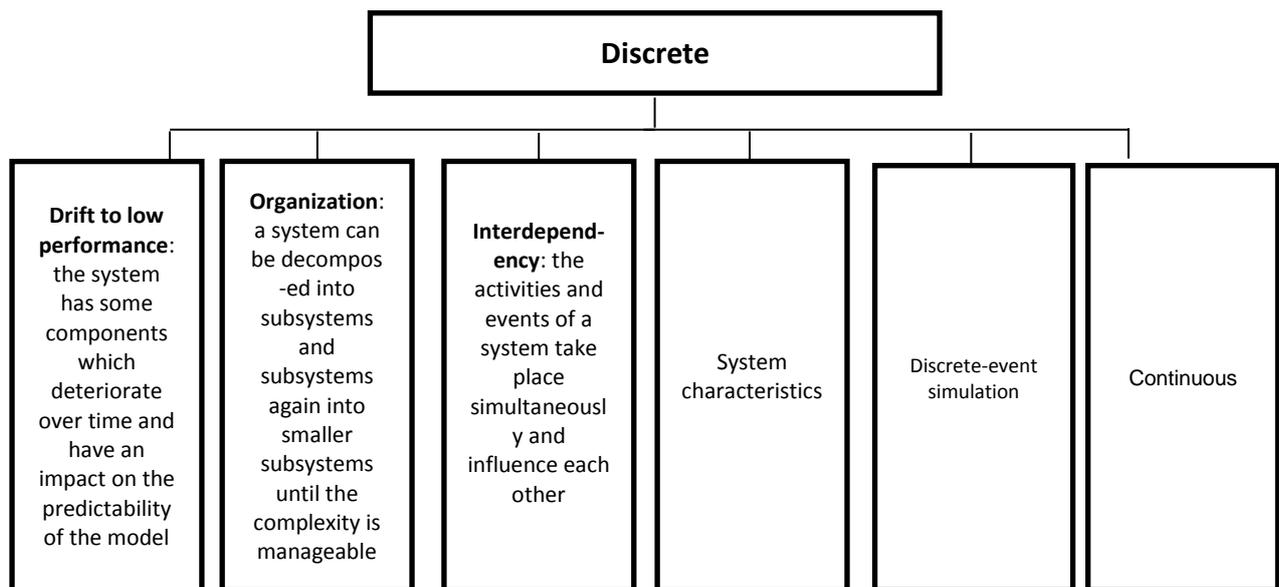


Figure 2-5: Communicative model forms by Nance [Nanc94]

The programmed model is translated from the communicative model, a process which involves programming. Commercial software packages such as Arena, Automod, Plant Simulation, Witness, Dosimis-3, Promodel, Flexsim or general purpose languages such as C# or Java can be used for the programming. The programmed model must then be verified (debugged) and validated.

After verification and validation, experiments can be designed and performed on the model. Design of experiments is a cyclical process. It starts with the planning of experiments for a hypothesis of interest. After the execution of these experiments the outputs are used for statistical analysis. The hypothesis will then be modified and the cycle of design of experiments will begin again.

The last step is to present the simulation results to the model user. If the user accepts the results, it can be implemented in a real system.

2.3 The model validation paradigms of Shannon, Law and Kelton, and Banks

Some other paradigms have been proposed for the simulation procedure, including those by Shannon [Shan75], Law and Kelton [LaKe00], Banks et al [Bank98]. (Figure 2-6). Shannon's paradigm is a model of the waterfall type. It begins with problem formulation and then goes straight to the implementation and documentation of the simulation model. There is no iteration of validation. Law and Kelton's model improves on this by realizing that the validation of a conceptual and a computerized model are cyclical processes, and explicitly mentions data collection. However, it does not mention conceptual model building as an

independent process. One of the most used simulation paradigms is that of Banks et al. This adds iteration at the simulation experimentation phase to permit non-sequential transitions between steps. The analysis of complete simulation runs determines whether additional runs are needed and what design those additional experiments should have.

The difference between these three paradigms, besides minor differences of activity, is obviously in the iterations between processes. This is called ‘model calibration’. When a model is calibrated, the constants and variable parameters are adjusted to make the results agree with the experimental data. Although model calibration is a common means of validating a model, this process can be very time consuming.

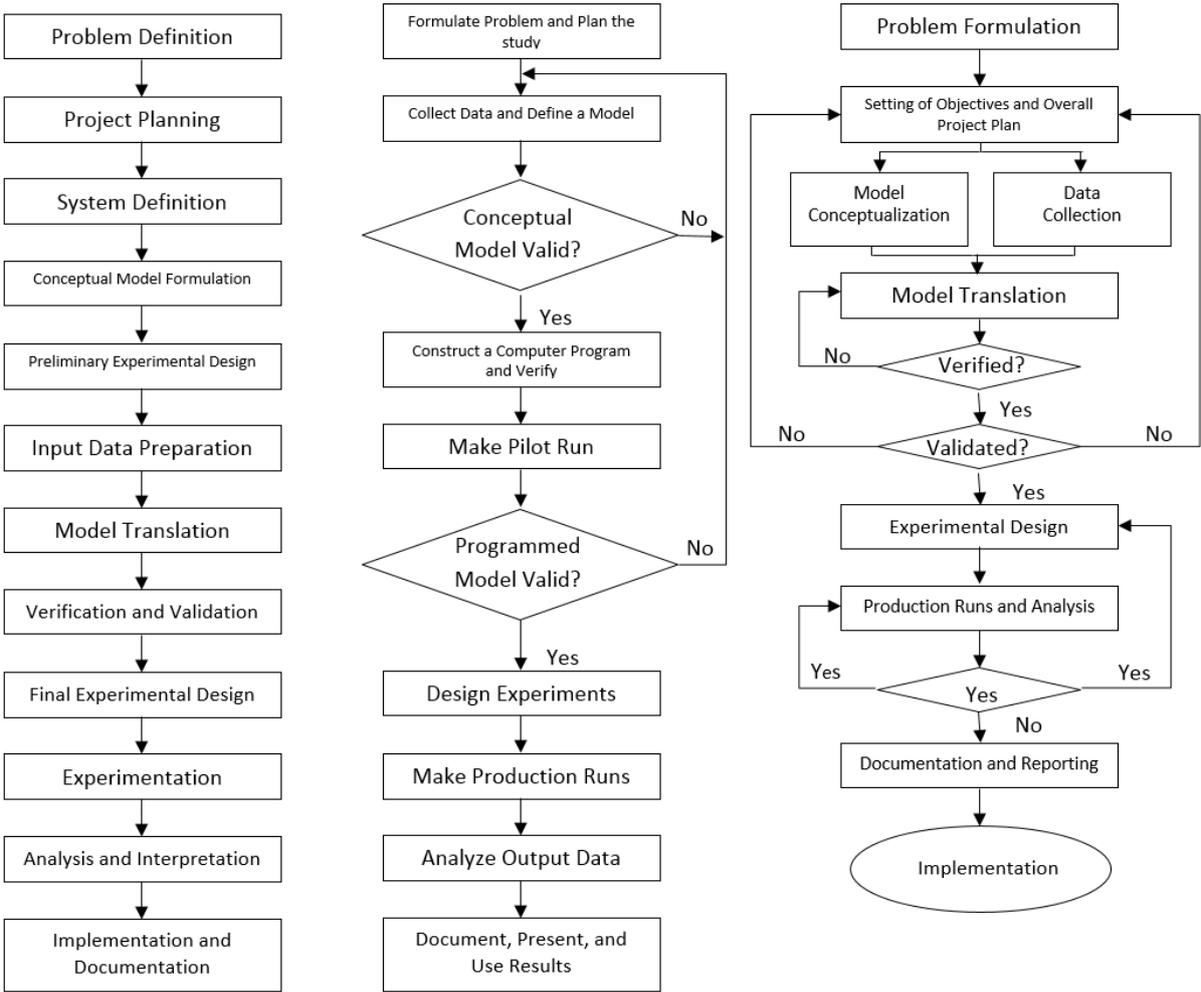


Figure 2-6: Simulation procedure paradigms by Shannon [Shan75], Law and Kelton [LaKe00], Banks et al [Bank98]

2.4 The model validation paradigm by the Arbeitsgemeinschaft Simulation (ASIM)

In Germany there are also some popular paradigms used by simulation practitioners. One is the simulation procedure proposed by the Arbeitsgemeinschaft Simulation (ASIM) [ASIM97],

shown here in Figure 2-7. Special attention should be paid to this paradigm if the model builder applies it, firstly because in this paradigm the conceptual model is not explicitly presented. Instead, it uses the process modelling and abstraction of the real or planned system to represent the conceptual model-building process. This lack of the conceptual model might make it hard to communicate with the model user, since the computerized model is complex and difficult to comprehend by people with little knowledge of computer modelling. Moreover, it can lead to low model credibility and acceptability on the side of the model user. Secondly, data validation is not mentioned in the early phase. No data validation at the modelling and abstraction phase would also leave any mistakes hidden to successive phases and thus lower the credibility of the model to the user. Thirdly, model validation is not explicitly mentioned in the paradigm nor any iteration of it. It is mentioned only in the explanation of the model: “after the implementation (of the simulation model on the computer) follows the validation. It will check whether the implemented model agrees with the real system.” This might lead to the problem that undetected mistakes could be discovered only when the model was built. It would be very expensive and time consuming to correct or modify the model in the event of (unavoidable) mistakes. Fourthly, there is no suggestion of documentation, either of the modelling process or in reporting to the model user. This would also make it difficult for the model user to understand how the model works. Further, this would be a great drawback if modification were needed to the model in the future.

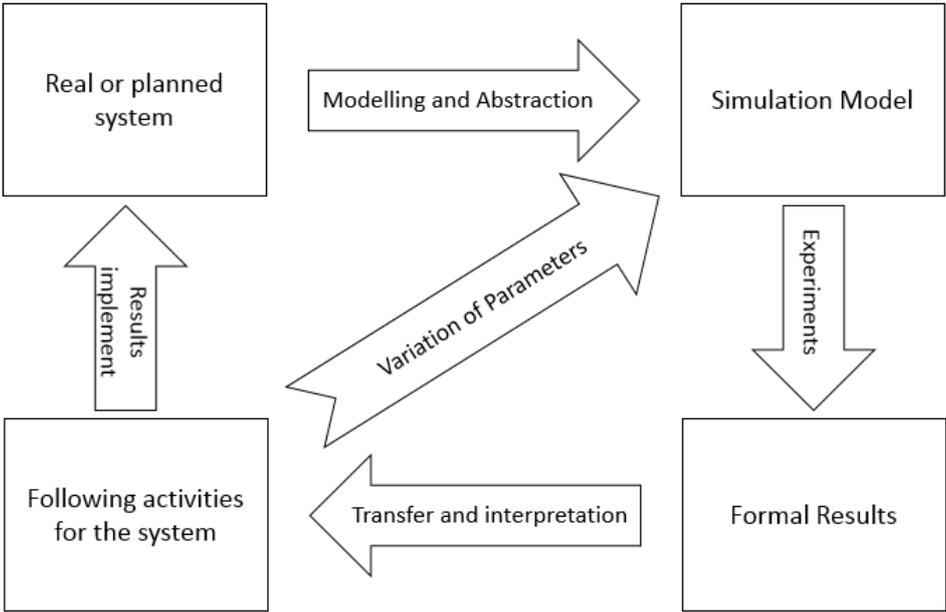


Figure 2-7: Simulation study procedure by ASIM [ASIM97]

2.5 The model validation paradigm of Rabe et al.

Another paradigm, proposed by Rabe et al. [RaSW08], is shown in Figure 2-8. Figure 2-8 (a) shows the main simulation procedure, which also begins with the needs of the study sponsor (model user) and puts the collection and preparation of data on a parallel with the main

activities of the model building. The reason for this, as they say, is that “raw data does not need to be completely collected before the elaboration of the formal model. The same applies to the prepared data, analogously”. However, normally the “raw data” should be collected and then analysed and prepared in the proper form in order for one to define the conceptual model. Even when the raw data is missing or it takes too much time to collect, in the conceptual modelling phase it should be at least estimated, because data availability, quality and randomness must always be considered, checked and validated by both the model builder and the model user in the conceptual phase. This model also lacks the validation iteration of the intermediate model. It is actually very similar to the Shannon model, which is of the waterfall type. Figure 2-8 (b) shows the procedure model for the verification and validation (V&V) of simulation in the production and logistics domain. This model, by Rabe et al., was modified from Brade’s model [Brad00]. In the original form, Brade does not consider the role of data in the validation process.

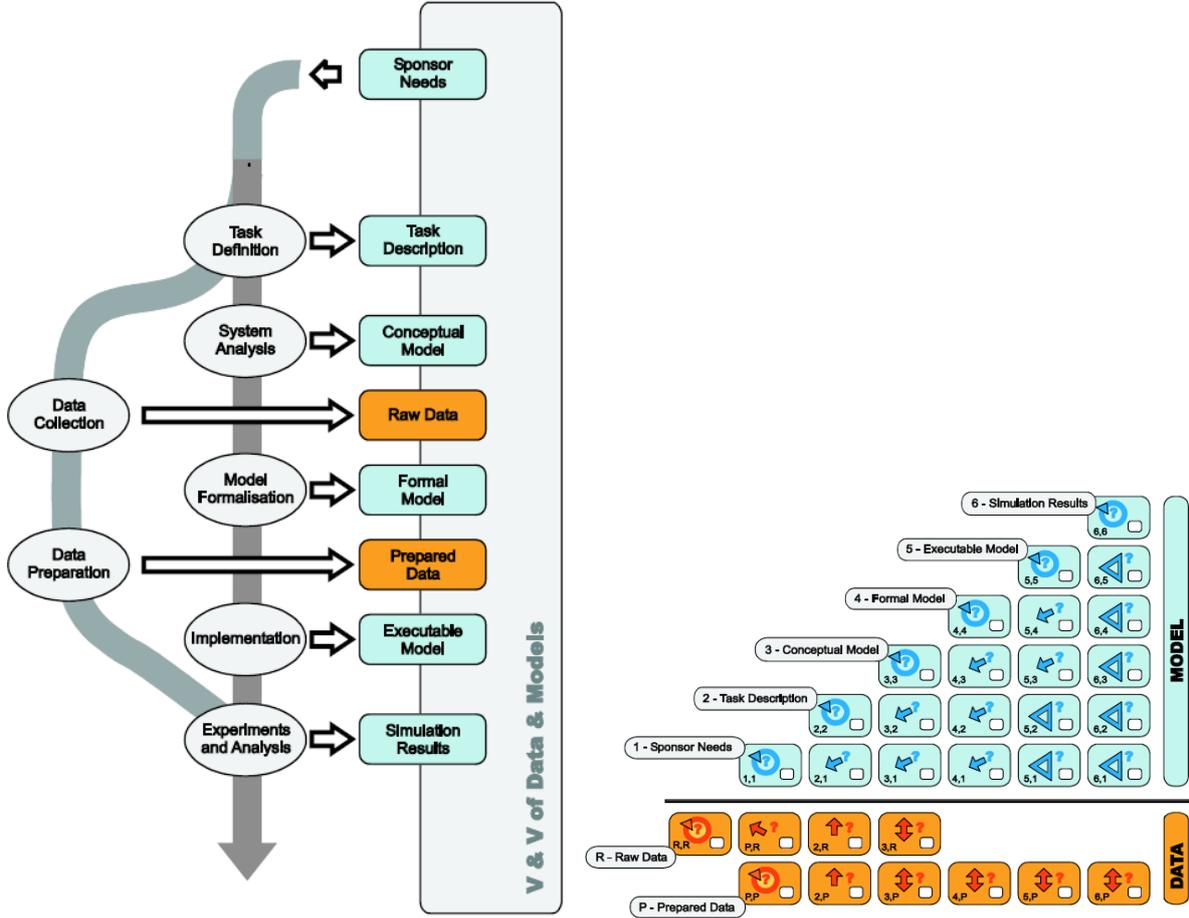


Figure 2-8: Simulation procedure proposed by Rabe, Spieckermann and Wenzel [RaSW08]

teacher who gives only a final exam to students. If they take only a final exam without having any tests or homework throughout the semester, it will be too late to help many of them and they will fail. Simulation studies should be conducted throughout the life cycle. It is not a one-way street: any deficiencies that emerge from this process in this version of the model can be fixed and the new version tested once again. Furthermore, it is better to discover the model’s deficiencies at an early stage rather than validate a defective model. As Banks described, and is shown in Figure 2-10, a model should always be compared with reality through various subjective and objective tests. This iterative process is carried out continuously until the model reaches the accuracy agreed between model builder and model user.

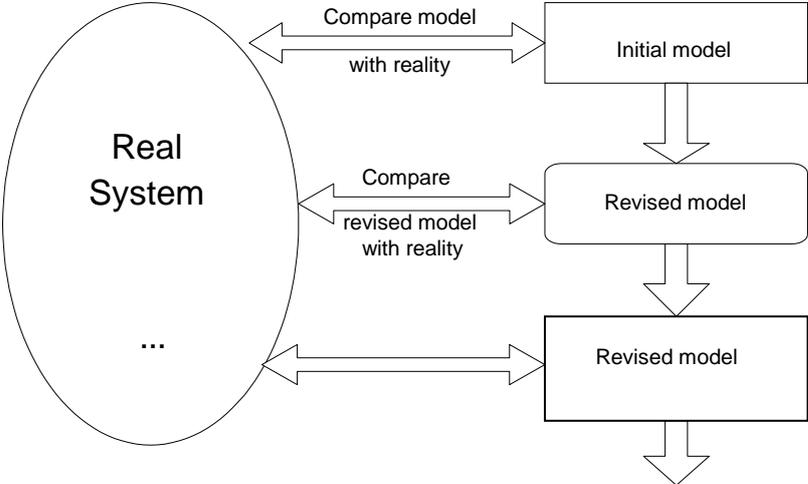


Figure 2-10: The iterative process of model validation by Banks [BCNN96]

The testing of a simulation model can be divided into four levels, from the lowest to the highest: element testing, component testing, sub-system testing and whole-model testing. It is important to recognize that the model should be validated as a whole and that is impossible ever to check the model completely. The validation of each part of a sub-model cannot guarantee the validation of the entire model, since there are interactions between the sub-models and the validity of the whole model is dependent on the interactions between the sub-models just as the validity of the sub-models is dependent on the interactions of the elements in them. Complete model testing under all possible conditions is not feasible, since infinite time and budgets are not available. The purpose of these tests is therefore to increase the credibility of the model and our confidence in it.

In the iterative process of model validation, data plays a vital role as a messenger between the real system, the model, the model builder and the model user. We must validate the input data that we collect for the simulation model from the real system, if it exists. We need data for comparison between the real system and the model. We need to analyse the comparisons to determine if they are the “same” – sometimes called the ‘double validation problem’. If the data sets agree with each other, then we present the results based on the comparison to the

model user at the validation procedure. If not, the model has to be revised and further comparisons made until the two data sets do agree.

However, during the validation process the model builder should never regard the simulation result as a binary value which is either absolutely correct or absolutely wrong. George Box expressed it this way: “All models are wrong, but some are useful.” Since a model is always an abstraction of the original, the only exact representation is the system itself. This issue relates to a model’s scope and level of detail. In order to augment credibility, the level of detail should be increased. But this will lead to higher development costs, as shown in Figure 2-11:

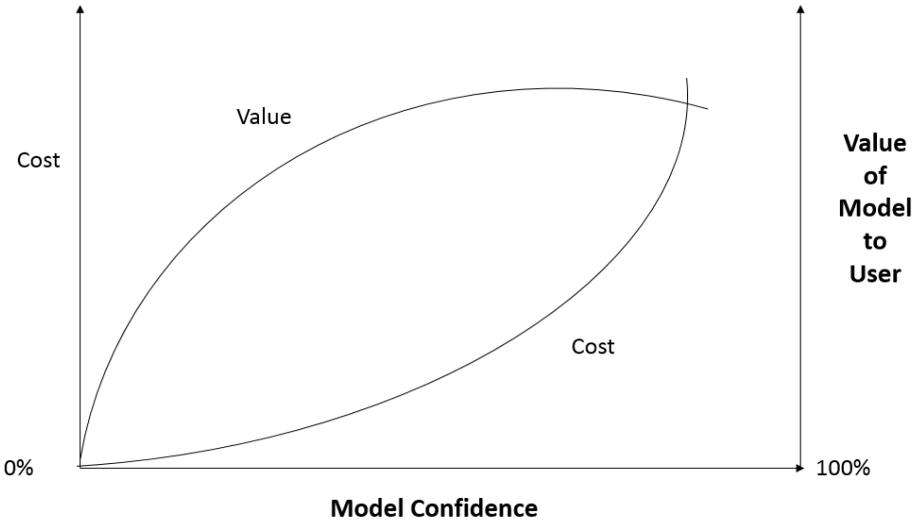


Figure 2-11: Model credibility and cost [Sarg03]

So, unless unlimited resources (both time and budget) can be devoted to the building of a model, the model will never be accurate. Thus, the outcome of the model will be accepted only with some degree of confidence. It is never a binary result of success or failure. Moreover, a model is developed to investigate some specific aspects of a system, so its validity is based on the assumptions and goals of the simulation study. An accurate specification of the simulation study objectives is the foundation of a successful study, no matter whether a model is built to simulate a new production line or an existing model needs to be modified, such as in the addition of a new machine to a production line. If the simulation objectives are set inadequately or deviate from the actual problem that is to be solved, then the simulation study is destined to fail. This is called the ‘Type III error’. Agreement between the model builder and model user about model confidence can be expressed on a scale from 0 to 10, with 0 as completely insufficient and 10 as absolutely sufficient.

Potential dangers that must be mentioned in accepting model results according to the agreement between the model builder and the user are the Type I error and the Type II error. If simulation results are sufficiently credible but rejected, this is the Type I error. Sometimes, it is also called the ‘model builder’s error’. If the simulation results are invalid but accepted as sufficient, this is the Type II error, sometimes called the ‘model user’s risk’. Type II and the previously mentioned Type III error are very dangerous. If one of these occurs, the

consequences for the model user can be catastrophic. How one handles this issue is really critical, since Type I and Type II errors are a trade-off of many factors. Balci and Sargent [BaSa84] discuss this issue using a statistical hypothesis test – the cost-risk analysis – for the validation of a multivariate response, self-driven, steady-state simulation model and a trace-driven, terminating simulation model. They give a methodology for constructing the relationships of Type I and Type II risks, an acceptable validity range agreed between model builder and model user, and the cost of data collection.

It should also be considered that a validated model still cannot be accepted by a model user because the model lacks credibility for him. Accreditation is a concept used by the U.S Department of Defence (DOD) [DOD98] as “the official certification that a model, simulation, or federation of models and simulations is acceptable for use for a specific purpose”. It is an official credibility assessment procedure of simulation results. The model builder is responsible not only for delivering a validated model but also, in the end, a model that is acceptable to the model user. Thus, the involvement of the model user in the model development phase is vital for the credibility and acceptability of the simulation results because the model user will gain an insight into the model and have the feeling of participation in it and of being part of the model-building process.

Documentation should be kept throughout the whole validation process. This includes the objectives of the simulation study, the scope of the simulation model, assumptions and testing data. The documentation can be used by the model user, project management and the model builder.

2.7 Model validation paradigms in production and logistics

There are indeed enough paradigms for model builders to follow when they wish to build and validate a model. Some are used intensively by model practitioners and work very well. The problem is that they are very general and can be applied to many simulation areas. Only the Rabe, Spieckermann and Wenzel [RaSW08] paradigm is said to be used for validating simulation models in production and logistics. However, it does not indicate any special characteristics of the production and logistics simulation model, which can be unique and quite helpful in the validation process.

For production and logistics systems, differing simulation models with various complexities should follow a variety of validation paradigms. The aim of a validation paradigm is to ensure a model's predictability and acceptability. Production and logistics simulation models have some characteristics that have an impact on model predictability. These appear in three aspects:

1. System complexity: the simulation object could be a small job shop with several machines or a large automobile factory with a kilometre-long assembly line, accumulating conveyors, suspended overhead conveyors, AGV, various testing stations, repairing stations, etc. The respective simulation models might contain

several model elements or lines of code, or thousands of elements or lines of code in their realizations of the material flows and related control logic in the real systems.

The level of system complexity in a production system can be divided into three. From the highest to the lowest these are: firstly, the rough planning, which includes the production principle, the material flow principle and the control principle. Secondly, the structural planning, which includes layout planning, machine functionality and transport means control. Thirdly, detail planning, detailed layout planning, processing system and personnel assignment planning.

The level of system complexity in logistics systems can be divided into five. From the highest to the lowest these are: the supply chain level, factory level, production line level, job shop level and machine level.

A system with high complexity involves not only more physical objectives, but also more people. More time will be spent collecting data from the physical objectives and communicating with the related people. Thus, a very complex system will result in a large simulation model and will further lead to difficulty in model validation. Sargent once said that “a number of simulation professionals, including myself, believe that it is impossible to verify and validate large-scale simulation models to a reasonable confidence level.” Therefore, system complexity is the first factor considered in a validation paradigm classification.

2. Model scope and level of model detail (model structure uncertainty and model formation uncertainty): a model's scope and level of detail depend on the intended use of the simulation model and the resources that people have. The major factors in determining the level of detail are the time and budget available. System complexity must also be consistent with the availability of data. It is seldom that one builds a model which represents every element in a system. A model should be as simple as possible, as long as the performance measurements for evaluation of the system in the model can be fulfilled. A very complex system can have a very simple model as long as the results of the model can fulfil the goal of the simulation study.

Model scope and level of detail can bring two uncertain factors to model validation: those of structure uncertainty and formation uncertainty. Different people will build different models with differing structures for the same system. They may use various modelling techniques or elements for the same process. They may apply varying amounts of detail to model the system. We cannot judge which model would be the best. This depends not only on the modelling requirements, but also on the skills of the model builders. Although it is impossible to quantify the uncertainty in model structure and the formation process, a classification of model type would help this process, with some guidance.

Even though performance measurements for production and logistics systems are quite standardized, there is no standard for determining a model's level of detail. The more detail there is in a model, the harder it is to validate it. SMEs and a sensitivity analysis

should be involved in order to decide the level of detail when building a model. If the model is large and complex with a great deal of detail built into it, then a black box validation might be better suited, since it would be enormously expensive to check every detail in such a model.

3. The third aspect concerns data. This can be further divided into three sub-categories: 1) data availability, 2) data quality, 3) data randomness.

A simulation model is a type of mathematical model. It is built on data, uses data to run, and is checked according to its output data. One of the most influential factors in model validation is data. Thus, the availability of data is crucial in the building, running and later validating of a model. Data availability in production and logistics systems is two-fold: endogenous and exogenous. Endogenous data is, for example, the processing time of a machine, downtimes of machines, conveying speeds of AGVs, working times of labour, etc. Endogenous data is relatively easy to obtain from an equipment manufacturer, machine operator or repairman. Exogenous data consists, for example, of customer order information, supplier shipment information, weather conditions for large distribution systems, etc. Exogenous data is sometimes completely missing or can only be estimated if a system is in the design phase. If the system already exists and needs to be modified, historical data (endogenous plus exogenous data) might suffice. Historical data can be used in the validation process in comparisons with the output data from the model. If the two data sets agree to an adequate confidence interval, then the model may be considered as validated. If the data, either exogenous or endogenous or both, can only be estimated, then the uncertainty in the model's predictability will be sharply increased.

Data quality should also be considered in model validation. Endogenous data in modern manufacturing and transportation systems can be easily recorded. However, it may not be in an appropriate format for input analysis or to make trace-driven simulations. Some data may be manipulated, modified or falsely recorded or measured. Some data used in the input is already an estimated value, e.g. weather forecasts, sales forecasts; the quality of this data can hardly be judged.

Data randomness can also be divided into endogenous and exogenous forms. Variation is the most influential factor in production and logistics systems, and cannot be eliminated from them. Sources of endogenous data randomness are, for example, the arbitrary breakdown of machines, the fluctuating processing time of workers, defective product rates, etc. Sources of exogenous data randomness are, for example, the number of lines of a customer order, the number of items per order line, etc. The more sources of data randomness in a model, the harder it is to validate the model.

In an essay from the Winter Simulation Conference in 2000 [SGKLMY00], several simulation experts from various areas were asked about the future of model validation. Glasow thought little research had been done in practice. Kleijnen considered that simulation practitioners were lacking in statistical skills. Law, with his thirty-three years of academic and

industrial experience in simulation, was utterly disillusioned about any objective approach to this issue and suggested that currently the subjective approach to model validation is the norm. Eleven years later, at the Winter Simulation Conference, Sargent [Sarg10] was using a paradigm developed thirty years ago and techniques that have also been used for years to advocate model validation methods. From graphical model representation to 3D models and, further, to Digital Factory, simulation models in production and logistics have been making huge progress over the years. The subject of model validation has merely been repeatedly mentioned and is evolving only slowly. There should be some new approach to this issue. I do believe that model validation in production and logistics should follow a different validation paradigm according to some criteria such as the three aspects discussed above. There are still many open issues. A methodology for establishing this validation classification and associated criteria is needed from future research.

Chapter 3 Complexity, uncertainty and data availability in the model validation of production and logistics systems

3.1 The three aspects of discrete-event simulation in model validation

Logistics and production systems are highly complex engineered systems. Simulation models are required to estimate system performance when other analytical methods reach their limits. The reasons why it is difficult to validate simulation models of production and logistics systems (below referred to as ‘models’) can be summarised into three aspects: model complexity, model uncertainty (general model uncertainty, environmental uncertainty; lack of knowledge, randomness, and volatility) and data (data quality, data availability). The purpose of this chapter is to answer the following questions, which is vital if we wish to achieve a deeper understanding of model validation: What is input uncertainty and how can we measure it? What is model complexity and why has it not been possible up to now to measure this in reality? How does the complexity of a model amplify the uncertainty and variability in the model’s output? How does input randomness alter bottlenecks and the behaviour of a model? How do input uncertainty and model complexity influence model validation? How do data quality and availability affect model complexity, input uncertainty, and even the model validation process?

As mentioned above, all models are wrong, but some are useful. Some, however, might be harmful!

Nowadays, logistics systems are interconnected with one another. Networks are getting bigger and becoming more and more complex. In order to increase system utilization, reduce lead time, raise delivery reliability, and at the same time keep inventories at a low level in logistics and production systems, logistics engineers are essentially treating a system with problems of two types: complexity and uncertainty. These properties are consequentially passed on to simulation models and lead to difficulties in model validation.

So what is model complexity? What is model uncertainty? Why is data availability vitally important in model validation? Why are complexity and uncertainty so essential in logistics systems, and how do they affect model validation?

‘Complexity’ is a term used to describe the difficulty of understanding or analysing a system, as opposed to a ‘simple’ system, in which causality can be easily identified. The degree of complexity depends on the number of elements, their interconnectedness and the number of different system states [UIPr88]. Complexity is increasing in production and logistics systems, owing to high product complexity, short product life cycles, small lot sizes and increasing

numbers of product variants. The components of complexity in a simulation model include the control logic and the interaction between entities.

