

Frank Wagner

Ein integriertes Verfahren zur robusten statistischen Auswertung von Simulationsergebnissen in der Produktion

In dieser »Schriftenreihe zu Arbeitswissenschaft und Technologiemanagement« werden die Dissertationen, die im Rahmen von Forschungs- und Entwicklungsarbeiten im Institut für Arbeitswissenschaft und Technologiemanagement IAT der Universität Stuttgart und im Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO entstanden sind, veröffentlicht.

An beiden Instituten wird universitäre Grundlagenforschung mit angewandter Auftragsforschung verknüpft und erfolgreich in zahlreichen Projekten praxisgerecht umgesetzt.

Technologiemanagement umfasst dabei die integrierte Planung, Gestaltung, Optimierung, Bewertung und den Einsatz von technischen Produkten und Prozessen aus der Perspektive von Mensch, Organisation, Technik und Umwelt. Dabei werden neue anthropozentrische Konzepte für die Arbeitsorganisation und -gestaltung erforscht und erprobt. Die Arbeitswissenschaft mit ihrer Systematik der Analyse, Ordnung und Gestaltung der technischen, organisatorischen und sozialen Bedingungen von Arbeitsprozessen sowie ihren humanen und wirtschaftlichen Zielen ist dabei zentral in die Aufgabe des Technologiemanagements eingebunden.

ISBN 978-3-8396-0509-7



ISSN 2195-3414

Fraunhofer Verlag



SCHRIFTENREIHE ZU ARBEITSWISSENSCHAFT UND TECHNOLOGIEMANAGEMENT

Herausgeber

Univ.-Prof. Dr.-Ing. Dr.-Ing. E.h. Dieter Spath

Univ.-Prof. Dr.-Ing. habil. Prof. e. h. mult. Dr. h. c. mult. Hans-Jörg Bullinger

Institut für Arbeitswissenschaft und Technologiemanagement IAT
der Universität Stuttgart, Stuttgart

Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart

Band 2

Frank Wagner

Ein integriertes Verfahren zur robusten statistischen Auswertung
von Simulationsergebnissen in der Produktion

Impressum

Kontaktadresse:

*Institut für Arbeitswissenschaft
und Technologiemanagement IAT
der Universität Stuttgart und
Fraunhofer-Institut für Arbeitswirtschaft
und Organisation IAO
Nobelstraße 12, 70569 Stuttgart
Telefon +49 711 970-01, Fax -2299
www.iat.uni-stuttgart.de
www.iao.fraunhofer.de*

*Schriftenreihe zu Arbeitswissenschaft
und Technologiemanagement*

Herausgeber:

*Univ. Prof. Dr.-Ing. E.h. Dieter Spath
Univ. Prof. Dr.-Ing. habil. Prof. e.h. mult.
Dr. h.c. mult. Hans-Jörg Bullinger*

*Institut für Arbeitswissenschaft
und Technologiemanagement IAT
der Universität Stuttgart und
Fraunhofer-Institut für Arbeitswirtschaft
und Organisation IAO*

*Bibliografische Information der
Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet
diese Publikation in der Deutschen National-
bibliografie; detaillierte bibliografische Daten sind
im Internet über www.dnb.de abrufbar.*

*ISSN 2195-3414
ISBN 978-3-8396-0509-7*

D 93

Zugl.: Stuttgart, Univ., Diss., 2012

*Druck und Weiterverarbeitung:
IRB Mediendienstleistungen
Fraunhofer-Informationszentrum
Raum und Bau IRB, Stuttgart*

*Für den Druck des Buches wurde chlor-
und säurefreies Papier verwendet.*

*© by FRAUNHOFER VERLAG, 2013
Fraunhofer-Informationszentrum
Raum und Bau IRB
Postfach 800469, 70504 Stuttgart
Nobelstraße 12, 70569 Stuttgart
Telefon +49 711 970-2500, Fax -2508
E-Mail verlag@fraunhofer.de
<http://verlag.fraunhofer.de>*

Alle Rechte vorbehalten

*Dieses Werk ist einschließlich aller seiner Teile ur-
heberrechtlich geschützt. Jede Verwertung, die über
die engen Grenzen des Urheberrechtsgesetzes hin-
ausgeht, ist ohne schriftliche Zustimmung des Verlages
unzulässig und strafbar. Dies gilt insbesondere für
Vervielfältigungen, Übersetzungen, Mikroverfilmungen
sowie die Speicherung in elektronischen Systemen.
Die Wiedergabe von Warenbezeichnungen und Han-
delsnamen in diesem Buch berechtigt nicht zu der
Annahme, dass solche Bezeichnungen im Sinne der
Warenzeichen- und Markenschutz-Gesetzgebung als
frei zu betrachten wären und deshalb von jedermann
benutzt werden dürften. Soweit in diesem Werk
direkt oder indirekt auf Gesetze, Vorschriften oder
Richtlinien (z.B. DIN, VDI) Bezug genommen oder aus
ihnen zitiert worden ist, kann der Verlag keine Ge-
währ für Richtigkeit, Vollständigkeit oder Aktualität
übernehmen.*

Geleitwort

Grundlage der Arbeiten am Institut für Arbeitswissenschaft und Technologiemanagement IAT der Universität Stuttgart und am kooperierenden Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO ist die Überzeugung, dass unternehmerischer Erfolg in Zeiten globalen Wettbewerbs vor allem bedeutet, neue technologische Potenziale nutzbringend einzusetzen. Deren erfolgreicher Einsatz wird vor allem durch die Fähigkeit bestimmt, kunden- und mitarbeiterorientiert Technologien schneller als die Mitbewerber zu entwickeln und anzuwenden. Dabei müssen gleichzeitig innovative und anthropozentrische Konzepte der Arbeitsorganisation zum Einsatz kommen. Die systematische Gestaltung wird also erst durch die Bündelung von Management- und Technologiekompetenz ermöglicht. Dabei wird durch eine ganzheitliche Betrachtung der Forschungs- und Entwicklungsthemen gewährleistet, dass wirtschaftlicher Erfolg, Mitarbeiterinteressen und gesellschaftliche Auswirkungen immer gleichwertig berücksichtigt werden.

Die im Rahmen der Forschungsarbeiten in den Instituten entstandenen Dissertationen werden in der »Schriftenreihe zu Arbeitswissenschaft und Technologiemanagement« veröffentlicht. Die Schriftenreihe ersetzt die Reihe »IPA-IAO Forschung und Praxis«, herausgegeben von H. J. Warnecke, H.-J. Bullinger, E. Westkämper und D. Spath. In dieser Reihe sind in den vergangenen Jahren über 500 Dissertationen erschienen. Die Herausgeber wünschen den Autoren, dass ihre Dissertationen aus den Bereichen Arbeitswissenschaft und Technologiemanagement in der breiten Fachwelt als wichtige und maßgebliche Beiträge wahrgenommen werden und so den Wissensstand auf ein neues Niveau heben.



Dieter Spath



Hans-Jörg Bullinger

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als Wissenschaftler am Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO in Stuttgart.

Mein herzlicher Dank gilt Herrn Prof. Dr.-Ing. Dr.-Ing. E.h. Dieter Spath, Leiter des Instituts für Arbeitswissenschaft und Technologie-management IAT der Universität Stuttgart und des Fraunhofer IAO, für die wissenschaftliche Unterstützung und die wohlwollende Förderung dieser Arbeit.

Herrn Prof. Dr.-Ing. Prof. E.h. Dr.-Ing. E.h. Dr. h.c. mult. Engelbert Westkämper, ehemaliger Leiter des Instituts für Industrielle Fertigung und Fabrikbetriebslehre IFF der Universität Stuttgart und des Fraunhofer-Instituts für Produktionstechnik und Automatisierung IPA in Stuttgart, danke ich für das Interesse an der Arbeit und der Übernahme des Mitberichts.

Meinen Kolleginnen und Kollegen am Fraunhofer IAO gilt mein Dank für die Zusammenarbeit sowie für die Motivation und Unterstützung beim Erstellen der Arbeit. Insbesondere bedanke ich mich bei Herrn Prof. Dr.-Ing. habil. Joachim Warschat, Dr. Sven Schimpf und Jacoba Haug für die kritische Durchsicht der Arbeit. Ein wichtiger Dank geht an die studentischen Mitarbeiterinnen und Mitarbeiter Jutta Grammel, Wolfgang Imig, Andreas Groß sowie besonders an Alina Zaretser.

Vor allem danke ich meiner lieben Rita für Motivation und Geduld während all der gemeinsamen Jahre.

Gerlingen im September 2012
Frank Wagner

Ein integriertes Verfahren zur robusten statistischen Auswertung von Simulationsergebnissen in der Produktion

Von der Fakultät Konstruktions-, Produktions- und Fahrzeugtechnik
der Universität Stuttgart
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung

Vorgelegt von
Dipl.-Ing. Frank Wagner
aus Stuttgart-Bad Cannstatt

Hauptberichter: Prof. Dr.-Ing. Dr.-Ing. E.h. Dieter Spath
Mitberichter: Prof. Dr.-Ing. Prof. E.h. Dr.-Ing. E.h. Dr.h.c. mult.
Engelbert Westkämper

Tag der mündlichen Prüfung: 5. September 2012

Institut für Arbeitswissenschaft und Technologiemanagement IAT
der Universität Stuttgart

Abbildungsverzeichnis	6
Tabellenverzeichnis	8
Verzeichnis der verwendeten Abkürzungen und Formelzeichen	9
1 Einleitung	13
1.1 Problemstellung.....	13
1.2 Notwendigkeit der statistischen Auswertung von Simulationsdaten ...	14
1.3 Auswahl geeigneter statistischer Verfahren	15
1.4 Abgrenzung der Problematik.....	17
1.5 Ziel und Aufbau der Arbeit.....	18
2 Stand der Technik	22
2.1 Analyse des Stands der Technik.....	22
2.2 Begriffsbestimmungen.....	22
2.3 Produktionssimulation, Digitale Fabrik und Digitale Produktion	24
2.4 Simulationstechnik im industriellen Einsatz.....	26
2.5 Statistische Auswertung und Interpretation der Ergebnisse.....	27
2.6 Ziele und Aufgaben der statistische Auswertung	28
2.6.1 Bestimmung der Stationarität.....	30
2.6.2 Autokorrelation und Unabhängigkeit der Stichproben.....	31
2.6.3 Schätzen von Kennzahlen mit Güteindikator	31
2.6.4 Vergleich von Systemalternativen.....	31
2.7 Stand der Forschung.....	32
2.7.1 Verfahren für Teilaufgaben	32
2.7.2 Auswertungsmethodiken.....	36
2.7.3 Rechnerunterstützte Auswertungswerkzeuge.....	43
2.8 Statistische Auswertung bei kommerziellen Werkzeugen	46
2.8.1 Simulationssysteme	48
2.8.2 Statistikwerkzeuge	51
2.9 Stand der Praxis	53
2.10 Ableitung des Anforderungsprofils	55
2.10.1 Robustheit	55
2.10.2 Plausibilität	56
2.10.3 Integration	56
2.10.4 Anforderungsprofil.....	57

2.11	Zusammenfassung der Bewertung des Stands der Technik.....	57
3	Konzeption des integrierten Verfahrens	60
3.1	Integriertes Vorgehensmodell	60
3.2	Trennung von Modell und Experimentierumgebung	61
3.3	Einbindung in die Simulationsmethodik.....	62
3.3.1	Modellierung der Eingangsgrößen	64
3.3.2	Verifikation eines Modells	64
3.3.3	Validierung der Simulation	64
3.3.4	Versuchsplanung	65
3.3.5	Optimierung.....	65
3.4	Einbindung des Anwenders.....	65
3.4.1	Visualisierung.....	66
3.4.2	Entscheidungskompetenz	66
3.4.3	Sensitivitätsanalyse.....	66
3.5	Konzeption des integrierten Auswertungsprozesses mit Teilaufgaben	67
3.5.1	Schätzung der Leistungskennzahl eines Modells mit Qualitätsangabe.....	67
3.5.2	Vergleich von Systemalternativen.....	71
3.6	Auswahl und Integration der Verfahren.....	75
3.6.1	Robustheit	75
3.6.2	Einfachheit und Plausibilität	76
3.6.3	Vorgehensweise zur Auswahl und Integration der Verfahren	77
3.6.4	Monte-Carlo-Verfahren	77
3.6.5	Die Monte-Carlo-Studien zur Evaluation der Verfahren.....	78
3.7	Konzept des Auswertungsprozesses	80
3.8	Sonstige Aufgaben bei der Auswertung	82
4	Vorgehensweise zur Auswertung eines Modells.....	84
4.1	Zielsetzung und Rahmenbedingungen.....	84
4.2	Bestimmung der stationären Phase	84
4.3	Eliminierung der Autokorrelation der Stichproben mit Batch Means ...	86
4.4	Schätzung einer Kennzahl und Bewertung der Ergebnisqualität	88
4.4.1	Lageparameter der Verteilung	89
4.4.2	Anzahl und Verteilung der Stichproben.....	89

4.4.3	Beurteilung und Klassifizierung der Verteilungsform	90
4.4.4	Geringe Abweichung von der Normalverteilung:.....	96
4.4.5	Symmetrische Abweichungen von der Normalverteilung.....	96
4.4.6	Schiefe Abweichungen von der Normalverteilung	97
4.5	Integration der Vorgehensweise zur Auswertung eines Modells.....	99
4.6	Zusammenfassung	102
5 Vorgehensweise zum Vergleich von mehreren Simulationsmodellen		103
5.1	Zielsetzung und Rahmenbedingungen.....	103
5.2	Auswahl der geeigneten Verfahren der Mehrfachen Mittelwertvergleiche	104
5.2.1	Anzahl der Stichproben	104
5.2.2	Varianzunterschiede	105
5.2.3	Abweichung von der Normalverteilung	106
5.3	Empfohlene Verfahren der Multiplen Mittelwertvergleiche	107
5.3.1	Vereinfachte Darstellung und Auswahl	110
5.3.2	Sensitivitätsanalyse und weitere Beobachtungen.....	111
5.4	Integration der Vorgehensweise zum Vergleich mehrerer Simulationsmodelle	111
5.5	Zusammenfassung	114
6 Anwendung des integrierten Verfahrens		115
6.1	Anwendung des integrierten Verfahrens im industriellen Umfeld.....	115
6.2	Beschreibung der Anwendungsszenarien	115
6.3	Simulations- und Auswertungsumgebung.....	116
6.4	Simulation einer Produktion für Fertigbauteile	117
6.4.1	Aufgabenstellung	117
6.4.2	Beschreibung der Anwendung des integrierten Verfahrens.....	118
6.4.3	Bewertung	123
6.5	Simulation einer Schuhproduktion.....	124
6.5.1	Aufgabenstellung	124
6.5.2	Beschreibung der Anwendung des integrierten Verfahrens.....	125
6.5.3	Bewertung	129
6.6	Zusammenfassung	129
7 Evaluation und Diskussion des integrierten Verfahrens		131

7.1	Evaluation.....	131
7.1.1	Robustheit	131
7.1.2	Plausibilität	131
7.1.3	Integration	132
7.1.4	Allgemeine Anforderungen.....	132
7.1.5	Zusammenfassung der Evaluation.....	132
7.2	Diskussion	133
7.2.1	Aufwand	133
7.2.2	Effizienz versus Plausibilität und Robustheit.....	134
7.2.3	Anwendungsbereich und Übertragbarkeit.....	134
7.2.4	Erweiterbarkeit	135
8	Ausblick.....	137
9	Zusammenfassung	139
10	Abstract	141
11	Literaturverzeichnis.....	143
12	Anhang.....	155
12.1	Methoden.....	155
12.2	Simulationsergebnisse	165