‘Uncertainty’ is described as a lack of predictability or the occurrence of events whose outcomes are not settled, are in doubt, or are dependent on chance. Uncertainty in a simulation model might be the uncertainty in the accuracy of data collection, input distribution models and their respective parameters, the uncertainty in the accuracy of a simulation model in describing a system or process, or the uncertainty caused by exogenous factors such as the volatility of customer demand. The components of uncertainty in a simulation model are the quantity and quality of input data, the probability distribution and its parameters, such as machine failure rate and repairing times on a production line, or the inter-arrival rate of customer orders in a distribution centre.

Data is essential in simulation modelling. A model is built on data, uses data as input and ideally uses system output data for validation. The availability of data is paramount in a simulation study. It can in the first place determine whether a simulation study should be made or not. If no data is available or cannot even be estimated, for example if a system does not yet exist, a simulation cannot be built. Data availability also has a great impact on model complexity and uncertainty. With little data, a model’s detail cannot be high and its scope cannot be large and the uncertainty concerning activity in the model will also reach a maximum.

An increase in complexity raises uncertainty in a production and logistics system, since more entities are involved. The requirement of data quantity and quality will consequently be raised when describing the behaviour of these entities. On the other hand, uncertainty in the simulation model is propagated by the complexity and the availability of data and introduces more uncertainty into the simulation output. Consequently, the uncertainty in the model output leads to further uncertainty in the model validation and an inability to predict system behaviour. The main difficulties of model validation are therefore caused by the interaction of these three factors. The core challenge of model validation is hence to understand this interaction.

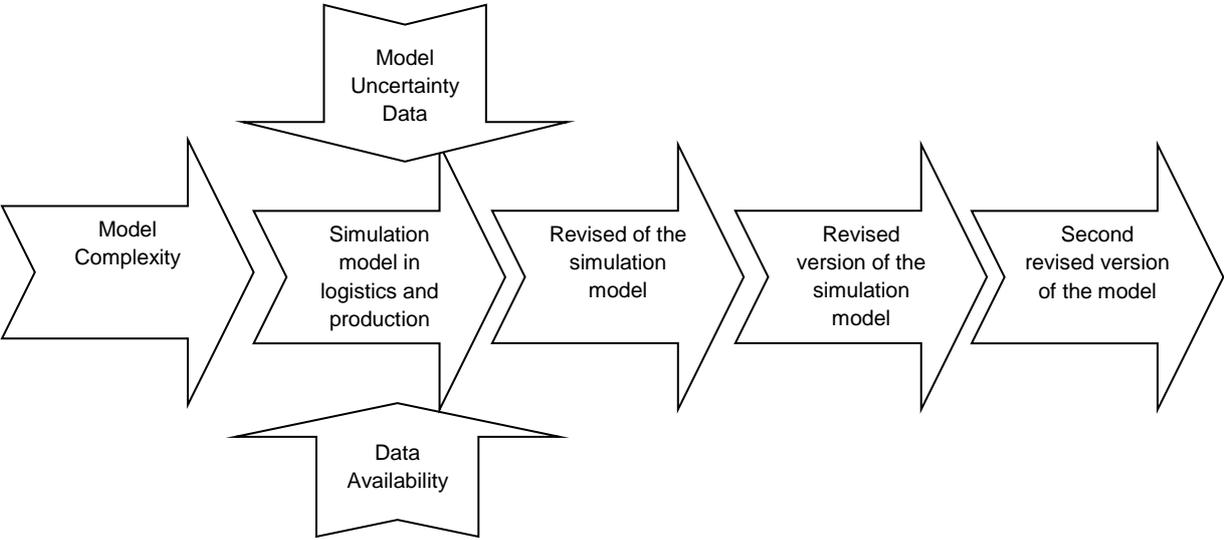


Figure 3-1: Complexity, uncertainty and low data availability lead to uncertainty in validation and prediction

3.1.1 Three aspects of model validation

- **Model complexity:** (definition of complexity: a great degree of interdependence between elements in a system).
 - Type 1 complexity: a variable is determined by its past value (autocorrelation)
 - Type 2 complexity: variables are influenced by one other (cross-correlation)
 - Type 3 complexity: a variable's value depends on the past value or values of another or other variables (cross-correlation with time lags).
- **Model uncertainty:** (system uncertainty and environmental uncertainty).
 - Aleatory uncertainty: the word 'aleatory' derives from the Latin 'alea', which means 'the rolling of dice'. It refers to variability, irreducible uncertainty, and is used to describe the inherent variation associated with a real system and its environment. It is assumed to be the intrinsic randomness of a system or processes.
 - Epistemic uncertainty: the word 'epistemic' derives from the Greek ἐπιστήμη (episteme), which means 'knowledge'. It refers to reducible uncertainty; model-form uncertainty is used to describe a lack of knowledge of a system or its surrounding environment, or incomplete information.
- **Data availability:**
 - Type 1 data: system input and output data
 - Type 2 data: system output data
 - Type 3 data: no data or theoretical system input data

Three factors have been identified. What are the connections between model validation and these factors?

3.1.2 Model complexity

Type 1 data can reveal a model's complexity by exposing different types of correlation between variables. The larger the amount of data collected from the input and output of a system, the easier the task of identifying correlations among variables.

When type 1 or type 2 data is not available, model complexity can hardly be revealed. However, by using sensitivity analysis or intelligent experiment design, the complexity of a model can be made clearer.

3.1.3 Model uncertainty

Model uncertainty is the stochastic nature of discrete-event simulation. Stochastic variables are the source of uncertainty. It is necessary to divide this uncertainty into two categories [Ober01]: aleatory uncertainty and epistemic uncertainty. They exist in most simulation models in production and logistics systems.

Data availability affects model uncertainty. Type 1 data can reduce epistemic uncertainty in both input and output. The more input and output data can be collected from a system, the greater the epistemic uncertainty is reduced.

Type 2 data cannot reduce uncertainty in input, but it reduces uncertainty in output and is very useful for the calibration procedure in model validation.

Type 3 data greatly increases the epistemic uncertainty, since neither input nor output can be set in some cases and can be only estimated. In particular, when no information concerning a machine failure is available, the best one can do is to determine the availability of a replacement machine from the equipment provider. In this situation, the value range and the choice of statistical distribution are both uncertain. The epistemic uncertainty reaches a peak.

On the other hand, aleatory uncertainty is related to stochastic variables. Aleatory uncertainty cannot be ignored in simulation models. Doing so will lead directly to model invalidity.

3.1.4 Data availability

Data availability is essential to model validation. Model complexity and model uncertainty are governed largely by data availability. [Klei98]

When type 1 data is available, then input-output transform validation (black box validation) can be performed. Trace-driven simulation can be used.

When type 2 data is available, statistical methods such as the paired t-test (for testing the mean) or the F-test (for testing variance) can be used.

When type 3 data is available, sensitivity analysis or experiment design can be used.

In the model validation procedure, the availability of data is essential. Model uncertainty and model complexity are built on it.

In the following part of this chapter, model complexity, model uncertainty and data availability in simulation models are investigated and their interactions and influence on the model validation process are discussed.

3.2 System complexity, and complexity in logistics and production systems

3.2.1 Complexity and system complexity

Weaver [WeWa48] roughly assigned systems three categories of order, according to their degree of randomness and complexity, namely, organized simplicity, organized complexity and disorganized complexity, as shown in Figure 3.2, below:

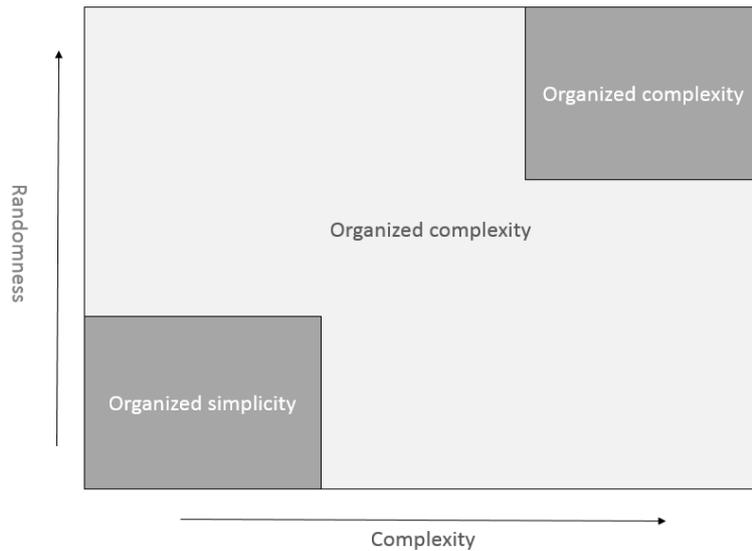


Figure 3-2: Three classes of system by Weaver [WeWa48]

The first set of systems falls under the heading ‘organized simplicity’. As described by Weaver, it involves over-simplified, two-variable (three or four at the most) problems which can be solved by nineteenth-century techniques.

At the other extreme there is the set of systems with disorganized complexity. These have huge numbers of variables, perhaps several billion. The problems of these systems can only be solved by statistical methods if one desires useful predictions. Moreover, the statistical methods apply with increasing precision when the numbers of variables increase. One can obtain results with a high degree of confidence in a large telephone network by predicting the average frequency of calls, or the probability of calls of the same number that overlap, despite lacking any knowledge of each individual.

The area between these two comprises systems with so-called ‘organized complexity’. These are systems with a moderate number of variables – larger than two but smaller than the number in the category of disorganized complexity. They include problems which involve dealing simultaneously with a considerable number of interrelated variables that make up a whole. They are too complicated for one to use the methods applied to systems with organized complexity, but also cannot be handled with the statistical methods that describe the average behaviour in systems of disorganized complexity. Such a system might be, for example, a supply chain system with some one hundred factors.

Logistics and production systems have organized complexity which originates from organized simplicity while the problems within it can be solved only by the statistical methods that we apply to disorganized complexity. They are a form of organized complexity because they normally have a sizeable number of factors and are interrelated with each other as a whole. They originate from organized simplicity because the basic theories of queuing models or operation research models, for example, can be applied to describe and solve problems in a basic form, while their measure of performance, such as average waiting time, mean

throughput per hour or machine utilization, can be calculated only by the statistical techniques used in solving disorganized complexity.

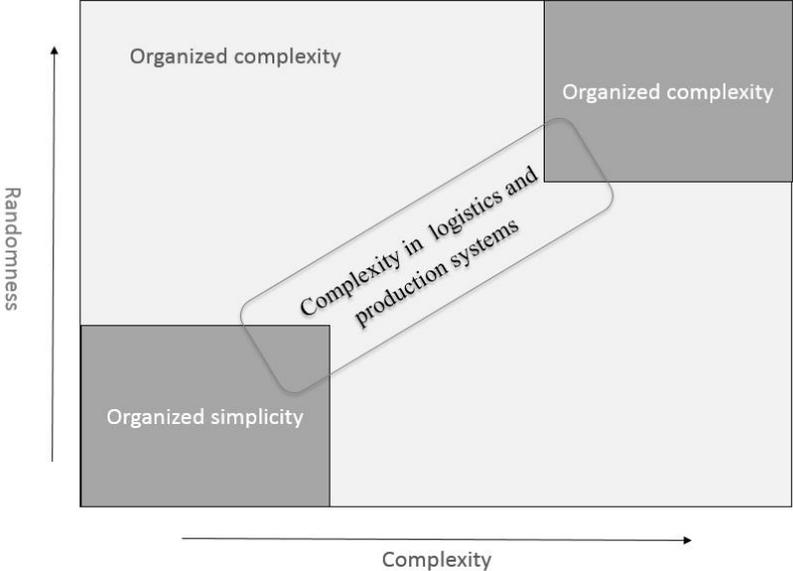


Figure 3-3: Complexity in logistics and production systems [Modified from WeWa48]

H. Simon defines a complex system as “one made up of a large number of parts that interact in a non-simple way. In such systems the whole is more than the sum of its parts, not in an ultimate, metaphysical sense, but in the important pragmatic sense that, given the properties of the parts and the law of their interaction, it is not a trivial matter to infer the properties of the whole”.

Complexity in model validation in production and logistics systems can be classified into two categories:

- System complexity
- Model complexity.

System complexity will to a high degree be transferred to the complexity of the model. The more variety and uncertainty a production and logistics system has, the harder it is to validate the corresponding simulation model.

The traditional way of dealing with complexity is to separate it into structural complexity and dynamic complexity. ‘Structural complexity’ often refers to the fixed nature of products, structures, processes, etc. ‘Dynamic complexity’ refers to variations in dates and quantities owing to material shortages, machine breakdowns, insufficient supplier reliability and so on. A more manufacturing-oriented classification is offered by Eimaraghy and Urbanic [UrEl06].

3.2.2 The measurement of system complexity

The measurement of system complexity by entropy was first achieved by Shannon [Shan48] in the area of information theory. He measured the amount of information associated with the occurrence of given states in a system. Frizelle [Friz96] borrowed the notion of entropy from Shannon and applied it to manufacturing systems. He proposed the following function:

$$H_s = - \sum_{i=1}^M \sum_{j=1}^S P_{ij} \log_2 P_{ij}$$

where:

P_{ij} : probability of resource i being in scheduled state j

S : number of scheduled states

M : number of stations,

to measure the structural complexity of a multi-station manufacturing system. The complexity consists of: 1) the number of processes, 2) the number of process states, 3) the frequency of variation of process states.

Yu and Efstathiou [YuEf06] modified Frizelle's function and applied it to more complex manufacturing systems to measure their variety and uncertainty. They call it 'network complexity'. The idea of network complexity is used to define structural complexity in a manufacturing network. It quantifies the effects of network shapes and the availability and working rates of servers and is determined by the interconnection of the stations in a system. The events at a station affect the other servers and even the rest of the whole system through the linkages between them, under the following assumptions: that the system has a constant input of material and a constant specified processing time; only one type of entity is processed and the material flow is conserved; and there is no buffer between the servers if they are connected in series.

System complexity can be calculated in four steps:

Step 1). State definition

In this step, the states of servers are defined as idle, working, failed or blocked. The working rate $W_{j,i}$ refers to server i in the state j ; server availability $P_{j,i}$ refers to the server i as working in state j .

Step 2). State generation

When servers are grouped into sub-systems, new states of the sub-systems need to be generated. In the case of two servers, if server 1 has m states and server 2 has n states, then the maximum number of states the sub-system can have is $m*n$. Further, they can be connected either in parallel or in series.

In serial connection, the working rate of the sub-system is the minimum working rate of server 1 and server 2:

$$W_{x,y} = \text{Min}(W_{1,x}, W_{2,y}).$$

The possibility that state x occurs in server 1 and state y occurs in server 2 simultaneously is:

$$P_{x,y} [W_{x,y}] = P_x[W_x] * P_y [W_y].$$

If two servers are connected in parallel, the working rate is the summation of the working rates of both servers:

$$W_{x,y} = \text{Sum}(W_{1,x}, W_{2,y})$$

The availability of these parallel servers is:

$$P_{x,y} [W_{x,y}] = P_x[W_x] * P_y [W_y]$$

Step 3). State reduction

Overlapped states are grouped together and the availability is the summation of the working rates of all grouped states.

Step 4). Measurement of performance

Complexity can be calculated using the following equation:

$$C = - \sum_{j=1}^s P_j \log_2 P_j$$

These techniques are difficult to apply to real-life systems which involve hundreds of entities and complicated control logic. The assumptions required for these techniques, such as of constant arrival rates or processing times, are also too unrealistic. Moreover, the results of the calculations offer little value for an understanding of the model or system.

3.2.3 Complexity in a simulation model: correlation

According to Shannon's definition [Shan48], a model is an abstract and idealized replication of a real system which reflects all of its relevant properties with sufficient accuracy with respect to an intended purpose.

Then, from the point of view of validation, what is complexity in a logistics and production system? It can be the number of entities in the system; it can be the number of logical connections between these entities; or it can be, as intended by Shannon, described and calculated by the status of service-providing elements, like that of machines, workers, etc. The number of entities is certainly one of the decisive factors in defining model complexity. However, it provides little help in the validation of a model. Just as in the calculation of

complexity using Shannon's formula, a plain number which describes the complexity of a whole system provides little help for the validation procedure. This means that an increase or decrease in this number is not helpful.

A better understanding of the complexity of a system, which would provide more insight into the relevant properties of its elements, can be achieved by measuring the correlations of those elements.

The most obvious correlation between model variables is in the model's input and output. For example, when the arrival rate is increased, then it is expected that the output rate will also rise, since these two variables are positively correlated with each other. However, such a black-box testing method is not enough, because it reveals little relation between variables in the model. If output results are counter-intuitive then there is no help for the simulation practitioner in detecting the problem inside the model.

Moreover, it is sometimes not only the actual values of the simulation that are important, but also their relationships to one another.

It might also be that the scale of variables is important. The relationship between the scale of one variable and that of another or all other variables is, however, not clear. For example, it is very common that in a production system model it is required that the throughput of the system be doubled. The requirement and the results of the simulation are clear. However, because there are so many variables in the model, and their quantity or their combination can affect the throughput, the correlations between them and the model throughput are very difficult to determine. In this case, the task of model validation is exactly to prove the correlation between these variables and the throughput.

Further, some occasions arise where the results of a simulation output are inexplicable, which is a serious problem.

From this point of view, model complexity in logistics and production systems can be regarded as the interdependence of variables. There are three types of complexity regarding the interdependence of variables:

- 1). A variable is determined by the past value of the variable itself
- 2). A variable is determined by the value of another variable (or other variables)
- 3). A variable is determined by the past value of another variable (or other variables).

Ignorance or inaccurate representation of any one of the above interdependences in a simulation model will lead to invalidity of the model.

The mathematical representation of interdependence is as follows:

- 1). Type I: autocorrelation if the variable is dependent on its past value.

The autocorrelation of a random variable or process is the correlation between values of the variable at different points in time. Let N denote the number of observations of a random

variable or process. The autocorrelation of the variable between observations separated by k time is

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

where r_k is the autocorrelation coefficient at lag k. The plot of the autocorrelation function as a function of lag is a correlogram.

There are two kinds of autocorrelation in simulation modelling: the first type is especially true when the variable is the simulation output, since essentially all output variables in a simulation are autocorrelated (and non-stationary); an example of this would be the waiting times of customers in a queue. The second type of autocorrelation exists in the simulation input, for example when a random variable is used to describe a learning curve, or in an inventory model the number of articles in successive orders is negatively lag-one autocorrelated.

2). Type II: correlation if the variable is dependent on the value of another variable.

Let N denote the number of pairs of observations of two variables x and y. Then, the correlation between x and y is:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2 \right]^{1/2} \left[\sum (y_i - \bar{y})^2 \right]^{1/2}}$$

where \bar{x} and \bar{y} are the mean values of x and y respectively, r is the correlation coefficient of x and y over N observations.

There are also two types of correlation in simulation modelling. The first is between the input variables. For example, in the repair process of a machine, the lead time of a spare part for the machine is positively correlated with the repair job. The second type is between input and output variables. For example, an increase in i will lead to longer waiting times in the system. The third type is between the output variables, for example, the occupancy of two successive buffers, or the system time and the number of entities in the system.

3). Type III: correlation of two variables with some time-offset or lag. Consider two random processes m and n with N pairs of observations. The cross-covariance function is as follows:

$$c_{mn}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (m_t - \bar{m})(n_{t+k} - \bar{n}), \text{ where } k = 1, 2, \dots, (N-1)$$

$$c_{mn}(k) = \frac{1}{N} \sum_{t=1-k}^N (m_t - \bar{m})(n_{t+k} - \bar{n}), \text{ where } k = -1, -2, \dots, -(N-1),$$

\bar{m} and \bar{n} are the sample means, and k is the lag. Since the functions are asymmetrical, e.g. values at lag k and at lag -k are not equal, both functions are needed here. The first equation

is used when n_t is shifted forward relative to m_t , or in other words n_t lags m_t . The second equation is used for the reverse situation.

The cross-correlation function is:

$$r_{mn}(k) = \frac{c_{mn}(k)}{\sqrt{c_{nn}(0)c_{mm}(0)}}$$

where $c_{nn}(0)$ and $c_{mm}(0)$ are the sample variances of m_t and n_t .

In this way, model complexity is divided into three more concrete categories. It can be understood, in principle, as the correlation of variables in time and space.

From this point of view, it reveals the important relationships among model elements and allows the judgment of the model builder and user to be based not only on intuition concerning the possible behaviour of a system. Some vital relationships that might be overlooked or remain hidden in highly complex systems may also be revealed.

This classification of model complexity is illustrated by the following example [Dosi12]:

- On the production line of an electrical manufacturer, consumer goods are produced. On this line there is a sub-area where the goods are inspected and even small repair jobs are done. The material flow is illustrated in Figure 3-4, below. A preliminary production area supplies the production line with two different kinds of product group (randomly distributed) in intervals of one minute. Both kinds of goods enter the system at a receipt buffer and are fed to the two work stations by shuttle. But there is a fixed assignment of each kind of goods to the individual work station, because full flexibility cannot be offered economically, owing to high investment needs.

The purpose of the simulation study of this system is primarily to find out whether it is possible to achieve the requested throughput of sixty parts per hour on average together with the utilization of the work stations.

The simulation model is shown in the figure below, which was built using Dosimis-3.

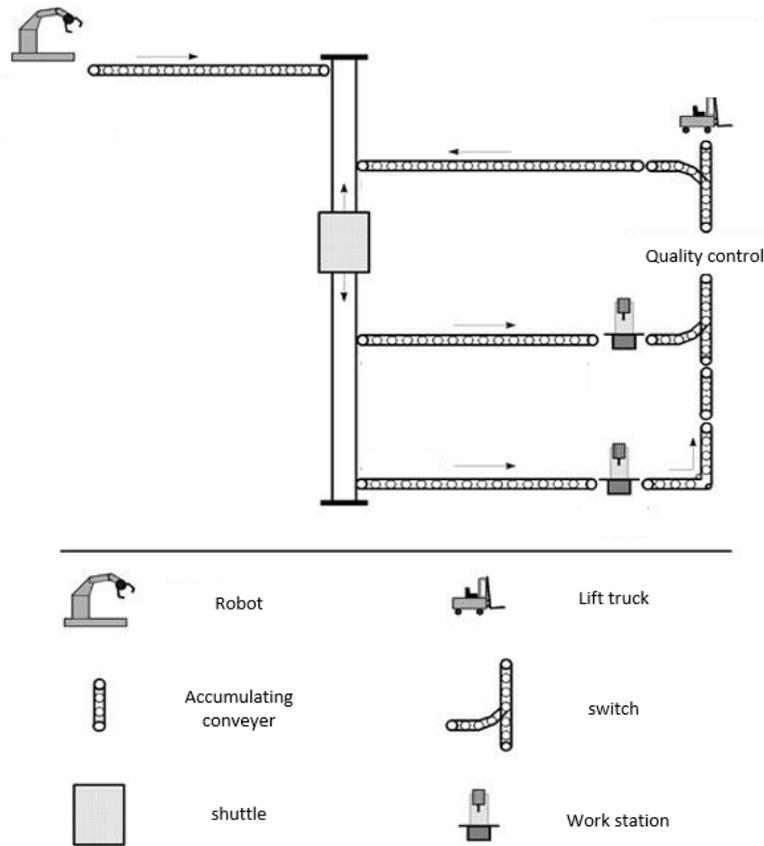


Figure 3-4: A small production system [Dosi12]

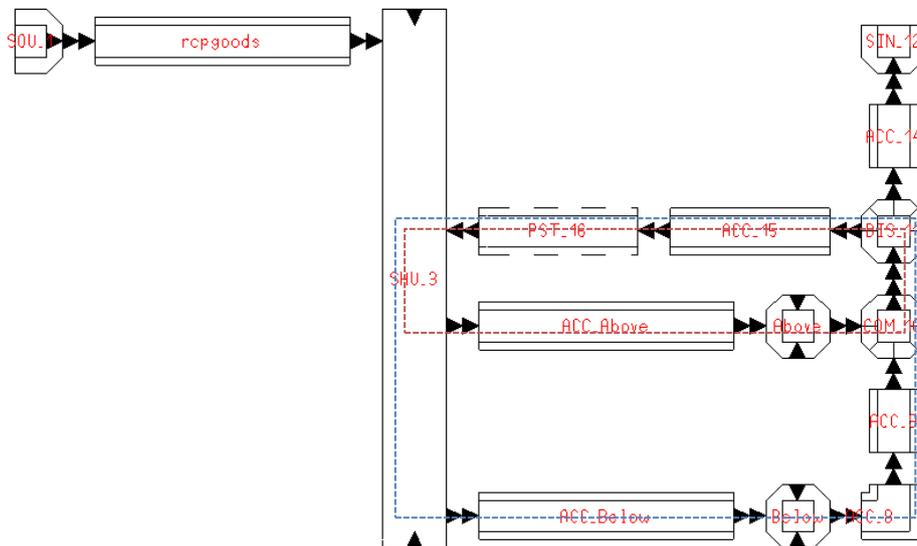


Figure 3-5: Simulation model of the small production system

The probability values of arriving entity types and rework proportions are listed in the following table:

Variable	Probability
Arrival of Entity 1	50%
Arrival of Entity 2	50%
Rework	15%

The probabilistic models in the simulation model are listed in the following table:

Variable	Probability distribution	Scale parameter	Shape parameter
Inter-arrival time	normal	60	5
Processing time work station 1	normal	80	5
Processing time work station 2	normal	80	5
Processing time forklift	normal	55	5

The complexity of this system cannot be calculated either by Frizelle's or Yu and Efstathiou's function, since, as can be seen, there are two rework loops in the system. The first rework loop is indicated by the red dotted line, the second by the blue dotted line. Furthermore, even if the complexity could be calculated and some numerical indicator given, the use of it in simulation validation would be every limited. So, here it is better to determine the relationships between the elements that are described above. In this way, any changes in assumptions or model input will be directly reflected.

As discussed previously, Type I autocorrelations exist in almost every output of a simulation study. So it can be expected that as the arrival rate increases the autocorrelation, such as in the waiting time of customers, will also rise in the model output. Autocorrelation in the simulation input, which can have a major impact on a simulation's results, can be dealt with using AR, ARMA or other time-series processes. This, however, is not the focus here.

Type II correlation between two variables exists in every manufacturing system. The two most basic elements are machines and buffers. The Type II correlation of buffer occupancy can normally be judged by a domain expert. For example, the occupancies of consecutive

buffers tend to be positively correlated, while the occupancies of buffers that compete for resources might be negatively correlated.

Type II correlation can be shown graphically by means of a correlogram, such as the one below:

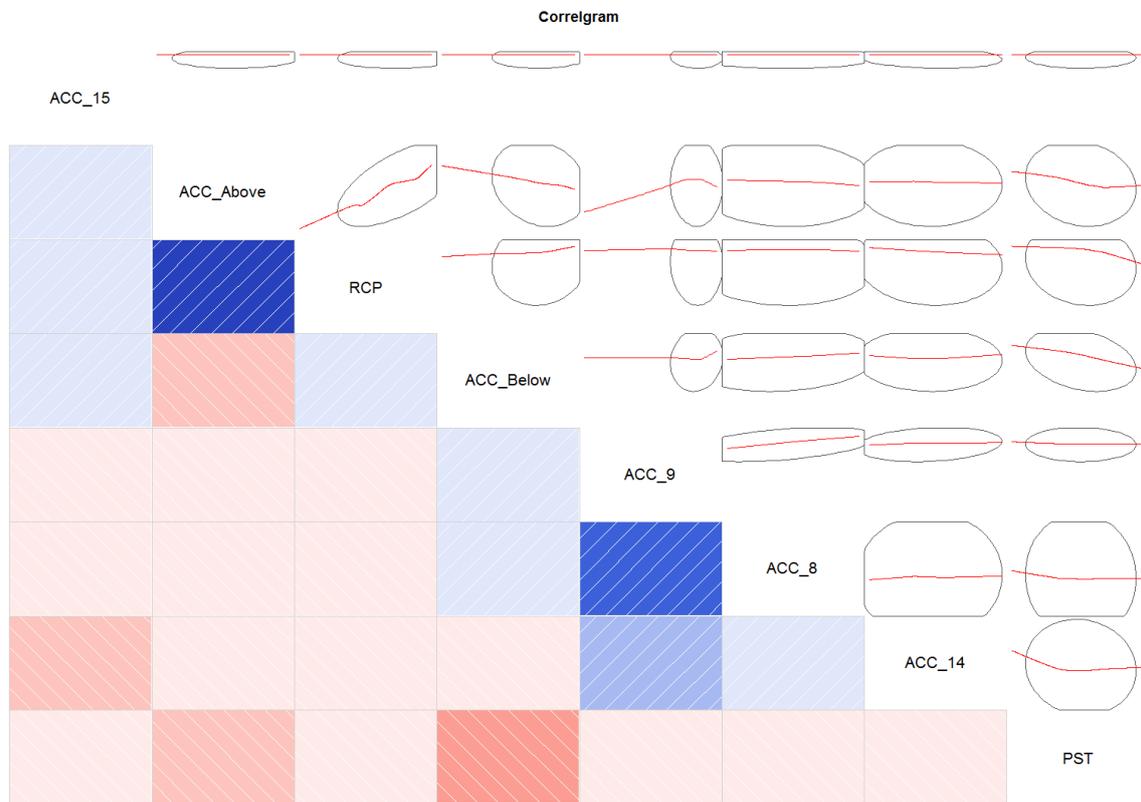


Figure 3-6: Type II correlation by correlogram

This is a correlogram of the occupancies of buffers. The names of the buffers are in the diagonal line. The lower part, in two colours, shows the various degrees of correlation of the buffers. The darker the colour is, the stronger is the correlation. The colour blue indicates a positive correlation, the colour red a negative correlation. For example, the buffers ACC_8 and ACC_9 are strongly positively correlated, since the colour is blue. The buffers ACC_Below and PST are strongly negatively correlated because the colour is red. On the upper part of the correlogram are the confidence ellipses and smoothed lines of the occupancies of the buffers.

The Type II correlation of the occupancy of buffers can be very helpful in the validation process, since it gives a first insight into the relationships and interactions of the basic elements of a simulation model. This type of correlation originates in a simulation model regardless of any other uncertainty it might contain. However, it is more useful if the interactions between more than two variables can be shown and, furthermore, a correlation between input and output variables can be found. Then a sensitivity analysis can be performed and used for model validation, e.g., if the model response is reasonable from an intuitive or theoretical perspective, or aligned with the correlation be sought, then the model may offer

some confidence with regard to its qualitative behaviour even if the quantitative precision or accuracy is unknown. This method is especially useful in situations in which only limited data or knowledge of a system is available.

Type III correlation can be shown graphically by the cross-correlation of two variables. For example, the figure below shows the cross-correlation between the occupancy of RCP and PST.

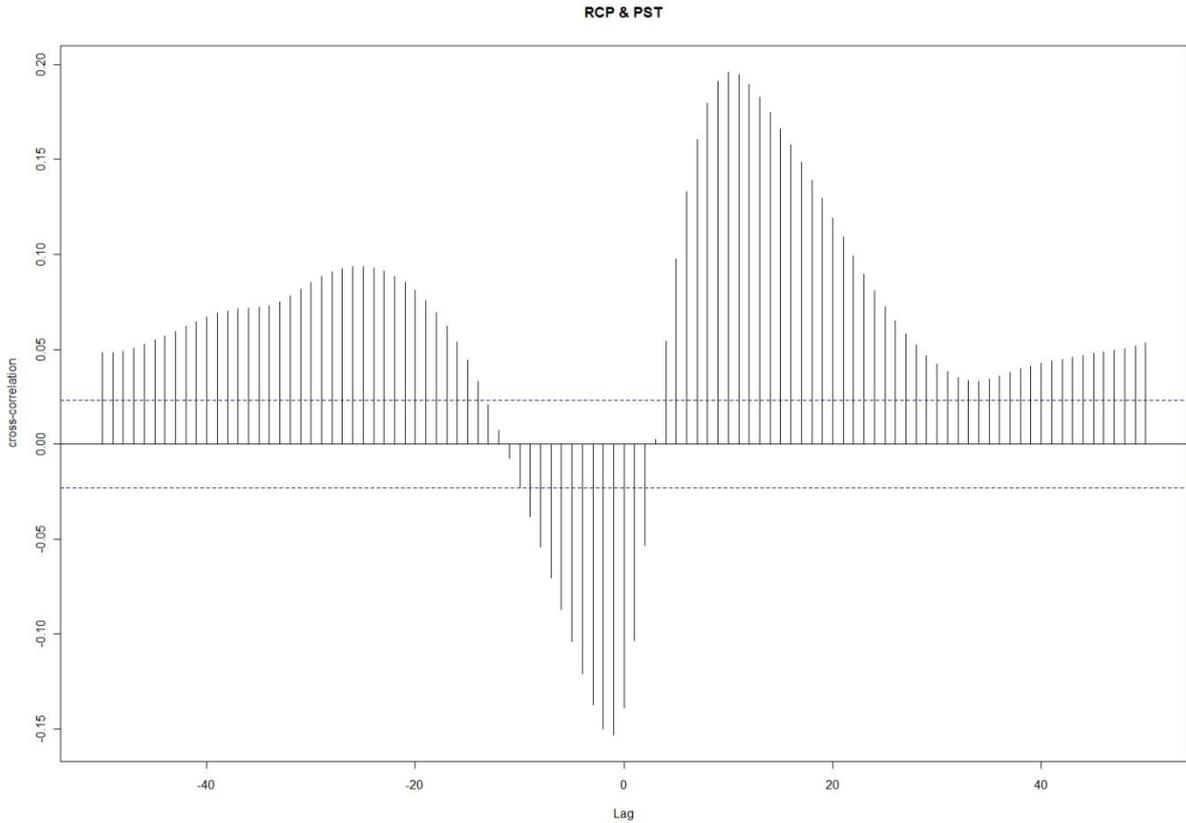


Figure 3-7: Type III correlation by cross-correlation plot

The two horizontal lines represent the approximate 95% confidence interval for statistical significance of the cross-correlation. This diagram shows the correlation between elements in delayed time periods. The longer the bar, the greater the correlation. The largest cross-correlations occur somewhere between $lag = -1$ and about $lag = 10$. It is difficult to read the lags exactly from the plot, so we might wish to give an object name to the cross-correlation factor and then list the object contents.

In a model in which only machines are directly connected with each other, the Type III correlation in the model has a very short time lag. For example, an increase in the arrival rate of parts will directly cause the first machine to increase its processing of those parts, while the second machine will immediately start processing the parts from the previous machine with the time lag of the service time of the first machine. When there are buffers in between, the

time lag of Type III correlation is increased by the delay of the buffers. And the correlation is also reduced by the existence of buffers, since the buffers decouple the interaction of the machines.

Another way to show graphically Type III correlation is in the scatterplots seen here, in Figure 3.8, below. In each plot, the occupancy of RCP is on the vertical and a past lag of PST is on the horizontal from lag zero to lag eleven. The correlation values are given on each plot in the upper right corner.

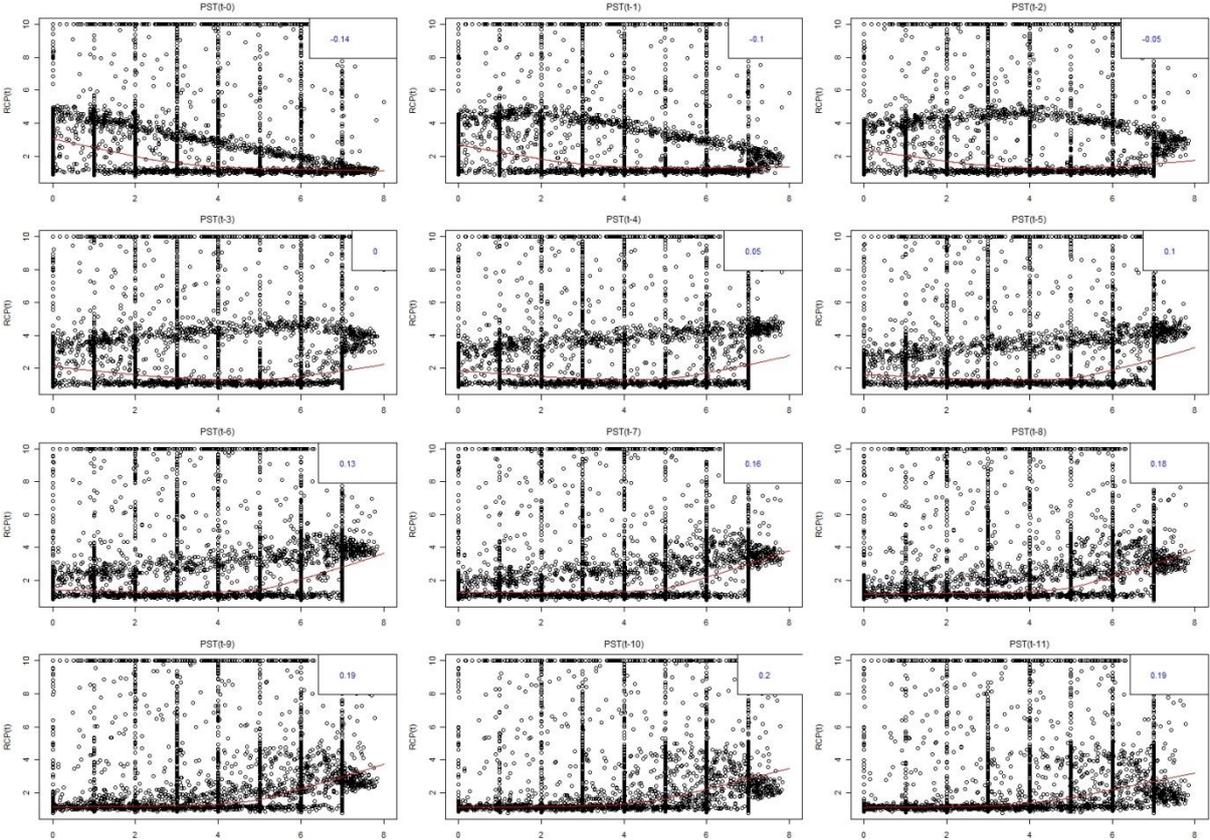


Figure 3-8: Type III correlations by scatterplot

Even though the three types of complexity can show only the relationship between a maximum of two parameters, they are still considered necessary to reveal the important relationships between vital elements. A multi-parameter correlation makes no sense because a sensitivity analysis can yield better results.

Actually, the parameters can also be the performance values of a sub-system. For example, it is not only the throughput of a buffer, but the throughput of a sub-system which is defined by a model builder. Taking the previous model as an illustration, all the entities within the grey dashed lines can be regarded as a sub-system. The Type II correlation between RCP goods and this subsystem can thus be revealed.

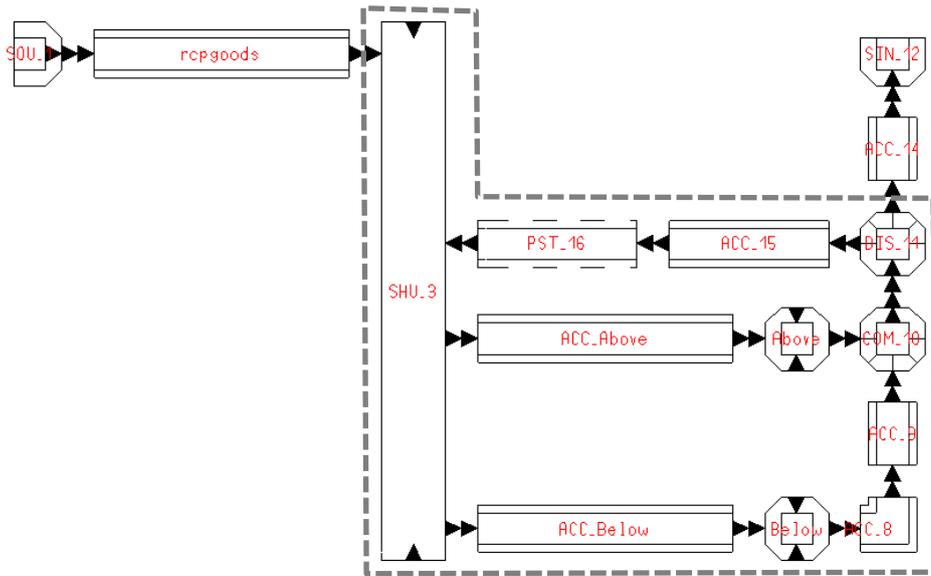


Figure 3-9: Simulation model of a small production system

3.3 Uncertainty in simulation models

3.2.1 Types of uncertainty

Uncertainty is traditionally classified into two groups: aleatory uncertainty and epistemic uncertainty [Ober01].

The term ‘aleatory uncertainty’ (also referred to as variability or irreducible uncertainty) is used to describe the inherent variation associated with a real system and its environment.

Epistemic uncertainty, which is also referred to as reducible uncertainty or model-form uncertainty, describes the lack of knowledge of a system or its surrounding environment, or incomplete information.

Aleatory uncertainty in production and logistics systems can be confined by an analysis of input data which will result in different types of input model, such as statistical distribution models, empirical distribution models or historical data for trace-driven simulation. Epistemic uncertainty is caused by: 1). A lack of knowledge of a system, for example when a new system such as a production line or a new warehouse is planned, the information about these systems and their behaviour is sparse; 2). A lack of knowledge about a system’s environment, for example customers’ demand for a new product and its effect on the demand for other serial products, which could be only estimated through market research; 3). Incomplete information can result when data is lacking or not enough data about processes or systems has been collected owing to time constraints. The collection of data on machine failure, for example, is time consuming. Such data may be available only from the providers of the equipment, as in the form of theoretical machine availability.

These two types of uncertainty complicate model validation, which is a process already made difficult by the complexity involved. The result of this is that the behaviour of neither the existing or designed system, nor of the simulated system, can be known for certain. The uncertainty in the measure of performance needed to validate systems can be so large that it cannot be judged.

Aleatory uncertainties cannot be reduced in a simulation model, since they are inherited from the system of interest. The only thing one can do is to guarantee that the right input models are chosen. By contrast, epistemic uncertainty can be reduced by obtaining more information about the system, for example, in extending the sampling period in order to gain sufficiently large sets of raw data from the system – if the system exists. However, data availability does indeed seriously affect the model-building and model validation processes. Kleijnen [Klei98] has suggested various methods for model validation regarding the availability of data, which will be discussed in detail later in this chapter. In general, epistemic and aleatory uncertainty should be addressed in the model assumptions and be agreed by both the model builder and the model user in advance.

Attention must also be paid to the fact that both epistemic and aleatory uncertainty change in the modelling process.

How aleatory uncertainty affects the model validation process will be shown in the following examples.

3.3.2 Examples of model uncertainty

To illustrate uncertainty in simulation modelling, a simple example is here presented. Law and Kelton [LaKe00] compared alternative system configurations with some modifications. The focus was not on the performance differences between the systems. The objective was to determine how changes in the perspectives of system complexity and variation influence model validation. So, the first comparison would be made without any randomness in the systems. Then different types of randomness were added in.

A factory planning to install a drilling machine must choose between buying one Zippy drill machine, two Klunky drill machines or two Oldy drill machines. One Zippy costs twice as much to purchase, install and operate as one Klunky, but the Zippy works twice as fast. Two serially connected Oldies (Oldy1 and Oldy2) have the same function as one Zippy or Klunky, and they work as fast as one Zippy with the same costs.

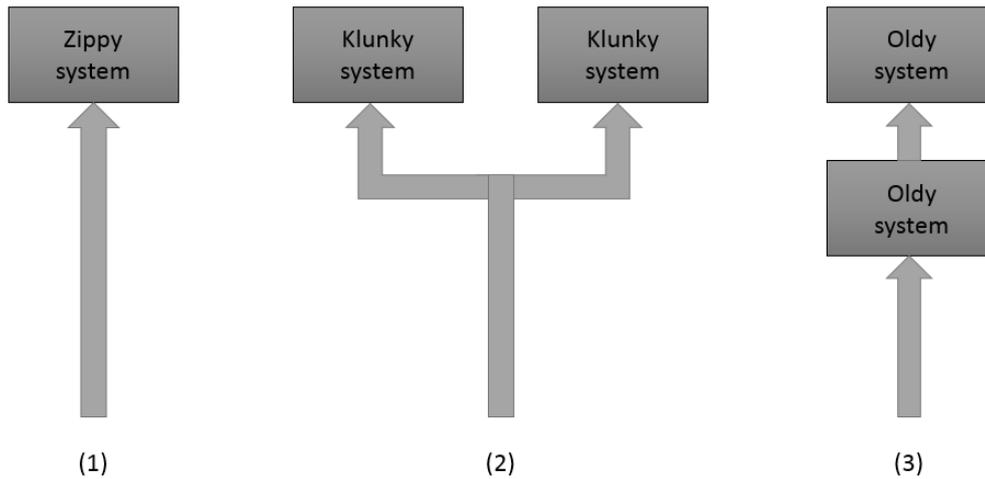


Figure 3-10: One-Zippy system (1), two-Klunkies system (2), and two-Oldies system (3)

Scenario 1

In Scenario 1, there is no aleatory uncertainty and no epistemic uncertainty in the model. Furthermore, the randomness in the model has been taken out.

Customers arrive one at a time at a rate of 1 per minute in all three systems. The Zippy is served at a constant service time of 0.9 minutes, the Klunky is served at a constant service time of 1.8 minutes, and the Oldy1 and Oldy2 are served at constant service times of 0.45 minutes.

The results are more than clear: no customer waiting time in all three systems; the throughput times of Zippy and Oldies are 0.9 minutes, for the Klunky system it is 1.8 minutes.

The Zippy system is a D/D/1 system. It has a deterministic arrival and a deterministic service time. The arrival rate $\lambda = 1$ is less than the service rate $\mu = 1/0.9 = 1.1$; thus, there is no waiting time.

The Klunkie system is a D/D/2 system. This, too, has a deterministic arrival and a deterministic service time. The arrival rate $\lambda = 1$ is less than the service rate $\mu = 1/1.8 * 2 = 1.1$; thus, there is also no waiting time.

The Oldy system is a production line system with two servers in series without buffers in between. It has an arrival rate of $\lambda = 1$ and a fixed cycle time of $0.45 + 0.45 = 0.9$. Therefore, it is a paced synchronous line. Thus, there is no waiting time.

Scenario 2

Customers arrive one at a time according to a Poisson process at a rate of 1 per minute for all three systems.

- The Zippy is served at an exponential service time with a mean of 0.9 minutes

- The Klunky is served at an exponentially distributed service time with a mean of 1.8 minutes
- The Oldy1 and Oldy2 are served at an exponentially distributed service time with a mean of 0.45 minutes.

In the Zippy system with Poisson arrival and exponential service time, an M/M/1 queue is formed. The theoretical mean time that customers spend in the queue can be calculated using the queuing model equation:

$$W_Q = \frac{\lambda}{\mu(\mu-\lambda)}$$

The theoretical mean time spent in the queue in the Zippy system is $W_Q = 8.1$ minutes.

Compare this to the simulation results (100 hours run length, 50 replications)

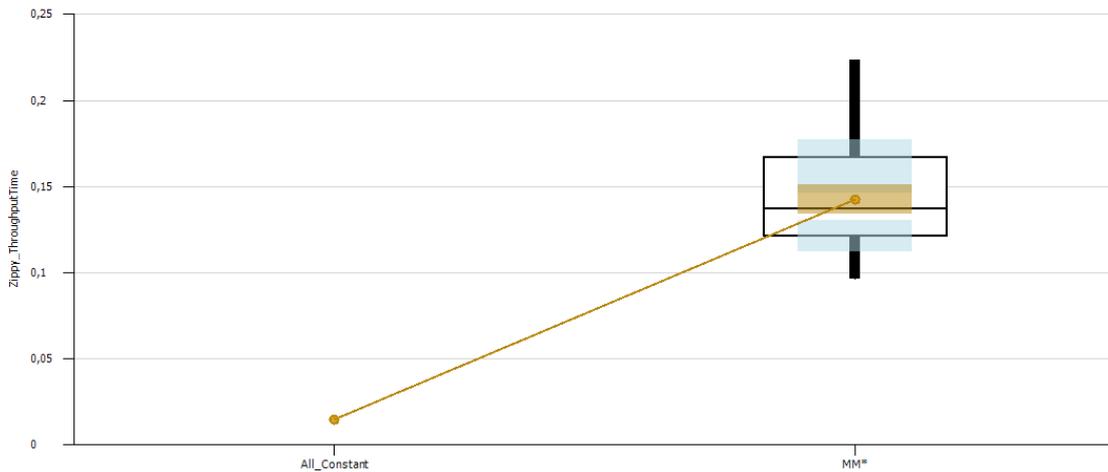


Figure 3-11: Comparing constant system throughput time and zippy system throughput time with Poisson arrival and exponential service time

It is quite clear that the result from the simulation model is no longer a constant. Firstly, the model’s mean waiting time is $W_{ModelQ} = 7.7$ minutes rather than the theoretical mean waiting time $W_Q = 8.1$ minutes above. Secondly, W_{ModelQ} has around half this value, at 0.48 minutes.

Ultimately, the result is a 95% confidence interval (7.22, 8.18), which covers $W_Q = 8.1$. In this way, it can be confirmed that the model is valid, since its 95% confidence interval covers the theoretical value, which is the “TRUE” value.

In the Klunky system, with the exponential inter-arrival time and the exponential service time, we get an M/M/2 queue. A queuing model can be applied to derive the theoretical mean time spent in the queue:

$$w_Q = \frac{L_Q}{\lambda}$$

in which L_Q is the average customer waiting time in the queue and can be calculated by means of the following equation:

$$L_Q = \frac{\left(\frac{\lambda}{\mu}\right)^2 \lambda \mu}{(2\mu - \lambda)^2} P_0$$

$$P_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\left(\frac{\lambda}{\mu}\right)^2}{2} \left(\frac{2\mu}{2\mu - \lambda}\right)\right)^{-1}$$

The theoretical mean time spent in the queue in the Klunky system is $W_Q = 7.67$ minutes

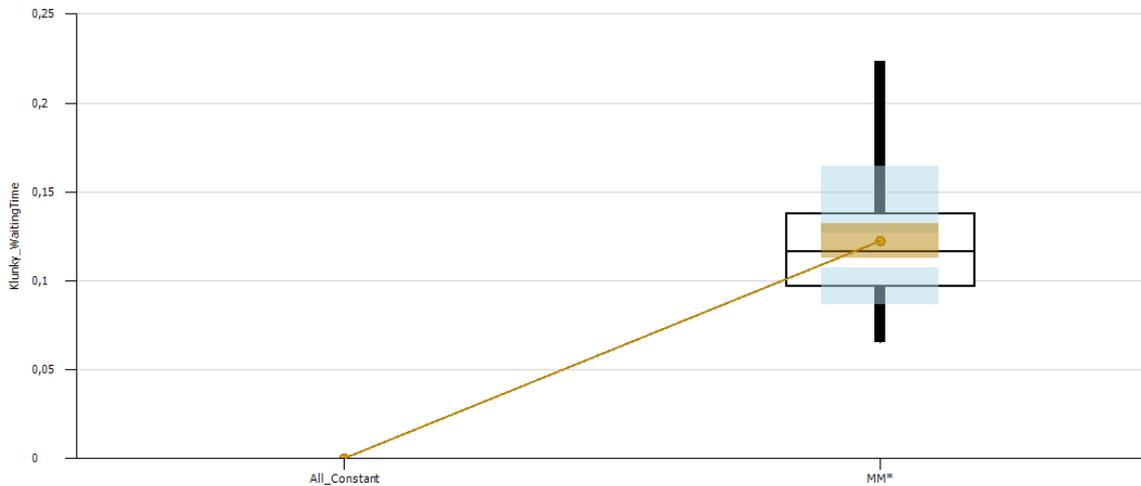


Figure 3-12: Comparing constant system waiting time and klunky system waiting time with exponential arrival time and exponential service time

The model's mean waiting time $W_{\text{ModelQ}} = 7.35$ minutes rather than the theoretical mean waiting time $W_Q = 7.67$ minutes above. W_{ModelQ} has half this value, at 0.54 minutes. Ultimately, the result gives a 95% confidence interval (6.81, 7.89), which covers $W_Q = 7.67$. In this way, it can also be confirmed that the model is valid, since its 95% confidence interval covers the theoretical value, which is the "TRUE" value.

In the Oldy system, with the exponential inter-arrival time and exponential service time, there should be an extra problem to consider, namely, the buffer size between Oldy1 and Oldy2. If it is assumed that the buffer size is infinite, then the system becomes an asynchronous line with exponential processing times. The mean production lead time (mean time spent in the system) is given by:

$$L = \sum_{i=1}^n \frac{1/\mu_i}{1 - \lambda_i/\mu_i}$$

By Burke’s theorem, the output process of an M/M/1 system is still a Poisson; the arrival rate at the second machine, Oldy2, is again a Poisson arrival with rate λ . So Oldy1 and Oldy2 are both M/M/1 with arrival rate λ and service rate μ (since the service rates for both Oldy1 and Oldy2 are identical). The mean time spent in the system can be written as:

$$L = 2 \frac{1/\mu}{1 - \lambda/\mu}$$

Thus, the mean time spent in the queue is:

$$W_Q = 2 \frac{\lambda}{\mu(\mu - \lambda)}$$

The theoretical mean time spent in the queue of the Oldy system with an infinite buffer between machines is $W_Q = 0.736$ minutes.

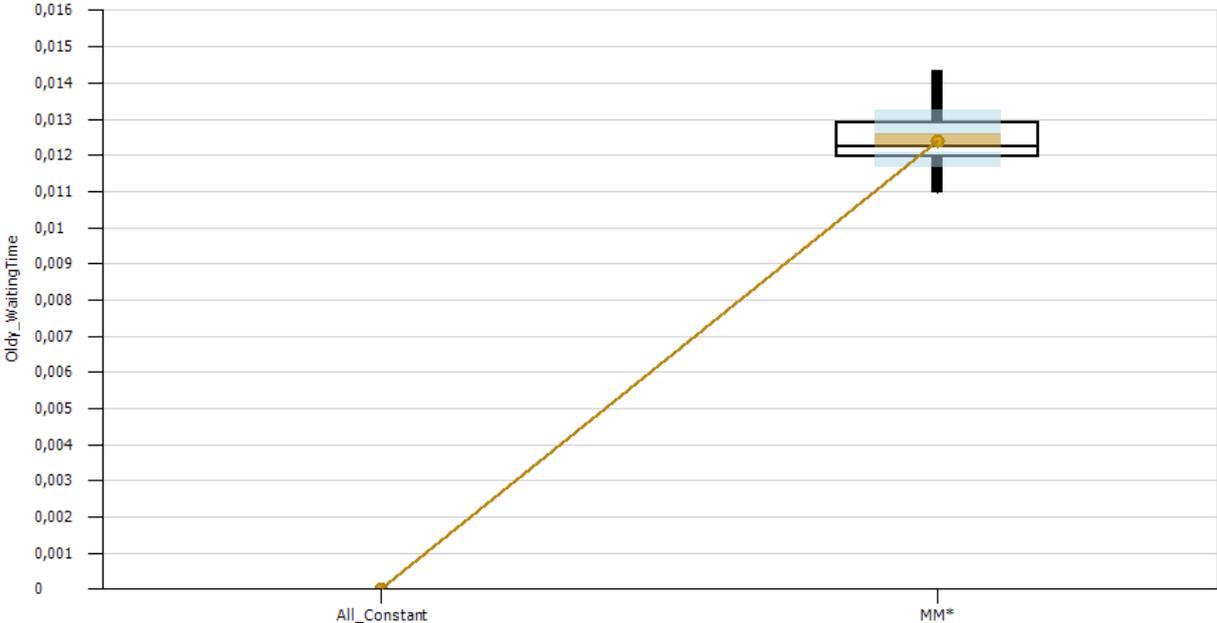


Figure 3-13: Comparing constant system waiting time and oldy system waiting time with exponential arrival time and exponential service time

The model’s mean waiting time $W_{ModelQ} = 0.738$ minutes, rather than the theoretical mean waiting time $W_Q = 0.736$ minutes above. W_{ModelQ} has half this value, at 0.012 minutes. Ultimately, the result is a 95% confidence interval (0.726, 0.750), which covers $W_Q = 0.736$. In this way, it can also be confirmed that the model is valid, since its 95% confidence interval covers the theoretical value, which is the “TRUE” value.

If the buffers between Oldy1 and Oldy2 are limited, for example if the buffer size is zero, then the Oldy system becomes an asynchronous line with finite buffer. In this case, the first machine, Oldy1, may be blocked if Oldy2 has a failure (failure of machines will be discussed in the following example) or a long operation time and thereby fills up the buffer. The buffer here is used to decouple the two machines. If it is assumed that the buffer size is zero, then

they have the largest coupling. If there are infinite buffers, as in the situation above, then they have the smallest coupling.

Assuming that the buffer size is zero, then the first machine, Oldy1, is never starved and the second machine is never blocked. Then, Oldy2 becomes an M/M/1/1 queue.

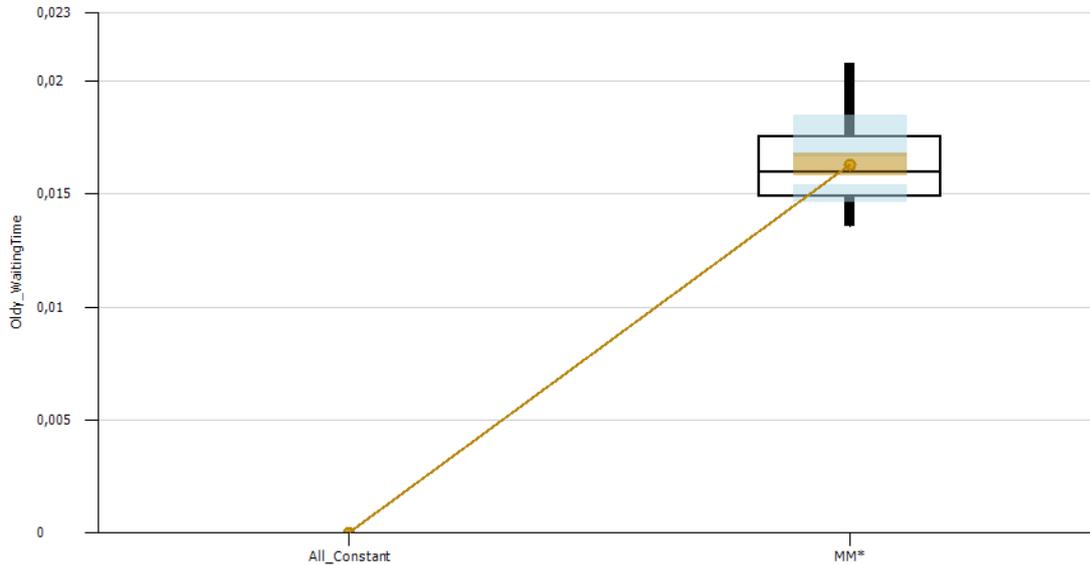


Figure 3-14: Comparing constant system waiting time and oldy system waiting time with exponential arrival time and exponential service time and no buffer

The model's mean waiting time $W_{ModelQ} = 0.978$ minutes rather than the theoretical mean waiting time $W_Q = 0.017$ minutes above. W_{ModelQ} has half the value, at 0.03 minutes. Ultimately, the result is a 95% confidence interval (0.975, 0.981).

The difference between the constant input model and the statistical distribution is caused by ignorance of the aleatory uncertainty.

There are several simulation software packages on the market that provide the function needed to disable the randomness and facilitate the verification and validation of simulation models. For example, in Simio, if 'disable randomness' is activated, the stochastic behaviour of any variable will be eliminated. The distribution will return the mean instead of a random variate, and any probabilistic selection in the logic will always select the choice that includes the 50% cumulative probability. This function is very useful when verifying simulation models. Since the randomness is taken out of the model, the behaviour of the simulation is much more predictable. However, this function should not be used as a validation tool.

Because the aleatory uncertainty in the model is irreducible, any ignorance of aleatory uncertainty will lead to the model being invalid. The essence of aleatory uncertainty is in the form of a statistical distribution model in this example.

Scenario 3:

In this scenario the epistemic uncertainty is demonstrated. It takes the example of incomplete information when there is a shortage of data or there is not enough time to collect data because of time constraints.

The inter-arrival time remains unchanged. The processing time is changed from an exponential distribution to 1) an Erlang distribution with $k=2$; 2) an Erlang distribution with $k=4$; and then 3) an exponential distribution with machine breakdowns; 4) an Erlang distribution with $k=2$ plus machine breakdowns; 5) an Erlang distribution with $k=4$ plus machine breakdowns.

The throughputs of all three systems are as follows.