Abbildungsverzeichnis

Abbildung 1.1: Typische Form einer Verteilungsdichte für Beobachtungen aus der Simulation eines Produktionssystems wie bei [WaWa94:632] und [Klei87:21].	16
Abbildung 1.2: Vorgehensweise bei Simulationsprojekten (vgl. [VDI93] [VDI00], [WaWa96:37], [BDAS86:200] und [Law07:67] mit einer Abgrenzung der Arbeit.	18
Abbildung 1.3: Problematik, Handlungsbedarf und Lösungsansatz der Arbeit.	19
Abbildung 1.4: Überblick über den Aufbau und die Vorgehensweise der Arbeit.	20
Abbildung 2.1: Das Betrachtungsfeld der „Digitalen Fabrik“ nach VDI Richtlinie 4499 [VDIF09].	25
Abbildung 2.2: Der Gegenstandsbereich der „Digitalen Produktion“ nach Spath [Spat06].	26
Abbildung 2.3: Vereinfachter Ablauf der statistischen Auswertung nach der VDI Richtlinie [VDIF95].	30
Abbildung 2.4: Anforderungsprofil an das integrierte Verfahren.	57
Abbildung 3.1: Weiterer Aufbau der Arbeit.	61
Abbildung 3.2: Modell und Experimentierumgebung.	62
Abbildung 3.3: Einbindung der Auswertung in die Simulationsmethodik nach VDI Richtlinie 3633 Blatt 1 [VDIF00].	63
Abbildung 3.4: UML Diagramm der Abläufe und Teilaufgaben des Auswertungsprozesses eines Modells.	68
Abbildung 3.5: Grafische Darstellungsformen zur Unterstützung der Auswertung (Zeitreihe, Histogramm, Dichtefunktion, Box-Plot).	70
Abbildung 3.6: UML Diagramm des Gesamtablaufs der integrierten Vorgehensweise zur Auswertung.	82
Abbildung 4.1: UML Diagramm der Vorgehensweise für die Bestimmung der stationären Phase.	86
Abbildung 4.2: UML Diagramm der Vorgehensweise für die Beseitigung der Autokorrelation mit dem Batch Mean Verfahren.	88
Abbildung 4.3: Histogramm und Dichteschätzung einer empirischen Verteilungsfunktion mit einer Normalverteilung als Referenz.	92
Abbildung 4.4: Box-Plots einer empirischen Verteilungsfunktion mit einer Normalverteilung als Referenz.	92
Abbildung 5.1: Geeignete Verfahren der mehrfachen Mittelwertvergleiche in Abhängigkeit von den drei Kriterien.	109

Abbildung 5.2: Geeignete Verfahren der mehrfachen Mittelwertvergleiche aufgeteilt nach Art der Abweichung von der Normalverteilung.....	109
Abbildung 5.3: Vereinfachte Auswahl geeigneter Verfahren der mehrfachen Mittelwertvergleiche.	110
Abbildung 5.4: UML Diagramm der Vorgehensweise zum Vergleich von mehreren Simulationsalternativen.	114
Abbildung 6.1: Konzept der mobilen Fabrik (aus [Manu08]).....	117
Abbildung 6.2: CAD-Skizze des Service Cores (aus [Manu08]).....	118
Abbildung 6.3: Simulationsmodell der Montage der „Service Core“ Komponente in der mobilen Fabrik (aus [Manu08]).....	119
Abbildung 6.4: Dichten und Box-Plots Verteilung der Durchlaufzeiten der simulierten Systemalternative (bei 20 Stichproben).	120
Abbildung 6.5: Dichten und Box-Plots Verteilung der Durchlaufzeiten der simulierten Systemalternative (bei 40 Stichproben).	122
Abbildung 6.6: Ausschnitt aus der integrierten Schuhfabrik in Vigevano (aus [CeC06]).....	124
Abbildung 6.7: Simulationsmodell der integrierten Schuhfabrik in Vigevano (aus [CeC06]).....	125
Abbildung 6.8: Geschätzte Dichten der Verteilung der Durchsatzes der vier Systemalternativen bei vierzig Stichproben.....	127
Abbildung 6.9: Geschätzte Dichten und Box-Plots der Verteilung der Durchsatzes der vier Systemalternativen bei vierzig Stichproben.....	127
Abbildung 12.1: Venn-Diagramm zur Bonferroni-Ungleichung für „vereinbare“ Ereignisse.....	156
Abbildung 12.2 : UML-Diagramm der Welch-Methode zur Bestimmung der stationären Phase.....	157
Abbildung 12.3: UML-Diagramm der Fischer-Prozedur zur Bestimmung der stationären Phase, vor allem für ein Anlaufen aus einem leeren Zustand.....	158
Abbildung 12.4: UML-Diagramm des Vassiliacopoulos-Tests zur Bestimmung der stationären Phase.....	161

Tabellenverzeichnis

Tabelle 2.1: Zusammenfassung der Bewertung der untersuchten Verfahren.....	59
Tabelle 3.1: Vergleich der Verfahren der Multiplen Mittelwertvergleiche mit Ranking & Selection Verfahren.....	74
Tabelle 4.1: Klassifikation der Abweichung von der Normalverteilung.....	95
Tabelle 5.1: Qualitative Klassifikation der Stichproben.....	104
Tabelle 5.2: Qualitative Klassifikation der Stichproben.....	106
Tabelle 6.1: Stationäre Eigenschaften der Zeitreihe der Durchlaufzeiten (bei 20 Stichproben).....	120
Tabelle 6.2: Mittelwerte der durchschnittlichen Durchlaufzeit mit Konfidenzintervall auf Basis der Verteilungsform bei 20 Stichproben (alternative inkorrekte Klassifizierung in Klammern).....	121
Tabelle 6.3: Stationäre Eigenschaften der Zeitreihe der Durchlaufzeiten (bei 40 Stichproben).....	122
Tabelle 6.4: Mittelwerte der durchschnittlichen Durchlaufzeit mit Konfidenzintervall auf Basis der Verteilungsform bei 40 Stichproben (alternative inkorrekte Klassifizierung in Klammern).....	123
Tabelle 6.5: Kennzahlen der vier Systemalternativen aus vierzig Stichproben.....	126
Tabelle 6.6: Testergebnisse der sechs Mittelwertvergleiche: p-Werte des Wilcoxon Tests.....	128
Tabelle 6.7: Testergebnisse der sechs Mittelwertvergleiche: p-Werte des Zweistichproben-t-Tests.....	128
Tabelle 7.1: Zusammenfassung der relativen Erfüllung der Anforderungen beider industriellen Anwendungen.....	133

Verzeichnis der verwendeten Abkürzungen und Formelzeichen

A_i	stochastisches Ereignis
Abb.	Abbildung
ANOVA	Analysis of Variance
ASIM	Arbeitsgemeinschaft Simulation der Gesellschaft für Informatik
B	Bonferonni-Ungleichung
Br	Breite eines Konfidenzintervalls
B_j	Batch Mean, d.h. Mittelwerte des j Batches
b	Anzahl der Batches beim Batch Mean Verfahren
bspw.	beispielsweise
C	C-Test nach Dunnett
c	Testgrössenvariable beim Vassiliacopoulos-Test
C_N	Testprüfgröße
CAD	Computer aided Design
CRN	Common random numbers
d^*	signifikanter Unterscheid bei R&S-Verfahren
d.h.	das heißt
DIN	deutsches Institut für Normung e.V.
DoE	Design of Experiments
e_{ij}	stochastische Störvariable in der Varianzanalyse
H_0	Nullhypothese
$E(X)$	Erwartungswert einer Zufallsvariablen
FEM	Finite-Elemente-Methode
FIFO	First In - First Out
GH	GH-Test nach Games und Howell
ISO	international organization for standardization
ISSTAS	Integriertes System zur Statistischen Auswertung von Simulationsdaten
k	Anzahl der Systemalternativen, Anzahl der Faktorstufen in der Varianzanalyse

Kap.	Kapitel
ln	natürlicher Logarithmus zur Basis der Eulerschen Zahl e
m	Größe der Batches beim Batch Mean Verfahren
MCB	Multiple Comparison with the Best
M/M/1	Kendall-Notation eines Wartesystems mit einer Servicestation und exponentiell verteilten Ankunfts- und Bedienzeiten
MMV	Mehrfache (Multiple) Mittelwertvergleiche
(N)	Normalverteilung
N(0,1)	standardisierte Normalverteilung
NAG	Numerical Algorithms Group
NID	Normally and Independent Distributed.
Pr(X)	Wahrscheinlichkeit des zufälligen Ereignisses X
$q'_{v,k,1-\alpha}$	Schranke der Studentized Augmented Range Verteilung beim Games-Howell-Test GH
r_i	Rangzahl mit Laufindex i
R	Raum der reellen Zahlen
R+	Raum der positiven reellen Zahlen
R&S	Ranking and Selection
RTD	Research and Technology Development
s	Schätzung der empirischen Standardabweichung, auch Rangzahl
s^2	Schätzung der empirischen Varianz
SAR	Studentized Augmented Range
s.o.	siehe oben
SR	Studentized Range
s.u.	siehe unten
S	Šidák-Ungleichung
t_{n-1}^α	α Quantil der t-Verteilung mit $n - 1$ Freiheitsgraden
\tilde{t}	Johnsons modifizierte t-Statistik
t	Zweistichproben-t-Test
T3	T3-Test nach Dunnett

T10	10%-getrimmte Schätzer
Tu	Tukey-Test
$\{U_N(k)\}$	auf Rängen basierender stochastischer Prozess beim Vassiliacopoulos-Test
u.a.	unter anderem; und andere(s)
UML	Unified Modelling Language
v	Variationskoeffizient
$Var(X)$	Varianz einer Zufallsvariablen, auch als σ^2 beschrieben
VDI	Verein Deutscher Ingenieure e.V.
W	Wilcoxon-Rangsummentest
W_k	Rangsumme der k Ränge
\bar{w}	mittlere Wartezeit bei Warteschlangensystemen
X	stochastische Merkmal, Kennzahl einer Simulation, Zufallsvariable
X_F	Abschneidegrenze bei der Bestimmung der stationären Phase nach Fischer
X_G	Abschneidegrenze bei der Bestimmung der stationären Phase einer Zeitreihe
X_W	Abschneidegrenze bei der Bestimmung der stationären Phase nach Welch
$x_{0,25}$	unteres Quartil
$x_{0,75}$	oberes Quartil
x_i	Beobachtung oder Stichprobe des stochastischen Merkmals X mit Laufindex i
$x_{(i)}$	geordnete Beobachtung mit Rang i
\bar{x}	arithmetischer Mittelwert
\tilde{x}	Median
x_{mod}	Modus, Modalwert
$\{x_k\}$	stochastischer Prozess
Y	stochastisches Merkmal, Kennzahl einer Simulation, Zufallsvariable
\bar{Y}_i	arithmetischer Mittelwert von Y_i

y_i	Beobachtung oder Stichprobe des stochastischen Merkmals Y mit Laufindex i
z^α	α Quantil der Standardnormalverteilung-Verteilung
z.B.	zum Beispiel
α	Signifikanzniveau, Fehlerrate
α_G	gesamtes Signifikanzniveau, Gesamtfehlerrate
α_k	individuelles Signifikanzniveau, Individuelle Fehlerrate
α_i	Effekt der i -ten Faktorstufe in der Varianzanalyse
$\hat{\alpha}$	Signifikanzniveau beim Vassiliacopoulos-Test
β	Trimm- oder Windsorfaktor
γ_1	Schiefe einer Verteilung
γ_2	Exzess, Kurtosis, Wölbung einer Verteilung
λ	Parameter der Exponentialverteilung der Ankunftsrate bei Warteschlangensystemen
μ	Erwartungswert einer Zufallsvariablen, auch als $E(X)$ bezeichnet
μ_3	drittes zentrales Moment
μ_4	viertes zentrales Moment
μ_i	Mittelwert der Systemalternative i
$\hat{\mu}_3$	Schätzung der Schiefe
ν	Anzahl der Freiheitsgrade
ρ	Verkehrsdichte bei Warteschlangensystemen
σ^2	Varianz einer Zufallsvariablen, auch als $Var(x)$ bezeichnet
ω	Parameter der Exponentialverteilung der Servicezeiten bei Warteschlangensystemen
χ^2	Chi-Quadrat (-Verteilung oder -Test)

1 Einleitung

1.1 Problemstellung

Die Kompetenz eines Unternehmens, seine Produkte und Dienstleistungen schnell und kostengünstig auf internationalen Märkten anbieten sowie flexibel auf veränderte Anforderungen reagieren zu können, ist ein wesentlicher und entscheidender Erfolgsfaktor. Wachsende Herausforderungen wie erhöhte Komplexität bei Produkten, Dienstleistungen und Prozessen, zunehmender Entwicklungs- und Produktionsaufwand sowie globale und dynamische Absatzmärkte erfordern leistungsfähige Methoden und Werkzeuge für die Planung, Realisierung und den Betrieb von Produktionssystemen. Nur wem es gelingt, Kundenanforderungen schnell und effizient in marktfähige Produkte zu überführen, kann sich Wettbewerbsvorteile sichern [SpWS10:4].

Die Simulationstechnik als eine Schlüsseltechnologie zur frühen und schnellen Planung, Entwicklung und Bewertung von Produkten, Dienstleistungen und Produktionsprozessen ist dabei von entscheidender Bedeutung. Die Abbildung von Wirkzusammenhängen in Technik und Organisation auf wissenschaftlich gesicherter Basis in Modellen kann den Aufwand erheblich reduzieren [West10:119].

Moderne Konzepte des Produktionsmanagements wie „Digitale Fabrik“ [West03:788] und „Digitale Produktion“ [Spat06:543], [Spat08] basieren auf einer virtuellen und simulierten Abbildung der Produktionssysteme und sind „Leitthemen“ der Produktionstechnik [West09:91].

Die Simulation von Produktionssystemen ist ein wichtiges Werkzeug für den gesamten Lebenszyklus von Planung, Realisierung, Betrieb sowie Neugestaltung und damit die Grundlage für weitreichende Entscheidungen. Durch eine Simulation des Produktionsablaufs kann die Produktion schon in einer frühen Planungsphase lange vor dem eigentlichen Anlauf abgesichert und, wenn notwendig, verbessert werden [SpLe08].

Mit der Verfügbarkeit komfortabler Simulationswerkzeuge für Produktion und Logistik wird ein Anwender bei der Vorgehensweise für den Aufbau und die Durchführung einer Simulationsstudie oder eines Simulationsprojekts fast durchgängig unterstützt. Allerdings sind bei der aufwändigen und komplexen statistischen Auswertung der Simulationsergebnisse noch große Defizite vorhanden, weshalb diese Aufgabe für industrielle Anwender verbessert und vereinfacht werden muss.

Diese Arbeit beschreibt ein **integriertes Verfahren zur robusten statistischen Auswertung von Simulationsergebnissen** in der Produktion. Dabei werden stochastische und ereignisdiskrete Modelle von Produktionssystemen simuliert.

Es wird ein einfaches, robustes und praxisnahes Verfahren zur statistischen Auswertung konzipiert, entwickelt und angewendet. Dieses Auswertungsverfahren soll ermöglichen, dass ein Simulationsanwender aus der Praxis ohne vertiefte Statistikenkenntnisse eine korrekte Ergebnisauswertung durchführen und daraus die richtigen Entscheidungen treffen kann. Die Anwendbarkeit des Verfahrens wird mittels Praxisbeispielen von Produktionssystemen aus der Bau- und Schuhindustrie aufgezeigt.

1.2 Notwendigkeit der statistischen Auswertung von Simulationsdaten

Bei der Simulation von Produktionssystemen und Fertigungsanlagen wird häufig ein *ereignisdiskretes Modell* mit *stochastischen* Einflüssen verwendet [HeNe06:3], [Law07:669].

Dabei werden mit den stochastischen (Teil-) Modellen Einflüsse und Prozesse abgebildet, welche nicht vollständig vorherbestimmt, sondern zufälliger Natur sind. Schwankende Bearbeitungszeiten, Ausfallzeiten, Prüfergebnisse oder andere zufallsbedingten Ergebnisse sind typische Beispiele hierfür [Fisc88:392]. Damit wird die Simulation eines Modells mit stochastischen Eigenschaften zu einem Zufallsexperiment, vergleichbar der Ziehung einer Stichprobe. Um anhand der Ergebnisse eines Zufallsexperiments Aussagen zu machen und Entscheidungen zu treffen, benötigt man eine „große“ Anzahl von unabhängigen Stichproben. Dies gilt auch umgekehrt: Bei wenigen oder gar nur einer Stichprobe erhält der Simulationsanwender mit einer recht großen Wahrscheinlichkeit falsche Aussagen über das untersuchte System.

Um Simulationen als Zufallsexperimente korrekt auszuwerten, sind die Methoden der mathematischen Statistik notwendig. Aus Zufallsexperimenten können keine Aussagen mit absoluter Sicherheit gewonnen werden. Die Statistik beinhaltet allerdings Methoden, welche nicht nur Schätzwerte für die Ergebnisse von Zufallsexperimenten angeben, sondern auch die Größe eines möglichen Fehlers (die „Genauigkeit“) und die Glaubwürdigkeit einer Aussage (die „Zuverlässigkeit“) beschreiben.