Throughput of the Zippy system

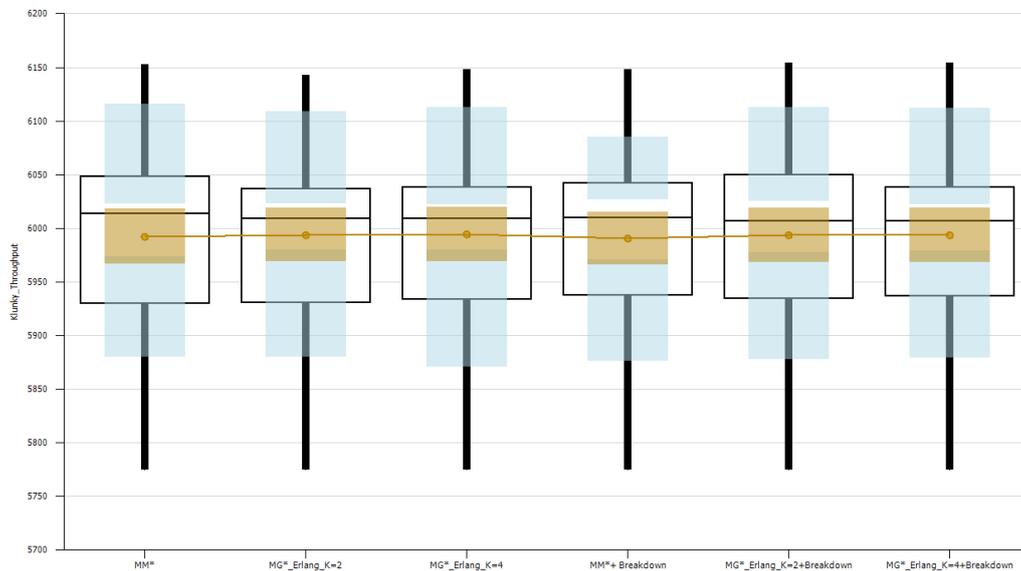


Figure 3-15: Comparing Zippy system throughput with different configurations

Throughput of the Klunky system

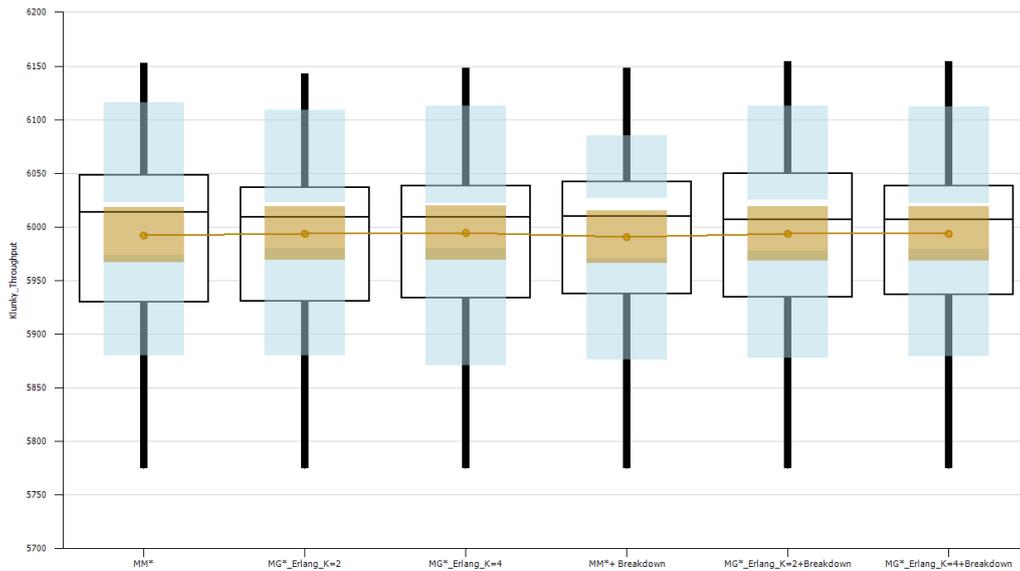


Figure 3-16: Comparing klunky system throughput with different configurations

Throughput of the Oldy system

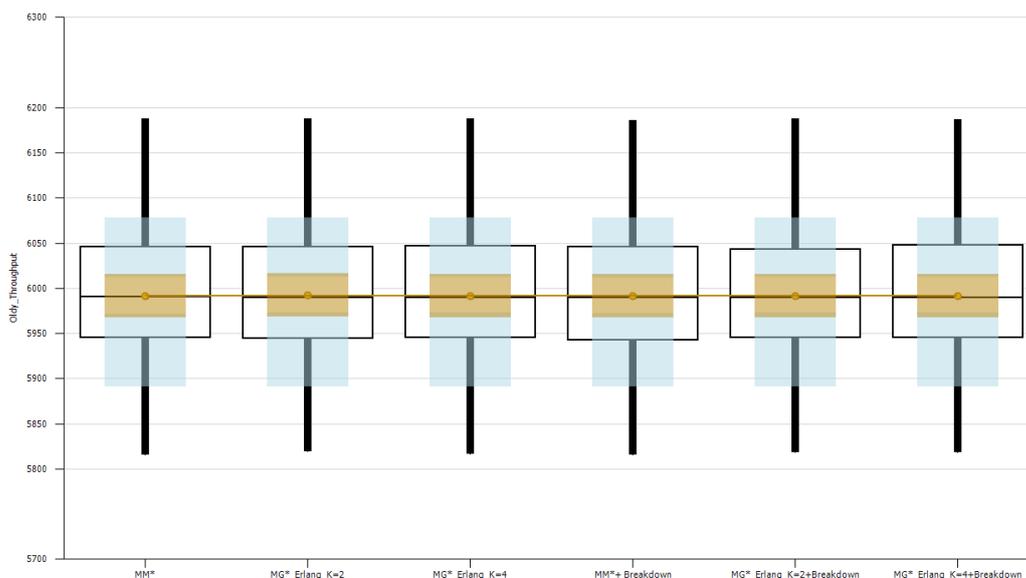


Figure 3-17: Comparing oldy system throughput with different configurations

As shown in the figures above, in each system the throughputs under different configurations are only slightly different from each other. The confidence intervals overlap with most of their parts, even the system configurations with breakdowns!

Moreover, the throughputs across the three different systems are also very close to one another (ca. 60/hour).

Why is the throughput not sensitive to the input model? If it is insensitive to the changes or “errors” in the input model, should it be used as an indicator in model validation?

One can apply Little’s law [LiSt08] in manufacturing systems. Hopp and Spearman [HoSp01] use it to define the relationship between WIP, cycle time and throughput as:

$$\text{Throughput} = \frac{\text{WIP}}{\text{Cycle Time}}$$

Throughput is defined as the average output of a production process (machine, workstation, line, plant) per unit time.

WIP (work in process) is defined as the inventory between the beginning and end points of a product route.

Cycle time is defined as the average time from the release of a job at the beginning of the routing until it reaches an inventory point at the end (the time the part spends as WIP).

Obviously, in order to increase the output, the work-in-progress inventory should be increased and/or the cycle time should be reduced. WIP is increased when 1) the inter-arrival rate is higher, 2) the buffer size is larger, 3) the service rate is reduced, 4) machine availability is reduced. The cycle time can be reduced when: 1) the service rate is higher, 2) the buffer size is smaller, 3) the arrival rate is reduced, or 4) machine availability is increased.

However, it is quite clear that the two variables are positively correlated with each other, since, when the WIP is increased and all other parameters remain unchanged, then the cycle time will also increase, and consequently the throughput will be barely affected.

3.3.3 The coefficient of variation

A measure of the variability in the distribution of model input is the ‘coefficient of variation’ (cv), expressed as the standard deviation as a percentage of mean. cv is defined by:

$$(cv)^2 = \frac{V(X)}{[E(X)]^2}$$

The larger the coefficient of variation, the more variable the distribution relative to its expected value. $V(X)$ is the variance, $E(X)$ is the mean.

Banks et al. report, for an M/G/1 system, that where the service time is constant, the variance of service time is zero and the cv is also zero. With Erlang distributed service times of order k , where the variance of service time is $1/k\mu^2$ and the expected value of service time is $1/\mu$, the cv is then $1/\sqrt{k}$. If the service time is exponentially distributed with a rate μ , then the expected service time is $1/\mu$ and the variance is $1/\mu^2$. So the cv is here 1. If the service time is hyperexponentially distributed, then the cv is always greater than or equal to 1.

If the server utilization is kept the same for all the service types mentioned above and we compare the server utilization with the mean number of customers in the queue, at the same level of server utilization, the smaller the cv of the server, the fewer the customers in the queue. According to the definition of the coefficient of variation, the larger the cv, the greater the variation.

The reason for this can be revealed by a simple transformation in terms of the coefficient of variation of the formula for the average number of customers in the queue of an M/G/1 system, since $(cv)^2 = \sigma^2/(1/\mu)^2 = \sigma^2\mu^2$. Then,

$$L_Q = \frac{\rho^2(1 + \sigma^2\mu^2)}{2(1 - \rho)} = \left(\frac{\rho^2}{1 - \rho}\right)\left(\frac{1 + (cv)^2}{2}\right)$$

The first part, $\frac{\rho^2}{1 - \rho}$ is the long-run average number of customers in the M/M/1 queue, while the second part, $\frac{1+(cv)^2}{2}$ and corrects the first part to account for a non-exponential service time distribution. It can also be understood as signifying that the larger the variation in the system is, the more is the waiting time.

What is the use of the coefficient of variation in model validation?

One validation method for production systems is to set the equipment's theoretical values without machine failures into one's model. In this way, all processing times are set as constant, the machines have no failures and the coefficient of variation of every machine is zero. Consequently, the output of the system reaches the highest level under these conditions.

If the service times of different machines are not constant and the arrival rate is not exponentially distributed, this rule is still valid.

However, the elimination of variation in simulation models does not simply infer that we use the mean, replacing the statistical distribution of random processes, and eliminate machine failure. In some logistics systems the variation may also be generated by the sequencing. For example, the classical "Perlenkettengüte" sequencing problem, which is a controlled sequencing problem, needs to be generated with great care in order to reach maximum performance. And the algorithm for ensuring that the possibilities in each "Perlenketten" are the same is also part of variation control.

Another kind of variation, which is often ignored, is the variation of proportion. The example of the small factory model is used again here to illustrate this problem. In the model, after the processing of machines above and below, 15% of the entities need to be reworked. In the programming, this is done not by changing any attributes of the entities at the workstation, but by a distribution element, DIS_11 (Figure 3-5). There, the element randomly sends 15% of all arriving entities, which are both Type 1 and Type 2 entities, to be reworked and the other 85% to the sink. This can cause a problem in the validation, because the entities making up the 15% for reworking no longer represent 15% of all Type 1 entities and 15% of all Type 2 entities. For validation purposes, an algorithm should be developed to control for this kind of variation.

Here, an experiment can be conducted to make this point clear. In Model 1, the reworking is done as described above; in Model 2, the reworking is done by defining two new entities, which are generated at work stations. Each new entity is 15% of the total processed entities and is sent for reworking by the distribution element. The throughput time of both systems can be seen in the following two figures:

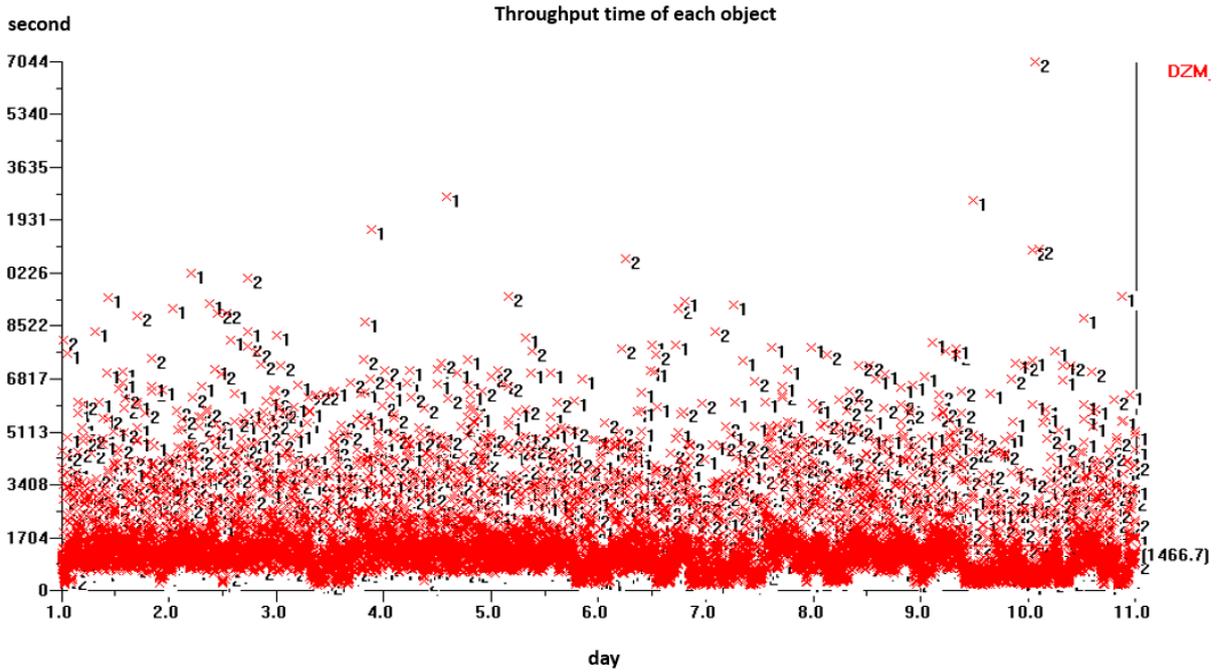


Figure 3-18: Throughput times of Model 1

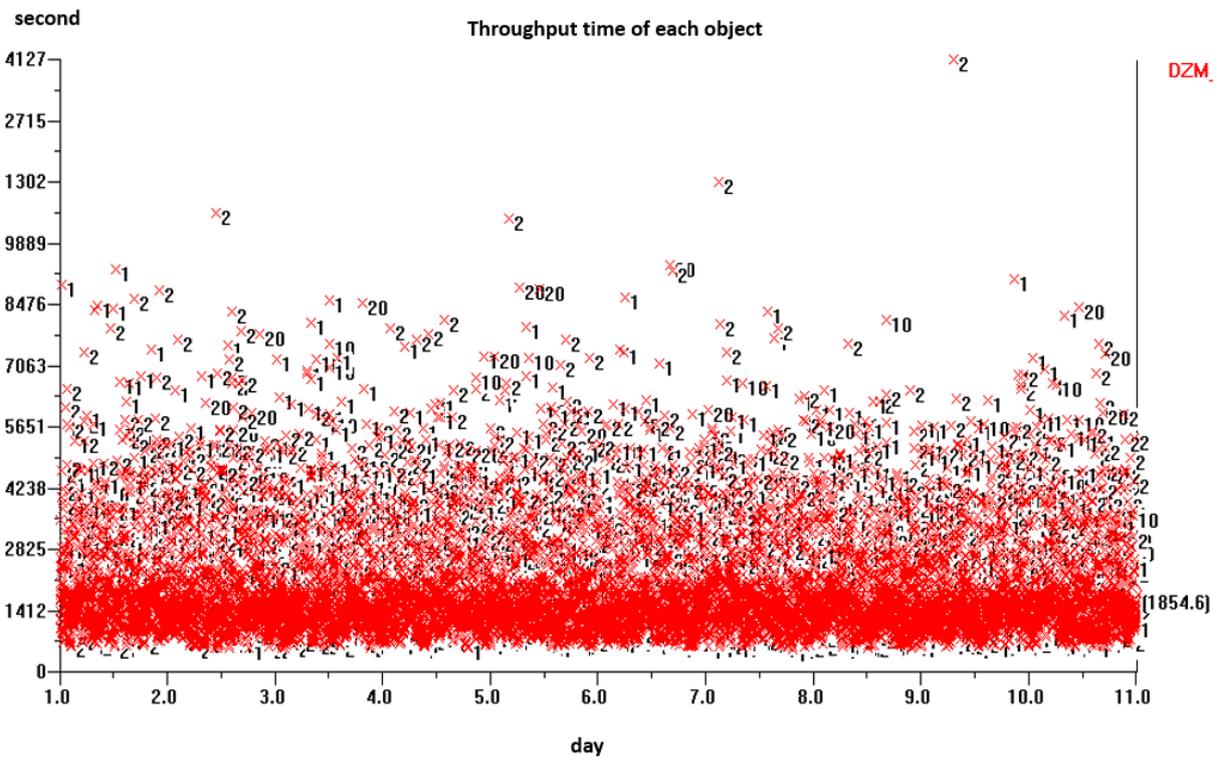


Figure 3-19: Throughput times of Model 2

It is quite obvious that in Model 2 the average throughput time is 1854.6 seconds, while in Model 1 it is 1466.7 seconds, even though the average throughput per hour in each model is approximately the same.

3.4 Data in model validation

Data is essential in any simulation study. It is also essential for model validation. Data is the fundament of system complexity and model uncertainty. The level of detail and the scope of a model, which have a large influence on model validation, depend on the purpose of a simulation study as well as on the quality and availability of data.

Sargent [Sarg10] discusses the idea of operational validation by setting data validity in the centre. He defines data validity as “ensuring that the data necessary for model building, model evaluation and test, and conducting the model experiments to solve the problem are adequate and correct.”

As shown in the Figure. 2-9, data is collected and used throughout the simulation study. It is collected for the model input. It is investigated for the building of the model. It is also collected, if the system exists, for the validation of the simulation model by statistical methods. It is then analysed as the output of the model experimentation.

In this section, data is understood as: 1) the number of variables, 2) the number of data points collected per variable, and 3) the number of constants.

The more variables there are in the model and the tighter the interdependence between them, the higher the model complexity and the epistemic uncertainty. By contrast, the more data points that can be collected for the variables, the less the model epistemic uncertainty will be. However, model complexity cannot be reduced by more data; it can only be better understood.

Data can be categorized into two groups from the validation perspective: data quality and data availability.

The quality of data depends on how it is collected and who collects it. Data that is collected electronically is essentially better than manually collected data. For example, when scanning technology is used with bar coding for point-of-sale (POS) and material tracking, the reading error is one per 2-15 million characters read. By comparison, manual data entry creates one error in every 150 keypunches. The second factor of data quality is that of the person collecting it. When data is provided by the process owner, for example the repairman, then care should be taken, since such data may be corrupted by him for his own interests. If this is the only way one can get such data because of time pressure, for example, then it is of course the only source that can be relied on. Furthermore, data being collected in this way has to be processed or “cleaned” and put into the proper format before it can be used in simulation modelling. If it is possible for the simulation analyst to collect data himself, the quality will be much better, provided there is enough time for system observation.

Data availability is another vital issue in model validation. Normally, data availability depends largely on whether a system is observable or not, e.g. whether it actually exists or is still in the design phase. If a system is only planned and still in the design phase, or if drastic modifications are planned for a system, then the amount of data is limited. Nevertheless, since logistics and production systems are highly engineered, input data from automated systems

can still be collected from equipment providers. This would include, for example, the processing times of machines, travelling speed of forklifts, pick cycle times of automated storage and retrieval system (ASRS or AS/RS), etc. Output data can still be calculated or be set in the form of system key performance indicators, even though it will have only theoretical values, for example the theoretical throughput per hour or the throughput time of a production line. In this way, input data usually comprises statistical models and output data point estimates. If a system is observable, then input data or output data or both, such as production or production planning data, customer order data, etc., can be collected.

For model validation the best situation is to have both input and output data from the existing system. If the input data sets are large enough, then trace-driven simulation can be performed. The output data of the model is then compared with the output data from the existing system. If they closely resemble each other, the model of the existing system is regarded as validated. This is the most effective and definitive test of a simulation model's validity. The method is called 'results validation'. Since the observed data from the systems is used to drive simulation, the method is also called the 'correlated inspection' approach.

Then the problem is how to compare the simulation data and historical data. Kleijnen reports that by using a scatter plot with historical data p and simulation output data q , one can fit a line $q = \alpha + \beta p$, and then see whether $\alpha = 1$ and $\beta = 0$, e.g. the line has a 45 degree slope and the zero intercept is totally wrong. This can be easily proved. Provided the real and simulated outputs have equal positive mean $\mu_p = \mu_q = \mu$, since the correlation between p and q cannot be perfect, e.g. $\rho_{pq} < 1$, then it leads to $0 < \alpha < \mu$, $0 < \beta < 1$.

If one has only output data, then the Student's t-test can be used for validation. However, this test is not very sensitive to non-normality. Because the output data of simulation models is almost always autocorrelated, output analysis should be used. For the termination of a simulation, the output data of each simulation run should be independent and identical.

In steady-state simulation, the repetition-deletion technique or batching method can be used to create identical, independently distributed output data.

If there is no data, which here can be understood as meaning that a system has yet to be built or will undergo drastic changes, knowledge of model complexity will be decisive in the validation procedure. As already explained, data on logistics equipment is almost always available. Changes of output in a simulation model can also be judged by changing certain inputs, since a change in direction can be predicted by one's knowledge of the model complexity. The vital procedure will be that of finding the most important input factors in the model.

3.5 The three factors of simulation model validation in production and logistics

The three factors of model validation, data availability, complexity and uncertainty, affect one another. However, data availability is normally seen as the most important factor in model validation, since only when data from a real system is available can objective statistical methods be used and model validation be truly performed. Data availability also affects complexity. In the extreme case, where no real data is available, the complexity can only be estimated. As in logistics systems data is almost always available, even if it has only theoretical values, the three types of complexity can help simulation practitioners check the plausibility of their models. Furthermore, how a change in one parameter will affect other parameters is made clear. The relationship between three or more parameters can be revealed only by sensitivity analysis; this will be discussed in the next chapter. An increase in data has an effect on model uncertainty, especially on the epistemic uncertainty. A decrease in epistemic uncertainty occurs in the process of data collection. The more data is obtained from a system, the less epistemic uncertainty there will be in the model, and the more plausible is the model from the point of view of validation.

Bearing these three factors in mind, attention should be paid to the following procedure:

1. Collect system data. Reduce epistemic uncertainty as much as possible. Depending on the type of data collected, the relevant statistical methods can be used, the one and only way to truly validate a model is to compare model output and system output when a real system already exists. If a real system does not exist, methods to prove each of the three model properties should be used.
2. Use the data on hand to run the model and check its complexity. Calculate the different types of complexity and determine any counter-intuitive correlations in the model. Since in realistic models the number of factors might be too large to be proved by people, experiment design should be used to check efficiently.
3. Reduce the uncertainty in the model to a minimum by eliminating failure and variation. In this way, the theoretical performance can be calculated. However, this will also cause the stochastic character of the discrete-event simulation model to be lost. A better way of reducing uncertainty is to determine which factors are the most important and have the largest impact on the model's performance. By ignoring the unimportant factors, the uncertainty in the model can be reduced and one's efforts can be focused on data collection and the reduction of uncertainty for the important factors.

In the next chapter, the methods for these suggestions will be discussed.

Chapter 4 Current techniques of model validation and a new paradigm including sequential bifurcation

In this chapter, the most widely used validation methods are discussed from the three aspects mentioned in Chapter 3: data availability, model complexity and model uncertainty.

However, by uncertainty we do not here mean from the input side, but from the output side, which means that the real uncertainty in model validation is the uncertainty of the output. No matter how vaguely the range of the input variables can be defined, as long as they do not affect the output their degree of uncertainty is not important. For example, if one factor in a complex system is changed, such as doubling the speed of a shuttle, if this does not lead to any change in the output of the model the uncertainty of the speed of the shuttle is not significant in the validation. This is also the purpose of sensitivity analysis. It is designed to determine the most important factors affecting model output, in other words the factors that most affect the uncertainty in model output.

4.1 Current techniques of model validation

There are both subjective and objective techniques of validation. Some are dynamic, while others are static.

4.1.1 Current validation techniques by Sargent

Sargent [Sarg03] speaks of fifteen model validation techniques, in the following section, these techniques are discussed with the three criterion:

a. Animation

In animation the graphical operation of a model is observed in order to expose any discrepancies in it. However, with the complexity of most even medium-sized models in production and logistics systems, this method can be used only as a rough check of the most obvious programming bugs rather than as a scientific method. With animation, one tries to detect the strong correlations between the entities in a model. For example, by observing the animation of an AS/RS system, the single play and double play of the SRM can be visualized and compared to the model design or the real system. Or the animation of an AGV system's behaviour at intersections can also be very helpful in identifying model discrepancies.

The model's complexity is partially revealed by the observation, although it is hard to quantify this complexity. The correlation between parameters can be shown by changing the input and observing the change in the animation to see whether the animation changes as expected. The model's uncertainty, both epistemic and aleatory, can be observed by animation only when it is combined with other tests, for example when a degeneracy test or extreme condition test are performed. Data availability is not important for the animation. Even if no data is available for validation – Type I data – the animation can still be made by creating assumption data as input. If Type III data is available, animation can be a powerful tool when combined with other testing instruments.

Animation should be used, however, with great caution, since it can reveal only obvious errors in a simulation model. As pointed out by Paul [Paul89], it cannot guarantee a model's correctness.

b. Comparison with other models

This method compares model output with the results of analytical methods or other already validated models. Most models are built for new systems or drastically modified ones. There may be no data about a new system. It is almost always impossible to get a comparable model. Furthermore, the analytical method can only be applied at a very high or very low level, which means that either the model is completely simplified as input-output transformation, that is, a black box test, or the model output is compared with analytical model output; or the model is decomposed to the lowest level, i.e. into queues. Otherwise, the analytical method is not helpful in the validation process. Validation on these two levels also provides little help.

A model's complexity is revealed only in the difference between models. The difference in the output from the models reflects the model's complexity if both models use the same input data and the common random number technique.

Uncertainty is difficult to expose by this method, since the difference in the output cannot be distinguished as the result of the modification of model structure or the changes in input.

If Type III data is available, then the models can be run with the same input data and the output from all models can be compared with the data from the system. If Type II data is available, the t-test can be used. The difference of the output data to the Type II data can then be used as a basis to judge the validity of the model.

c. Degeneracy test

This test selects appropriate input and model variables to evaluate model degeneracy. For example, with an increase in the arrival rate more customers are expected to be waiting in the queue. The degeneracy test should be applied by a system expert, since some degeneracy effects are hard to predict because of the high complexity a system can have. Furthermore, some system effects are cyclical, which means that responses increase and decrease at certain periods as the input changes. Input parameters also affect one another. The degeneracy effects would cancel one another as the input parameters change. Therefore, a more systematic approach, such as sensitivity analysis, is recommended.

With a degeneracy test one tries to understand a model's complexity through its uncertainty. Input factors are selected and changed accordingly in order to observe their effects on the output. Whatever type of data is available, the degeneracy test can always be performed, since it is intended only to show trends at the output side.

d. Event validity

This test compares for similarity the events in a model with those in a real system. In discrete-event simulation, an event is defined as a change of system state and occurs when activity is finished. The event list from a small model running for a long time has many events on it. Besides this, because of the stochastic nature of the model, any event would occur at a completely different moment than in real life, so it would be impossible to compare the two for similarity. One way of using this method is to make the comparison more specific. For example, the event might be the breakdown of a major piece of equipment which significantly reduces the performance of the system. Now the influence of the behaviour of the model and the system which has the equivalent event can be compared. It still cannot guarantee, however, that all other conditions are similar, both in the model and in the real system. That the events in the model and the system have occurred under exactly the same conditions is difficult to guarantee unless the simulation is trace-driven.

Event validity requires high data availability; more exactly, it requires Type III data – input and output data from the system. Even though a comparison cannot reveal a model's complexity and uncertainty, if there are discrepancies between model behaviour and system behaviour one should investigate why differences have occurred in the model and try to improve it to make its behaviour similar to the system. Type II data provides little help in event validity.

e. Special input test

The special input test checks model validity by giving input parameters various values, such as boundary values, extreme values, invalid values, or using trace data as input.

There are the following kinds of special input test:

- (1) Boundary values test: this divides the values of input parameters into equivalence classes, which are generated for each piece of specification and labelled as “valid” or “invalid”. The test data generated is then used as input to cause the model to produce values on the boundaries of output equivalence classes. The reason for doing this test is, as Ould and Unwin [OuUn86] declare, that most error-prone test cases lie along the boundaries.
- (2) Extreme value test: the model will be tested with extreme and unlikely input values. For example, what will happen in a production system if a large oven with hundreds of entities is broken and all the entities need to be taken out of the oven in thirty minutes? The results of such tests should be foreseeable or capable of being estimated. If the system behaviour even cannot be estimated, this test should not be used in validation. The extreme condition test is only one type of special input testing.

- (3) Invalid input testing: wrong input data is used as input in running a model to assess whether the model's behaviour is as expected.
- (4) Self-driven input testing: one changes the type of probabilistic models used for model input to test model validity.
- (5) Stress test: this tests the model under extreme conditions and increases the congestion in the model. Unexpected behaviours can reveal model discrepancies.

It is best if Type I data is available; if not, at least Type II should be available to make these tests effective, since the values on the input side should be reflected by those on the output side with a reasonable degree of confidence. This means that the model output from these condition tests should still be the "same" with the required degree of freedom.

Model uncertainty, or to be precise, model epistemic uncertainty is, or at least should be, exploited in these tests. When the input data is given values sometimes beyond the known possible values, the behaviour of the model should be as expected. And if the model's behaviour is abnormal, then the reason for this should be investigated and carefully documented.

Model complexity is also considered in these tests. An increase or reduction in a parameter value should be reflected by the parameter most impacted by it.

f. Face validity

The model and its behaviour are judged subjectively by comparing model and system output under similar environmental conditions by the people involved in the project, model users and domain experts. One potential problem with this test is that the people involved may have never seen the model before. They might focus only on the parts of the model that they are familiar with or that they understand. Also, the opinions of the experts in group discussions can have a strong influence on one another unless the test is made with complete anonymity. Furthermore, face validity can only provide a subjective estimation of model validity. It should be used only at the beginning of each model validation phase as an initial test. Animation can, perhaps, be combined with this test. If the model passes this test, it can only be said that there is no obvious logic error or counter-intuitive behaviour.

Face validity is based on human judgment and is completely subjective if conducted in isolation. Model complexity can only be observed from different aspects by different people regarding their knowledge domains. No specific types of data are required for this test. Model uncertainty is even more ambiguous owing to the involvement of human decision making.

g. Historical data validation

This test is used to run a model with trace data from a real system and then compare the behaviour of the model with the system. When data is abundant the model can be driven just by the trace; when only small sample data sets are available, then statistical distributions based on these sets should be used to drive the model.

This test requires Type III data – both input and output data from the system are needed. If there is only Type II data, the model can be run by input based on statistical distribution.

Model uncertainty and model complexity are not considered in this test.

h. Historical methods (rationalism, empiricism and positive economics)

‘Historical methods’ means merely that these methods are used before the multistage validation proposed by Naylor and Finger. Rationalism uses logical deductions to develop a valid model with the assumptions being regarded as common truths. Empiricism checks every assumption and ensures that these assumptions are valid. And positive economics emphasises the capacity of prediction of a model and ignores the model’s assumptions and internal structure. These methods were discussed in Chapter 2.

Historical methods are based largely on data availability and its corresponding importance to model validation. If the assumptions are granted as truths and do not need to be tested, then the current data about the system or model will be useless. If the assumptions are regarded as the core of the validation procedure, then data must be available in order to perform the test.

i. Internal validity

Internal validation is designed to check the variability of the model. It makes the model run for several replications with different random number seeds, then checks the consistency of the outputs from these replications. A lack of consistency will lead to further validation tests on the model.

Internal validity tests model uncertainty. Data availability and model complexity are not considered here.

j. Multistage validation (Naylor and Finger) [NaFi67]

Multistage validation is a combination of the three historical methods. The first stage is to develop model assumptions from existing theories, observations and available knowledge. The second stage is to validate model assumptions empirically where possible. The third stage is then to compare the input and output from the model with those from the real system. This method has been widely used and recommended in the classic book by Banks et al. However, a problem eventually arises if the data required for the third stage is not available, for example in the situation where the system is newly designed or undergoing severe modification. This will make the third stage, which is also the most important and most objective step, difficult to justify. So this method can be used, for example, as a tool for the further development of a legacy simulation model for an existing production line when the line needs to be upgraded or new products need to be produced by it.

Multistage validation requires Type III data. It is not designed to explain how a model works or the relationships between the entities in it, but to study the model as a black box or an input-output transformation. What happens in the model is not important in this test. Input data is required not only to run the model, but also to check its reasonableness and correctness. This test is the most widely used in model validation. However, as already

mentioned several times, it can rarely be conducted in real life because it is seldom that Type III data is available in logistics and production systems.

k. Operational graphics

Details of system performance, for example server utilization changes or WIP changes in a model constantly over time, are shown graphically when the model is running. The operational graphics can be designed as a dash board and used as a system cockpit for model validation supervision. However, if the model is large, perhaps containing hundreds of machines, and complex, with complicated control rules, the graphical display should focus on the most important elements in the model or the system KPIs. Otherwise, the overview of the model will be lost. This method is better used together with a sensitivity analysis or extreme conditions test. In combination with a sensitivity analysis, since the parameter values in the input are changed systematically, the output can also be estimated. The focus of the operational graphics can then be adjusted to the changes and show the desired performance of the elements. When used with the extreme conditions test, it can show the values of KPIs under these conditions.