Möchte man, wie häufig in Simulationsprojekten, nicht nur ein Modell betrachten, sondern mehrere Modelle unterschiedlicher Systemalternativen oder -varianten miteinander vergleichen und bewerten, so müssen die entsprechenden statistischen Verfahren verwendet werden, um vertretbare Aussagen treffen zu können. Mit diesen statistischen Verfahren lässt sich erkennen, ob Unterschiede im Verhalten der Simulationsmodelle von Systemalternativen

signifikant (d.h. durch unterschiedliche Strukturen und Parameter verursacht) sind oder durch zufällige Schwankungen entstehen.

Die Notwendigkeit der statistischen Auswertung lässt sich somit auf zwei Hauptmotivationen zurückführen. Bevor in Simulationsstudien vergleichende Aussagen gemacht werden können, muss mit den geeigneten statistischen Verfahren beurteilt werden:

1. mit welcher Genauigkeit die Schätzung der Eigenschaften der jeweiligen Simulationsmodelle zu bewerten ist,
2. wie sich die Eigenschaften der verschiedenen Simulationsmodelle der einzelnen Varianten signifikant unterscheiden.

Nur auf Basis einer korrekten statistischen Auswertung der Simulationsergebnisse lassen sich Aussagen und Entscheidungen über die Planung von Produktionssystemen aufbauen.

1.3 Auswahl geeigneter statistischer Verfahren

Viele der bekannten statistischen Verfahren zur Schätzung und zum Vergleich von Eigenschaften aus einer Anzahl von Stichproben setzen bestimmte Annahmen über diese Stichproben voraus. Die Stichproben müssen

- normalverteilt,
- mit demselben Mittelwert und derselben Varianz sowie
- voneinander unabhängig sein.

Kleijnen [Klei87:10] bezeichnet dies als die „**klassischen Voraussetzungen**“ (engl. „classical assumptions“) NID, “normally and independent distributed“. Leider sind diese Voraussetzungen bei Simulationsergebnissen häufig verletzt [Klei87:16], [Law07:485], [HaNe06:512], so dass auf spezielle robuste statistische Verfahren und Vorgehensweisen ausgewichen werden muss und sich damit die statistische Aufbereitung und Auswertung der Ergebnisse nicht mehr trivial gestaltet.

Diese Problematik wird auch von der ASIM Arbeitsgruppe „Qualitätskriterien“ [WWCP08:144] beschrieben:

„Oft werden z.B. normalverteilte unkorrelierte Werte vorausgesetzt. Dies trifft aber bei Ergebnissen von Simulationsexperimenten so gut wie nie zu.“

So kann man bei Simulationsstudien im Allgemeinen **nicht** von einer Normalverteilung der Stichproben von beobachteten Variablen ausgehen [Klei87:16]. Beispielsweise sind die Beobachtungen von Warte- oder Durchlaufzeiten häufig schief verteilt mit einer rechtslastigen Verteilung und einem entsprechenden

oberen „Schwanz“. Abb. 1.1 zeigt eine solche für Beobachtungen aus Simulationen typische Verteilungsfunktion mit unterschiedlichen Werten für Modus, Median und arithmetischem Mittelwert.

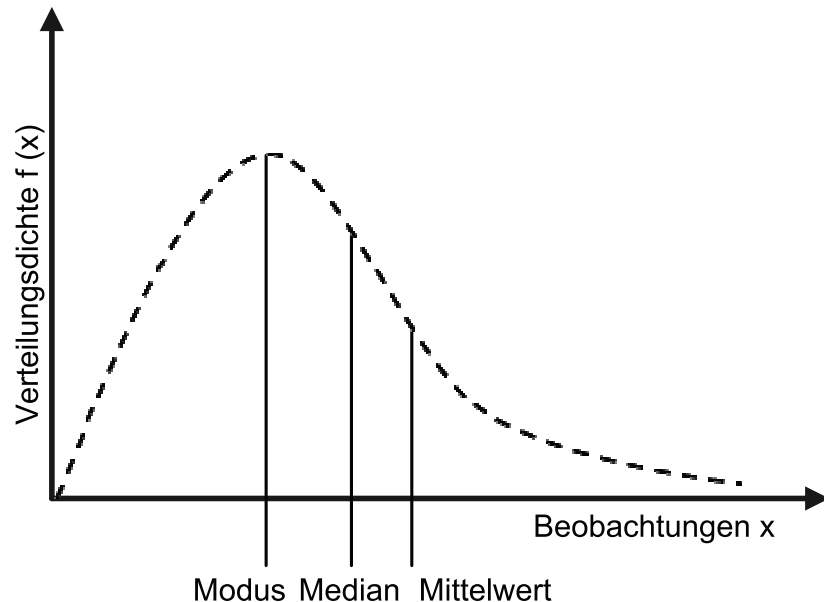


Abbildung 1.1: Typische Form einer Verteilungsdichte für Beobachtungen aus der Simulation eines Produktionssystems wie bei [WaWa94:632] und [Klei87:21].

Auf Grund dieser häufig nicht normalverteilten Stichproben der Szenarien und den unterschiedlichen Varianzen der Szenarien sind die Voraussetzungen der meisten bekannten statistischen Verfahren nicht erfüllt und die Verfahren liefern keine korrekten Ergebnisse.

Wenn Stichproben aus einem einzelnen Simulationslauf zur Schätzung von Eigenschaften verwendet werden, ist zusätzlich durch die Autokorrelation die Voraussetzung der Unabhängigkeit verletzt [Fisc88:402]. **Law** [Law07:485] schreibt dazu:

„A second reason for inadequate statistical analysis is that the output process of virtually all simulations are nonstationary and autocorrelated.“

Prinzipiell ist die statistische Auswertung von Simulationsergebnissen ein komplexer Prozess, der meist aus einer Kombination von exakten Verfahren und heuristischen Überlegungen besteht [IAO98:12]. Oft ist ein mehrstufiges, iteratives Vorgehen erforderlich. Die Auswahl der zur Verfügung stehenden Methoden und Prozeduren wird dabei sowohl von der Art der durchgeführten Simulation als auch von den statistischen Eigenschaften der Simulationsergebnisse bestimmt.

Dabei bestehen folgende wesentliche Probleme:

- In der Regel gilt ein statistisches Verfahren exakt nur unter bestimmten Voraussetzungen. Diese Voraussetzungen sind häufig nur aufwändig nachprüfbar.
- Die Auswahl der statistischen Verfahren, die Durchführung der Simulationen, die Interpretation der Daten und die Aufbereitung der Ergebnisse sind allesamt komplexe Vorgänge und erfordern ein hohes Maß an jeweiligem Expertenwissen. Häufig ist ein Simulationsanwender damit überfordert.
- Software-Werkzeuge wie einfache Tabellenkalkulationsprogramme (z.B. Excel) stellen viele der notwendigen Verfahren nicht bereit und spezielle Statistik-Programme (z.B. SPSS, SAS/STAT) werden nur selten eingesetzt.

Damit wird die systematische und korrekte statistische Auswertung von Simulationsergebnissen zu einer wesentlichen und erfolgskritischen Teilaufgabe im Rahmen der Durchführung einer Simulationsstudie von Produktionssystemen.

Zusammenfassend bedeutet dies, dass statistische Standardverfahren meist nicht zur Simulationsauswertung geeignet sind. Ein mehrstufiges iteratives Vorgehen mit robusten, einfachen und nachvollziehbaren statistischen Verfahren ist notwendig, um die industriellen Anwender zu unterstützen und nicht zu überfordern.

1.4 Abgrenzung der Problematik

Die vorliegende Arbeit konzentriert sich auf die statistische Auswertung von Ergebnisdaten aus stochastisch-ereignisdiskreten Simulationen, wie sie häufig bei der Simulation von Produktionssystemen eingesetzt werden. Andere Arten der Simulation, vor allem kontinuierliche Modelle und Simulationen bei fertigungstechnischen Bearbeitungsprozessen und den Bewegungen von Menschen und Maschinen sowie Finite-Elemente-Methoden (FEM) Simulation von Produkten wie bei **Vajna, Weber, Bley** und **Zeman** beschrieben [VWBZ09:242], werden nicht weiter betrachtet.

Die statistische Auswertung ist nur ein Teilaspekt und eine Teilaufgabe in der iterativen Vorgehensweise bei der Durchführung von Simulationsstudien (siehe Abb. 1.2) nach den Vorgehensmodellen von **Bullinger u.a** [BDAS86:200], **Law** [Law07:67] und der aktuellen **VDI Richtlinie** [VDIF00]. Die angrenzenden und ebenso wichtigen Problembereiche bei der Durchführung von Simulationsstudien wie

- generelle konzeptionelle Vorgehensweise,
- Modellierung der stochastischen Eingangsdaten,
- Verifikation und Validierung der Simulationsmodelle,

- Versuchsplanung der einzelnen Simulationsexperimente und
- Optimierung der Simulationsmodelle

werden in dieser Arbeit ebenfalls nicht weiter detailliert. Die Einbindung der statistischen Auswertung in die Vorgehensweise sowie die Methoden- und Werkzeugkette von Simulationsprojekten wird in Kapitel 3 aufgezeigt, bei Bedarf wird auf die entsprechenden Verfahren verwiesen und die weiterführende und vertiefende Literatur referenziert.

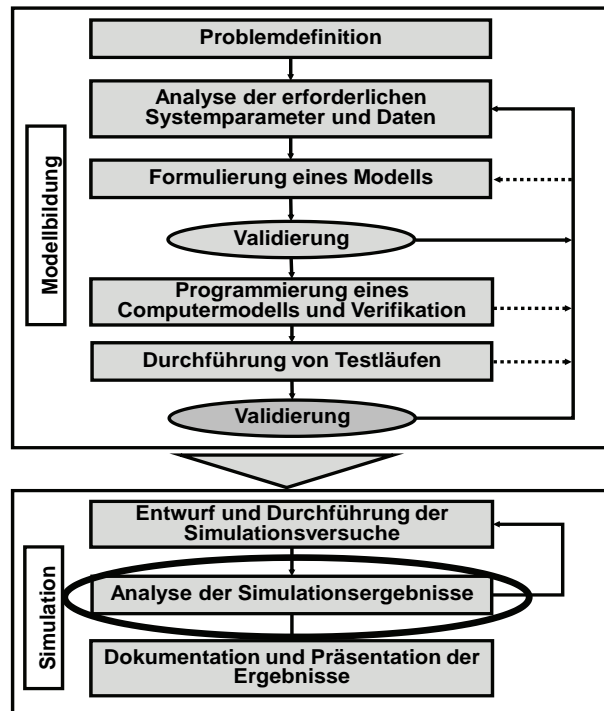


Abbildung 1.2: Vorgehensweise bei Simulationsprojekten (vgl. [VDI93] [VDI00], [Wa-Wa96:37], [BDAS86:200] und [Law07:67] mit einer Abgrenzung der Arbeit.

1.5 Ziel und Aufbau der Arbeit

Auf der Grundlage der beschriebenen Motivation soll als Ziel dieser Arbeit **ein integriertes Verfahren zu robusten statistischen Auswertung von Simulationsergebnissen in der Produktion** entwickelt werden, das Simulationsanwender ohne vertiefte Statistikkenntnisse bei der korrekten Auswertung ihrer Simulationsergebnisse unterstützt.

Als Basis hierfür wird ein integriertes einfaches und robustes Vorgehensmodell konzipiert, welches eine strukturierte Vorgehensweise der statistischen Auswertung von Simulationen mit stochastischen ereignisdiskreten Modellen beschreibt.

Die grundlegende Problematik sowie der Lösungsansatz und Handlungsbedarf der Arbeit sind in Abb. 1.3 zusammengefasst.

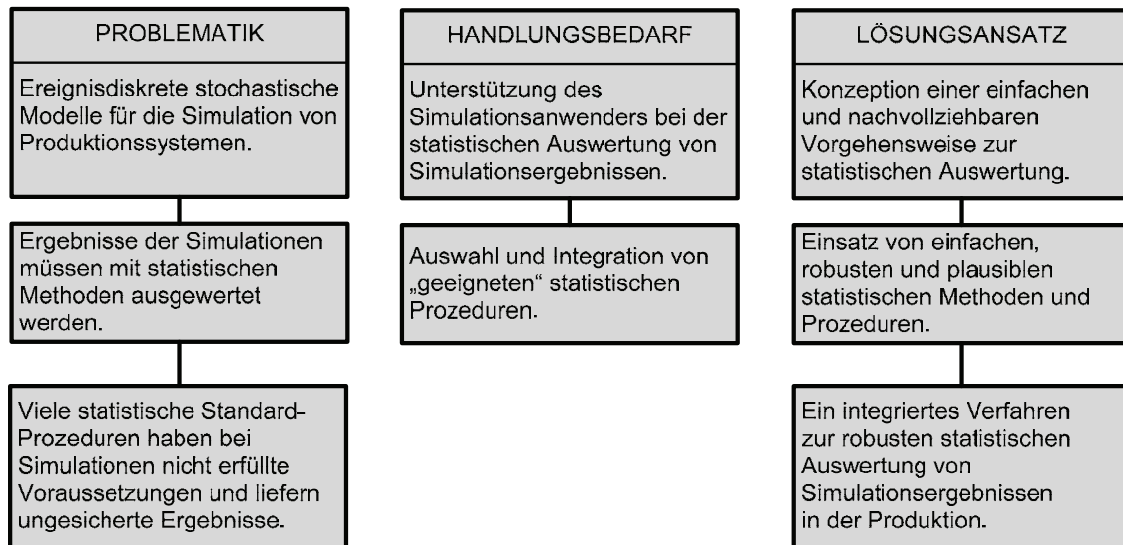


Abbildung 1.3: Problematik, Handlungsbedarf und Lösungsansatz der Arbeit.

Die Arbeit leistet damit einen Beitrag zu einer methodisch und statistisch korrekten Vorgehensweise bei der Durchführung von Simulationsstudien und -projekten, speziell auch für die Simulation von Produktionssystemen. Das Verfahren und das Vorgehensmodell werden für Anwender aus einem industriellen Umfeld entwickelt, bei denen keine vertieften Statistikkenntnisse vorausgesetzt werden dürfen. Der konzeptionelle Aufbau der Arbeit ist in Abb. 1.4 dargestellt.

In **Kapitel 2** werden aufbauend auf der industriellen Ausgangssituation und Problematik der Stand der Technik dargestellt und die existierenden Forschungsansätze identifiziert und bewertet. Ein Anforderungsprofil für das integrierte und robuste Verfahren zur statischen Auswertung der Simulationsdaten wird abgeleitet.

Der Schwerpunkt von **Kapitel 3** bildet die Konzeption eines integrierten Verfahrens als ein strukturiertes Vorgehensmodell mit Teilaufgaben. Die Auswahl und Integration der Verfahren wird dargestellt.

Das **Kapitel 4** beschreibt die Auswahl und Integration der Verfahren zur Auswertung eines Simulationsmodells und zur Schätzung einer Leistungskennzahl mit Qualitätsangabe.

Kapitel 5 zeigt die Auswahl und Integration der Verfahren zum Vergleich von mehreren Modellen und Systemalternativen hinsichtlich signifikanter Unterschiede.