Operational graphics has no requirement of data availability, though it works better when more realistic data is available.

l. Sensitivity analysis

This test is intended to uncover the most sensitive parameters that can significantly change a model's behaviour or output by systematically changing model input values over certain ranges. If unexpected effects appear in the model, then these parameters should be investigated. The input values can be either qualitative or quantitative. If the values are qualitative, the output can be shown as an increasing or decreasing trend. If the values are quantitative, the output can be shown in both direction and exact magnitude of changes. Unexpected behaviour in the model when the parameter values change can reveal discrepancies in the model. Also, input values can be deliberately changed to induce errors in order to determine the sensitivity of the model's behaviour to these errors.

m. Predictive validation

The test known as 'predictive validation' requires both input and output data from a real system. The input data is used to drive the model and then the model output is compared to the output of the system. If the model can "predict" the historical output data correctly, then the model passes the predictive validation. The prediction is ultimately retrospective rather than prospective. It requires the largest quantities of data from the system – input and its corresponding output data, which is not always available.

Predictive validation requires Type III data. Model uncertainty and model complexity are not explicitly considered in this test.

n. Top-down test

The top-down test decomposes a model into sub-models and then thoroughly tests each of the sub-models. In the next step, the sub-models are added incrementally together up to the highest level. In this way, any modelling errors can be localized at the sub-models. However, this test is very expensive, especially if the model is large and contains many

sub-models. Furthermore, it requires high data availability, since each sub-model requires input data. But, owing to the complexity of the model, data and control strategies can sometimes be hard to obtain.

o. Traces

Traces are used to record the behaviour of pre-specified entities through model replications in order to check the model's logic and accuracy. This method is widely discussed and many people have tried to use it, since almost all simulation packages provide this function for tracing entities when a model is running. The problem is, which entities should be selected, how many should be selected, and how can one prove whether they represent the behaviour of the model? If too many entities are traced, the output will produce trace data that is too large and complex to analyse, while, if too few entities are traced, the results will not represent the true behaviour of the model.

Simulation trace data can sometimes be very helpful. There have been many studies that tried to use trace data to check model validity.

The solution for handling large trace data can be to use a relational database, such as MS Access, SQL Server, etc., to store the trace data and then use SQL language to analyse it. In this way, the model's complexity can be investigated.

As an example, the daily operation of a factory needs to be monitored. With the help of Manufacturing Execution Systems and the Advanced Planning and Scheduling System, data on the release of orders to the shop floor, WIP size, machine utilization, production cycle time, takt time, etc. can be collected. Then a histogram of these performance indicators can be produced. The trace data of the simulation model can then be compared with this data to check model validity.

p. Turing test [Tur48]

The Turing test is also based on subjective decisions by experts who know the system under study. The output from both model and system under the same input conditions is obtained and shown to the experts. If the experts cannot differentiate them, the model passes the Turing test and the confidence in the model is increased. If the experts can successfully identify the model results, the evaluation becomes very valuable in the improvement of the model.

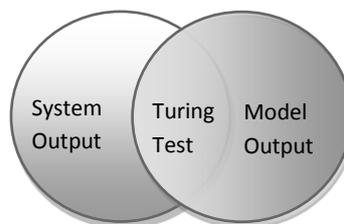


Figure 4-1: Turing test

The Turing test treats a simulation model as an input-output transformation. Only the plausibility of model output is accessed. Data availability is sometimes crucial in a Turing test, since the comparison is made between the model output and the system output. So at

least Type II data must be available, if not Type I. Model uncertainty and model complexity can scarcely be checked with this test.

4.1.2 Current validation techniques by Balci [Balc94] [Bank98]

Balci has also reviewed validation techniques in detail and divided them into four categories: informal, formal, static and dynamic.

Most validation techniques are informal. The decision-making approaches are normally dependent on subjective reasoning and a lack of exact mathematical calculation.

On the other hand, formal validation techniques are based on stringent mathematical reasoning and calculation. These formal validation techniques are, however, incapable of being applied even to a reasonably complex simulation model, the reason being that formal validation techniques require large quantities of system data and the amount of data required in real simulation projects is hardly available. As Law [Law00] says, “the problem with applying this idea in practice is that one is usually fortunate to have even one good set of data from the existing system. As a matter of fact, I have only seen one simulation project (either my own or published by someone else) where there was more than one set of system data available and where this idea was actually applied. Thus, in practice the comparison of the model and system performance measures is typically done in an informal manner.” Some techniques that do not need lots of data also have drawbacks, such as being hard to apply. For example, predicate calculus would be too complex to apply to real-life models in production and logistics. Also, if it is possible to use mathematical equations to solve problems in production and logistics, there will be no need for a simulation in the first place.

Static validation techniques are intended to evaluate the structure of a model, or the entity flow or logic control within it. They require no model execution.

Dynamic validation techniques require the execution of a simulation model. They evaluate the behaviour of the model execution over time. This is one of the most important properties of discrete simulation. Balci declares that most dynamic validation requires the insertion of additional code into the model for the collection of information with the purpose of understanding model behaviour, which is then called model instrumentation. The application of dynamic validation follows three steps: first, the extra code is inserted into the model. The location is decided by the structure of the model. Then the model is executed. Finally, the model output is analysed and the model behaviour evaluated.

validation techniques	Type	Technique description
Acceptance test	Dynamic	Operational test of the model with actual hardware and data to determine whether specified requirements are met.
Alpha test	Dynamic	Operational testing of initial, complete version of the model at an in-house site uninvolved in the model development.
Audit	Subjective	Assessment of the application of M&S with respect to established policies, standards, guidelines.
Beta test	Dynamic	Developer's operational testing of first release version of complete model at a beta user site.
Boundary value test	Dynamic	Examination of accuracy using test cases on the boundaries of input data classes.
Cause-effect graph	Static	Identification of causes/effects of modelled system, creation of decision table and conversion into test cases with which the model is tested.
Comparison test	Dynamic	Comparison of results from models of the same system.
Debugging	Dynamic	Iterative process to uncover and correct errors or misconceptions.
Equivalence partitioning test	Dynamic	Test of the accuracy of a model with a representative value from each input data class.
Execution monitoring	Dynamic	Gathering and examination of activity- and event-oriented (low-level) information resulting from model execution.
Execution profiling	Dynamic	Examination of high-level information (profiles) on activities and events during model execution.
Execution trace	Dynamic	Exposure of errors by review of line-by-line execution of a simulation.
Extreme input test	Dynamic	Assessment of model at minimum or maximum values for the model inputs.
Face validation	Subjective	Subjective comparison of model and system behaviours; preliminary approach to validation in early stages of development.
Fault/failure insertion test	Dynamic	Insertion of a fault or failure to observe whether the model produces the expected invalid behaviour.
Field test	Dynamic	Placement of the model in an operational situation to collect information for validation.
Functional test	Dynamic	Assessment of the accuracy of model input-output transformation.
Graphical comparison	Dynamic	Comparison of graphs of model variables over time with system variables.
Inspection	Subjective	Formalized five-step process, checklist approach to uncover errors.
Invalid input test	Dynamic	Examination of model performance using incorrect input data.
Model interface analysis	Static	Assessment of whether interface structure and behavior are accurate.
Partition test	Dynamic	Analysis of model's functional partitions by comparing partitions of the model specification and implementation and testing model with test data.
Predictive validation	Dynamic	Application of past input data and subsequent comparison of model outputs with past output data.
Product test	Dynamic	Preparation for acceptance testing.
Real-time input test	Dynamic	For simulations representing embedded real-time systems, assessment of model accuracy using real-time input data.
Review	Subjective	Evaluation relative to specifications and standards by management level team.
Self-driven input test	Dynamic	Assessment of model using input data sampled from probabilistic models representing random input conditions for a real system.
Sensitivity analysis	Dynamic	Identification of variables/parameters to which model behaviour is very sensitive.
Statistical techniques	Dynamic	Model and system outputs are compared using multivariate statistical techniques to capture correlation.
Stress test	Dynamic	Assessment of model validity under extreme workload conditions.
Sub-model/module test	Dynamic	Decomposition of model into sub-models. Comparison of the behavior of each sub-model to the behavior of the corresponding sub-system of the actual system.
Traceability assessment	Static	Used to match elements of one form of the model to another, such as requirements specification elements to model design specification.
Trace-driven input test	Dynamic	Refinement of raw trace data collected from a real system for testing of a model.
Turing test	Subjective	Examination by experts of output data, one set from the model and one from the system, for feedback in correcting model representation.
User interface analysis	Static	Examination of user-model interface to prevent errors during user's interaction with model.
Visualization/animation	Dynamic	Display of graphical images of model's dynamic behavior during execution.
Walkthroughs	Subjective	Meeting to detect and document faults; less formal than inspections.

Table 4-1: Model validation techniques [Bank98] [BaNa85]

These validation techniques focus on various aspects of a model regarding its validity. As discussed for each technique, from the point of view of the three essential characteristics of a

DES model, most of these techniques touch at least one of them. However, whichever validation techniques are used and however many of them need to be applied, the decision as to whether a model is valid or not will still be made by people. So the question is, how many criteria are required to verify a model's validity? In the techniques belonging to black box testing, the criteria focus mainly on system KPIs and normally require Type III data, such as from trace-driven validation or the Turing test, whether it is subjective or objective. As long as the KPIs of the model conform to those of the real system, the model is regarded as valid. These KPIs are chosen according to the purposes of the simulation study. Normally, in a logistics or production system, these KPIs will be, for example, the throughput per unit time, the production takt time, buffer occupancies, machine utilizations, throughput time, or cycle time. For the techniques that aim to understand the internal relationships of a model for the purpose of judging validity, there are distinct criteria according to the objectives of the technique. These techniques, such as cause-effect graphs or event validation, require various different types of data, but most of them are used to understand the complexity of a model.

4.1.3 Current statistical techniques in simulation model validation

The best way to validate a simulation model is to compare its output with the system output. This requires the use of statistical techniques and the collection of large amounts of data from both the model and the system.

Before applying statistical validation techniques, a null hypothesis and an alternative hypothesis can be made:

Null hypothesis H_0 : model is valid under the testing environment for the required range of accuracy

Alternative hypothesis H_1 : model is invalid under the testing environment for the required range of accuracy

Usually, from the model validation point of view, the focus is more on determining whether the difference between the expected value of a model output μ^m and the system output μ^s is within an acceptable range and to a required level of confidence. The intention is to decide whether $D = \mu^m - \mu^s$ falls into the range (L, U), in which L is the lower boundary and U is the upper boundary. For the model to be valid, the interval (L, U) must include zero. Then the null hypothesis becomes

$$H_0: L \leq D \leq U$$

and the alternative hypothesis becomes

$$H_1: D < L \text{ or } D > U.$$

Regardless of which kinds of validation technique are applied to a model, the results fall into one of the following four categories:

1. H_0 is not rejected when H_0 is true
2. H_0 is rejected when H_1 is true
3. H_0 fails to be rejected when H_1 is true – the so-called Type II error
4. H_0 is rejected when H_0 is true – the so-called Type I error

Attention must be paid to the fact that here the model itself is not the object of validation. Since a model is only an imitation of an original system, saying a model is equivalent to a system is in any case a wrong proposition.

If multiple outputs need to be compared between the model and the system output variables, then the confidence intervals of each pair of output variables from model and system make up a so-called model range of accuracy.

Assuming that there are k output variables from both model and system, let $(\mu^m)' = [\mu_1^m, \mu_2^m, \dots, \mu_k^m]$ and $(\mu^s)' = [\mu_1^s, \mu_2^s, \dots, \mu_k^s]$ be the k dimensional vectors of population means of the model and system output variables, respectively [BaSa84].

If a specific range of accuracy is given, then the null hypothesis becomes:

$$H_0: L \leq \mu^m - \mu^s \leq U$$

where $L = [L_1, L_2, \dots, L_k]$ and $U = [U_1, U_2, \dots, U_k]$ are the lower and upper boundaries of the acceptable difference.

However, when several outputs are compared simultaneously, then the multiple-comparisons problem must be considered. Assuming that U_s is a $100(1-\alpha_s)\%$ confidence interval for the measure of performance μ_s , then the probability that all k confidence intervals simultaneously contain their respective true measures is:

$$P(\mu_s \in U_s, \text{ for all } s = 1, 2, \dots, k) \geq 1 - \sum_{s=1}^k \alpha_s$$

whether the U_s s are independent or not. This is known as the Bonferroni inequality. This is a very useful inequality that has a large impact on simulation analysis. For example, for an M/M/1 queue, confidence intervals are constructed for the time in the system, server utilization, time in the queue, with $\alpha_s = 0.1$ for each of them. So $s = 3$. Then the probability that each of the three confidence intervals contains its true measure is at least 70%. This will seriously undermine the conclusion. The solution obviously increases the overall confidence interval for every output variable, so that $\sum_{s=1}^k \alpha_s = \alpha$. However, the difficulty here is that the confidence intervals will be larger than they originally were if a fixed-sample-size procedure is used. So it is recommended that the number of variables should be no larger than ten. Hence, when ten output variables are selected, ten 99% confidence intervals are constructed and the overall confidence level is still at least 90%, which means that if simultaneous multiple-interval estimates are desired with an overall confidence coefficient of $1 - \alpha$, one can construct each interval with the confidence coefficient $(1 - \alpha/g)$, and the Bonferroni inequality guarantees that the overall confidence coefficient will be at least $1 - \alpha$. Moreover, the Bonferroni inequality [Bonf36] is also very general and does not depend on

how the individual confidence intervals are formed; it does not require results from the same number of replications of models, nor are they independent.

A better statistical validation method than the traditional hypothesis, which is to test whether $\mu_x = \mu_y$ as H_0 , is to compare the model with the system by constructing a confidence interval for $\delta = \mu_x - \mu_y$, because a confidence interval provides more information than the corresponding hypotheses tests. A confidence interval shows not only whether the means are equal or not, but also gives indications about the magnitude of the difference.

If Type II data is available, a self-driven simulation can be conducted. The model output data is then independent of the system output data, although they are from the same populations. The two data sets are independently and identically distributed. In this case, the model's range of accuracy is determined by the $100(1 - \alpha)\%$ simultaneous confidence interval for $\mu^m - \mu^s$, as

$$[\delta - \tau]$$

where $\delta' = [\delta_1, \delta_2, \dots, \delta_k]$ represent lower boundaries and $\tau' = [\tau_1, \tau_2, \dots, \tau_k]$ represent the upper boundaries of the simultaneous confidence interval. The confidence interval is $100(1 - \alpha)\%$, which means there can be $100(1 - \alpha)\%$ confidence that the true difference between the population means of the model and system output variables are simultaneously contained within $[\delta - \tau]$.

Another method is to first construct a $100(1 - \alpha^m)\%$ simultaneous confidence interval for when μ^m is

$$[\sigma^m, \tau^m]$$

where $(\sigma^m)' = [\sigma_1^m, \sigma_2^m, \dots, \sigma_k^m]$ and $(\tau^m)' = [\tau_1^m, \tau_2^m, \dots, \tau_k^m]$. Then the $100(1 - \alpha)\%$ simultaneous confidence interval for μ^s is

$$[\delta^m, \tau^m]$$

where $\delta' = [\delta_1, \delta_2, \dots, \delta_k]$ and $(\tau^s)' = [\tau_1^s, \tau_2^s, \dots, \tau_k^s]$. Now, using the Bonferroni inequality, the model range of accuracy can be determined by the simultaneous confidence interval $[\delta^m - \tau^s, \tau^m - \delta]$ for $\mu^m - \mu^s$ with a confidence level of at least $(1 - \alpha^m - \alpha^s)$ when the output data is independent, when the model and system output data sets are dependent and with a level of at least $(1 - \alpha^m - \alpha^s + \alpha^m \alpha^s)$ [Klei75].

The second approach can also be used when Type I data is available. The trace-driven simulation – running the simulation with the same input data and operational conditions as for the real system – can be performed. The output of the model is then dependent on and identical to the system output. The outputs from model and system are in pairs and the model's range of accuracy is determined by a $100(1 - \alpha)\%$ simultaneous confidence interval for μ^d , the population means for the difference of paired observations, as

$$[\sigma^d, \tau^d]$$

where $(\sigma^d)' = [\sigma_1^d, \sigma_2^d, \dots, \sigma_k^d]$ and $(\tau^d)' = [\tau_1^d, \tau_2^d, \dots, \tau_k^d]$.

The statistical methods available are shown below:

a. Analysis of variance [Klei75]

Analysis of variance (ANOVA) is used to test the hypothesis that the means of model output and system output are equal with the condition that the two data sets are normally distributed. If only one factor needs to be analysed, a one-way ANOVA is performed. For example, in an M/M/1 only the mean time customers spend in a queue is studied. If two factors need to be analysed, then a multiple ANOVA should be performed.

Since the assumption is of homogeneity of variance, in order to apply ANOVA the data sets should be tested. For example, the Bartlett's test can be used.

However, when comparing two groups like the output data from the model and the output data from the system by ANOVA, the method is equal to the variance independent t test.

b. Confidence interval/region [Klei75]

Confidence intervals are used widely in model validation when real data is available. If Type I data is available, then let $X_{s1}, X_{s2}, \dots, X_{sn}$ be a sample of IID observations from the system and $X_{m1}, X_{m2}, \dots, X_{mn}$ be a sample of IID observations from the model. If more data can be obtained from the simulation, the extra data can be discarded in order to make the two data sets the same size. Then one can define $Z_i = X_{si} - X_{mi}$, for $i = 1, 2, \dots, n$. Now the Z_i 's are independently and identically distributed random variables. The mean of Z_i is set to δ . Let

$$\bar{Z}(n) = \frac{\sum_{i=1}^n Z_i}{n}$$

and

$$\widehat{Var}[\bar{Z}(n)] = \frac{\sum_{i=1}^n [Z_i - \bar{Z}(n)]^2}{n(n-1)}.$$

If Z_i s are normally distributed, the $100(1-\alpha)\%$ confidence interval is:

$$\bar{Z}(n) \pm t_{n-1, 1-\alpha} \sqrt{\widehat{Var}[\bar{Z}(n)]}.$$

Otherwise, one should rely on the central limited theorem, the result of which is that the coverage probability is near $1 - \alpha$ for large n .

For the confidence interval, the X_{sn} and X_{mn} are positively correlated. However, this can be a desired property, because it reduces the variance of Z_i and leads to a smaller confidence interval. The reason for this is

$$\widehat{Var}[\bar{Z}(n)] = \widehat{Var}[X_{sn} - X_{mn}] = \widehat{Var}[X_{sn}] + \widehat{Var}[X_{mn}] - 2\widehat{Cor}[X_{sn}, X_{mn}].$$

If X_{sn} and X_{mn} are positively correlated, then

$$\widehat{Cor}[X_{sn}, X_{mn}] > 0.$$

If X_{sn} and X_{mn} are independent, then

$$\widehat{Cor}[X_{sn}, X_{mn}] = 0.$$

Thus, $\widehat{Var}[\bar{Z}(n)]$ is smaller when X_{sn} and X_{mn} are positively correlated. And when $\widehat{Var}[\bar{Z}(n)]$ is smaller the $100(1 - \alpha)\%$ confidence interval will also be smaller.

However, sometimes only Type II data is available. Since Type II data is much less than the data generated by a simulation model, which means that this data does not pair up, the paired-t confidence interval used above can be used only with some modification. In the paired-t method above, it is assumed that the variances of system data and model data are equal: $\text{Var}(X_{sn}) = \text{Var}(X_{mn})$. However, in this case, using the paired-t method will show serious coverage degradation in the confidence interval. Welch proposes a method for comparing two data sets with unequal and unknown variances, where $s \neq m$ and $\text{Var}(X_{sn}) \neq \text{Var}(X_{mn})$. This is also called the Behrens-Fisher problem. However, according to Scheffé [Sche59], the two data sets should be normally distributed. This can be problematic if the data needed to be collected from a system is of rare events, such as the MTTR and MTBF of a relatively reliable machine. To get sufficiently large data sets of MTTR and MTBF a lot of time is needed, and in this case many statistical distributions used to describe MTBF will be skewed. To ensure a more normal distribution, a larger number of data points will be collected according to how skewed the underlying distribution of the data set is. Another reason for $\text{Var}(X_{sn}) \neq \text{Var}(X_{mn})$ is that, according to Kelton and Law, the simulation output data sets are almost always correlated, so the variance $\text{Var}(X_{mn})$ is no longer an unbiased estimator. According to Andersen,

$$E[\text{Var}(X_{mn})] = \sigma^2 \left[1 - 2 \frac{\sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \rho_i}{n-1} \right].$$

Since the output data from the model is positively correlated, $\rho_i > 0$. So $\text{Var}(X_{mn})$ has a negative bias and $E[\text{Var}(X_{mn})] < \sigma^2$.

If indeed the Type II data can be approximately normally distributed (here, since the amount of data from the simulation model is normally relative large, the major concern is on the data from the system), then the following method can be used:

Let

$$\bar{X}_s(n_i) = \frac{\sum_{j=1}^{n_i} X_{sj}}{n_i} \text{ and } \bar{X}_m(n_i) = \frac{\sum_{j=1}^{n_i} X_{mj}}{n_i}$$

and

$$S_s^2(n_i) = \frac{\sum_{j=1}^{n_i} [X_{sj} - \bar{X}_s(n_i)]^2}{n_i - 1} \text{ and } S_m^2(n_i) = \frac{\sum_{j=1}^{n_i} [X_{mj} - \bar{X}_m(n_i)]^2}{n_i - 1}$$

for $i = 1, 2$, and $n_1 \neq n_2$. The estimated degree of freedom is

$$\hat{f} = \frac{[\frac{S_s^2(n_1)}{n_1} + \frac{S_m^2(n_2)}{n_2}]}{[\frac{S_s^2(n_1)}{n_1}]^2 / (n_1 - 1) + [\frac{S_m^2(n_2)}{n_2}]^2 / (n_2 - 1)}$$

and the approximate $100(1-\alpha)\%$ confidence interval is

$$\bar{X}_s(n_1) - \bar{X}_m(n_2) \pm t_{\hat{f}, 1-\alpha} \sqrt{\frac{S_s^2(n_1)}{n_1} + \frac{S_m^2(n_2)}{n_2}}$$

This is also called the Welch confidence interval. Because \hat{f} is in general not an integer, its value needs to pass through the interpolation of a printed t table or one needs to use statistical software.

c. The t-test [Klei75]

The t-test is used when the means of validation-relevant output variables need to be compared with those of system output data or a standard value. When the mean value of the model is compared with a standard value, this is called the one-sample t-test. If there are two data sets, one from the system and one from the model, then it is called the two-sample t-test. The purpose of a t-test is to assess the likelihood that the means of system and model are the same.

The t-test is a type of hypothesis test; it has the form:

$$t_n = [\widehat{\mu}^m - \widehat{\mu}^s] / \sqrt{\widehat{\sigma}/n}$$

and t_n has t distribution with $n-1$ degree of freedom. The confidence interval of $100(1-\alpha)$ is:

$$\widehat{\mu}^m \pm t_{n-1, 1-\alpha/2} \sqrt{\widehat{\sigma}/n}$$

This is actually the same as the (2) confidence intervals. In the t-test, this equation can be used to calculate L and U, the lower and upper boundaries. The t-test is designed to decide whether $D = \mu^m - \mu^s$ falls into the range (L, U), in which L is the lower boundary and U is the upper boundary. In order for the model to be valid, the interval (L, U) must include zero. Then the null hypothesis is:

$$H_0: L \leq D \leq U$$

and the alternative hypothesis is:

$$H_0: D < L \text{ or } D > U$$

d. Hotelling's T² Test [BaSa84]

If several output variables need to be compared and at least Type II data is available from an observable system, then a generalized Student's t-test, known as Hotelling's T² test, can be applied, given that three assumptions are met. It can be used in the testing of multiple output variables, in particular when these variables are correlated. For example, regarding throughput time and machine utilization, since these are correlated, a univariate t-test will neglect the covariance among output variables and increase the possibility of Type I error.

The three assumptions that need to be satisfied for Hotelling's T² test are those of independence, multivariate normality and equality of variance-covariance matrices. There are two independence assumptions: the first is that the model and system output data matrices are independent, while the second assumes that the vectors in the data matrix from the model and the system are independent. The first assumption can be met by using self-driven simulation. The second can be met by using a batch mean method for a steady-state simulation and replication techniques to terminate simulation. Multivariate normality requires that the model output variables and system output variables have a multivariate normal distribution. The third assumption is that the variance-covariance matrix of all response variables from the model and the system are the same. If this assumption is not satisfied, the sample sizes of both systems must be made equal and fairly large. This means that more time will need to be spent for Type II data collection and more replications of the model.

Balci and Sargent report a two-sample Hotelling's T^2 test. The test method is straightforward. However, in order to meet the three assumptions above, many related techniques for checking these assumptions are called for. The following is a short description of this test, provided that the three assumptions are met:

- (1) Set the range of accuracy

$$|\mu^m - \mu^s| \leq \delta_j, j = 1, 2, \dots, k,$$

where $(\mu^m)' = [\mu_1^m, \mu_2^m, \dots, \mu_k^m]$ and $(\mu^s)' = [\mu_1^s, \mu_2^s, \dots, \mu_k^s]$ are the k dimensional vectors of the population means of the model and system output variables, and δ is the vector of the largest acceptable differences.

- (2) If the sizes of both data sets are approximately equal $n \cong N$, then we calculate the T^2 statistics by

$$T^2 = \frac{nN}{n-N} (\bar{x} - \bar{y})' S^{-1} (\bar{x} - \bar{y}),$$

where (\bar{x}) and (\bar{y}) and S are the estimates of μ^m and μ^s and the common variance-covariance matrix. When $\mu^m = \mu^s$, the T^2 statistic has the central T^2 distribution. And, further, if

$$T^2 \leq \frac{k(n+N-2)}{n+N-k-1} F_{\alpha; k, n+N-k-1},$$

then we can accept the model validity on level of confidence α , with the degree of freedom k and $n + N - k - 1$. Otherwise, we reject the validity of the model.

- (3) If the model is found invalid, then one should determine which variables have caused the invalidity by testing the equality of each output variable separately on the level of accuracy.

e. Multivariate analysis of variance [Bank98]

1). Standard MANOVA

The purpose of a t-test is to evaluate the likelihood that the means of two output variables of a system and a model are the same. The purpose of an ANOVA is to test whether the means of two or more variables of system and model are the same. The multivariate equivalent of the t-test is Hotelling's T^2 . This tests whether the two vectors of means of variables of system and model are the same. MANOVA is the multivariate analogue to Hotelling's T^2 . The purpose of MANOVA is to test whether the vectors of means of output variables for system and model are the same. Just as Hotelling's T^2 will provide a measure of the likelihood of picking two random vectors of means for one

time, MANOVA gives a measure of the overall likelihood of picking two or more random vectors of means of output variables for one time.

2). Permutation methods

Permutation tests are a type of statistical significance test based on permutation resamples drawn at random from system and model output data. Permutation resamples are drawn without replacement. At least Type II data is required. As an example, $X_{m1}, X_{m2}, \dots, X_{mi}$ are model output data with c.d.f. F_m and $X_{s1}, X_{s2}, \dots, X_{si}$ are the system output data with c.d.f. F_s . The means, then, are μ_m and μ_s respectively, and the intention is to test whether there are significant differences between them at the 5% significant level. Let n_s and n_m be the numbers of data points in two data sets. The permutation tests can determine whether the difference is large enough to reject the null hypothesis H_0 .

H_0 : $F = G$, the two data sets have identical probability distributions.

H_1 : $F \neq G$, the two data sets have different probability distributions.

In the first step, the difference between the means of model output and system output $d = \mu_m - \mu_s$ is calculated. Then, the two data sets are pooled together. In the third step, two data sets with the same sizes as n_s and n_m are randomly resampled N times (by Monte Carlo sampling, and N should be sufficient, around one thousand) from the pooled data set, which is called the permutation resample, and assigned to two groups, then the difference of means is calculated for each pair. The one-sided p-value of the test is calculated as the proportion of sampled permutations where the difference in means was greater than or equal to d . The two-sided p-value of the test is calculated as the proportion of sampled permutations where the absolute difference was greater than or equal to the absolute value of d . Small p-values are evidence against the null hypothesis. The P value can be calculated as

$$p = \frac{1 + \sum_{i=1}^N I(t_i \geq d)}{N+1}.$$

Compared to the t-test, the permutation test is more general. The null hypothesis is the “have identical probability distribution”, which implies the null hypothesis of the t-test “no difference between the means”. The t-test requires normality of data sets from system and model outputs and relies on a central limit theorem if the data sets are large. It is robust, especially when the two data sets are similar.

The permutation test completely removes the normality condition. The resampling moves data points between the two data sets, which requires that the two populations are identical when the null hypothesis is true — not only are their means the same, but also their spreads and shapes [MMDS03].

3). Non-parametric ranking methods

Non-parametric ranking methods can be used for model validation when the system is observable, Type I or Type II data is available, and the data set is small, independent, strongly skewed and deviates from normality. One assumption for these methods is that all the output variables are continuously distributed. The continuity assumption is a technical one that eliminates the possibility of ties and simplifies the definition of median.

The methods test hypotheses as:

H_0 : the output data from system and model has the same distribution.

H_1 : the output data from the model is systematically lower or higher.

If the assumption is met and Type II data is available, then the Wilcoxon rank sum test can be performed. Let $X_{s1}, X_{s2}, \dots, X_{sm}$ be a sample of IID observations from the system and $X_{m1}, X_{m2}, \dots, X_{mn}$ be a sample of IID observations from the model, where $m \neq n$. Let $N = m + n$, then rank all N data points from smallest to largest $W_{(1)}, W_{(2)}, \dots, W_{(N)}$. When the data points tie, then the average of their ranks is assigned to each of them. The sum of the ranks for the model output is the Wilcoxon rank sum statistic. If the model is valid, then W has the mean

$$\mu_W = \frac{n(N + 1)}{2}$$

and standard deviation

$$\sigma_W = \sqrt{\frac{nm(N + 1)}{12}}$$

The Wilcoxon rank sum test rejects the null hypothesis when the Wilcoxon rank sum statistic W strongly deviates from the mean. If the model is valid, then all data points from both system and model are randomly assigned labels, and values in the two different data sets should be somewhat equally distributed between the two.

P-values for the Wilcoxon test are based on the sampling distribution of the rank sum statistic W when the null hypothesis (no difference in distributions) is true. The rank sum W is either too big or too small. The P-value for the test is the probability of falling into the tail of the distribution closest to W and doubling it (two-sided). The P-values can be calculated from the Wilcoxon

rank sum table, from statistical software, or through a normal approximation (with continuity correction) if the data sets are large enough.

When the Type II data sets are small and independent, and show strong skewness, the Wilcoxon test is more reliable than the two-sample t-test. Permutation tests have the advantage of flexibility. They allow a wide choice of the statistic used to compare model and system output variables, which is their advantage when compared to t- and Wilcoxon tests. Both rank and permutation tests are non-parametric. They require no assumptions about the shape of the distribution.