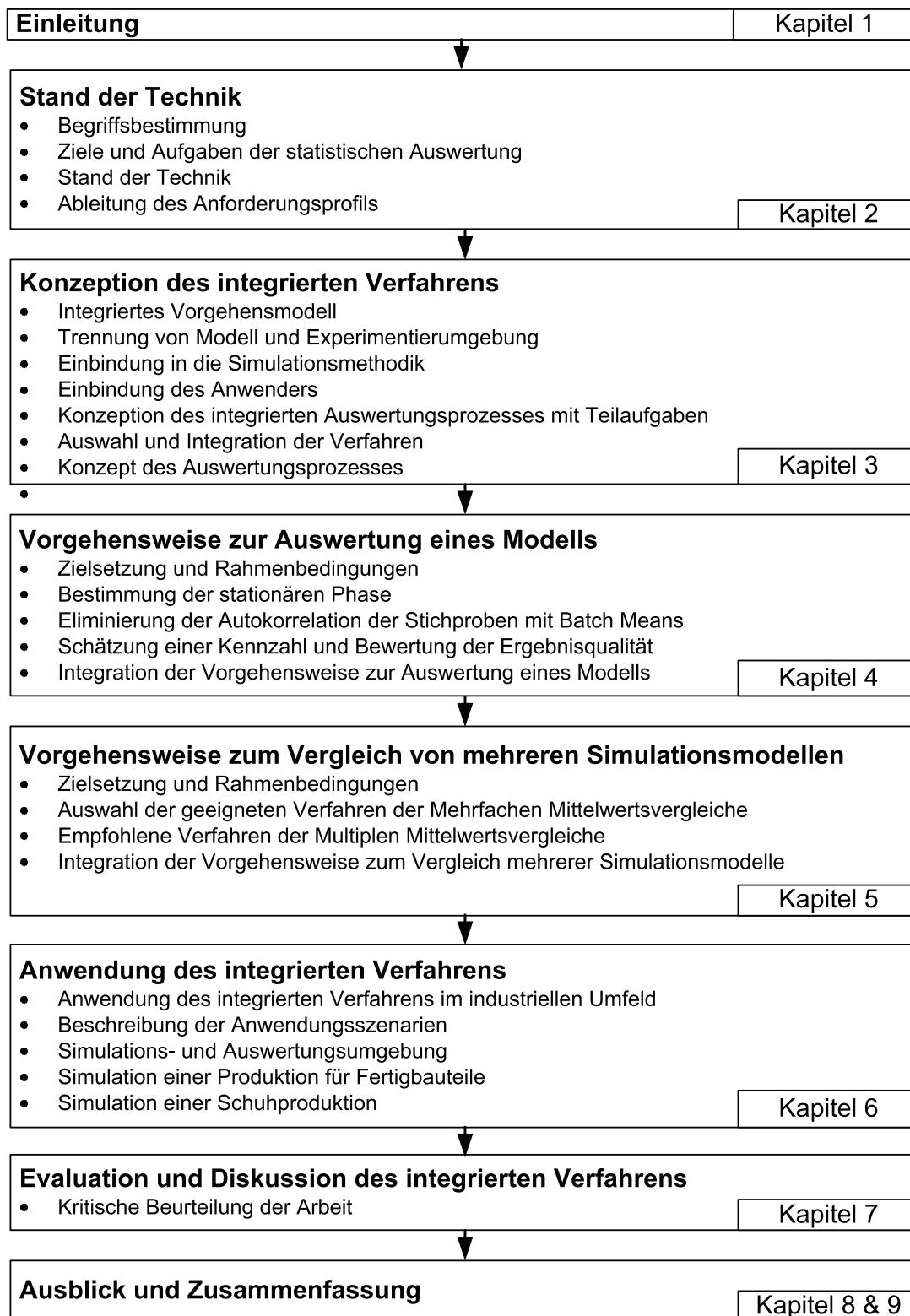


Abbildung 1.4: Überblick über den Aufbau und die Vorgehensweise der Arbeit.

Kapitel 6 beschreibt die Anwendung und Validierung des integrierten Verfahrens und der Zielsetzung an Hand zweier industrieller Simulationsstudien aus den Bereichen der Bau- und Schuhindustrie.

Die Arbeit endet mit einer Evaluation und Diskussion des entwickelten Verfahrens in **Kapitel 7**.

2 Stand der Technik

2.1 Analyse des Stands der Technik

Ausgehend von der Motivation für den Einsatz der Simulationstechnik und den Aufgaben bei der Auswertung und Interpretation der Ergebnisse beschreibt dieses Kapitel den Stand der Technik in Forschung und Praxis bei einzelnen Methoden als auch bei integrierten Verfahren. Aus den identifizierten Defiziten der derzeitigen Methoden werden die Anforderungen an **ein integriertes Verfahren zur robusten statistischen Auswertung von Simulationsergebnissen in der Produktion** abgeleitet. Die statistische Methoden, welche in den weiteren Kapiteln der Arbeit relevant sind, werden im Anhang detailliert erläutert und beschrieben.

2.2 Begriffsbestimmungen

Im folgenden Abschnitt werden die für die vorliegende Arbeit wesentlichen Begriffe der Simulationstechnik definiert. Dabei werden soweit wie möglich die Begriffsdefinitionen der VDI Richtlinie 3633 [VDIF96] verwendet bzw. auf diese verwiesen. Dieser Richtlinienentwurf definiert auch weitere Begriffe aus dem Themengebiet der Simulation.

Die Begriffe

- *Methode*
- *Verfahren*
- *Vorgehensweise*
- *Prozedur*
- *Simulation*
- *System*
- *Modell*
- *Replikation*

werden im dem hier verwendeten Sinn beschrieben und erläutert.

Methode

Unter einer Methode oder einer Methodik versteht man „*ein nach Mittel und Zweck planmäßiges (= methodisches) Verfahren, welches zu technischer Fertigkeit bei der Lösung theoretischer und praktischer Aufgaben führt*“ [Mitt95:876ff].

Verfahren

Ein Verfahren unterscheidet sich in dieser Arbeit von einer Methode durch eine ausgeprägte Prozesseigenschaft und deutlich höhere Durchgängigkeit und Komplexität. Ein Verfahren kann sich aus mehreren Methoden zusammensetzen. Das Ergebnis eines Verfahrens sollte quantitativ und reproduzierbar sein [Bopp01:15]. Die Qualitätsrichtlinie DIN ISO 9000:2005 [DIN05] versteht unter einem Verfahren eine „festgelegte Art und Weise, eine Tätigkeit auszuführen.“

Die Begriffe der **Vorgehensweise** und der **Prozedur** werden im Rahmen der Arbeit als Synonyme für ein Verfahren verwendet.

Simulation

Nach VDI-Richtlinie 3633 [VDIF00] ist die

„Simulation das Nachbilden eines Systems mit seinen dynamischen Prozessen in einem experimentierfähigen Modell, um zu Erkenntnissen zu gelangen, die auf die Wirklichkeit übertragbar sind.“

Im weiteren Sinne wird unter Simulation das Vorbereiten, Durchführen und Auswerten gezielter Experimente mit einem Simulationsmodell verstanden.“

Soweit nicht explizit anders beschrieben, wird in dieser Arbeit die Kategorie der stochastischen ereignisdiskreten Simulationen betrachtet.

System

Ein System ist ebenfalls in der VDI-Richtlinie 3633 [VDIF00] definiert: Ein System ist eine abgegrenzte Anordnung von Komponenten, die miteinander in Beziehung stehen. Es ist gekennzeichnet durch

- die Festlegung seiner Grenze gegenüber der Umwelt (Systemgrenze), mit der es über Schnittstellen Materie, Energie und Informationen austauschen kann (System-Ein- und -Ausgangsgrößen),
- die Komponenten, die bei der Erhöhung der Auflösung selbst wiederum Systeme darstellen (Subsysteme) oder aber als nicht weiter zerlegbar angesehen werden (Systemelemente),
- die Ablaufstruktur in den Komponenten, die durch spezifische Regeln und konstante oder variable Attribute charakterisiert wird,
- die Relationen, die die Systemkomponenten miteinander verbinden (Aufbaustruktur), so dass ein Prozess ablaufen kann,
- die Zustände der Komponenten, die jeweils durch Angabe der Werte aller konstanten und variablen Attribute (Zustandsgrößen) beschrieben werden, von denen im Allgemeinen nur ein kleiner Teil untersuchungsrelevant ist sowie

- die Zustandsübergänge der Komponenten als kontinuierliche oder diskrete Änderungen mindestens einer Zustandsgröße aufgrund des in dem System ablaufenden Prozesses.

Modell

Ein Modell ist in Anlehnung an die VDI-Richtlinie 3633 [VDIF00] die zweckgebundene vereinfachte reduzierte Abbildung eines geplanten oder existierenden Systems mit seinen Prozessen in einem anderen begrifflichen oder gegenständlichen System. Es unterscheidet sich hinsichtlich der untersuchungsrelevanten Eigenschaften nur innerhalb eines vom Untersuchungsziel abhängigen Toleranzrahmens vom Vorbild.

Replikation

Das einmalige Ausführen eines Simulationsmodells mit einem oder mehreren eindeutigen Startwerten für die Zufallszahlengeneratoren des Simulationsprogramms wird als Replikation oder Simulationslauf bezeichnet.

2.3 Produktionssimulation, Digitale Fabrik und Digitale Produktion

Die moderne Informationstechnik prägt mehr und mehr die Arbeitsweise produzierender Unternehmen [Spat06]. Die Verfügbarkeit von leistungsfähigen Werkzeugen sowie durchgängigen und integrierten Produktionssystemplanungs-Softwarepaketen, wie z.B. von Dassault Delmia und Siemens-Tecnomatix mit den eingebundenen Simulationswerkzeugen QUEST [Delm08] und Plant Simulation [Siem10] steigern die Akzeptanz und die Häufigkeit der Simulationstechnik im industriellen Einsatz. In Ergänzung zur virtuellen Produktentwicklung werden die Methoden und Werkzeuge der „Digitalen Fabrik“ und damit auch die Simulation von Produktionssystemen verstärkt eingesetzt. Der Begriff der „Digitalen Fabrik“ mit der Methode und dem Werkzeug der Simulation wird in der VDI Richtlinie 4499 [VDIF09] definiert:

„Die Digitale Fabrik ist der Oberbegriff für ein umfassendes Netzwerk von digitalen Modellen, Methoden und Werkzeugen – u.a. der Simulation und der dreidimensionalen Visualisierung –, die durch ein durchgängiges Datenmanagement integriert werden. Ihr Ziel ist die ganzheitliche Planung, Evaluierung und laufende Verbesserung aller wesentlichen Strukturen, Prozesse und Ressourcen der realen Fabrik in Verbindung mit dem Produkt.“

Der Fokus der Digitalen Fabrik liegt auf einer mit allen Unternehmensprozessen frühzeitig und sorgfältig abgestimmten Produktionsplanung und Gestaltung der Fabrik.“

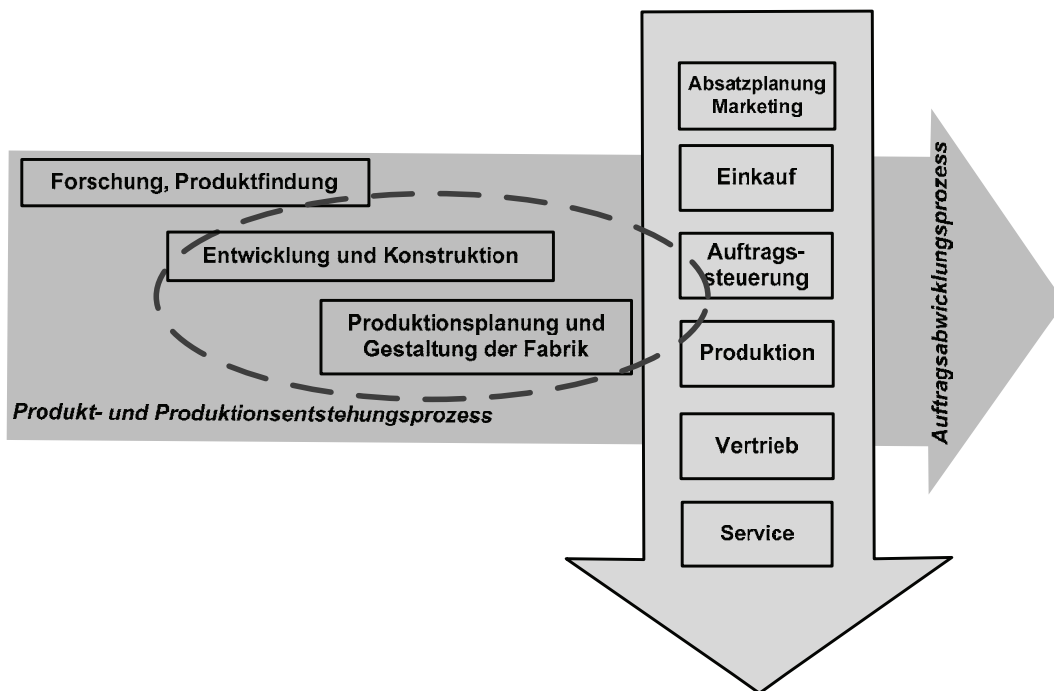


Abbildung 2.1: Das Betrachtungsfeld der „Digitalen Fabrik“ nach VDI Richtlinie 4499 [VDIF09].

Das Stuttgarter Innovationscluster „Digitale Produktion“ [SpLe06] erweitert den Fokus der „Digitalen Fabrik“ um den kompletten Produktlebenszyklus und Auftragsabwicklungsprozess und „umfasst in einem ganzheitlichen Ansatz Inhalte der Produkt- und Produktionsentstehung wie die Entwicklung und Konstruktion, Arbeitssystemplanung, Arbeitsablaufplanung sowie Produktionsplanung, -simulation und -steuerung“ [SpLR07].

Die Abb. 2.1 und 2.2 zeigen die Gemeinsamkeiten und Unterschiede der Gegenstandsbereiche der beiden verwandten Konzepte.

Im Verständnis der „Digitalen Fabrik“ nach VDI 4499 [VDIF09] beinhaltet die Simulation „*unterschiedliche Facetten der Geschäftsprozesssimulation, der Materialfluss-, Logistik- und Produktionsablaufsimulation, der Roboter-, der Personal- und Ergonomiesimulation sowie der Personaleinsatz- und der Prozesssimulation (beispielsweise von maschinennahen Fertigungsprozessen)*“.

Allerdings beschränkt sich diese Arbeit (wie in Kapitel 1 beschrieben) auf die Simulation von ereignisdiskreten und stochastischen Modellen, welche bei der Roboter-, Personal- und Ergonomiesimulation sowie bei der Simulation der maschinennahen Fertigungsprozesse praktisch nicht verwendet werden.

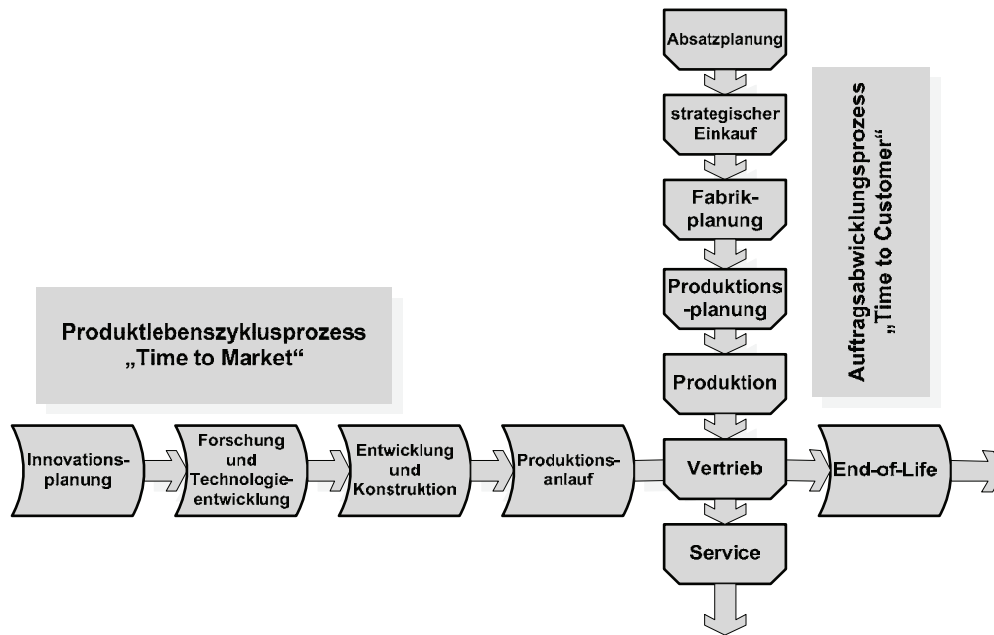


Abbildung 2.2: Der Gegenstandsbereich der „Digitalen Produktion“ nach Spath [Spat06].

2.4 Simulationstechnik im industriellen Einsatz

Unterstützt von den Konzepten der Digitalen Fabrik ist die Simulationstechnik ein wichtiges Werkzeug des Ingenieurs bei der Planung, Realisierung und dem Betrieb von Produktionssystemen [VDIF00]. Durch die Verfügbarkeit von leistungsfähiger und preiswerter Computerhardware und vor allem durch Simulationswerkzeuge mit komfortabler grafischer Benutzeroberfläche und attraktiven Animationsfunktionen werden Simulationen als Analyse- und Problemlösungsmethode verstärkt in der industriellen Produktionsplanung eingesetzt [RaSW08:8], [WWCP08:1].

Die entscheidenden Vorteile der Simulationstechnik werden vielfach beschrieben [Law07:76] [VDIF00]:

- Häufig ist bei der Kombination von Realsystem und Fragestellung keine mathematische Modellierung und analytische Bewertung möglich oder viel zu aufwändig.
- Andere Zeitabläufe und -konstanten sind möglich, Prozesse lassen sich schneller oder langsamer simulieren.
- Das Verhalten von vorhandenen Systemen lässt sich unter veränderten Bedingungen überprüfen.
- Es kann eine (theoretisch) unbeschränkte Anzahl von alternativen Systemvarianten untersucht werden.

- Die Simulation ermöglicht eine vollständige Kontrolle der Versuchsbedingungen und damit auch eine komplette Reproduzierbarkeit der Versuche.

Diesen Vorteilen stehen meist nur wenige Nachteile gegenüber [Wagn10]:

- Die Modellierung ist per Definition nur eine reduzierte Abbildung der Realität.
- Der Aufwand der Modellierung kann verhältnismäßig hoch sein.
- Die Qualität der Modellierung ist abhängig von der Fachkompetenz der Anwender und Autoren des Modells.
- Die Interpretation der Ergebnisse ist nicht immer einfach.

Diese generellen Vor- und Nachteile gelten direkt auch für die Simulation von Produktionssystemen im industriellen Einsatz, wo sie sich als wichtige Analysemethode etabliert hat. Eine Absicherung der Produktionssystemplanung durch die Simulation wird häufig eingesetzt, wenn wichtige Entscheidungen getroffen werden müssen und keine geeigneten analytischen Hilfsmittel zur Verfügung stehen [RSW08].

Mittlerweile gibt es eine Vielzahl von komfortablen Werkzeugen zur Modellierung und Simulation von Produktionssystemen. Einen Überblick über das Angebot an Simulationswerkzeugen im deutschsprachigen Raum zeigen **Wenzel** und **Noche** [WeNo00] in ihrem Konferenzbeitrag. Vergleichbare internationale Übersichten über industrielle Anwendungen und neue Softwareentwicklungen sind in den englischsprachigen Zeitschriften „**Industrial Engineer**“ [IoIE11] und „**Simulation**“ [SiCo11] zu finden.

2.5 Statistische Auswertung und Interpretation der Ergebnisse

Im Kontrast zu dem hohen Aufwand bei der Modellierung wird die statistisch korrekte Auswertung und Beurteilung der Qualität der Simulationsergebnisse bei vielen Simulationsstudien häufig nur unzureichend durchgeführt [MoGO93], [PiRD90], [Fisc88:393]. In einem Standardwerk zur Simulation schreibt **Law** [Law07:485]:

„In many simulation studies a great deal of time and money is spent on model development and “programming”, but little effort is made to analyze the simulation output data appropriately. As a matter of fact, a very common mode of operation is to make a single simulation run of somewhat arbitrary length and then to treat the resulting simulation estimates as the „true“ model characteristics. Since random samples from probability distributions are typically used to drive a

simulation through time, these estimates are just particular realizations of random variables that may have large variances. As a result, these estimates could, in a particular simulation run, differ greatly from the corresponding true characteristics for the model. The net effect is, of course, that there could be a significant probability of making erroneous inferences about the system under study.“

Diese Diskrepanz zwischen in der Literatur beschriebenen Methoden zur statistischen Auswertung und dem realen Vorgehen bei praktischen Simulationsstudien von und mit Industrieunternehmen lässt sich u.a. nach **Law** [Law07:485], **Mellichamp und Young** [MeYo89] sowie **Deslandres und Pierreval** [DePi91] auf mehrere Ursachen zurückführen:

- Das fehlende Bewusstsein für die Notwendigkeit der statistischen Auswertung. Viele Simulationsanwender sehen die Herausforderung vor allem in der Modellierung und der Fehlerbeseitigung.
- Die meist unzureichenden Kenntnisse der Anwender über die geeigneten statistischen Verfahren und ihre Leistungsfähigkeit. Dies ergibt Probleme mit der Komplexität der Vorgehensweisen, da bekannte Verfahren nicht direkt anwendbar sind.
- Die fehlende komfortable Unterstützung durch die eingesetzten Simulations- oder Auswertungswerkzeuge.
- Eine Scheu vor dem erhöhten Personen- und Rechenaufwand, obwohl Rechenleistung zu geringen (Mehr-)Kosten verfügbar ist.

Wenn bei Simulationsprojekten eine statistische Auswertung der Ergebnisse durchgeführt wird, so wird meistens von einer Normalverteilung der Stichproben ausgegangen. Dies *kann* bei einer additiven Eigenschaft der beobachteten Größe nach dem Zentralen Grenzwertsatz berechtigt sein [ChIn98], *muss* aber nicht notwendigerweise der Fall sein. Besonders **Kleijnen** [Kleij87:16] zeigt die Risiken der nicht erfüllten Voraussetzungen der statistischen Verfahren, ihrer verfälschten Ergebnisse und der damit verbundenen irrtümlichen Entscheidungen auf.

2.6 Ziele und Aufgaben der statistische Auswertung

Im Rahmen einer Simulationsstudie möchte der Anwender Aussagen über die Leistungsfähigkeit des untersuchten Systems sowie über Leistungsunterschiede bei den alternativen System- und Modellvarianten erhalten. Daraus ergeben sich für die statistische Auswertung der Simulationsergebnisse die beiden Hauptziele:

- Schätzung von Leistungskennzahlen eines Modells mit einem Güteindikator und
- Überprüfung der Systemalternativen auf signifikante Unterschiede hinsichtlich der Leistungskennzahlen mit der Identifikation des „besseren“ Systems.

Je nach Fragestellung, Art der Leistungskennzahlen und den zu simulierenden Systemen müssen dazu unterschiedliche (Teil-)aufgaben entlang eines Auswertungsprozesses erfüllt und miteinander integriert werden.

Der Ablauf der statistischen Auswertung von Simulationsergebnissen wird u.a. bei **Law** [Law07:491] als auch, leicht vereinfacht, in der VDI Richtlinie 3633 Blatt 3 [VDIF95] und bei der ASIM Arbeitsgruppe „Qualitätskriterien“ [WWCP08:139] beschrieben.

Aus den beiden oben beschriebenen Hauptzielen und den beiden, bei entsprechender Ausgangslage notwendigen vorbereitenden Eigenschaften der Stationarität und Unabhängigkeit ergeben sich prinzipiell die relevanten vier (Teil-) Aufgaben:

- Bestimmung der Stationarität der Zeitreihe¹,
- Sicherung der Unabhängigkeit der Stichproben,
- Schätzen der Kennzahlen mit Güteindikatoren,
- Vergleich der Systemalternativen,

die im Folgenden begründet und erläutert werden.

Diese Ziele und Aufgaben gelten für die generelle ereignis-diskrete Simulation genauso wie für die Simulation von Produktionssystemen im Speziellen.

¹ Eine Zeitreihe ist eine Folge zeitlich sortierter Daten (vgl. [SaHe06:364]). Die Ergebnisse eines Simulationslaufs werden häufig als eine Zeitreihe der Leistungskennzahl dargestellt.

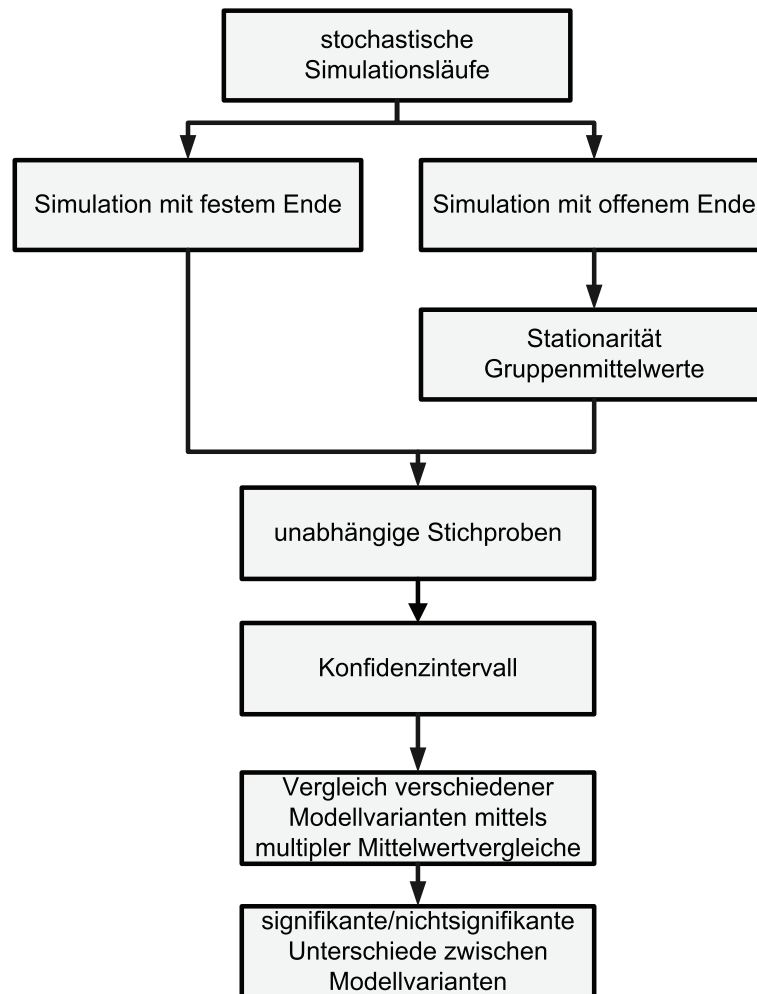


Abbildung 2.3: Vereinfachter Ablauf der statistischen Auswertung nach der VDI Richtlinie [VDIF95].

2.6.1 Bestimmung der Stationarität

Eine häufige Aufgabe bei Simulationsstudien ist die Bestimmung der stationären (engl. „steady state“) bzw. der gegenteiligen, transienten Phase des beobachteten Systemverhaltens. Bei vielen Simulationsstudien ist der Anwender an dem stationären, zeitinvarianten und eingeschwungenem Verhalten des untersuchten Systems interessiert. Das Durchlaufen von transienten Einschwingphasen der Zeitreihen von Beobachtungen, z.B. das Anlaufen einer Produktionsanlage, beeinträchtigt häufig die Beurteilung der Eigenschaften des stationären Verhaltens durch eine schlechte bzw. verfälschte Schätzung der Kennzahlen.

Die vorbereitende Teilaufgabe bei der statistischen Auswertung besteht aus der Identifikation des Übergangs der transienten in die stationäre Phase der Zeitreihe der Beobachtungen. Das Ziel ist das Abschneiden bzw. das Unterdrücken der transienten Stichproben für eine unverfälschte Schätzung des stationären Verhaltens.

2.6.2 Autokorrelation und Unabhängigkeit der Stichproben

Bei der Simulation von Warteschlangensystemen erhält man meist eine positive Autokorrelation der einzelnen Beobachtungen. Das bedeutet, dass die Beobachtung x_i mit hoher Wahrscheinlichkeit ähnliche Werte wie die nachfolgende Beobachtung x_{i+1} aufweist. Wie **Kleijnen** [Klei87:59] und **Fischer** [Fisc89:382] zeigen, führt diese positive Autokorrelation zu einer Unterschätzung der Stichproben-Varianz und damit zu fälschlich kleinen Güteindikatoren, z.B. Konfidenzintervallen, bei der Schätzung der Erwartungswerte der Leistungskennzahlen.

Speziell bei Simulationen mit offenem Ende, bei denen die Stichproben aus einem „langen“ Simulationslauf benutzt werden sollen, muss die Autokorrelation beseitigt werden, um die Unabhängigkeit der Beobachtungen zu gewährleisten.

Die vorbereitende Aufgabe definiert sich damit als Beseitigung der Autokorrelation für die Unabhängigkeit der Stichproben aus einem Simulationslauf.

2.6.3 Schätzen von Kennzahlen mit Güteindikator

Für das Schätzen von Kennzahlen aus den Stichproben der Simulationsergebnisse werden erwartungstreue Punktschätzer für den Mittelwert benutzt. Wegen der stochastischen Natur der Kennzahlen werden Güteindikatoren wie bspw. Konfidenzintervalle, auch Vertrauensbereiche genannt, für die Beurteilung der Qualität der Punktschätzung benötigt.

Bei dem Ziel der Schätzungen von Leistungskennzahlen aus Simulationsergebnissen muss daher immer ein Konfidenzintervall oder ein vergleichbarer Güteindikator mit der Punktschätzung angegeben werden, um die Qualität und Aussagekraft dieser Angabe beurteilen zu können.

2.6.4 Vergleich von Systemalternativen

Der Vergleich von simulierten Systemalternativen auf signifikante Unterschiede bei konkreten Leistungskennzahlen mit der Bestimmung des „besseren“ Systems ist das häufige (nachfolgende) Ziel von Simulationsstudien. Das Erkennen eines statistisch signifikanten Unterschieds² im Gegensatz zu zufälligen Einflüssen des Modellverhaltens ist die wesentliche Grundlage für einen Vergleich der Systemalternativen für nachfolgende Entscheidungen oder weiterführende Optimierungsansätze.

² Die Bedeutung der statistischen Signifikanz wird in Kap. 3 erläutert.

2.7 Stand der Forschung

Im Folgenden wird der Stand der Forschung beschrieben und kritisch diskutiert. Besonderer Schwerpunkt der Diskussion sind die Defizite der bekannten Verfahren und Vorgehensweisen mit den Problematiken der nichtnormalen Verteilung und der evtl. fehlenden Unabhängigkeit der Stichproben. Besonderer Wert wird auf die Plausibilität, Robustheit, Anwendbarkeit und Integrationsfähigkeit der Verfahren gelegt.

Es werden zunächst die (in der Literatur teilweise intensiv untersuchten) Einzelmethoden für die oben beschriebenen vier Aufgaben diskutiert, um anschließend den Stand der Forschung bei mehr oder weniger durchgängigen Auswertungsmethodiken und Auswertungswerkzeugen zu erörtern.

2.7.1 Verfahren für Teilaufgaben

2.7.1.1 Stationarität der Zeitreihe

Die Bestimmung der stationären Phase ist ein in der Literatur häufig untersuchtes und diskutiertes Problem. Viele Autoren untersuchen Verfahren, welche einen möglichst guten Kompromiss für Effizienz und Robustheit darstellen und damit für möglichst viele Anwendungsfälle einsetzbar sind. Diese Verfahren lassen sich dabei in zwei prinzipielle Kategorien einteilen:

- direkte Verfahren, welche versuchen, die frühesten stationären Beobachtungen innerhalb der Zeitreihe zu erkennen und
- indirekte Verfahren, d.h. statistische Tests, welche die Beobachtungen nach einem angenommenen Zeitpunkt und Phasenübergang auf Stationarität überprüfen.

Ein weitverbreitetes, häufig referenziertes und leistungsfähiges Verfahren [Law07:509] ist die Methode von **Welch** [Welc83], die von **Linton** und **Harmonosky** [LiHa02] mit vier anderen Prozeduren verglichen und empfohlen wird.

Einen Überblick über acht direkte und indirekte Verfahren geben **Mellichamp und Yuong** [MeYo89] bei der Beschreibung eines Auswertungswerkzeugs. Dabei werden die Verfahren von **Gafarian, Conway, Gordoner** und **Tocher**, die zwei Methoden von **Fishman** sowie die Aufsätze von **Emshoff** und **Sisson** evaluiert.

Im Rahmen des **ISSTAS Projekts** [IAO98] wurden u.a. die drei Verfahren von **Fischer** [Fisc89], **Vassiliacopoulos** [Vass89] sowie von **Schruben, Singh** und **Thierney** [ScST83][Klei87:110] auf die Eignung für praktische Anwendungen untersucht.

Mathematisch aufwändig und weniger für den Praktiker geeignet, bestimmen **Gallagher, Bauer und Maybeck** [GaBM96] die stationäre Phase mit Kalman Filtern. **White, Cobb und Spratt** [WhCS00] bewerten vier weitere heuristische Verfahren und empfehlen das Verfahren der „marginal standard error rules MSER-5“ kombiniert mit Batch Means. In einer sehr vollständigen Übersicht klassifiziert **Robinson** [Robi02] 26 Verfahren und empfiehlt eine neue SPC-Prozedur auf Grundlage der statistischen Prozesskontrolle.

Marajal und Ingalls [Maln04] untersuchen sieben Verfahren mit stark, mäßig und wenig ausgelasteten Systemen³ und empfehlen dabei das MSER-5 Verfahren sowie je nach Auslastung auch andere Verfahren. Das MSER-5 Verfahren wird auch von **Mokashi, Tejada, Yousefi, Xu, Wilson, Tafazzoli und Steiger** [MTYX10] mit dem, ebenfalls Match Means basierten, „N-Skart“ Verfahren verglichen, ohne dabei signifikante Vorteile zu erzielen.