If the assumption is met and Type I data is available, then the Wilcoxon signed rank test can be performed. Let $X_{s1}, X_{s2}, \dots, X_{sn}$ be a sample of IID observations from the system and $X_{m1}, X_{m2}, \dots, X_{mn}$ be a sample of IID observations from the model. Rank the absolute values of the difference of each pair $(X_{si} - X_{mi})$, where $i = (1, n)$. Wilcoxon signed rank statistic W^+ is the sum of ranks for the positive difference. If the pairs are tied, then they are assigned the average ranks. The difference between the pair is zero. The usual procedure simply drops such pairs from the rank. This will eventually favour the null hypothesis (no difference). If there are many ties, the test may be biased in favour of the alternative hypothesis.

If the model is valid, then W^+ has the mean

$$\mu_{W^+} = \frac{n(N + 1)}{4}$$

and standard deviation

$$\sigma_{W^+} = \sqrt{\frac{n(n + 1)(2n + 1)}{24}}$$

The Wilcoxon signed rank test rejects the null hypothesis when the rank sum W^+ is far from its mean. The P-values can be calculated using statistical software, or through a normal approximation (with continuity correction) if the data sets are large enough.

f. Non-parametric goodness-of-fit tests [Bank98]

1). The Kolmogorov-Smirnov test

Let $X_{s1}, X_{s2}, \dots, X_{sm}$ be a sample of IID observations from the system and have the empirical cdf $S_n(x)$. Let $X_{m1}, X_{m2}, \dots, X_{mn}$ be a sample of IID observations from the model and have the empirical cdf $M_n(x)$. The test statistic D is the maximum deviation between the empirical cumulative distribution functions of the system and model data.

First, rank the m data points from the system and n data points from the model and assign each a subscript corresponding to its rank. Then calculate d_i

$$d_i = \frac{s_i}{m} - \frac{m_i}{n}$$

and

s_i is the number of system data points less than the i th order statistic

m_i is the number of model data points less than the i th order statistic.

The test statistic D is the maximum $|d_i|$, $i = 1, 2, \dots, m+n$. Under the assumption that the data from system and model are from the same distribution, the distribution of D is known and can be calculated.

When D is plotted, it is the largest vertical distance between the two empirical distributions.

If Type I data is available, then the two data set sizes are equal. If Type II data is available, then the two data set sizes are unequal. The critical values for both situations can be found from tables of the Smirnov statistic for equal size and the Smirnov statistic for different sizes.

If D is larger than the critical value, then the null hypothesis is rejected.

2). Cramér-von-Mises test

An alternative to the Kolmogorov-Smirnov test is the Cramér-von-Mises test. This is also one of the distribution-free two-sample tests. Let $X_{s1}, X_{s2}, \dots, X_{sm}$ be a sample of IID observations from the system and $X_{m1}, X_{m2}, \dots, X_{mn}$ be a sample of IID observations from the model. Then let S_m and M_n be their empirical distribution functions, respectively. The test statistic T_2 is defined as:

$$T_2 = \frac{mn}{(m+n)^2} \left\{ \sum_{i=1}^m [S_m(X_{Si}) - M_n(X_{Si})]^2 + [S_m(X_{mi}) - M_n(X_{mi})]^2 \right\}$$

The test rejects the null hypothesis H_0 if the value statistic T_2 is larger than the tabulated values [Ande62].

3). Chi-square test

To perform the Chi-square test, the data needs to be grouped. Type I or Type II data should be available. The grouping of both system and model data should be the same. The Chi-square test can check whether the number of model data points in each group is similar to that of the system data, which would prove

that the two data sets are from the same distribution and hence confirm the validity of the model.

Firstly, the data is divided into n groups or ‘bins’. The test statistic is defined as follows:

$$\chi^2 = \sum_{i=1}^n \frac{(N_1 M_i - N_2 S_i)^2}{M_i + S_i}$$

where the summation is for group 1 to n , S_i is the frequency for group i of the system Type I or Type II data and M_i is the frequency for group i of the model data. N_1 and N_2 are scaling constants used to adjust for unequal sizes of data sets from system and model.

$$N_1 = \sqrt{\frac{\sum_{i=1}^n S_i}{\sum_{i=1}^n M_i}}$$

$$N_2 = \sqrt{\frac{\sum_{i=1}^n M_i}{\sum_{i=1}^n S_i}}$$

Care should be taken when using the Chi-square test, since it is sensitive to the number of groups n . The results would be different with a different choice of n . The reason for choosing the number-of-group value n is that the group should be equal to $0.3 \cdot S$, where S is the data set standard deviation. The upper and lower limits are equal to the data set mean $\hat{\mu}$ plus or minus $6 \cdot S$. The groups with zero frequency are abandoned. Besides this, in each group there should be at least five data points.

The significant value is generally chosen as $\alpha = 0.05$ or 0.10 . The degree of freedom is $(n-c)$, where n is the number of groups. If two data set sizes are equal, which means that Type I data is available, $c = 1$; if not, $c = 0$.

The null hypothesis is rejected if the test statistic is larger than $\chi^2_{(1-\alpha, n-c)}$, which means that the differences in model and system are not due to random variation or chance alone.

g. Regression analysis

Regression analysis focuses on the linear regression of scatter plots of system and model output data. Model output data is normally plotted along the x axis, the system output data along the y axis. If the model is valid, then the data points of two data sets should fall on one line with a slope of one and an interception of zero. However, this approach has been criticised by Kleijnen in recent papers. The magnitude of the

standard error of the slope depends on the number of data points that are not on the regression line. The more scattered the data points, the more likely it is that it will not be rejected. Another argument is that since there is almost always model uncertainty, there will certainly be errors in the data. Hence, the model output data should not be plotted along the x axis in a linear regression.

h. Theil's inequality coefficient (TIC) [Thei61]

Theil's inequality coefficient is defined as

$$U = \frac{\sqrt{\frac{1}{n} \sum (\mu_m - \mu_s)^2}}{\sqrt{\frac{1}{n} \sum \mu_m^2 + \frac{1}{n} \sum \mu_s^2}}$$

The inequality coefficient U is between 0 and 1. Perfect prediction is achieved if it equals 0; 1 is the worst prediction. The drawback here, however, is that the validity of the model is fuzzy with the U value. For example, if U is smaller than 0.4, then the question arises as to when the model is valid. Rowland and Holmes indicate that "...one may arbitrarily identify TIC values above 0.7 as corresponding to rather poor models, TIC values between 0.4 and 0.7 for average-to-good models, and TIC values below 0.4 as very good or excellent models." But what is an "average-to-good model"? What is the difference between an "average-to-good model" and a "good model" or an "excellent model"? Thus, it is better to use Theil's inequality coefficient only as one indicator of model validation rather than as the decisive one.

4.2 The sequential bifurcation method for simulation model validation in production and logistics

Logistics and production systems are highly complex engineered systems. Simulation models are required to estimate system performance when other analytical methods reach their limits. The reasons why it is difficult to validate simulation models of production and logistics systems (below referred to as 'models') can be summarised into three aspects: model complexity, model uncertainty and data. The method is to answer the following questions, which is vital if we wish to achieve a deeper understanding of model validation: What is input uncertainty and how can we measure it? What is model complexity and why has it not been possible up to now to measure this in reality? How does the complexity of a model amplify the uncertainty and variability in the model's output? How does input randomness alter bottlenecks and the behaviour of a model? How do input uncertainty and model complexity influence model validation? How do data quality and availability affect model complexity, input uncertainty, and even the model validation process?

Nowadays, logistics systems are interconnected with one another. Networks are getting bigger and becoming more and more complex. In order to increase system utilization, reduce lead time, raise delivery reliability, and at the same time keep inventories at a low level in logistics and production systems, logistics engineers are essentially treating a system with problems of three types: complexity, data availability, and uncertainty. These properties are consequentially passed on to simulation models and lead to difficulties in model validation.

So what is model complexity? What is model uncertainty? Why is data availability so important in model validation? Why are complexity and uncertainty so essential in logistics systems, and how do they affect model validation?

‘Complexity’ is a term used to describe the difficulty of understanding or analysing a system, as opposed to a ‘simple’ system, in which causality can be easily identified. The degree of complexity depends on the number of elements, their interconnectedness and the number of different system states. Complexity is increasing in production and logistics systems, owing to high product complexity, short product life cycles, small lot sizes and increasing numbers of product variants. The components of complexity in a simulation model include the control logic and the interaction between entities.

‘Uncertainty’ is described as a lack of predictability or the occurrence of events whose outcomes are not settled, are in doubt, or are dependent on chance. Uncertainty in a simulation model might be the uncertainty in the accuracy of data collection, input distribution models and their respective parameters, the uncertainty in the accuracy of a simulation model in describing a system or process, or the uncertainty caused by exogenous factors such as the volatility of customer demand. The components of uncertainty in a simulation model are the quantity and quality of input data, the probability distribution and its parameters, such as machine failure rate and repairing times on a production line, or the inter-arrival rate of customer orders in a distribution centre.

Data is the fundament of simulation modelling. A model is built on data, uses data as input and ideally uses system output data for validation purposes. The availability of data is paramount in a simulation study. It can in the first place determine whether a simulation study should be made or not. If no data is available or cannot even be estimated, for example if a system does not yet exist, a simulation cannot be built. Data availability also has a great impact on model complexity and uncertainty. With little data, a model’s detail cannot be high and its scope cannot be large and the uncertainty concerning activity in the model will also reach a maximum.

An increase in complexity raises uncertainty in a production and logistics system, since more entities are involved. The requirement of data quantity and quality will consequently be raised when describing the behaviour of these entities. On the other hand, uncertainty in the simulation model is propagated by the complexity and the availability of data and introduces more uncertainty into the simulation output. Consequently, the uncertainty in the model output leads to further uncertainty in the model validation and an inability to predict system behaviour. The main difficulties of model validation are therefore caused by the interaction of

these three factors. The core challenge of model validation is hence to understand this interaction.

A new way of thinking in the validation of simulation model methods in production and logistics systems is presented. The focus is on finding the most important parameters for the validation of simulation models, the relationships between these parameters in the form of correlations which influence the complexity of the models, the required availability of data for these parameters, and the uncertainty which is largely dependent on data availability. Further, a short assessment is made of sequential bifurcation as a method for model validation in the context of finding important parameters, reducing complexity, and exploring uncertainties.

- **Model complexity:** definition of complexity: a great degree of interdependence between elements in a system.
 - Type 1 complexity: a variable is determined by its past value (autocorrelation)
 - Type 2 complexity: variables are influenced by one other (cross-correlation)
 - Type 3 complexity: a variable's value depends on the past value or values of another or other variables (cross-correlation with time lags).

Type 1 data can reveal a model's complexity by exposing different types of correlation between variables. The larger the amount of data collected from the input and output of a system, the easier the task of identifying correlations among variables.

When type 1 or type 2 data is not available, model complexity can hardly be revealed. However, by using sensitivity analysis or intelligent experiment design, the complexity of a model can be made clearer.

- **Data availability:**
 - Type 1 data: system input and output data
 - Type 2 data: system output data
 - Type 3 data: no data or theoretical system input data

Data availability is essential to model validation. Model complexity and model uncertainty are governed largely by data availability.

When type 1 data is available, then input-output transform validation (black box validation) can be performed. Trace-driven simulation can be used.

When type 2 data is available, statistical methods such as the paired t-test (for testing the mean) or the F-test (for testing variance) can be used.

When type 3 data is available, sensitivity analysis or experiment design can be used.

In the model validation procedure, the availability of data is essential. Model uncertainty and model complexity are built on it.

- **Model uncertainty:** system uncertainty and environmental uncertainty. [HJOS06]
 - Aleatory uncertainty: the word 'aleatory' derives from the Latin 'alea', which means 'the rolling of dice'. It refers to variability, irreducible uncertainty, and is used to describe the inherent variation associated with a real system and its

environment. It is assumed to be the intrinsic randomness of a system or processes.

- Epistemic uncertainty: the word ‘epistemic’ derives from the Greek ἐπιστήμη (episteme), which means ‘knowledge’. It refers to reducible uncertainty; model-form uncertainty is used to describe a lack of knowledge of a system or its surrounding environment, or incomplete information.

Model uncertainty is the stochastic nature of discrete-event simulation. Stochastic variables are the source of uncertainty. It is necessary to divide this uncertainty into two categories: aleatory uncertainty and epistemic uncertainty. They exist in most simulation models in production and logistics systems.

Data availability affects model uncertainty. Type 1 data can reduce epistemic uncertainty in both input and output. The more input and output data can be collected from a system, the greater the epistemic uncertainty is reduced.

Type 2 data cannot reduce uncertainty in input, but it reduces uncertainty in output and is very useful for the calibration procedure in model validation.

Type 3 data greatly increases the epistemic uncertainty, since neither input nor output can be set in some cases and can be only estimated. In particular, when no information concerning a machine failure is available, the best one can do is to determine the availability of a replacement machine from the equipment provider. In this situation, the value range and the choice of statistical distribution are both uncertain. The epistemic uncertainty reaches a peak.

On the other hand, aleatory uncertainty is related to stochastic variables. Aleatory uncertainty cannot be ignored in simulation models. Doing so will lead directly to model invalidity.

Consider the three factors of simulation validation: model uncertainty, model complexity and data availability. When model complexity is reduced, the uncertainty in a model will be at least reduced and the required data for validation will also automatically decrease, i.e. the requirement for data availability will also decrease. So it is important to determine which factors are important in model performance. Then, model uncertainty and data availability can be focused on according to those factors, which means the interaction of model input factors and the data needed for those factors. Sequential bifurcation can be used to detect these factors and thereby provide invaluable insight into features which would perhaps significantly affect model validation, while insignificant parameters are removed from further study. Also, the interactions between parameters will be revealed by the end of the study, if there are any.

Sequential bifurcation is a screening technique. It is a method developed by Bettonvil and Kleijnen in 1996 which can detect the most important factors from a large number with a relatively small number of simulation runs. The result from a sequential bifurcation is a list of important factors. The simulation models of logistics and production systems normally contain large numbers of control factors. Many techniques have been proposed to identify important features with reasonable numbers of control factors and few simulation runs.

These reductions can be vital in most situations in industrial projects, especially when systems are large and have many input parameters. When the most important variables are screened

out, then the limited time and resources can be applied to them for the collection and analysis of data. According to the sparsity-of-effects principle, the performance of a system, however simple or complex, is influenced largely by main effects and lower-order interactions. The main effects are the effects of an independent variable on a dependent variable averaging across the levels of any other independent variables. The lower-order interactions here are two factorial interactions. So the interactions of more than two factors are very rare. C.F. Jeff Wu and Hamada [WuHa00] call this the hierarchical ordering principle. They state that the sparsity-of-effects principle implies that only a small number of input parameters in a factorial design experiment will be statistically significant.

Sequential bifurcation belongs to group screening. It was developed in the doctoral dissertation of Bettonvil [Bett90], later summarized by Bettonvil and Kleijnen [BeKl96], and further developed by Campolongo et al. [CaKA00], with discussions of applications. Later, in 1997, Cheng and Holland developed it further for stochastic simulations. In 2006, Wan et al. provided a control of Type I and Type II error probabilities and a method for counting the interactions between factors in discrete-event simulation. In the same year, Ankenman et al. and Kleijnen et al. used SB in discrete-event simulations to solve the practical supply chain problem for Ericsson in Sweden. They used the method to reduce input factors from 92 to 10 in only 19 combinations.

The sequential bifurcation method is very suitable for model validation because it meets the requirements that have been discussed in the previous chapters. The reduction of input variables is essential in model validation.

Since there is no way of checking the output of a simulation model that does not yet exist, the only way to objectively check the output results of a simulation model is to reduce the model to a meta-model or some kind of analytical model, e.g. a queuing model, Little's formula, etc. However, in doing this, the information from a simulation model is drastically lost and only some performance measures can be checked. The validation of the model cannot be checked thoroughly.

The simulation model in industry normally contains hundreds or even thousands of input variables. It is computationally not feasible to check the complexity, correlation and variation of these variables. It is indeed impossible to check some of them without a real system, since we can only have confidence based on what we already know rather than what we think we know, and focus on a limited number of things, of which the variations can be controlled, the correlations detected, and the complexity comprehended.

The aim of sequential bifurcation is to find the most important input parameters from a large number of them. "Most important" means when an experiment has a single output and the input parameters have only additive effects. The model is observed as a black box, an input-output transformation, and is modelled by a first-order polynomial or main effect only. The "most important" input parameter has the largest absolute value for the first-order polynomial or main effect. By contrast, the most unimportant factor has an effect almost equal to zero. In a model, there are frequently more than only input parameters that have a main effect and are all important. Sequential bifurcation can give a list of important factors. Then the validation will focus on these factors and their relationships. All other factors not on the list can be ignored and will not affect the validation procedure or other validation techniques. In this way, validation expenses can be saved.

The following paradigm is suggested for simulation model validation without the prerequisite of the existence of a real-world system. The amount of knowledge in the real system depends largely on whether the system exists or not.

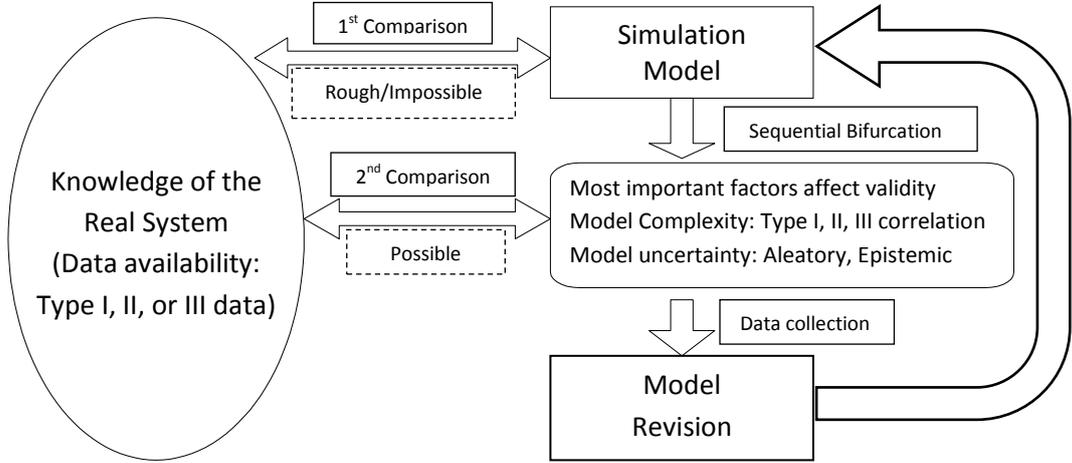


Figure 4-2: Model validation paradigm with sequential bifurcation

Data availability is decisive in the validation procedure. The power of the validation procedure depends largely on whether a real system exists or not. The existence of a system will provide much more data and knowledge for model validation than if one does not yet exist.

If a system exists, any knowledge about it will help simulation experts using statistical methods to validate their simulation model. However, when a simulation model fails the tests, it is difficult to modify it using the test information because almost all statistical methods treat the simulation model as a black box. The real system in industry contains thousands of variables or even more, and the randomness of some is impossible to investigate, while the correlation among them is difficult to detect. It is hard to revise a model, because it is difficult to locate the mistakes in it for the reasons given above. Even when a system exists, sometimes the model validation procedure can be difficult, since the amount of data available is computationally formidable. The first comparison of simulation results and Type III data cannot, therefore, lead to any conclusion other than the yes or no question of validation.

Hereafter, it is suggested, for the further validation procedure, that one should use the sequential bifurcation method to find the most important variables and their relationships in a simulation model. Since model validation should focus on the most important variables or factors and their relationships to one another, the next task, when revising the model, is to concentrate on data collection. A second comparison is made possible by this procedure, since the number of variables is manageable. This will save much time and effort in locating mistakes in the model. The revised model will again make comparisons with historical data and if the model again fails the tests, it will be further tested using sequential bifurcation, and the procedure repeated until the model does pass the statistical tests.

In logistics and production systems, in many cases the real system does not exist or will undergo drastic modification. This makes the validation procedure even more difficult because information about the system and its environment is limited, which means that the

aleatory and epistemic uncertainties are both high and only Type III data is available. In a logistics and production system even if the system is only in the planning phase some data can be gained or reasonable assumptions made, since, in most cases, many comparable systems already exist. But there will still be no output data from the real system. A comparison of output data from system and model is impossible. In this case, other validation techniques should be used. The factors that are found to be important should first be compared to a real system, to decide whether they are really important in it.

The critical part of the diagram is the sequential bifurcation procedure.

The procedure of sequential bifurcation is very straightforward. First of all, the input parameters are grouped together and whether the group has an important effect or not is tested. If indeed there is an important effect, the group is split into two sub-groups and each of these is tested for the important effect. This procedure is repeated and the unimportant groups are eliminated while the groups with important effects are further split into sub-groups. In this way, all the factors with important effects will be found and the unimportant ones eliminated.

There are two assumptions for sequential bifurcation. [KIPe03]

The first assumption is that the model is a first-order polynomial, possibly augmented with two-factor interactions. This can be shown as follows:

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \beta_{1,2} x_1 x_2 + \dots + \beta_{(k-1),k} x_{k-1} x_k + \varepsilon$$

where

Y is the response for the metamodel

K is the total number of factors in the experiment

β_j is the first-order or main effect of factor j with $j = 1, \dots, k$

$\beta_{j',j}$ is the interaction effect of the factors j' and j with $1 \leq j' \leq j \leq k$

x_j is the value of the j th factor, standardized to lie in $[-1, +1]$

ε is the noise variable arising from both the use of pseudorandom numbers and an approximation error.

This metamodel is linear in β_j and $\beta_{j',j}$, but not linear in x_j . The advantage of this low-order polynomial is that it is simple and can be easily applied to discrete-event simulation. Actually, all classic group screening designs assume the first-order polynomial. When using sequential bifurcation, only half the number of runs are required, as compared to an approximation with two-factor interactions.

The most efficient way to estimate the parameters of the metamodel above is to experiment with only two values (levels) per factor. The values of these factors should be as realistic as possible.

The second assumption is that all main effects are known and are non-negative. This requires that the signs or directions of the effects should be known. The reason for this is that the individual effects do not cancel one another. This assumption is reasonable, since most of the time the sign of a factor can be estimated by a system domain expert or the model builders of logistics and production systems, even though the magnitude of the change is unknown. For example, for the output variable throughput time, it can be assumed that the reduction in setup time has non-negative effects on the output variable. This also applies to a reduction in buffer size in production systems.

If the effects of some factors are hard to estimate, they should be treated individually and should not be grouped with other factors in the procedure.

Bettonvil [Bett90] suggested that factors should be ranked in increasing order of importance to make the sequential bifurcation method more efficient. This means:

$$\beta_{j'} \leq \beta_j \text{ when } j' < j$$

This may involve prior knowledge of the effects of factors in the simulated system. Besides this, Kleijnen [KlPe03] also suggests that, in order to increase efficiency, prior knowledge should also be used to keep similar factors together as “clusters”. In this way, it can be assumed that if a factor in one “cluster” is not important, then all the factors in this “cluster” are unimportant. Bettonvil [Bett90] also suggests dividing the factors into groups, as the number of sub-groups is a power of two.

Sequential bifurcation begins with the simulation of two extreme scenarios. In the first scenario, all factors are set to their low levels; in the second scenario all factors are set to their high levels.

The simulation experiments are run with common pseudorandom numbers in order to reduce the variances in the outputs, since positive correlations between the responses at the various factor combinations are created.

For $i = 1, 2$, let $X_{i1}, X_{i2}, \dots, X_{in_i}$ be a sample of n_i IID (identical and independent data) observations from system S , and let $\mu_i = E(X_{ij})$ be the expected response of interest; a confidence interval can be constructed for $\zeta = \mu_1 - \mu_2$. If the confidence interval does not contain 0, there is a statistical difference between the two scenarios. Whether or not X_{1j} and X_{2j} are independent depends on how the simulations are executed, and could determine which of the following two confidence-interval approaches should be used.

- Paired-t confidence interval
- Two-sample-t confidence interval

Since common pseudorandom numbers are used in the running of different scenarios, the numbers of output data points are expected to be the same. If there are more data points in one scenario than the other, they can be discarded. So one can pair X_{1j} and X_{2j} to define $Z_j = X_{1j} - X_{2j}$, for $j = 1, 2, \dots, n$ (it is not necessary that for a fixed j , X_{1j} and X_{2j} are independent). Then the Z_j s are IID random variables and $E(Z_j) = \zeta$, the quantity for which we want to construct a confidence interval. And let,

$$\bar{Z}(n) = \frac{\sum_{j=1}^n Z_j}{n}$$

and

$$\widehat{Var}[\bar{Z}(n)] = \frac{\sum_{j=1}^n [Z_j - \bar{Z}(n)]^2}{n(n-1)}$$

and form the (approximate) 100(1- α)% confidence interval:

$$\bar{Z}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\widehat{Var}[\bar{Z}(n)]}$$

It is said to be approximate, since, if the Z_j s are normally distributed, this confidence interval is exact, i.e. it covers ζ with the probability $1 - \alpha$; otherwise, we rely on the central limit theorem, which implies that this coverage probability will be near $1 - \alpha$ for large n .

In the following case studies, sequential bifurcation is applied to two examples of manufacturing systems. The first example, already used in Chapter 3, is an academic model with limited complexity (model 1). The second is a real example from industry of high complexity (model 2).

This example was used in the previous chapter. Now it will be described in detail.

- On the production line of an electrical manufacturer consumer goods are produced. On this line there is a sub-area where the goods are inspected and even small repair jobs are done. The material flow is illustrated in Figure 3-5. A preliminary production area supplies our production line with two different kinds of product groups (randomly distributed) at intervals of one minute. Both kinds of goods enter the system at a receipt buffer and are fed to the two work stations by shuttle. But there is a fixed assignment of each kind of goods to the individual work station, because full flexibility cannot be operated economically owing to high investment needs.
- The outputs are the throughputs of the system, the inputs (factors) are the inter-arrival times of products 1 and 2; the capacities of conveyors $rcpgoods$, ACC_Above , ACC_Below , ACC_8 , ACC_9 , ACC_15 , PST_16 , ACC_14 ; conveyor speed; shuttle speed; processing time and setup time of machine above and machine below; the rework rate; leaving time of the sink.

The first step in the sequential bifurcation is to sort all the identified input variables from smallest to largest according to their effect on the output variable. This requires knowledge of the system. For the system above, there are fifteen variables, as follows: the inter-arrival times of products 1 and 2; the capacities of conveyors $rcpgoods$, ACC_Above , ACC_Below , ACC_8 , ACC_9 , ACC_15 , PST_16 , ACC_14 ; conveyor speed; shuttle speed; processing time and setup time of machine above and machine below; the rework rate; leaving time of the sink. The sorted sequence of the variables is:

ACC_8	ACC_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods	Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
-------	-------	-----------	-----------	--------	--------	--------	----------	----------------	---------------	-----------------	------------	-------------	--------------------	-----------

Table 4-2: Sorted sequence of the variables of model 1

This sequence puts all the conveyors in the front, from ACC_8 to repgoods; the values will change according to their capacities. In the second part, all are related to time, except the work rate. From conveyor speed to sink time, the values will change in a way that the time a part spends in the system increases or decreases.

The first experiment contains two scenarios. In the first scenario, all the values of the input variables are set to their low value. In the second scenario, all the values of the input variables are set to their high value. The values are in the following table:

Experiment1	AC C_8	AC C_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	repgoods	conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
Scenario 1	1	4	2	2	2	4	5	10	0.2	1	N(80,5)	60	0.15	N(60,5)	E(55)
Scenario 2	2	8	10	10	10	8	10	20	0.4	2	N(70,5)	30	0.05	N(50,5)	N(55,5)

Table 4-3: Experiment 1 - two scenarios of input variables of model 1

Both scenarios are run in a simulation model. The simulation time is ten days. The warm up time is one day. The results are as follows:

Experiment 1 replicate	Scenario 1 Y(0-15)	Scenario 2 Y(0-15)	Difference β_{1-15}
1	47.40	65.40	18.00
2	44.60	61.80	17.20
3	47.20	65.40	18.20
4	47.60	65.40	17.80
5	43.90	65.30	21.40
Average	46.14	64.66	18.52
Standard Error	0.70	0.64	0.66

Table 4-4: results of experiment 1 of model 1

Each scenario is run five times with different random number seeds. However, for each replicate of both scenarios, the common random number technique is used to guarantee the maximum correlation.

The result is very clear. In scenario 1 Y(0-15), the throughput per hour is 46.14 with standard error 0.70. In scenario 2 A(0-15), the throughput per hour is 64.66 with standard error 0.64. The difference between the two scenarios is 18.52 with standard error 0.66.

The increase in the input variable values has a very obvious effect on the throughput per hour.

In the second step, the sequence is cut into two parts, the so-called bifurcation. The first part contains all the conveyors. The second part contains all other input variables. These are shown in the following table.

Part 1	ACC_8	ACC_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods
Part 2	conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time	

Table 4-5: First bifurcation of variables of model 1

Experiment 2, with the values of the input variables, is as follows.

Experiment2	AC C_8	AC C_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods	conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
Scenario 1	1	4	2	2	2	4	5	10							
Scenario 2	2	8	10	10	10	8	10	20	0.4	2	N(70,5)	30	0.05	N(50,5)	N(55,5)
									0.2	1	N(80,5)	60	0.15	N(60,5)	E(55)

Table 4-6: Experiment 2 – scenario 1 and scenario 2 after first variables bifurcation of model 1

In Experiment 2, the first scenario keeps all the values of the first part at a low value and raises the values of the second part. The second scenario is the exact contrary of the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table:

Experiment 2	Scenario 1	Scenario 2	Difference
replicate	Y(1-8)	Y(9-15)	β_{1-15}
1	50.60	65.20	14.60
2	50.50	65.10	14.60
3	50.60	65.20	14.60
4	50.60	65.10	14.50
5	50.80	64.90	14.10
Average	50.62	65.10	14.48
Standard Error	0.04	0.05	0.09

Table 4-7: results of experiment 2 of model 1

The experiments are run under the same conditions as the first, as are all the following experiments. Both scenarios are run in a simulation model. The simulation time is ten days. The warm-up time is one day. Each scenario is run five times with different random number seeds. However, for each replicate of both scenarios, the common random number technique is used to guarantee the maximum correlation.

The results show that in the first scenario the throughput per hour decreased to 50.62 with a standard error of 0.04. This means that the increase in the capacities of the conveyors does indeed have a positive effect on the throughput of the system. However, in the second scenario, the throughput per hour keeps the same 65.10 with a standard error of 0.05, which is the same as the second scenario of Experiment 1. This shows that the main effect can be in the second scenario. In this case, the first part and the second part are investigated separately in the following steps of this method.

In Experiment 3, the main effect is investigated. In this experiment all the values in the first part of Experiment 2 are fixed at their low value.

ACC_8	ACC_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods
-------	-------	-----------	-----------	--------	--------	--------	----------

Table 4-8: experiment 3 – part 1 - main effect investigation of model 1

The second part of Experiment 2 is divided further into two parts.

Part 1	conveyor speed	shuttle speed			
Part 2	processing time	setup time	rework rate	inter-arrival time	sink time

Table 4-9: experiment 3 – part 2 - main effect investigation of model 1

The reason for the grouping is that all the speed-related variables are in Part 1. In the second group the variables are all time-related.

Experiment 3, with the values of the input variables, is as follows:

Experiment3	AC_C_8	AC_C_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods	conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
Scenario 1	1	4	2	2	2	4	5	10							
Scenario 2	1	4	2	2	2	4	5	10	0.2	1					
									0.4	2	N(70,5)	30	0.05	N(50,5)	N(55,5)
											N(80,5)	60	0.15	N(60,5)	E(55)

Table 4-10: Experiment 3 – scenario 1 and scenario 2 after second variables bifurcation of model 1

In Experiment 3, the first scenario keeps the values of all the first part at a low value and raises the values of the second part. The second scenario is exactly contrary to the first. It raises the values of the first part to their high value and keeps the second part at their low value. The result of the second experiment is in the following table.

Experiment 3	Scenario 1	Scenario 2	Difference
replicate	Y(9-10)	Y(11-15)	β_{1-15}
1	53.20	53.30	0.10
2	53.30	53.20	-0.10
3	53.30	52.60	-0.70
4	53.20	52.90	-0.30
5	53.40	53.50	0.10
Average	53.28	53.10	-0.18
Standard Error	0.03	0.14	0.13

Table 4-11: results of experiment 3 of model 1

This time one can see that the throughputs of both Scenario 1 and Scenario 2 are unable to remain above 60. They are 53.28 and 53.10 respectively. This indicates two points: firstly, the interaction between the two groups is important; secondly, the main effect can be in either of the groups.

Since there are only two input variables, conveying speed and shuttle speed, in the first part of Experiment 3, their effects can be tested individually.

With the shuttle speed being set to a high value: 2, the throughput per hour is 51.5 with standard error 0.5. When the conveying speed is only doubled to 0.4, the throughput per hour is 46.2 with standard error 0.6, which is the same as when all the variables are set to their low values. So the main effect in this part is the shuttle speed.

In order to find the main effect of the second part of Experiment 3, a further bifurcation in Experiment 4 is needed.

In Experiment 4, all the values in the first part of Experiment 2 and Experiment 3 are fixed at their low value.

ACC_8	ACC_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods	conveyor speed	shuttle speed
-------	-------	-----------	-----------	--------	--------	--------	----------	----------------	---------------

Table 4-12: Experiment 4 – variables of first part of experiment 2 and 3 fixed to low values

The second part of Experiment 3 is divided further into two parts:

Part 1	processing time	setup time	
Part 2	rework rate	inter-arrival time	sink time

Table 4-13: Experiment 4 - Third bifurcation of variables in the second part of experiment 3

The reason for the grouping is that in Part 1 all the variables are machine-related times. All the external times except the rework rate are in the second part.

Experiment 4 with the values of the input variables is as follows.

Experiment4	ACC_8	ACC_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods	Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
Scenario 1	1	4	2	2	2	4	5	10	0.2	1					
Scenario 2	1	4	2	2	2	4	5	10	0.2	1	N(80,5)	60	0.05	N(50,5)	N(55,5)
											N(70,5)	30	0.15	N(60,5)	E(55)

Table 4-14: Scenario 1 and scenario 2 after third variables bifurcation of model 1

In Experiment 4, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is the exact opposite of the first. This raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 4	Scenario 1	Scenario 2	Difference
replicate	Y(11-12)	Y(13-15)	β_{1-15}
1	52.30	49.40	-2.90
2	52.20	49.30	-2.90
3	52.30	49.10	-3.20
4	52.40	49.10	-3.30
5	52.20	49.30	-2.90
Average	52.28	49.24	-3.04
Standard Error	0.05	0.05	0.08

Table 4-15: Results of experiment 4 of model 1

The results show that the main effect is in Scenario 1, since the throughput is larger by 3.04 with standard error 0.08.

In order to find the main effect of the second part of Experiment 4, we need to further bifurcate in Experiment 5.

In Experiment 5 all the values from the first parts of Experiments 2, 3, and 4 are fixed at their low values.

ACC_8	ACC_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods	Conveyor speed	shuttle speed	processing time	setup time
-------	-------	-----------	-----------	--------	--------	--------	----------	----------------	---------------	-----------------	------------

Table 4-16: Experiment 5 - variables of first part of experiment 2, 3 and 4 fixed to low values

The second part of Experiment 4 is divided further into two parts:

Part 1	rework rate	
Part 2	inter-arrival time	sink time

Table 4-17: Experiment 5 – fourth bifurcation of variables in the second part of experiment 4

It is obvious that the rework rate should be in a different group from the time-related variables.

Experiment 5 with the values of the input variables is as follows.

Experiment5	ACC_8	ACC_9	ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgoods	Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
Scenario 1	1	4	2	2	2	4	5	10	0.2	1	60	60	0.15	N(80,5)	N(55,5)
Scenario 2	1	4	2	2	2	4	5	10	0.2	1	N(80,5)	60	0.05	N(50,5)	N(60,5)

Table 4-18: Scenario 1 and scenario 2 after fourth variables bifurcation of model 1

In Experiment 5, the first scenario keeps all the values of the first part at the low value and raises the values of the second part. The second scenario is the exact contrary of the first. This raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table

Experiment 5	Scenario 1	Scenario 2	Difference
replicate	Y(11-12)	Y(13-15)	β_{1-15}
1	47.90	51.80	3.90
2	47.60	51.70	4.10
3	47.80	51.40	3.60
4	47.70	51.60	3.90
5	47.70	51.60	3.90
Average	47.74	51.62	3.88
Standard Error	0.06	0.06	0.07

Table 4-19: Results of experiment 5 of model 1

The results show that in Scenario 1 the main effect is the rework rate, since the throughput is larger by 3.88 with standard error 0.07.

To sum up the important input variables for the second part of Experiment 2:

Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
----------------	---------------	-----------------	------------	-------------	--------------------	-----------

Table 4-20: important input variables in the second part of experiment 2 of model 1

They are the shuttle speed and the rework rate, and the interactions between shuttle speed and rework rate.

Looking back to the first part of Experiment 2, further bifurcations need to be carried out in order to find the main effects in the conveyors. Although they are not able to improve the throughput individually, they can interact with the input variables of shuttle speed and rework rate.

In Experiment 6 all the values from the first parts of Experiments 2, 3, 4 and 5 are fixed at their low values.

Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
----------------	---------------	-----------------	------------	-------------	--------------------	-----------

Table 4-21: Experiment 6 - variables of first part of experiment 2, 3, 4 and 5 fixed to low values

The first part of Experiment 2 is divided further into two parts:

Part 1	ACC_8	ACC_9	ACC_Above	ACC_Below
Part 2	ACC_14	ACC_15	PST_16	rcpgoods

Table 4-22: Experiment 6 - fifth bifurcation of variables in the first part of experiment 2

Experiment 6 with the values of the input variables is as follows:

Experiment6									Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
Scenario 1					ACC_14	ACC_15	PST_16	rcpgoods	0.2	1	N(80,5)	60	0.15	N(60,5)	E(55)
Scenario 2	ACC_8	ACC_9	ACC_Above	ACC_Below	10	8	10	20	0.2	1	N(80,5)	60	0.15	N(60,5)	E(55)
	1	4	2	2	2	4	5	10							
	2	8	10	10											

Table 4-23: Scenario 1 and scenario 2 after fifth variables bifurcation of model 1

In Experiment 6, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is the exact opposite of the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 6	Scenario 1	Scenario 2	Difference
replicate	Y(5-8)	Y(1-4)	β_{1-15}
1	48.00	50.60	2.60
2	48.30	50.80	2.50
3	47.90	50.50	2.60
4	48.30	50.90	2.60
5	48.30	50.70	2.40
Average	48.16	50.70	2.54
Standard Error	0.08	0.06	0.04

Table 4-24: Results of experiment 6 of model 1

The results show that the main effect is in Scenario 2, since the throughput is larger by 2.54 with standard error 0.10.

In order to find the main effect of the second part of Experiment 6, we need a further bifurcation in Experiment 7.

In Experiment 7 all the values of the first part of Experiments 2, 3, 4, 5, and 6 are fixed at their low value.

ACC_14	ACC_15	PST_16	rcpgoods	Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time
4	5	6								

Table 4-25: Experiment 7 - variables of first part of experiment 2, 3, 4, 5 and 6 fixed to low values

The second part of Experiment 6 is divided further into two parts:

Part 1	ACC_8	ACC_9
Part 2	ACC_Above	ACC_Below

Table 4-26: Experiment 7 - sixth bifurcation of variables in the second part of experiment 6

Experiment 7, with the values of the input variables, is as follows:

Experiment7								Conveyor speed	shuttle speed	processing time	setup time	rework rate	inter-arrival time	sink time	
Scenario 1								0.2	1	N(80,5)	60	0.15	N(60,5)	E(55)	
Scenario 2			ACC_Above	ACC_Below	ACC_14	ACC_15	PST_16	rcpgo ods	0.2	1	N(80,5)	60	0.15	N(60,5)	E(55)
	ACC_8	ACC_9	10	10	2	4	5	10							
	1	4	2	2											
	2	8													

Table 4-27: Scenario 1 and scenario 2 after sixth variables bifurcation of model 1

In Experiment 7, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is the exact contrary of the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 7	Scenario 1	Scenario 2	Difference
replicate	Y(1-2)	Y(3-4)	β_{1-15}
1	47.80	50.60	2.80
2	48.10	50.80	2.70
3	47.60	50.50	2.90
4	47.90	50.70	2.80
5	47.50	50.30	2.80
Average	47.78	50.58	2.80
Standard Error	0.10	0.08	0.03

Table 4-28: Results of experiment 7 of model 1

The results show that in Scenario 2, the capacities of both conveyors before machine ACC_Above and ACC_Below are the main effect, since the throughput is larger by 2.80 with standard error 0.03.

Therefore, the most important input variables of the conveyors are the conveyors before machine ACC_Above and ACC_Below.

It can be concluded that by using sequential bifurcation, not only the most important parameters of the system can be found, but also the most significant interactions between these variables. This makes model validation much easier, since the effort needed for data collection and the investigation of model complexity and model uncertainty decrease exponentially.

As can be seen, long run performance such as throughput, machine utilization or throughput time can be achieved only through steady-state simulation, and most performance measurements in logistics and production systems need to be estimated in this way.

Batch means have the advantage of using only one experiment design which is based on a single and long replication. In this way, the number of replications for each experiment is starkly reduced. This is an advantage especially for projects in real life in which the systems

are large and require a long time to run. Even though there are no universal rules for choosing batch sizes or the numbers of batches, some guidelines are still available: 1). The number of batches should be between 10 and 30; 2). The lag-1 autocorrelation should be near 0; 3). If the sample size needs to be increased, the number of batches should also be increased by the square root of the sample size after finding a batch size at which the lag-1 autocorrelation is near zero.

The batch mean is calculated by the equation:

$$\bar{Y}_j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} Y_{i+d}$$

for $j = 1, 2, 3, \dots, k$.

The variance of the sample mean is estimated by the equation:

$$\frac{S^2}{k} = \frac{\sum_{j=1}^k \bar{Y}_j^2 - k\bar{Y}^2}{k(k-1)}$$

The sample lag-1 autocorrelation of the batch mean is calculated by the equation:

$$\widehat{\rho}_1 = \frac{\sum_{j=1}^{k-1} (\bar{Y}_j - \bar{Y})(\bar{Y}_{j+1} - \bar{Y})}{\sum_{j=1}^k (\bar{Y}_j - \bar{Y})^2}$$

if the $\widehat{\rho}_1 \leq 0.2$, then a confidence interval uses $k-1$ degrees of freedom for the t-distribution, otherwise, the replication should be extended by 50% to 100%. If the length of replication cannot be further extended, then re-batch the data into $k=10$ batches and form the confidence interval using $k-1$ degrees of freedom for the t-distribution.

An additional check on the confidence interval is the following test statistic:

$$C = \sqrt{\frac{k^2 - 1}{k - 2}} \left(\widehat{\rho}_1 + \frac{(\bar{Y}_j - \bar{Y})^2 + (\bar{Y}_k - \bar{Y})^2}{2 \sum_{j=1}^k (\bar{Y}_j - \bar{Y})^2} \right)$$

if $C < z_\beta$, then the independence of the batch mean can be accepted, otherwise the replication should be extended by 50% to 100%. If the length of replication cannot be further extended, then re-batch the data into $k=10$ batches and form the confidence interval using $k-1$ degrees of freedom for the t-distribution.

In the next example, a realistic production and transportation system is shown with the use of sequential bifurcation. The scenarios of each experiment are analyzed using the batch mean method.

The example has already been mentioned in Chapter 3. The purpose of the simulation study is to determine how to increase the throughput per hour by 50% in a production system.

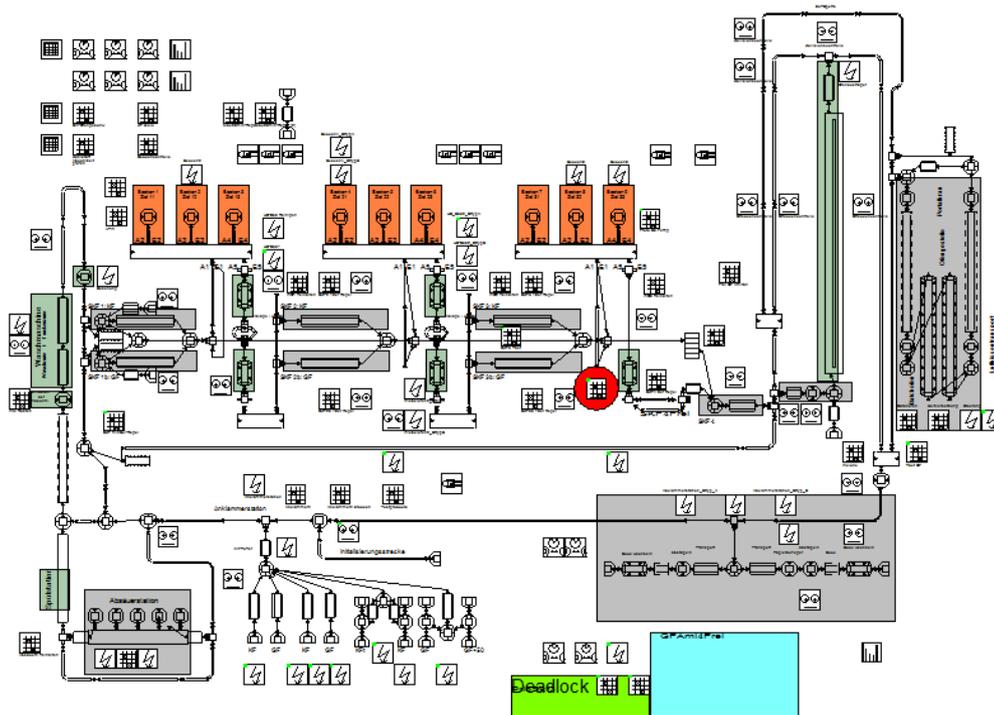


Figure 4-3: A realistic production and transportation system with the use of sequential bifurcation

This production system produces two groups of product. Product group Small has six types of product, 1, 2, 3, 4, 5, and 6. Product group Large has ten types of product, 7-1, 7-2, 8, 9, 11, 13, 15, 16, 19, and 21. The raw materials are conveyed to the system and brought on work-piece trays. The trays carrying the raw materials travel through the acid station, flushing station and washing station, then go through a basin and an oven, and are finally sent out of the production system.

Because of the manufacturing technique, the semi-products in the system need some cooling time before they are moved to each next station.

Owing to complex control logic and the strong influence that the elements in the system have on one another, it is difficult to judge how to increase the throughput of the system by 50%, and to experiment on the real system would affect the current customer order and also damage the products if an incident occurred during the experiment. So a simulation model is built for this purpose.

The validation of the model is obviously vital, since the results from all experiments being performed will be based on the model's accuracy. Since this is an existing system, real production data should have been available for the statistical tests mentioned in the early part of this chapter. However, owing to the client's reluctance to provide this data, for confidentiality reasons, no data is available for statistical model validation. Therefore, the sequential bifurcation method is used to find the most important input variables and their relationships in the system. In this way, the model's complexity and uncertainty are reduced.

Forty-two input variables that can affect the system throughput are identified.

1	LIM_SKF_KF	11	LF_ABKLAMMERN	21	v_SchwO	31	ANZ_INIT_EHB_GF
2	LIM_SKF_GF	12	MAX_Wartezeit	22	v_Waschen	32	ANZ_INIT_BAD_KF
3	LIM_UTW_AMIRAN	13	limRaeumen	23	FW_Amiran	33	ANZ_INIT_SWO_KF
4	LIM_UTW_SKF4	14	Brenndauer_AMIRAN	24	vWasche2	34	ANZ_INIT_EHB_KF
5	Leistungsfaktor	15	Brenndauer_AMIRAN2	25	T_ABKLAMMERN	35	SKF_ini_KF
6	LF_AbkuehlenGF	16	Brenndauer_AMIRAN3	26	T_ABKLAMMERN2	36	SKF_ini_GF
7	LF_Waschen	17	Brenndauer_AMIRAN4	27	EHB_V	37	SKF_Q_KF
8	LF_SwOfen	18	Brenndauer_AMIRAN_5	28	T	38	MIN_SKF_KF
9	LF_AMIRAN	19	Brenndauer_AMIRAN_6	29	ANZ_INIT_BAD_GF	39	MIN_SKF_GF
10	LF_VTWBAD	20	Brenndauer_AMIRAN_7	30	ANZ_INIT_AMO_GF	40	KAP_SKF_GF
41	KAP_SKF_KF						
42	KAP_LIFO						

Table 4-29: forty two input variables in model 2

Some input variables are dependent on other input variables:

- $Brenndauer_AMIRAN = 60 \cdot 60 \cdot 10 / LF_AMIRAN$
- $Brenndauer_AMIRAN2 = 60 \cdot 60 \cdot 18.25 / LF_AMIRAN$
- $Brenndauer_AMIRAN3 = 60 \cdot 60 \cdot 11 / LF_AMIRAN$
- $Brenndauer_AMIRAN4 = 60 \cdot 60 \cdot 14 / LF_AMIRAN$
- $Brenndauer_AMIRAN_5 = 60 \cdot 60 \cdot 7.5 / LF_AMIRAN$
- $Brenndauer_AMIRAN_6 = 60 \cdot 60 \cdot 10.3 / LF_AMIRAN$
- $Brenndauer_AMIRAN_7 = 60 \cdot 60 \cdot 13.3 / LF_AMIRAN$
- $v_Waschen = 1.53 / 60 \cdot LF_Waschen$
- $T_ABKLAMMERN = 150 / LF_ABKLAMMERN$
- $T_ABKLAMMERN2 = 300 / LF_ABKLAMMERN$

Further, the values of LIM_SKF_KF. So there are in fact only twenty-nine input variables whose values can be changed. Following this, all input variables are sorted from smallest to largest according to their effect on the output variables. This requires a knowledge of the system.

1	LIM_SKF_KF	11	KAP_SKF_KF	21	LF_AMIRAN
2	LIM_SKF_GF	12	MAX_Wartezeit	22	LF_VTWBAD
3	LIM_UTW_AMIRAN	13	limRaeumen	23	LF_ABKLAMMERN
4	LIM_UTW_SKF4	14	v_SchwO	24	ANZ_INIT_BAD_GF
5	SKF_ini_KF	15	FW_Amiran	25	ANZ_INIT_AMO_GF
6	SKF_ini_GF	16	vWasche2	26	ANZ_INIT_EHB_GF
7	SKF_Q_KF	17	LF_Abkuehlen	27	ANZ_INIT_BAD_KF
8	MIN_SKF_KF	18	LF_AbkuehlenGF	28	ANZ_INIT_SWO_KF
9	MIN_SKF_GF	19	LF_Waschen	29	ANZ_INIT_EHB_KF
10	KAP_SKF_GF	20	LF_SwOfen		

Table 4-30: sorted sequence of twenty-nine input variables affect the system throughput in model 2

The first experiment comprises two scenarios. In the first scenario, all the values of the input variables are set to their low value. In the second scenario, all the values of the input variables are set to their high value. The values are in the following table:

No.	Experiment1	Scenario 1	Scenario 2
1	LIM_SKF_KF	3	12
2	LIM_SKF_GF	3	7
3	LIM_UTW_AMIRAN	3	10
4	LIM_UTW_SKF4	4	8
5	SKF_ini_KF	2	12
6	SKF_ini_GF	2	10
7	SKF_Q_KF	1	10
8	MIN_SKF_KF	1	6
9	MIN_SKF_GF	1	6
10	KAP_SKF_GF	3	6
11	KAP_SKF_KF	4	8
12	MAX_Wartezeit	10	40
13	limRaeumen	9	20
14	v_SchwO	0.05/60	0.2/60
15	FW_Amiran	6	24
16	vWasche2	0.45/60	1.84/60
17	LF_Abkuehlen	1	2
18	LF_AbkuehlenGF	1	4
19	LF_Waschen	1	2
20	LF_SwOfen	1	2
21	LF_AMIRAN	1	4
22	LF_VTWBAD	1	4
23	LF_ABKLAMMERN	1	2
24	ANZ_INIT_BAD_GF	10	20
25	ANZ_INIT_AMO_GF	38	49
26	ANZ_INIT_EHB_GF	12	14
27	ANZ_INIT_BAD_KF	24	44
28	ANZ_INIT_SWO_KF	30	40
29	ANZ_INIT_EHB_KF	8	24

Table 4-31: Experiment 1 - two scenarios of input variables of model 2

Both scenarios are run in a simulation model. The simulation time is thirty-three days. The warm-up time is three days, that is, the statistical data collection time is ten times longer than the warm-up period. The results are as follows:

Experiment 1	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	104.1	40.50	248.70	81.30	144.60	40.80
Standard Error	1.94	0.41	7.93	1.11		

Table 4-32: Results of experiment 1 of model 2

The second step is cut into two parts, the so-called bifurcation: the first part contains all the conveyors; the second part contains all the other input variables. They are shown in the following table.

Part 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	LIM_SKF_KF	LIM_SKF_GF	LIM_UTW_AMIRAN	LIM_UTW_SKF4	SKF_ini_KF	SKF_ini_GF	SKF_Q_KF	MIN_SKF_KF	MIN_SKF_GF	KAP_SKF_GF	KAP_SKF_KF	MAX_Wartezeit	limRaeumen	v_SchwO	FW_Amiran	vWasche2
Part 2	17	18	19	20	21	22	23	24	25	26	27	28	29			
	LF_Abkuehlen	LF_AbkuehlenGF	LF_Waschen	LF_SwOfen	LF_AMIRAN	LF_VTWBAD	LF_ABKLAMMERN	ANZ_INIT_BAD_GF	ANZ_INIT_AMO_GF	ANZ_INIT_EHB_GF	ANZ_INIT_BAD_KF	ANZ_INIT_SWO_KF	ANZ_INIT_EHB_KF			

Table 4-33: First bifurcation of variables of model 2

Experiment 2 with the values of the input variables is as follows.

No.	Experiment 2	Scenario 1	Scenario 2
1	LIM_SKF_KF	3	12
2	LIM_SKF_GF	3	7
3	LIM_UTW_AMIRAN	3	10
4	LIM_UTW_SKF4	4	8
5	SKF_ini_KF	2	12
6	SKF_ini_GF	2	10
7	SKF_Q_KF	1	10
8	MIN_SKF_KF	1	6
9	MIN_SKF_GF	1	6
10	KAP_SKF_GF	3	6
11	KAP_SKF_KF	4	8
12	MAX_Wartezeit	10	40
13	limRaeumen	9	20
14	v_SchwO	0.05/60	0.2/60
15	FW_Amiran	6	24
16	vWasche2	0.45/60	1.84/60
17	LF_Abkuehlen	2	1
18	LF_AbkuehlenGF	4	1
19	LF_Waschen	2	1
20	LF_SwOfen	2	1
21	LF_AMIRAN	4	1
22	LF_VTWBAD	4	1
23	LF_ABKLAMMERN	2	1
24	ANZ_INIT_BAD_GF	20(10)	10
25	ANZ_INIT_AMO_GF	49	38
26	ANZ_INIT_EHB_GF	14(12)	12
27	ANZ_INIT_BAD_KF	44(24)	24
28	ANZ_INIT_SWO_KF	40	30
29	ANZ_INIT_EHB_KF	24	8

Table 4-34: Experiment 2 – scenario 1 and scenario 2 after first variables bifurcation of model 2

Because of the control logic in the system some combinations of the variables could lead to a deadlock. To make the sequential bifurcation work and avoid any deadlock in the system, some variables' values are purposely kept at their low value. This includes variables 24, 26 and 27. This will also be the case in the other experiments in the following. However, this is not expected to cause any degradation of the results. The reason for this is that the main effect will still be found even though some variables' values are kept the same, and at least these input variables will not pull down the main effect.

Experiment 2	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	143.8	43.80	154.00	55.50	10.20	11.70
Standard Error	2.22	0.38	3.77	0.56		

Table 4-35: Results of experiment 2 of model 2

The results from Experiment 2 show that in both scenarios the throughputs are not able to reach the level of Scenario 2 in Experiment 1. However, both results are essentially larger than the throughput in Scenario 1 of Experiment 1, in which all the values of the input

variables are kept at their low levels. This indicates two things: firstly, there are important interactions between variables in Part 1 and Part 2 of Experiment 2. Secondly, there are main effects in input variables in both parts.

In the next step of the sequential bifurcation, the variables from Part 1 of Experiment 2 will first be studied further. Then Part 2 will be studied using the same procedure. In these procedures, all the values of the other parts are kept at their low levels. For example, when Part 1 has been bifurcated, the values of Part 2 are kept at their low levels, and vice versa.

The variables in Part 1 of Experiment 2 are:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Part 1	LIM_SKF_KF	LIM_SKF_GF	LIM_UTW_AMIRAN	LIM_UTW_SKF4	SKF_ini_KF	SKF_ini_GF	SKF_Q_KF	MIN_SKF_KF	MIN_SKF_GF	KAP_SKF_KF	KAP_SKF_GF	MAX_Wartezeit	limRaeumen	v_SchwO	FW_Amiran	vWasche2

Table 4-36: Experiment 3 – part 1 - main effect investigation of model 2

The first part of Experiment 2 is divided further into two parts:

Part 1	1	2	3	4	5	6	7										
	LIM_SKF_KF	LIM_SKF_GF	LIM_UTW_AMIRAN	LIM_UTW_SKF4	SKF_ini_KF	SKF_ini_GF	SKF_Q_KF										
Part 2	8	9	10	11	12	13	14	15	16								
	MIN_SKF_KF	MIN_SKF_GF	KAP_SKF_GF	KAP_SKF_KF	MAX_Wartezeit	limRaeumen	v_SchwO	FW_Amiran	vWasche2								

Table 4-37: Experiment 3 – part 2 - main effect investigation of model 2

Experiment 3 with the values of the input variables is as follows.

No.	Experiment3	Scenario 1	Scenario 2
1	LIM_SKF_KF	3	12
2	LIM_SKF_GF	3	7
3	LIM_UTW_AMIRAN	3	10
4	LIM_UTW_SKF4	4	8
5	SKF_ini_KF	2	12
6	SKF_ini_GF	2	10
7	SKF_Q_KF	1	10
	8	MIN_SKF_KF	6
	9	MIN_SKF_GF	6
	10	KAP_SKF_GF	6
	11	KAP_SKF_KF	8
	12	MAX_Wartezeit	40
	13	limRaeumen	20
	14	v_SchwO	0.2/60
	15	FW_Amiran	24
	16	vWasche2	1.84/60
17	LF_Abkuehlen	1	1
18	LF_AbkuehlenGF	1	1
19	LF_Waschen	1	1
20	LF_SwOfen	1	1
21	LF_AMIRAN	1	1
22	LF_VTWBAD	1	1
23	LF_ABKLAMMERN	1	1
24	ANZ_INIT_BAD_GF	10	10

25	ANZ_INIT_AMO_GF	38	38
26	ANZ_INIT_EHB_GF	12	12
27	ANZ_INIT_BAD_KF	24	24
28	ANZ_INIT_SWO_KF	30	30
29	ANZ_INIT_EHB_KF	8	8

Table 4-38: Experiment 3 – scenario 1 and scenario 2 after second variables bifurcation of model 1

In Experiment 3, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is the exact contrary of the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 3	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	94.2	36.50	152.6	56.30	58.40	19.80
Standard Error	1.68	0.35	2.91	0.38		

Table 4-39: Results of experiment 3 of model 1

The results of Experiment 3 show very clearly that the main effect is in the first part of Experiment 3, since the throughputs of Scenario 2 for both KF and GF are much larger than those of Scenario 1 and the input variable values of the second part of Scenario 2 are all kept at their low level.

In Experiment 4, the first part of Experiment 3 is further studied.

1	2	3	4	5	6	7
LIM_SKF_KF	LIM_SKF_GF	LIM_UTW_AMIRAN	LIM_UTW_SKF4	SKF_ini_KF	SKF_ini_GF	SKF_Q_KF

Table 4-40: Experiment 4 – variables of first part of experiment 2 and 3 fixed to low values

This is further bifurcated into two groups.

Part 1	1	2	3	4
	LIM_SKF_KF	LIM_SKF_GF	LIM_UTW_AMIRAN	LIM_UTW_SKF4
Part 2	5	6	7	
	SKF_ini_KF	SKF_ini_GF	SKF_Q_KF	

Table 4-41: Experiment 4 - third bifurcation of variables in the second part of experiment 3

Experiment 4 with the values of the input variables is as follows:

No.	Experiment4	Scenario 1	Scenario 2	
1	LIM_SKF_KF	3	12	
2	LIM_SKF_GF	3	7	
3	LIM_UTW_AMIRAN	3	10	
4	LIM_UTW_SKF4	4	8	
		5	SKF_ini_KF	12
		6	SKF_ini_GF	10
		7	SKF_Q_KF	10
8	MIN_SKF_KF	1	1	
9	MIN_SKF_GF	1	1	
10	KAP_SKF_GF	3	3	
11	KAP_SKF_KF	4	4	
12	MAX_Wartezeit	10	10	
13	limRaeumen	9	9	
14	v_SchwO	0.05/59	0.05/60	
15	FW_Amiran	5	6	
16	vWasche2	0.45/59	0.45/60	
17	LF_Abkuehlen	1	1	
18	LF_AbkuehlenGF	1	1	
19	LF_Waschen	1	1	
20	LF_SwOfen	1	1	
21	LF_AMIRAN	1	1	
22	LF_VTWBAD	1	1	
23	LF_ABKLAMMERN	1	1	
24	ANZ_INIT_BAD_GF	10	10	
25	ANZ_INIT_AMO_GF	38	38	
26	ANZ_INIT_EHB_GF	12	12	
27	ANZ_INIT_BAD_KF	24	24	
28	ANZ_INIT_SWO_KF	30	30	
29	ANZ_INIT_EHB_KF	8	8	

Table 4-42: Scenario 1 and scenario 2 after third variables bifurcation of model 2

In Experiment 4, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is exactly opposite to the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table

Experiment 4	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	104.1	40.50	152.4	56.10	48.30	15.60
Standard Error	1.94	0.41	4.48	0.75		

Table 4-43: Results of experiment 4 of model 2

The results of Experiment 4 show clearly that the main effect is in the first part of Experiment 4, since the throughputs of Scenario 2 for both KF and GF are much larger than those of

Scenario 1, and the input variable values of the second part of Scenario 2 are all kept at their low level.