Einen aktuellen und sehr vollständigen Überblick über die Vielzahl der Konzepte und Verfahren zur Bestimmung des Übergangs von der transienten in die stationäre Phase bieten **Pasupathy und Schmeiser** [PaSc10].

Als wesentliche Defizite all dieser Verfahren für den praktischen Einsatz lassen sich zwei Eigenschaften bestimmen:

Eine starke Effizienzorientierung, welche die Kriterien der Robustheit und Plausibilität verdrängt sowie die fehlende Kombination und Integration der Verfahren, um evtl. vorhandene Schwächen zu kompensieren. Speziell die unzureichende Nachvollziehbarkeit und Plausibilität bei Anwendern ohne vertiefte Statistikkenntnisse verhindert den Einsatz der weniger intuitiven und komplexeren Verfahren.

2.7.1.2 Autokorrelation und Unabhängigkeit

Zur Beseitigung der Autokorrelation bei der Auswertung der Stichproben aus einem Simulationslauf gibt es in der Literatur eine Vielzahl von Verfahren.

Die Methode der „Batch Means“⁴ ist ein bekanntes, verbreitetes und häufig diskutiertes, variiertes und modifiziertes Verfahren zur Reduktion bzw. Beseitigung der Autokorrelation und Autokovarianz der Stichproben aus einer Replikation einer nichtterminierenden Simulation. **Alexopoulos** und **Goldman** [AlGo03]

³ Marajal und Ingalls [Maln04] bezeichnen in ihrer Studie eine durchschnittliche Auslastung der Ressourcen bei 90% als „high“, bei 70% als „moderat“ und bei 50% als „low“.

⁴ Auf die Übersetzung von „Batch Means“ in „Gruppenmittelwerte“ wurde verzichtet, da dieser Begriff im Rahmen der Varianzanalyse anderweitig verwendet wird.

vergleichen die Batch Means Methode mit dem Konzept mehrerer unabhängiger Replikationen.

Die meisten Methoden basieren auf den grundlegenden Arbeiten von **Fishman** [Fish78] und **Schmeiser** [Schm82], welcher auch eine obere Grenze von 10-20 für die Anzahl der Gruppen bzw. Batches empfiehlt. **Yeh und Schmeiser** [YeSc00] beschreiben ein speichereffizientes Verfahren von dynamischen Batch Means.

Die asymptotischen Eigenschaften für eine große Anzahl von Batch Means und Stichproben in einer Gruppe analysieren **Chien, Goldman und Melamed** [ChGM97], während **Sherman** [Sher98] ein Verfahren zur Bestimmung einer sinnvollen Gruppengröße vorstellt. Als Variante schlagen **Andradóttir und Argon** [AnAr01] die Methode der replizierten Batch Mean („replicated batch means“) vor.

Effizienzsteigernde Weiterentwicklungen entstehen, wenn **Steiger und Wilson** [StWi00] die Leistungsfähigkeit ihrer ASAP Prozedur mit den Verfahren ABATCH und LBATCH von **Fishman** [Fish96:554] vergleichen. Hierauf empfiehlt **Fishman** [Fish01] fünf Jahre später die Weiterentwicklung LABATCH.2 und **Steiger, Lada, Wilson, Alexopoulos, Goldsman und Zouaoui** [SLWA02] reagieren mit der verbesserten Version von ASAP2. **Law** [Law07:522] bietet dazu einen Vergleich und Überblick. Eine derzeit aktuelle Entwicklung ist die „spaced batch means“ SBatch Prozedur von **Lada und Wilson** [LaWi07].

Alternative Verfahren, die auf Konzepten wie der Spektralanalyse oder der Analyse der Kovarianzstruktur der Zeitreihe basieren, sind wegen der komplexeren und anspruchsvolleren Algorithmen für Praktiker schlecht nachzuvollziehen und damit wegen der fehlenden Plausibilität nach **Fischer** [Fisc88] weniger geeignet.

Dies gilt auch für die Variante der Methode der überlappenden und teilweise überlappend Batch Means („overlapping and partially overlapping batch means“), welche nach **Yeh und Schmeiser** [YeSc01] keine wesentlichen Vorteile bietet.

Als relevante Defizite ergeben sich auch hier die unzureichende Plausibilität und fehlende Nachvollziehbarkeit für Anwender der komplexen und mathematisch anspruchsvolleren Varianten des Batch Means Verfahren wie teilweise oder vollständig überlappend, LABATCH, ABATCH, ASAP(2) und SBatch, welche den eventuell vorhandenen Gewinn an Effizienz nicht aufheben.

2.7.1.3 Konfidenzintervalle bei Punktschätzern

Der am meisten verbreitete Güteindikator für eine Punktschätzung bei der Auswertung von Simulationsergebnissen ist ein Konfidenzintervall, auch als Vertrauensbereich bezeichnet. In der Simulationsliteratur wird praktisch immer ein auf Basis der Student-t Statistik berechnetes Konfidenzintervall verwendet. Alternativen zum Konfidenzintervall wie der von **Schmeiser** [Schm01], [Schm04] vorgeschlagene „geschätzte Standardfehler“ („estimated standard error“) der Punktschätzung haben weniger Relevanz und werden selten benutzt. Dies gilt auch für alternative Intervalle wie z. B. „Wald's Intervall“ und das „Clopper-Pearson Intervall“, welche von **Buchholz** und **Müller** [BuMü09] beschrieben, aber für praktische Simulationsanwendungen nicht verwendet werden.

Bei der Berechnung des Konfidenzintervalls verlassen sich viele Anwender nach **Law** [Law07:496] und **Kleijnen** [Klei87:23] auf die angebliche Robustheit der Student-t Statistik. **Schmeiser** und **Yeh** [ScYe02] zeigen dazu die Problematik und die Grenzen der Genauigkeit der Student-t Statistik-basierten Konfidenzintervalle auf.

Die alternativen Verfahren zur Berechnung eines Konfidenzintervalls wie z.B. die modifizierte Johnson-t-Statistik oder nicht-parametrische Verfahren, wie von **Kleijnen** [Klei87:24] vorgeschlagen, werden sehr selten verwendet. Dies gilt auch für rechenintensive Methoden, die auf Resampling-Konzepten wie „Bootstrapping“ oder „Jackknifing“ ([Klei87:78], [Chen06] und [ChCu09]) basieren.

Die Defizite bei den Verfahren der Berechnung eines Güteindikators für die Punktschätzung einer Leistungskennzahl bestehen darin, dass für die Berechnung des Konfidenzintervalls praktisch nur die Student-t Statistik verwendet wird, ohne ihre Robustheit und Gültigkeit bei der Abweichung von den „Klassischen Voraussetzungen“ NID (vgl. Kapitel 1.3) zu überprüfen. Alternative und robustere Verfahren sind wenig bekannt und werden selten eingesetzt.

2.7.1.4 Vergleich von Systemalternativen

Beim Vergleich von zwei oder mehreren Systemalternativen gibt es zwei prinzipielle Vorgehensweisen mit leicht unterschiedlicher Zielsetzung, die sich bei Bedarf gut kombinieren lassen: Verfahren der Multiplen Mittelwertvergleiche (MMV) und Ranking and Selection Verfahren (R&S)⁵. **Bechhofer**, **Santner** und **Goldman** [BeSG95] zeigen die generellen Grundlagen und Anwendungen der

⁵ Die Verfahren der Multiplen Mittelwertvergleiche sowie Ranking and Selection werden in Kapitel 3 gegenübergestellt und verglichen.

beiden Konzepte und Verfahren während **Goldsmann, Nelson** und **Schmeiser** [GoNS91] die Vor- und Nachteile der Verfahrensprinzipien beim Vergleich von Simulationsalternativen darstellen.

In der Literatur und Praxis der Auswertung von Simulationsergebnissen liegt ein eindeutiger Schwerpunkt auf R&S-Verfahren. **Goldsmann** und **Nelson** [GoNe98], **Law** [Law07], **Banks, Carson II, Nelson** und **Nicol** [BCNN10] sowie **Kim** und **Nelson** [KiNe06], [KiNe07] geben einen guten Überblick und zeigen die letzten Entwicklungen. Prinzipiell basieren nach **Chen** [Chen07] viele R&S-Verfahren direkt oder indirekt auf den Prozeduren von **Dudewicz** und **Dalal** [DuDa75] oder **Rinott** [Rino78]. **Kim** und **Nelson** [KiNe06] führen die Herkunft der R&S-Verfahren auf die Arbeiten von **Bechhofer** [Bech54] und **Gupta** [Gupt65] zurück und verweisen bei der Voraussetzung der Normalität der Verteilungen der Systemalternativen nur auf den Zentralen Grenzwertsatz

Einen Einstieg mit Grundlagen und Anwendungen der MMV-Verfahren bieten die Arbeiten von **Hochberg** und **Tamhane** [HoTa87] sowie **Hsu** [Hsu96], ohne jedoch speziell auf die Anwendung bei der Simulationsauswertung einzugehen. Allerdings wird bei einem großem Teil der Standardwerke, wie bei **Law** [Law07:561], einerseits auf die „multiple comparison with the best (MCB)“ Verfahren als „Verbindung“ (nach **Kim** und **Nelson** [KiNe06]) zwischen den R&S- und den MMV-Verfahren verwiesen, andererseits nur die Bonferroni-Ungleichung als MMV-Verfahren für die robuste Abschätzung der Alpha-Fehler-Kumulierung empfohlen.

Die relevanten Defizite der Verfahren zum Vergleich von simulierten Systemalternativen sind einerseits der einseitige Fokus auf die R&S-Verfahren, wobei meist für die Voraussetzung der Normalität nur auf den Zentralen Grenzwertsatz verwiesen wird. Außer der Bonferroni-Ungleichung gibt es wenig Anwendung der MMV-Verfahren und keine Kombination von R&S- mit MMV-Verfahren.

2.7.2 Auswertungsmethodiken

Als Methodik wird hier die Gesamtheit der Methoden einer spezifischen Vorgehensweise zur statistischen Auswertung bezeichnet.

Law und **Kelton** [LaKe00] bzw. **Law** [Law07] gehen in der dritten und vierten Auflage des Standardwerks zur Simulationsmethodik „*Simulation Modeling and Analysis*“ explizit auf die richtige Vorgehensweise zur statistischen Auswertung ein. Bei der Ergebnisanalyse eines Systems werden die Vorgehensweisen und Verfahren für Simulation mit und ohne festes Ende beschrieben. Es wird anhand einer Vielzahl von Verfahren und Beispielen aufgezeigt,

- wie Mittelwerte und die zugehörigen Konfidenzintervalle bzw. Stichprobenumfänge geschätzt werden, ebenso Schätzungen für Wahrscheinlichkeiten und Quantile,
- wie bei Simulationen ohne festes Ende die Einschwingphase bestimmt und die Autokovarianz bzw. Autokorrelation eliminiert wird.

Bei den Verfahren zum Vergleich von Systemalternativen

- werden das gepaarte-t- („paired-t“) und das Zweistichproben-t Konfidenzintervall mit Freiheitsgraden nach **Welch** [Welc83] für zwei Alternativen beschrieben,
- wird für mehrfache Vergleiche die Kombination mit der Bonferroni-Ungleichung empfohlen sowie
- werden die „multiple comparison with the best (MCB)“ Verfahren als Alternative erwähnt und auf die entsprechende Literatur von **Hsu** [Hsu96] sowie **Hochberg** und **Tamhane** [HoTa87] referenziert sowie
- mehrere Verfahren für „Ranking and Selection“ vorgestellt.

Die Defizite bei der Vorgehensweise nach Law (und Kelton) bestehen erstens in dem Verweis auf die angebliche Robustheit der t-Statistik bei Abweichungen von der Normalverteilung, solange keine Stichproben nach Art einer Weibull-Verteilung vorliegen. Zweitens werden beim Vergleich von Alternativen nur Verfahren für den paarweisen Vergleich sowie die Bonferroni-Ungleichung beschrieben und ansonsten die R&S-Verfahren empfohlen. Weitere geeignete MMV-Verfahren werden nicht dargestellt.

In einem anderen Grundlagenwerk, dem „*Handbook of Industrial Engineering*“, beschreibt **Nelson** [Nels01] in dem Kapitel „*Statistical Analysis of Simulation Results*“ Richtlinien und Verfahren für den Entwurf und die Auswertung von Simulationsexperimenten. Kurz, praxisnah und mit Beispielen wird die Auswertung von einzelnen Simulationsmodellen sowie der Vergleich von alternativen Systemen beschrieben. Nützlich sind dabei die prägnanten „*Design and Analysis Principles*“.

Nelson empfiehlt zur statistischen Auswertung die Nutzung spezieller Statistikwerkzeuge als Alternative zu den von Simulationswerkzeugen generierten Ergebnissen. Diese spezialisierten Auswertungswerkzeuge

- bieten mehr und speziellere statistische Verfahren zur Auswertung an,

- benutzen dabei die „rohen“, nicht aggregierten Daten für eine Überprüfung von Bedingungen, bspw. die der Normalverteilung, und
- können bei einer iterativen Vorgehensweise die Daten aus mehreren Simulationsläufen effizient nutzen.

Im Gegenzug dazu werden Probleme beim Datenmanagement und der Schnittstelle zu Simulationswerkzeugen aufgeführt.

Die Vorgehensweise nach Nelson beinhaltet

- Verfahren für Punktschätzer wie Mittelwerte, Wahrscheinlichkeiten und Quartile,
- beschreibt Methoden zum Bestimmen und Eliminieren des Einflusses des Einschwingverhaltens, besonders bei Simulationen ohne festes Ende,
- zeigt Verfahren zur Bestimmung von Konfidenzintervallen zur Bestimmung der Ergebnisqualität,
- nutzt ein Verfahren der Mehrfachen Mittelwertvergleiche (MMV) für den Vergleich alternativer System von **Hsu** [Hsu96] und zur Vorbereitung von Metamodellen und Optimierungen und endet mit
- dem „Common Random Number (CRN)“ Verfahren als einer Methode zur Varianzreduktion.

Nelsons Auswertungsmethodik nutzt einfache und relativ leicht nachvollziehbare Verfahren zur statistischen Auswertung. Als entscheidendes Defizit geht er allerdings von einer Normalverteilung der Stichproben aus. Ist diese Voraussetzung nicht gegeben, verweist er auf eine Vergrößerung der Stichproben, um nach dem zentralen Grenzwertsatz die Verteilung der Normalverteilung anzugleichen. Ähnlich setzt er bei der Prozedur der MMV gleiche Varianzen (Homoskedastizität) voraus und empfiehlt bei der Verletzung der Bedingung entsprechende varianzreduzierende Methoden, speziell CRN-Verfahren zur Angleichung der Varianzen.

Kleijnen [Klei87], [KlvG92], [Klei98] und [Klei08] beschreibt v.a. in „*Statistical Tools for Simulation Practitioners*“ sehr detailliert und praxisnah die Vorgehensweise zur statistischen Auswertung von Simulationsergebnissen. Dabei geht Kleijnen nicht automatisch von den „klassischen Voraussetzungen“ aus, sondern empfiehlt

- bei symmetrischen Abweichungen verteilungsfreie Verfahren und

- bei schiefen Abweichungen die modifizierte Johnsons t-Statistik [John78]

zur Berechnung von Konfidenzintervallen. Eine einfache, vor allem grafische und auf dem Schruben, Singh und Tierney-Test basierende Vorgehensweise zur Bestimmung der Stationarität wird vorgestellt. Die Gruppenmittelwerte werden nach den Empfehlungen von **Schmeiser** [Schm82] berechnet und mit dem von Neuman-Test auf Unabhängigkeit überprüft. Zusätzlich werden anspruchsvollere Verfahren wie der Renewal-Ansatz oder das Jackknifing-Prinzip diskutiert. Beim Vergleich von mehreren Systemalternativen stellt Kleijnen das Konzept der Metamodelle auf Basis der Regressionsanalyse vor. Zum einfachen Vergleich von zwei Modellen schlägt er die gedoppelte t-Statistik vor und rät zur Bonferroni-Ungleichung für mehrfache Vergleiche, da diese ohne besondere Voraussetzungen gültig, dafür aber (im statistischen Sinne) effizient ist. Als einen praktischen Nachteil der Bonferroni-Ungleichung erwähnt er die eventuell ungewöhnlichen Fehlerwerte. Als alternative MMV-Verfahren werden die Prozeduren von Scheffé und der verteilungsfreie Wilcoxon-Vorzeichen-Rangtest beschrieben.