In Experiment 5, the first part of Experiment 4 is further studied.

	1	2	3	4
Part 1	LIM_SKF_KF	LIM_SKF_GF	LIM_UTW_AMIRAN	LIM_UTW_SKF4

Table 4-44: Experiment 5 - variables of first part of experiment 2, 3 and 4 fixed to low values

It is further bifurcated into two groups.

	1	2
Part 1	LIM_SKF_KF	LIM_SKF_GF
	3	4
Part 2	LIM_UTW_AMIRAN	LIM_UTW_SKF4

Table 4-45: Experiment 5 – fourth bifurcation of variables in the second part of experiment 4

Experiment 5 with the values of the input variables is conducted as follows.

No.	Experiment5	Scenario 1	Scenario 2		
1	LIM_SKF_KF	3	12		
2	LIM_SKF_GF	3	7		
		3	LIM_UTW_AMIRAN	10	3
		4	LIM_UTW_SKF4	8	4
5	SKF_ini_KF	2	2		
6	SKF_ini_GF	2	2		
7	SKF_Q_KF	1	1		
8	MIN_SKF_KF	1	1		
9	MIN_SKF_GF	1	1		
10	KAP_SKF_GF	3	3		
11	KAP_SKF_KF	4	4		
12	MAX_Wartezeit	10	10		
13	limRaeumen	9	9		
14	v_SchwO	0.05/59	0.05/60		
15	FW_Amiran	5	6		
16	vWasche2	0.45/59	0.45/60		
17	LF_Abkuehlen	1	1		
18	LF_AbkuehlenGF	1	1		
19	LF_Waschen	1	1		
20	LF_SwOfen	1	1		
21	LF_AMIRAN	1	1		
22	LF_VTWBAD	1	1		
23	LF_ABKLAMMERN	1	1		
24	ANZ_INIT_BAD_GF	10	10		
25	ANZ_INIT_AMO_GF	38	38		
26	ANZ_INIT_EHB_GF	12	12		
27	ANZ_INIT_BAD_KF	24	24		
28	ANZ_INIT_SWO_KF	30	30		
29	ANZ_INIT_EHB_KF	8	8		

Table 4-46: Scenario 1 and scenario 2 after fourth variables bifurcation of model 2

In Experiment 5, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is the exact opposite of the first. This raises the values from the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 5	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	101.7	49.00	123.3	36.80	21.60	-12.20
Standard Error	2.17	0.56	2.65	0.58		

Table 4-47: Results of experiment 5 of model 2

Experiment 5 shows very interesting results. For the product group KG, the throughput has increased by 21.6. However, for the product group GF the throughput has decreased by 12.2. The results clearly differ from the previous experiments' results. It is not clear now in which part the main effect is. So further experiments need to be carried out to determine which input variable or variables are the main effects. These experiments would have the traditional experiment design. However, the number of variables has already decreased to four and the complexity, uncertainty and the amount of data required have all been reduced.

Now we come back to the second part of Experiment 2 to find the main effect in this part. The variables here are:

17	18	19	20	21	22	23	24	25	26	27	28	29
LF_Ab kuehlen	LF_Abku ehlenGF	LF_W aschen	LF_S wOfen	LF_A MIRAN	LF_VT WBAD	LF_ABKL AMMERN	ANZ_INIT _BAD_GF	ANZ_INIT _AMO_GF	ANZ_INIT _EHB_GF	ANZ_INIT _BAD_KF	ANZ_INIT _SWO_KF	ANZ_INIT _EHB_KF

Table 4-48: Important input variables in the second part of experiment 2 of model 2

While all the values of input variables in the first part are kept at their low level, the second part of Experiment 2 can be further bifurcated into two groups:

Part 1	17	18	19	20	21	22	23
	LF_Abkuehlen	LF_AbkuehlenGF	LF_Waschen	LF_SwOfen	LF_AMIRAN	LF_VTWBAD	LF_ABKLAMME RN
Part 2	24	25	26	27	28	29	
	ANZ_INIT_BAD_ GF	ANZ_INIT_AMO_ GF	ANZ_INIT_EHB_ GF	ANZ_INIT_BAD_ KF	ANZ_INIT_SWO_ KF	ANZ_INIT_EHB_ KF	

Table 4-49: Experiment 6 - fifth bifurcation of variables in the first part of experiment 2

Experiment 6 with the values of the input variables is conducted as follows:

No.	Experiment6	Scenario 1	Scenario 2
1	LIM_SKF_KF	3	3
2	LIM_SKF_GF	3	3
3	LIM_UTW_AMIRAN	3	3
4	LIM_UTW_SKF4	4	4
5	SKF_ini_KF	2	2
6	SKF_ini_GF	2	2
7	SKF_Q_KF	1	1
8	MIN_SKF_KF	1	1
9	MIN_SKF_GF	1	1
10	KAP_SKF_GF	3	3
11	KAP_SKF_KF	4	4
12	MAX_Wartezeit	10	10
13	limRaeumen	9	9
14	v_SchwO	0.05/60	0.05/61
15	FW_Amiran	6	7
16	vWasche2	0.45/60	0.45/61
17	LF_Abkuehlen	1	2
18	LF_AbkuehlenGF	1	4
19	LF_Waschen	1	2
20	LF_SwOfen	1	2
21	LF_AMIRAN	1	4
22	LF_VTWBAD	1	4
23	LF_ABKLAMMERN	1	2
24	ANZ_INIT_BAD_GF	10	10
25	ANZ_INIT_AMO_GF	49	38
26	ANZ_INIT_EHB_GF	12	12
27	ANZ_INIT_BAD_KF	24	24
28	ANZ_INIT_SWO_KF	40	30
29	ANZ_INIT_EHB_KF	24	16

Table 4-50: Scenario 1 and scenario 2 after fifth variables bifurcation of model 2

In Experiment 6, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is the exact contrary of the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 6	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	103.3	32.80	145.1	43.90	41.80	11.10
Standard Error	1.22	0.31	1.88	0.28		

Table 4-51: Results of experiment 6 of model 2

The results of Experiment 6 show clearly that the main effect is in the first part of Experiment 4, since the throughputs of Scenario 2 for both KF and GF are much larger than those of Scenario 1 and the input variable values of the second part of Scenario 2 are all kept at their low level.

In Experiment 7, the first part of Experiment 6 is further studied.

17	18	19	20	21	22	23
LF_Abkuehlen	LF_AbkuehlenGF	LF_Waschen	LF_SwOfen	LF_AMIRAN	LF_VTWBAD	LF_ABKLAMMERN

Table 4-52: Experiment 7 - variables of first part of experiment 2, 3, 4, 5 and 6 fixed to low values

While all the values of input variables of the first part are kept at their low level, the second part of Experiment 2 can be further bifurcated into two groups.

Part 1	17	18	19	
	LF_Abkuehlen	LF_AbkuehlenGF	LF_Waschen	
Part 2	20	21	22	23
	LF_SwOfen	LF_AMIRAN	LF_VTWBAD	LF_ABKLAMMERN

Table 4-53: Experiment 7 - sixth bifurcation of variables in the second part of experiment 6

Experiment 7 with the values of the input variables is conducted as follows:

No.	Experiment6	Scenario 1	Scenario 2
1	LIM_SKF_KF	3	3
2	LIM_SKF_GF	3	3
3	LIM_UTW_AMIRAN	3	3
4	LIM_UTW_SKF4	4	4
5	SKF_ini_KF	2	2
6	SKF_ini_GF	2	2
7	SKF_Q_KF	1	1
8	MIN_SKF_KF	1	1
9	MIN_SKF_GF	1	1
10	KAP_SKF_GF	3	3
11	KAP_SKF_KF	4	4
12	MAX_Wartezeit	10	10
13	limRaeumen	9	9
14	v_SchwO	0.05/60	0.05/61
15	FW_Amiran	6	7
16	vWasche2	0.45/60	0.45/61
		17 LF_Abkuehlen	1 2
		18 LF_AbkuehlenGF	1 4
		19 LF_Waschen	1 2
		20 LF_SwOfen	2 1
		21 LF_AMIRAN	4 1
		22 LF_VTWBAD	4 1
		23 LF_ABKLAMMERN	2 1
	24 ANZ_INIT_BAD_GF	10	10
	25 ANZ_INIT_AMO_GF	38	38
	26 ANZ_INIT_EHB_GF	12	12
	27 ANZ_INIT_BAD_KF	24	24
	28 ANZ_INIT_SWO_KF	30	30
	29 ANZ_INIT_EHB_KF	16	16

Table 4-54: Scenario 1 and scenario 2 after sixth variables bifurcation of model 2

In Experiment 7, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is exactly the contrary of the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 7	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	112.8	36.90	126.70	37.90	13.90	1.00
Standard Error	2.80	0.27	1.98	0.41		

Table 4-55: Results of experiment 7 of model 2

The results of Experiment 7 show clearly that the main effect is in the first part of Experiment 4, since the throughputs of Scenario 2 for both KF and GF are much larger than those of Scenario 1 and the input variable values of the second part of Scenario 2 are all kept at their low level.

So, in Experiment 8, the main effect is in the input variables.

	17	18	19
Part 1	LF_Abkuehlen	LF_AbkuehlenGF	LF_Waschen

Table 4-56: Experiment 8 - main effect variables

While all the values of input variables in the first part are kept at their low level, the second part of Experiment 2 can be further bifurcated into two groups

	17	18
Part 1	LF_Abkuehlen	LF_AbkuehlenGF
	19	
Part 2	LF_Waschen	

Table 4-57: Experiment 7 - seventh bifurcation of variables in the second part of experiment 2

Experiment 8 with the values of the input variables is conducted as follows.

No.	Experiment6	Scenario 1	Scenario 2		
1	LIM_SKF_KF	3	3		
2	LIM_SKF_GF	3	3		
3	LIM_UTW_AMIRAN	3	3		
4	LIM_UTW_SKF4	4	4		
5	SKF_ini_KF	2	2		
6	SKF_ini_GF	2	2		
7	SKF_Q_KF	1	1		
8	MIN_SKF_KF	1	1		
9	MIN_SKF_GF	1	1		
10	KAP_SKF_GF	3	3		
11	KAP_SKF_KF	4	4		
12	MAX_Wartezeit	10	10		
13	limRaemen	9	9		
14	v_SchwO	0.05/60	0.05/61		
15	FW_Amiran	6	7		
16	vWasche2	0.45/60	0.45/61		
		17	LF_Abkuehlen	1	2
		18	LF_AbkuehlenGF	1	4
		19	LF_Waschen	2	1
		20	LF_SwOfen	1	1
		21	LF_AMIRAN	1	1
		22	LF_VTWBAD	1	1
		23	LF_ABKLAMMERN	1	1
24	ANZ_INIT_BAD_GF	10	10		
25	ANZ_INIT_AMO_GF	38	38		
26	ANZ_INIT_EHB_GF	12	12		
27	ANZ_INIT_BAD_KF	24	24		
28	ANZ_INIT_SWO_KF	30	30		
29	ANZ_INIT_EHB_KF	16	16		

Table 4-58: Scenario 1 and scenario 2 after seventh variables bifurcation of model 2

In Experiment 8, the first scenario keeps all the values of the first part at their low value and raises the values of the second part. The second scenario is the exact opposite of the first. It raises the values of the first part to their high value and keeps those of the second part at their low value. The result of the second experiment is in the following table.

Experiment 8	Scenario 1		Scenario 2		Difference	
	KF	GF	KF	GF	KF	GF
Average	102.2	32.50	126.90	37.90		
Standard Error	2.69	0.34	1.98	0.41		

Table 4-59: Results of experiment 8 of model 2

The results show that the input variables LF_Abkuehlen and LF_AbkuehlenGF are the main effects in the second part of Experiment 2.

To sum up, the total main effects in this model are the input variables 1, 2, 3, 4 and 22, 23; and also their interactions.

The focus of model validation will be on these variables and their relationships. This means that data availability, uncertainty and the correlation between these variables are the vital aspects of the validation procedure.

The sequential bifurcation method is very useful in validating simulation models in logistics and production systems. It can identify important factors in a simulation model with a large number of parameters. Our two examples have demonstrated the use of this method in validating models.

The next step, following sequence bifurcation, would be a sensitivity analysis for each of the variables identified by the sequential bifurcation method.

The definition of a sensitivity analysis is not universal. In the previous section, the procedure of sensitivity analysis and its purposes were briefly described. In this section, it will be described in detail.

Sensitivity analysis is the systematic evaluation of the influence of changes in the model inputs on the model output. It can also be very useful as a validation method, particularly in the case of Type III data. It takes the model as a black box. It considers, for example, a very simple M/M/1 queue as:

$$\bar{p} = p(\lambda, \mu, R_0)$$

where λ is the arrival rate, μ is the service rate, R_0 is the RNG seed and $p(\cdot)$ is the functional of the simulation model. Further, assuming that $\rho = \lambda/\mu < 1$, then the steady-state mean time of delay for a customer in the queue can be calculated by:

$$d(\lambda, \mu) = \frac{\lambda}{\mu^2 - \lambda\mu}$$

and the change of the expected mean, which is the partial derivative of two-vector d and shows the sensitivity of the simulation output to very small changes in the input side, is:

$$\nabla d(\lambda, \mu) = \begin{bmatrix} \frac{\partial d}{\partial \lambda} \\ \frac{\partial d}{\partial \mu} \end{bmatrix} = \begin{bmatrix} \frac{1}{(\mu - \lambda)^2} \\ \frac{-\lambda(2\mu - \lambda)}{\mu^2(\mu - \lambda)^2} \end{bmatrix}$$

The vector is called the gradient of the expected response function; the dimensionality equals the number of input parameters, which in this case is two.

The sensitivity analysis is generally performed in the following steps:

- 1) Select the input parameters and output variables from the model. In real-life models of logistics and production systems, there are normally a large number of input parameters. So it is very important to find the most important ones which have the largest influence on the model output. Output variables, by contrast, are normally dependent on the type of system. For example, in a manufacturing system the common measures of performance of the systems in a simulation study include throughput, production cycle time, time spent in a queue, time spent in transport, WIP size, machine and personnel utilization, etc.
- 2) Determine how to generate the values for each input parameter. The method here would be to use probability density functions, trace-data or other statistical methods such as bootstrapping, jack-knifing, etc., which depend on data availability.
- 3) Run the model according to an input matrix which has different values for all selected input parameters generated by the simulation packages, and then evaluate the model output.
- 4) Evaluate the influences of input parameters on the output variables.

A great deal of sensitivity analysis is performed by changing only the values of input parameters one at a time, for example, changing the base values by $\pm 10\%$ and/or the standard deviation of an input parameter by $\pm 10\%$, while holding all other parameter values at the same value. The ratio of model outputs is measured accordingly. However, there is here no interaction between parameters – i.e. whether the value of one input parameter depends on the values of the others, which dangerously ignores the model's complexity. Factorial designs introduced by Box cannot be practically applied. In particular, the model is more realistic and large, while even the fractional factorial method is implemented.

A factorial experiment is defined as merely an adequately chosen fraction of the treatment combinations required for the complete factorial experiment selected to be run. Even the number of input variables is here small: for example, k , a full factorial design for two levels, still requires 2^n runs, which is extremely simple, since the values for each input variable are set to either plus or minus and, moreover, have equal probability. In this way, both the most important factors and all other possible combinations of input variables can be found. However, when n is large, the factorial design requires large resources of time and computational capacity. As an example, when $k = 50$, then 2^{50} runs need to be performed just for a single replicate design, and normally it will require multiple replications.

Before further discussion, some concepts should first be cleared up. In a real-life simulation project there are normally a large number of input variables. In factorial design, each input variable is called a factor. Not all factor values in a simulation model can be changed. Only those factors which can be changed are valuable to the sensitivity analysis and model validation. These factors are called the control factors.

So, in a simulation model, first, only the controllable variables are considered as factors. If the number of these factors is still too large for full factorial design, then some factors need to be

considered as having no effect on the system or no significant interactions with other factors, and ignored. By this means it is only a fraction of 2^k .

With the identification from the sequential bifurcation method, the number of experiments that need to be performed will be sharply reduced. This will lead to extra time and a focus on the essential factors affecting the model's validity.

Chapter 5 Conclusion

A simulation model can be validated first and foremost by its assumptions; it can be validated by following various validation paradigms; a simulation model can only be "truly" validated by comparing the consistencies between its output and the output of a real system.

In this dissertation, the fundamental problems of validation have been discussed, and various different model validation philosophies presented. The most widely accepted method is adopted. The predictive power of a model in relation to the future and past of a real system can be seen as the most important criterion of model validity. Since the data of the future system is impossible to obtain accurately, historical data is generally used for this purpose. For a comparison of the data from the system and the model, the first part of Chapter 4 listed twelve statistical techniques that can be applied, according to the types of data availability.

This was followed by a discussion of the methodology, in which three properties were identified for model validation in production and logistics systems: model complexity, model uncertainty and data availability. These factors should be thoroughly checked if one wishes to ensure the validity of a model. Model complexity is considered as the degree of interdependence of elements in a system and is further categorized into Type I, Type II and Type III complexity, according to correlations of different variables. Model uncertainty is divided into aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty is the intrinsic randomness dwelling in a system and cannot be reduced, while epistemic uncertainty is reducible and is caused by a lack of knowledge of the system. The last property, data availability, is categorized into three types according to the quantity of data.

We use data on hand to run a model and check its complexity. We calculate the different types of complexity and find any counter-intuitive correlations in the model. Since, in realistic models the number of factors might be too large to be proved by people, design of experiment should be used for efficient checking. Reducing the uncertainty reveals which factors are the most important and have the largest impact on model performance. By ignoring the unimportant factors, the uncertainty of the model is reduced and one's efforts can be focused on the uncertainty reduction and data collection of the important ones. We reduce epistemic uncertainty as much as possible. According to the type of data collected, the appropriate statistical methods can be applied. Since the only way to truly validate a model is to compare model output and system output when a real system already exists, if a real system does not exist, methods to prove each of the three properties of models should be used.

Statistical validation techniques are discussed according to data availability. When no system exists, the sequential bifurcation method is suggested. This can reduce model uncertainty by detecting the most important factors in a model. At the same time, it can expose the interactions between factors. In this way, a model's complexity can be revealed. This method is very useful, since simulation in logistics and production often deals with system designs without real systems or with drastic modifications of a system, and therefore where no historical data is available.

A further research focus would be on:

1. More elementary factors other than uncertainty, complexity and data availability.
2. Different grouping logic for simulation models of different kinds of production and logistics systems.
3. More sophisticated technique for deciding the scale of influence of factors so that the number of experiment runs can be reduced further.
4. How sensitivity analysis can be applied to the factors selected out by the sequential bifurcation method.

BIBLIOGRAPHY

- [AIAA98] American Institute of Aeronautics and Astronautics, Guide for the verification and validation of computational fluid dynamics simulations, AIAA-G-077-1998, Reston, VA, 1998.
- [Ande62] Anderson, T. W. On the distribution of the two-sample cramer–von Mises criterion. The Annals of Mathematical Statistics (Institute of Mathematical Statistics), 1962.
- [ASIM97] ASIM, Leitfaden für Simulationsbenutzer in Produktion und Logistik. Heft Nr. 58, 1997.
- [Balc94] Balci, O., Validation, verification, and testing techniques throughout the life cycle of a simulation study, Annals of Operations Research 53, 121-174, 1994.
- [BaNa85] Balci, O., and Nance, R. E., Formulated problem verification as an explicit requirement of model credibility," Simulation, Vol.45, No. 2, pp 75-86, 1985.
- [Bank98] Banks, J., Handbook of simulation, Wiley & Sons Inc, New York, P345 – 354, 1998.
- [BaSa84] Balci, O., Sargent, R.G., Validation of simulation models via simultaneous confidence intervals. American Journal of Mathematical and Management Sciences, 4, pp. 375–406, 1984.
- [BCNN05] Banks, J., Carson, J. S., Nelson, B. L., and Nicol, D. M., Discrete-event system simulation. 4th ed. Upper Saddle River, N.J.: Prentice-Hall, 2005.
- [BCNN96] Banks, J., Carson, J. S., Nelson, B. L. Discrete event system simulation. 2d ed. Upper Saddle River, New Jersey: Prentice-Hall, 1996.
- [BeKl96] Bettonvil, B. and Kleijnen, J.P.C., Searching for important factors in simulation models with many factors: sequential bifurcation. European Journal of Operational Research, 96, 180–194, 1996.
- [Bett90] Bettonvil, B., Detection of important factors by sequential bifurcation. Tilburg University Press, Tilburg, 1990.
- [Blau78] Blaug, M., Economic theory in retrospect. : Cambridge University Press, Cambridge, UK, 1978 [1962].
- [BIRH62] Blaug M., Richard D. Irwin, Homewood, Ill., Economic Theory in Retrospect, 1962.
- [BNAO00] Balci, O., Nance,R. E., Arthur,J. D., and Ormsby, W. F., Expanding our horizons in verification, validation, and accreditation research and practice, Proceedings of the 2002 Winter Simulation Conference, San Diego, CA, IEEE, pp. 653-663, 2002.
- [Brad00] Brade, D., Enhancing modeling and simulation accreditation by structuring verification and validation results. In: Proceeding of the 2000 Winter Simulation Conference, 2000.
- [Bonf36] Bonferroni, C. E., Teoria statistica delle classi e calcolo delle probabilità, Pubbl. d. R.

Ist. Super. di Sci. Econom. e Commerciali di Firenze (in Italian) 8: 1–62, 1936.

- [CaKA00] Campolongo, F., Kleijnen, J.P.C. and Andres, T., Screening methods. In: Sensitivity analysis. Editors: A. Saltelli, K. Chan, and E.M. Scott, Wiley, Chichester, England, 65–89, 2000.
- [Carv91] Carvalho, G.F., Evaluating computerized business simulators for objective learning validity. *Simulation & Gaming*, 22(3), 328-348, 1991.
- [DOD98] DoD Instruction 5000.61, DoD modeling and simulation (M&S) verification, validation, and accreditation (VV&A), May 13, 2003.
- [Dosi12] Example taken from Dosimis-3 Tutorial, Bernd Noche with modification, SimulationsDienstleistungsZentrum GmbH (SDZ), Dortmund, 2012.
- [EiIn38] Einstein, A., Infeld, L., *The evolution of physics*, Cambridge University Press, 1938.
- [FeCa06] Feinstein, Andrew Hale, Cannon, Hugh M., Fidelity, verifiability, and validity of simulation: constructs for evaluation. Wayne State University Marketing Department Working Paper 2001-006.
- [Frie53] Friedman, M. *The methodology of positive economics. Essays in positive economics.* University of Chicago press, Chicago, IL, 3-43, 1953.
- [Friz96] Frizelle, G., Getting the measure of complexity. *Manufacturing Engineer* 75(6): 268-270, 1996.
- [GDCV90] Goldberg, J., Dietrich, R., Chen, J., Valenzuela, T., Criss, E.. A simulation model for evaluating a set of emergency vehicle base locations development, validation, and usage. *Socio-Economic planning science* 24 125-141, 1990.
- [Glas98] Glasow, P. A., Characteristics, selection, and use of subject matter experts. Tech. Report 99W0000048, POET WG2K IV&V Paper No. 25. McLean, VA: The MITRE Corporation, 1998.
- [Goss08] Gosset, W. S., The probable error of a mean. *Biometrika* 6 (1): 1–25. March 1908.
- [Gree72] Greenberg, S., *GPSS Primer*. Wiley-Interscience, New York, 1972.
- [Hetz84] Hetzel, W., *The complete guide to software testing*, QED Information Sciences, Wellesley, MA, 1984.
- [HJOS06] Helton, J.C., Johnson, J.D., Oberkampf, W.L., Storlie, C.B., A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory, Technical Report 5557, Sandia National Laboratories, 2006.
- [HoBr01] Holton G., Brush, S.G., *Physics, the human adventure: From Copernicus to Einstein and Beyond*, 3rd ed. (Rutgers Univ. Press, New Brunswick, NJ, p.526, 2001.
- [HoSp01] Hopp, W; Spearman, M.: *Factory physics*. McGraw-Hill Higher Education, New York,

2001.

- [Kant71] Kant, Immanuel, Critique of pure reason, Meiner Verlag, Hamburg, 1998.
- [Klei75] Kleijnen, J.P.C., Statistical techniques in simulation (Vol. 2), New York: Marcel Dekker, 1975.
- [Klei98] Kleijnen, J.P.C., Validation of simulation, with and without Real Data, Discussion Paper, Tilburg University, Center for Economic Research 1998-22, Tilburg University, Center for Economic Research, 1998.
- [KIPe03] Kleijnen, J.P.C., Persson Fredrik, Screening for the Important Factors in Large Discrete-Event Simulation Models: Sequential Bifurcation and Its Applications, ssrn electronic journal · January 2003
- [KIGr92] Kleijnen, J., Groenendaal, W.J., Simulation: a statistical perspective, Wiley & Sons, New York, P207 - 212, 1992.
- [LaKe00] Law, A.M, Kelton W.D., Simulation modeling and analysis, McGraw-Hill, 2000.
- [LaMc97] Law, A.M., McComas, M. G., Simulation of manufacturing systems, Proceedings of the 1997 Winter Simulation Conference, 1997.
- [Law00] Law, A., Strategic directions in verification, validation, and accreditation research, Proceedings of the 2000 Winter Simulation Conference, 2000.
- [LiSt08] Little, J.D.C., Stephen C. G., Chapter 5 in: Building Intuition: Insights from basic operations management models and principles, Dilip Chhajed and Timothy J. Lowe, eds. Springer Science+Business Media, LLC, New York, 2008.
- [MMDS03] Moore, D. S., McCabe, G. P., Duckworth, W. M. and Sclove, S. L., The practice of business statistics, W. H. Freeman, New York, 2003.
- [NaFi67] Naylor, T. H., Finger, J. M., Verification of computer simulation models, management science, Vol. 2, pp. B92– B101, 1967.
- [Nanc94] Nance, R.E., The conical methodology and the evolution of simulation model development, annals of operation research, Vol.53, pp. 1-45, 1994.
- [NaWi10] Richard E., Nance, Wilson, J. R., A brief history of simulation revisited, David Goldsman, Proceedings of the 2010 Winter Simulation Conference, 2010.
- [Noch98] Noche, B., Simulation in Produktion und Materialfluß: Entscheidungsorientierte Simulationsumgebung, 1998.
- [NoWe91] Noche, B, Wenzel, S., Marktspiegel Simulationstechnik in Produktion und Logistik. Köln: TÜV Rheinland, 1991.
- [Ober01] Oberkampf, W.L., et, al. Challenge problems: uncertainty in system response given uncertain parameters, 2001.

- [OuUn86] Ould, M. A., Unwin, C., Testing in software development, Cambridge University Press, Great Britain, 1986.
- [Pace98] Pace, D. K., Verification, validation, and accreditation. In Applied Modeling and Simulation: An Integrated Approach to Development and Operation. ed. D. J. Cloud and L. B. Rainey, 369-410. New York, NY: McGraw-Hill, 1998.
- [Pace99] Pace, D. K., Development and documentation of a simulation conceptual model. Proceedings of the 798 robinson 1999 fall simulation interoperability workshop, 1999.
- [Paul89] Paul, R. J., Visual simulation: seeing is believing?, in R. Sharda, B.L. Golden, E. Wasil, O. Balci, & W. Stewart (Eds.), Impacts of recent computer advances on operations research, pp. 422–432, Elsevier, NY, 1989.
- [Pegd86] Pegden, C. D., Introduction to SIMAN, Proceedings of the 18th conference on Winter simulation, p.95-103, Washington, D.C., United States, December 08-10, 1986.
- [RaSW08] Rabe M., Spieckermann S., Wenzel S., Verifikation und Validierung für die Simulation in Produktion und Logistik - Vorgehensmodelle und Techniken (in German). Springer-Verlag, Berlin, 2008.
- [Reic51] Reichenbach, H., The rise of scientific philosophy, University of California Press, Berkeley, pp. 73-74, 1951.
- [Robb35] Robbins, L., An essay on the nature and significance of economic science, Mac-millan, London, 1935.
- [Robi04] Robinson, S. Simulation: The Practice of Model Development and Use. Wiley, Chichester, UK. 2004.
- [Russ12] Russell, B., The problems of philosophy. London, Williams and Norgate. 1912.
- [Sarg03] Sargent, R. Verification and validation of simulation models, Proceedings of the 2003 Winter Simulation Conference. 2003.
- [Sarg10] Sargent, R. G., A new statistical procedure for validation of simulation and stochastic models. Technical Report SYR-EECS-2010-06, Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, New York. 2010.
- [Sche59] Scheffé, H. The analysis of variance. Wiley, New York. 1959.
- [SGKLMY00] Sargent, R. G., Glasow, P. A., Kleijnen, J. P. C., Law, A. M., McGregor, I., and Youngblood, S., Strategic directions in verification, validation, and accreditation research. In Proceedings of the 2000 Winter Simulation Conference, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 909-916. Piscataway, New Jersey: IEEE, 2000.
- [Shan75] Shannon, R. E., Systems simulation: the art and the science, Prentice-Hall, 1975.
- [Shan48] Shannon, C. E., A mathematical theory of communication, Bell Systems Tech J, 1948.
- [Stac73] Stachowiak, H., Allgemeine Modelltheorie. Springer-Verlag, Wien and New York, 1973.

- [Stan08] Stanford Encyclopedia of Philosophy, Rationalism vs. Empiricism, First published Thu Aug 19, 2004; substantive revision Wed Aug 6, 2008.
- [Thei61] Theil, H., Economic forecasts and policy, North-Amsterdam Holland, 1961.
- [Toch63] Tocher, K. D., The art of simulation, The English Universities Press Ltd, London, 1963.
- [Tur48] Turing, A., Machine intelligence, in Copeland, B. Jack, The Essential Turing: The ideas that gave birth to the computer age, Oxford: Oxford University Press, 1948.
- [UIPr88] Ulrich, Probst, Anleitung zum ganzheitliche Denken und Handeln. Bern/Stuttgart, Haupt, 1988.
- [UrEI06] Urbanic, R. J., ElMaraghy, W.H., Modeling of manufacturing process complexity vol. VII, 2006.
- [VDI36] VDI 3633 Blatt 1: Simulation von Logistik-, Materialfluss- und Produktionssystemen, Grundlagen (Gründruck). 2010.
- [Wall03] Walley, P., Designing the accident and emergency system: lessons from manufacturing, UK, 2003.
- [WeWa48] Weaver, W., "Science and complexity". American Scientist 36 (4): 536–44. PMID 18882675. Retrieved 2007-11-21, 1948.
- [WoPi81] Woolley, R.N., Pidd M., Problem structuring – A literature review. Journal of the Operational Research Society 32: 197-206, 1981.
- [WuHa00] Wu, C. F. Jeff and Hamada, Michael, Experiments: Planning, analysis, and parameter design optimization, New York: Wiley, ISBN 0-471-25511-4. 2000.
- [YuEf06] Yu, S. B. and Efstathiou, J., An introduction of network complexity, University of Oxford. 2006.
- [Zeig76] Zeigler, B.P., Theory of modeling and simulation. Wiley, New York, 1976.