Als einer der wenigen Autoren warnt Kleijnen vor der fehlenden Robustheit eines Teils der Verfahren bei der Verletzung ihrer Voraussetzungen, wie bspw. der Normalverteilung oder der gleichen Varianzen bei Verfahren der mehrfachen Vergleiche, speziell bei der Scheffé-Prozedur ([Klei87:341]). Ein anderes Beispiel ist die Empfindlichkeit der F- und X^2 -Statistik ([Klei87:36]) bei Abweichungen von der Normalverteilung.

Kleijnens Auswertungsmethodik ist relativ praxisnah bei der Bewertung der Ergebnisse einzelner Modelle. Ein Defizit ist allerdings, dass der Einsatz der Metamodelle der Regressionsanalyse für die Bewertung alternativer Simulationsmodelle in der täglichen Praxis zu komplex ist.

Banks, Carson II, Nelson und Nicol [BCNN10] beschreiben in der fünften Auflage von „*Discrete-Event System Simulation*“ eine einfache Vorgehensweise zur Ergebnisanalyse einzelner Modelle sowie zum Vergleich von Modellalternativen. Bei der Analyse einzelner Modelle wird die t-Statistik für Konfidenzintervalle und das Verfahren der Batch Means bei autokorrelierten Stichproben aus Simulationen ohne festes Ende anhand von einfachen Beispielen beschrieben. Zum Vergleich von Systemalternativen wird bei zwei Systemen ein Zwei-Stichproben-t Konfidenzintervall mit Freiheitsgraden nach Welch, bei mehreren Systemen die Bonferroni Ungleichung für multiple Vergleiche und als Grundlage für eine R&S-Prozedur zur Auswahl der besten Alternative empfohlen. Dazu wird beim Vergleich von Systemen das Varianzreduktionsverfahren der Common Random Numbers (CNR) beschrieben.

Die wesentlichen Defizite in dieser Vorgehensweise ergeben sich durch fehlende Verfahren zur Bestimmung der Einschwingphase (außer der Empfehlung der Glättung und grafischen Aufbereitung) sowie der stets und grundsätzlich angenommenen Voraussetzung einer Normalverteilung der Stichproben.

Alexopoulos und **Seila** [AlSe98] beschreiben mit ihrem Kapitel „Output Data Analysis“ im „*Handbook of Simulation*“ von **Banks** [Bank98] sowie leicht modifiziert und aktualisiert in einem Paper der „Winter Simulation Conference 2007“ von **Alexopoulos** [Alex07] ausführlich die Ergebnisanalyse eines Simulationsmodells. Bei Simulationen mit festem Ende wird dargestellt, wie Mittelwerte und Quantile geschätzt werden. Zur Bestimmung des Übergangs von der transienten in die stationäre Phase bei Simulationen ohne festes Ende wird neben Literaturverweisen die grafisch interaktive Prozedur von Welch [Welc83] empfohlen. Für die Bestimmung bzw. Eliminierung der Autokovarianz werden verschiedene Verfahren der Gruppenmittelwerte sowie anspruchsvollere Verfahren der Standardisierten Zeitreihen und der Spektralanalyse beschrieben. Multivariate Schätzungen mit Hilfe der Bonferroni-Ungleichung werden kurz erläutert.

Die Veröffentlichung von **Alexopoulos und Kim** [AlKi02] zeigt ein prinzipiell gleiches Vorgehen. Simulationen mit und ohne festem Ende („finite-horizon and steady-state“) werden unterschieden und es wird ausdrücklich darauf hingewiesen, dass Daten aus einem einzelnen Simulationslauf meist autokorreliert und nicht normal verteilt sind („...*dependent and not quite normal*“). Bei der Vorgehensweise zum Vergleich von Alternativen wird die obige Vorgehensweise von **Alexopoulos** und **Seila** [AlSe01] mit zwei auf dem „Indifference Zone“ Ansatz basierenden R&S-Prozeduren sowie der Empfehlung für das „Common Random Number“ Verfahren zur Varianzreduktion ergänzt.

Als Defizit wird bei allen Verfahren stets die Annahme einer Normalverteilung der Stichproben vorausgesetzt. Nichtnormale und autokorrelierte Daten sollen mit der Methode der Gruppenmittelwerte so aufbereitet werden, bis sie den „klassischen“ Anforderungen an Unabhängigkeit und Normalverteilung entsprechen. Beim Vergleich von Alternativen werden nur R&S-Prozeduren beschrieben.

Goldman und Nelson [GoNe98] beschreiben ebenfalls im „*Handbook of Simulation*“ [Bank98] den Vergleich von Systemalternativen. Dabei werden sowohl Verfahren für Ranking and Selection (R&S) als auch Mehrfache Mittelwertvergleiche (MMV) dargestellt, teilweise in Verbindung mit der Varianzreduktionsmethode der „Common Random Numbers“. Empfohlen werden das gepaarte t-Konfidenzintervall sowie das Tukey-Verfahren für Mehrfache Mittelwertvergleiche. Zusätzlich werden drei, teilweise zweistufige, Prozeduren als „multiple comparison with the best“ (MCB) Verfahren erläutert.

Als wichtigstes Defizit vertrauen Goldsman und Nelson bei den vorgestellten Verfahren auf die *häufig* („frequently“) erfüllte Voraussetzung der Normalverteilung beim Zusammenfassen eine großen Anzahl von Stichproben nach dem zentralen Grenzwertsatz.

In einem eher betriebswirtschaftlichen Umfeld beschreibt **Liebl** [Lieb95] recht pragmatisch die statistische Auswertung von Simulationen. Er beschreibt die Problematik der transienten Anlaufphase und beschreibt die Verfahren von Conway, Gafarian und Fishman. Konfidenzintervalle und das Verfahren der „Batch means“ werden empfohlen. Bei dem Vergleich unterschiedlicher Systemversionen folgt er den Empfehlungen von **Law und Kelton** [LaKe00] und befürwortet bei Normalverteilung und gleichen Varianzen nach dem F-Test die Konfidenzintervalle zum doppelten t-Test, bei ungleichen Varianzen den Test von Welch. Entsprechend werden für mehrfache Vergleiche die Verfahren der einfachen Varianzanalyse oder die Bonferroni-Ungleichung empfohlen.

Bei der Auswertungsmethodik nach Liebl ist vor allem die Voraussetzung der Normalverteilung der Stichproben als wesentliches Defizit zu sehen. Die Robustheit der teilweise einfach verständlichen Verfahren wird nicht diskutiert.

Fishman [Fish01] dokumentiert in seinem Grundlagenbuch zur ereignisdiskreten Simulation mehrere Verfahren für die Bestimmung der transienten Phase („warm-up intervall“) und geht detailliert auf unterschiedliche Batch Means Verfahren ein, um anschließend seinen LABATCH.2 Algorithmus zu empfehlen. Für den Vergleich von zwei Alternativen wird das Zwei-Stichproben-Konfidenzintervall mit Freiheitsgraden nach Welch beschrieben. Die Bonferroni-Ungleichung wird für multivariate Aussagen erwähnt und für mehrfache Systemvergleiche wird nur auf die Literatur von **Banks, Carson II, Nelson** und **Nicol** [BCNN09] sowie **Law** [Law07] (bzw. **Law und Kelton** [LaKe00] in jeweils älteren Ausgaben) verwiesen.

Die Defizite sind eindeutig die fehlende Erwähnung der Problematik der Normalverteilung sowie die hohe Komplexität der Batch Mean Verfahren. Für eine Integration in der Vorgehensweise der Simulationsauswertung wird nur auf zwei Standardwerke verwiesen.

Nakayama [Naka08] beschreibt in einem Artikel der Winter Simulation Conference geeignete Methoden für die Ergebnisanalyse von stochastischen ereignisdiskreten Simulationen mit und ohne festes Ende. Er zeigt die Unterschiede der transienten und stationären Phase bei der Schätzung von Leistungskennzahlen. Die Problematik der nicht normalverteilten Stichproben wird erwähnt („...output data from simulations are often not normal...“), allerdings ist der Verweis auf den Zentralen Grenzwertsatz der einzige Lösungsansatz. Konfidenzintervalle werden mit der Student-t-Statistik berechnet, mit dem Hinweis,

dass schiefe Verteilungen das Intervall verfälschen. Zum Erkennen der stationären Phase verweist er auf die Methoden von Law [Law07] und bei multivariaten Aussagen auf die Bonferroni-Ungleichung.

Bei dieser Vorgehensweise sind Nakayama zwar die Defizite bei der oft auftretenden Abweichung von der Normalverteilung bewusst, allerdings bietet er keine Lösung dafür an. Die Methoden sind nachvollziehbar und mittels Beispielen dargestellt. Eine Integration wird nicht aufgezeigt.

Auch **Ross** [Ross02] verweist auf den Zentralen Grenzwertsatz, um ggf. eine Annäherung der Beobachtungen an die Normalverteilung zu erreichen und benutzt die Student-t-Statistik für die Berechnung der Konfidenzintervalle. Zusätzlich wird ein Bootstrapping-Verfahren für die Kalkulation der mittleren Fehlerquadrate („mean square error“) der geschätzten Leistungskennzahlen vorgestellt und bei den Verfahren zur Modellvalidierung auf den Wilcoxon-Zweistichprobentest verwiesen. Weitere Verfahren für den Vergleich von Systemalternativen werden nicht erwähnt.

Als Defizite der Vorgehensweise fehlen Verfahren für die Einschwingproblematik sowie für den Vergleich der Systemalternativen. Abweichungen von den Voraussetzungen der Student-t-Statistik bzw. ihre Robustheit werden nicht erwähnt. Das Resampling-Verfahren ist relativ komplex und eine Integration in eine durchgängige Vorgehensweise ist nicht erkennbar.

Im deutschsprachigen Raum bildet vor allem die **VDI Richtlinie 3633 Simulation von Logistik- Materialfluss- und Produktionssystemen, insbesondere Blatt 3 Experimentplanung und -auswertung** [VDIF97], den Leitfaden für die praxisrelevante korrekte Durchführung und Auswertung von Simulationsexperimenten⁶. Blatt 3 der VDI Richtlinie beschreibt die Vorgehensweise zur Auswertung eines Simulationsmodells einschließlich der Bestimmung von Konfidenzintervallen sowie die Vergleiche von Alternativen mit den Verfahren der mehrfachen Mittelwertvergleiche. Bei Simulationen mit offenem Ende werden grafische Verfahren zur Bestimmung der Stationarität und Gruppenmittelwerte für unabhängige Stichproben empfohlen. Zur Berechnung von Konfidenzintervallen bei nicht normalverteilten Stichproben rät die Richtlinie zur Johnson-t-Statistik bei schiefen bzw. zu verteilungsfreien Verfahren bei symmetrischen Abweichungen. Eine ähnliche Unterscheidung wird bei den Verfahren der Mehrfachen Vergleiche empfohlen: Der Tukey-Test wird nur bei leichten Abweichungen empfohlen,

⁶ Durch die Mitwirkung bei der Gestaltung der VDI-Richtlinie sind frühe Teilergebnisse der vorliegenden Arbeit, speziell zur Methodenauswahl, bereits in stark vereinfachter Form in die VDI Richtlinie 3633 Blatt 3 eingegangen.

ansonsten soll bei stärkeren Abweichungen von der Normalverteilung der verteilungsfreie Wilcoxon-Test benutzt werden.

Die VDI-Richtlinie beschreibt in kurzer und leicht vereinfachter Form eine Vorgehensweise zur statistischen Auswertung. So wird bei der Bestimmung der Stationarität nur auf eine heuristische, grafische Vorgehensweise verwiesen und die Auswahl der Verfahren der Multiplen Mittelwertvergleiche auf zwei Fälle reduziert. Die Problematik von unterschiedlichen Varianzen der zu vergleichenden Stichproben wird nicht diskutiert.

Die **Arbeitsgemeinschaft Simulation ASIM** veröffentlichte im Rahmen der ASIM-Mitteilung Nr. 102 von **Wenzel, Weiß, Collisi-Böhmer, Pitsch und Rose** [WWCP08] eine Empfehlung für eine „qualitätssichernde Vorgehensweise“, welche sich an die VDI-Richtlinie 3633 sowie an das Vorgehensmodell und die einzelnen Methoden von **Law** [Law07] anlehnt.

Die von der ASIM Arbeitsgruppe empfohlene Vorgehensweise bestätigt die fehlende Voraussetzung von unabhängigen und normalverteilten Stichproben (vgl. Kap.1). Als Defizite dieser Verfahren wird bei der Normalverteilung nur auf eine grobe Näherung verwiesen, falls die Stichproben aus Mittelwerten von unabhängigen Simulationsläufen bestehen. Dabei wird auch hier mit dem „Zentralen Grenzwertsatz“ und einer „großen“ Anzahl von Stichproben argumentiert. Bei autokorrelierten Daten sollen die Konfidenzintervalle nur als Näherungswerte betrachtet werden. Die Problematik des Vergleichs von Systemalternativen wird nicht angesprochen.

Die Untersuchung des Stands der Forschung bei Auswertungsmethodiken zeigt kein einzelnes eindeutig richtiges Vorgehen zur statistischen Auswertung von Simulationsergebnissen⁷. Die jeweiligen Autoren beschreiben unterschiedliche Schwerpunkte und vernachlässigen andere meist wesentliche Teilaspekte. Notwendige Robustheitsuntersuchungen für die Gültigkeit der Verfahren sind selten. Viele der entwickelten Verfahren sind von komplexer Struktur und lassen sich für industrielle Anwender nur schwer nachvollziehen.

2.7.3 Rechnerunterstützte Auswertungswerkzeuge

In den letzten Jahren gab es mehrere Initiativen, welche die Aufgabenstellung der statistischen Auswertung von Simulationsergebnissen mit rechnerbasierten Verfahren, Methoden und Werkzeugen unterstützten. Die wichtigsten Ansätze dieser Auswertungswerkzeuge werden vorgestellt und diskutiert.

⁷ Einen guten und aktuellen Überblick über den internationalen Stand der Forschung in Bezug auf Methodik, Verfahren und Werkzeuge zur statistischen Auswertung von Simulationsergebnissen bietet die alljährliche „**Winter Simulation Conference**“ [WSC10] in den USA.

Scheifele [Sche86] beschreibt als einer der ersten einen Postprozessor für die Simulationssprache SLAM, der unter der Voraussetzung der Normalverteilung einen mehrfachen Vergleich von alternativen Szenarien mit den MMV-Verfahren von Tukey, Scheffé oder Bonferroni durchführt.

Allerdings werden keine Verfahren für die Schätzung von Parametern und den Vergleich von Mittelwerten aus nicht normalverteilten Stichproben angeboten, sondern als zukünftige Erweiterung vorgeschlagen. Der Anwender wird zumindest ansatzweise bei der Beurteilung der Varianzunterschiede in den Auswertungsprozess eingebunden.

Haddock [Hadd87] zeigt die Möglichkeiten der statistischen Ergebnisanalyse in einem integrierten System auf. Dabei nutzt das System den SIMAN Output Processor und integriert diesen mit einer Fortran-basierten Applikation in ein sogenanntes „Expert System Framework“. Die in Flussdiagrammen beschriebene Vorgehensweise orientiert sich stark an der Methodik der ersten Ausgabe des Buchs von **Law und Kelton** [LaKe00].

Es werden bei Haddock nur „Standardverfahren“ für die „klassischen Bedingungen“ in einem komplett automatisierten Ablauf eingebettet. Die Benutzereinbindung ist relativ einfach: es werden nur rudimentäre grafische Darstellungen angeboten, eine Organisation der Daten ist nicht vorhanden. Es wird eine Normalverteilung der Ergebnisstichproben vorausgesetzt, besondere robuste Verfahren werden nicht angeboten. Der Vergleich der Ergebnisse mehrerer Modelle wird nicht unterstützt.

Die Veröffentlichungen von **Ramachandran, Kimbler und Naadimuthu** [RaKN88] sowie von **Taylor und Horrión** [TaHo88] beschreiben nur konzeptionelle Ansätze für die Unterstützung eines Simulationsanwenders und die Notwendigkeit einer Sensitivitätsanalyse, ohne auf konkrete Vorgehensweisen, Methoden oder Verfahren einzugehen.

Mellichamp und Young [MeYo89] demonstrieren mit dem „*Statistical Expert System for Simulation Analysis*“ (SESSA) ein frühes regelbasiertes System für die statistische Simulationsauswertung. Das System wurde in einer Kombination aus einer regelbasierten Umgebung und der Programmiersprache Basic implementiert. Es bietet verschiedene Methoden zur Auswertung der Ergebnisse an, der Benutzer muss entscheiden, welche Voraussetzungen erfüllt sind und das entsprechende Verfahren auswählen. Der Vergleich von Alternativen mit der Student-t-Statistik wird unterstützt.

Trotz der Vielzahl der angebotenen Algorithmen (u.a. acht Verfahren zur Bewertung des Einschwingverhaltens, 22 Verfahren zur Bestimmung der Stichprobengröße) wird nur die Voraussetzung der Normalverteilung überprüft. Verfahren für nichtnormalverteilte Stichproben werden nicht angeboten. Eine

Einbindung des Anwenders oder eine durchgängige Gesamtvorgehensweise werden nicht dargestellt.

Pierreval und Deslandres [PiRD90], [DePi91] implementierten mit ihrem KB3S (Knowledge Based System for Simulation Statistics) mit Hilfe einer Expertensystem-Shell ein komfortables Unterstützungswerkzeug zur statistischen Auswertung. KB3S hat eine Schnittstelle zu dem Simulationswerkzeug SIMAN und nutzt teilweise dessen „Output Processor“.

Der Schwerpunkt des Forschungsprototyps KB3S liegt eher bei der Konzeption des wissensbasierten Systems. Der Anwender wird bei KB3S nur wenig in die Entscheidungen des stark automatisierten Systems eingebunden, er hat bei einer Erläuterungskomponente keinen Einfluss auf die Auswahl der Methoden und der getroffenen Entscheidungen. Der Forschungsprototyp bietet nur reduzierte grafische Auswertungen.

Mollamustafaoglu, Gurkan und Ozge [MoGO93] konzipieren ein objektorientiertes Auswertungswerkzeug für Simulationsergebnisse und lehnen sich ebenfalls dabei an die Auswertungsmethodik von **Law und Kelton** [Lake00] an. Dabei werden die Möglichkeiten eines objektorientierten Designs wie Kapselung und Vererbung intensiv genutzt.

Der Schwerpunkt der Arbeiten liegt auf dem objektorientierten Entwurf des Werkzeugs als auf der Auswahl und dem Einsatz der „richtigen“ statistischen Verfahren. Wichtig ist v. a. die Auswertung einzelner Simulationsmodelle, wobei eine Versuchsplanung unterstützt wird.

Tao und Nelson [TaNe97] formulieren mit dem „Simulation Experiment Design and Analysis (SEDA)“ ein Rahmenwerk („Framework“) für die computerunterstützte Simulationsanalyse und den damit verbundenen Problemlösungsprozess. Dabei wird ein hierarchisches Klassifikationsmodell entwickelt, welches Probleme und Aufgaben der Simulationsauswertung sowie ein Auswahl von Auswertungsprozeduren beinhaltet. Der Schwerpunkt der Auswertungsverfahren liegt beim Erkennen der transienten Phase und dem Vergleich von Systemalternativen.

Die beschriebene Hierarchie der Auswertungsprozeduren beinhaltet die Überprüfung der Annahmen der Unabhängigkeit und Normalverteilung der Beobachtungen. Es werden allerdings keinerlei Verfahren beschrieben, welche bei einer Verletzung dieser Voraussetzungen geeignet sind. Der Simulationsanwender wird nur minimal eingebunden und bei der Vorgehensweise besteht eine Anbindung nur an die Aufgabe der Versuchsplanung.

Mit dem Werkzeug *ToolSuite* entwickelte **Le-Min Yin** [Yin98] einen „*analytischen Werkzeugkasten für die Analyse von stochastischen Simulationsexperimenten*.“ *ToolSuite* besteht aus einem Teil zur Schätzung der Anzahl von notwendigen Stichproben bzw. von Konfidenzintervallen. Das zweite Teilwerkzeug stellt „*Ranking and Selection*“ Prozeduren bereit, wozu ein weiterer Teil eine Sensitivitätsanalyse nach der *Surface Response* Methode unterstützt. *ToolSuite* nutzt die entsprechenden Prozeduren von Law und Kelton [LaKe00] und baut auf dem Statistikwerkzeug SAS auf. Der Vorteil von *ToolSuite* ist seine Unabhängigkeit von einem Simulationswerkzeug, über einfache Routinen können Simulatoren eingebunden werden. Eine Visualisierung der Ergebnisse erfolgt mit den Routinen von Gnuplot.

Die in *Toolsuite* implementierten Prozeduren basieren auf der Erfüllung der „klassischen Voraussetzungen“ und benötigen SAS als Basis. Insgesamt ist *Toolsuite* als einfacher Forschungsprototyp konzipiert und implementiert.

Das Projekt „AutoSimOA - Automating Simulation Output Analysis“ von **Hoad, Robinson und Davies** [HoRD09] an der Universität Warwick in Zusammenarbeit mit der Simulationssoftwareunternehmen Simul8 beschreibt die Entwicklung eines automatisierten Simulationsergebnisanalyse-Werkzeugs. Das Softwarewerkzeug und die implementierten statistischen Methoden konzentrieren sich auf die Bestimmung der transienten Phase und der Anzahl der Replikationen bzw. der Simulationsdauer aus einer vorgegeben Breite des Konfidenzintervalls.

Die Defizite des Forschungsprototyps sind dabei im Wesentlichen das Fehlen von Methoden für eine Abweichung von der Normalverteilung und von robusten Verfahren. Der Vergleich von Systemalternativen wird nicht unterstützt. Interessanterweise wird bei dem Werkzeug AutoSimOA ein expliziter „Black-Box“ Ansatz ohne Entscheidungsoptionen für den Anwender („Automatisierung“) umgesetzt.

2.8 Statistische Auswertung bei kommerziellen Werkzeugen

Die statistische Auswertung der Ergebnisse wird von den meisten kommerziellen Simulationswerkzeugen nur unzureichend unterstützt. Schmeiser [Schm01] schreibt hierzu:

„Similarly, most commercial simulation environments support analysis of the output data only minimally; usually calculation of the sample mean and sometimes calculation of the standard deviation and sometimes a histogram or empirical cumulative distribution function.“

Wenigstens ein Teil der kommerziellen Simulationswerkzeuge bietet eine zumindest partielle Unterstützung bei der statistischen Auswertung der Ergebnisse an. Dabei gibt es drei prinzipielle Ansätze:

1. Es sind komfortable Ergebnisanalyse-Module für eine statistische Auswertung verfügbar.
2. Es werden nur einfache deskriptive Statistiken wie Mittelwert, Varianz, Minimal- und Maximalwerte sowie teilweise Konfidenzintervalle zur Auswertung einzelner Replikationen angeboten.
3. Es werden Schnittstellen zu Tabellenkalkulationen und speziellen Statistikwerkzeugen, wie SPSS oder SIMSTAT, bereitgestellt.

Vor dem Hintergrund dieser groben Einteilung werden beim Stand der Technik die Eignung zur statistischen Auswertung von Simulationsergebnissen sowohl von Simulations- als auch von speziellen Statistikwerkzeugen diskutiert und die vorhandenen Defizite aufgezeigt.

Eine beispielhafte Bewertung der statistischen Ergebnisanalyse von vier Simulations- und einem Statistikwerkzeug (AutoMod / AutoStat, ARENA, ProModel und Taylor II⁸ sowie SIMSTAT) beschreibt **Banks** [Bank96], indem er dazu eine Liste von zu erfüllenden Anforderungen erstellt. Die von Banks aufgelisteten funktionalen Anforderungen sind in die Kategorien

- Statistiken für einzelne Variablen,
- Statistiken zum Vergleich von mehreren Variablen,
- Statistische Grafiken,
- Business Grafiken und
- Datenmanagement

gegliedert. Als Ergebnis der Studie von Banks ergeben sich wesentliche Defizite der Simulationswerkzeuge

- bei den Verfahren zur Bestimmung der Stationarität,
- beim Vergleich von mehreren Variablen sowie
- bei den statistischen Grafiken.

Allerdings geht Banks wie bei der Auswertungsmethodik (s. o.) stets von der Annahme der Normalverteilung der Stichproben aus und bindet die bei einer

⁸ mittlerweile als Enterprise Dynamics bezeichnet.

abweichenden Verteilung notwendigen, alternativen Verfahren nicht in seine Anforderungsliste mit ein.

2.8.1 Simulationssysteme

Bei der Diskussion der Simulationssysteme fokussiert die Analyse auf in der industriellen Praxis bekannte und verbreitete Simulationspakete. Dabei kommt den Programmen „Plant Simulation“ (früher bekannt als em-Plant bzw. Simple++) und „QUEST“ als (ereignisdiskrete) Simulationsmodule der integrierten Fabrikplanungssystemen der beiden führenden Anbieter Dassault-Delmia und Siemens-Tecnomatix eine besondere Bedeutung zu. Des Weiteren werden noch die verbreiteten Werkzeuge Arena, AutoMod, Dosimis-3, Enterprise Dynamics, ProModell, SimProcess und Witness betrachtet. Einen guten Überblick über die Simulationswerkzeuge und insbesondere über die teilweise angebotenen Zusatzmodule zur Auswertung wie „ExpertFit oder Stat::Fit“ bietet **Banks** [Bank98] in dem Kapitel „Software for Simulation“ des „Handbook of Simulation“.

ARENA [KeSS98] als Weiterentwicklung der bekannten Simulationssprache **SIMAN** bietet ab der Version 3.0 und dem enthaltenen **Output Analyser** eines der derzeit leistungsfähigsten Werkzeuge zur statistischen Auswertung. Der Arena Output Analyser erfüllt einen erheblichen Teil der Anforderungen und bietet entsprechende Verfahren zur Bestimmung der Stationarität sowie für Autokorrelation und Stichprobenumfang mit der Methode der Gruppenmittelwerte. Konfidenzintervalle können mit der t -Statistik berechnet werden. Zur Erkennung signifikanter Unterschiede zweier Szenarien werden der „gedoppelte“ (auch in Kombination mit Common Random Numbers Verfahren) und der „Zweistichproben“ t -Test (auch bei ungleicher Stichprobenanzahl) (engl. „paired-t Test“ bzw. „Two-Sample-t Test“) angeboten. Zum Vergleich mehrerer Szenarien ermöglicht Arena sogar eine einfache ANOVA mit den post-hoc Verfahren der MMV von Bonferroni, Scheffé und Tukey.

Als wesentliches Defizit setzen die von ARENA Output Analyser angebotenen Verfahren allerdings stets eine Normalverteilung der Stichproben voraus bzw. verlassen sich auf die Robustheit der t -Statistik bei nur geringen Abweichungen. Verfahren für Stichproben mit einer deutlichen Abweichung von einer Normalverteilung werden nicht angeboten.

AutoStat [Cars96], [Bank00] als ein Auswertungsmodul für das Simulationswerkzeug **AutoMod** bietet Unterstützung bei der statistischen Auswertung von einzelnen Replikationen und sogenannten Szenarien. Es unterstützt die Berechnung von Konfidenzintervallen mit der t -Statistik, bietet eine Ranking-and-Selection Prozedur nach Law sowie das Verfahren zur Bestimmung der

stationären Phase nach Welch. Die Ergebnisse können in Tabellenkalkulationsprogramme oder andere Analysewerkzeuge exportiert werden.

Die relevanten Defizite bei AutoStat sind die stets vorausgesetzte Normalität der Ergebnisstichproben sowie die fehlenden Optionen beim Vergleich von Systemalternativen.

Plant Simulation (früher als **em-Plant** bzw. **Simple++** bezeichnet) [Siem10] wurde am Fraunhofer IPA entwickelt und ist mit der objektorientierten Architektur eines der modernsten Simulationswerkzeuge und Bestandteil der Siemens PLM Software.

In der Version 9 beinhaltet Plant Simulation [Siem08] mehrere statistische Werkzeuge als „Add-Ins“ zur Auswertung der Simulationsergebnisse. Diese Zusatzmodule bieten Unterstützung, um sowohl empirische Eingangsdaten zu modellieren („Distribution Fitting mit DataFit“) als auch bei der statistischen Auswertung mit der Berechnung eines Konfidenzintervalls, der Parameterbestimmung von beschränkten Verteilungen, einem Unabhängigkeitstest sowie Verfahren für die Regressions- und Varianzanalyse.

Dabei wird die Berechnung der Konfidenzintervalle mit einem Kommentar ergänzt: *„Die Berechnung der Konfidenzintervalle setzt voraus, dass die Stichprobe von einer normalverteilten Zufallszahl stammt. Diese Voraussetzung ist für die häufigsten Anwendungen bei der Auswertung von Simulationsergebnissen erfüllt.“*

Zur Beurteilung, ob eine empirische Verteilung als normalverteilt betrachtet werden kann, wird bei Plant Simulation ein Histogramm empfohlen und die Lagparameter Skewness und Kurtosis der Verteilung können berechnet werden. Zur „Validierung“ des Konfidenzintervalls wird ein Bootstrapping-Verfahren angeboten.

Des Weiteren wird eine einfache Varianzanalyse (ANOVA) für eine Hypothese der gleichen Mittelwerte und ein X^2 -Test für die Untersuchung der Unabhängigkeit von zwei Gruppen von Stichproben zur Verfügung gestellt.

Für Simulationen ohne festes Ende wird eine „Sequentielle Stichprobenentnahme“ offeriert. Mit einer vorgegebenen Abschneidegrenze für die transiente Phase werden die Beobachtungen mit Batch Means gruppiert. Die Simulation wird angehalten, wenn eine vorgegebene KI-Breite errichtet wird.

Trotz der guten Ansätze der Unterstützung der Auswertung durch die angebotenen statistischen Verfahren mit einer guten Einbindung in die Simulationsumgebung und der Vorgehensweise ergeben sich folgende wesentliche Defizite:

- keine Verfahren für nichtnormale Verteilungen sowie fehlende bzw. unklare Robustheit der angebotenen Verfahren,

- die Varianzanalyse für den Vergleich von Alternativen beinhaltet keine Post-Hoc Tests bzw. MMV,
- eine teilweise unzureichende Nachvollziehbarkeit für den Anwender und
- verbesserungsfähige Einbindung des Anwenders, da die Voraussetzungen und Konsequenzen der Entscheidungen nicht transparent sind.

Das Simulationswerkzeug **QUEST** [DELM08] als Teil des Fabrikplanungssystems („Digital Manufacturing Solutions“) von Delmia wird in der industriellen Praxis häufig eingesetzt. Einerseits bietet QUEST eine gute Unterstützung bei der statistischen Versuchsplanung („Design of Experiments“), andererseits aber nur wenig Hilfe bei der statistischen Auswertung. Die Berechnung von Konfidenzintervallen (mit der t-Statistik) für vorgegebene Wahrscheinlichkeiten sowie alternativ die Anzahl der Stichproben für eine gewünschte Genauigkeit der Schätzung wird unterstützt. Interessanterweise werden die Konfidenzintervalle dabei als „Alternative“ zu Punktschätzern bezeichnet. Für einen Vergleich von Alternativen wird auf die Anbindung an OptQuest für Optimierungsaufgaben verwiesen.

Als wesentliches Defizit fehlt bei QUEST, abgesehen von der Berechnung der Konfidenzintervalle, eine weitere Unterstützung der statistischen Auswertung.

Dosimis-3 [SDZ07] als eines der bekannten Simulationswerkzeuge für Materialfluss und Logistik im deutschsprachigen Raum bietet praktisch keine Unterstützung für die statistische Auswertung. Es werden nur einfache Mittelwerte und Histogramme dargestellt. Als Alternative wird auf eine Auswertung über die Schnittstelle zu MS-Excel verwiesen, welche aber entsprechend konfiguriert werden muss.

Enterprise Dynamics von Incontrol Simulation Solutions [Inco04] bietet Konfidenzintervalle auf Basis der t-Statistik sowie die Berechnung des Korrelationskoeffizienten und eines Korrelationsplots.

Das Simulationswerkzeug **ProModel** [ProM10] und die Variante **ProcessModel** verweisen bei der statistischen Auswertung auf die drei Zusatzmodule „Stat::Fit“, „OptQuest“ und „SimRunner“. Allerdings wurde Stat::Fit für die Eingangsdatenmodellierung und Anpassung von Verteilungen entwickelt und ist für die statistische Auswertung nur am Rande nutzbar.

Das Simulationssystem **SIMPROCESS** [CACI08] verweist für die statistische Auswertung auf die zwei Werkzeuge „ExpertFit“ und „OptTek“ (jeweils vergleichbar mit „OptQuest“ und „Stat::Fit“, s.o.), welche allerdings ebenfalls für die Eingangsdatenmodellierung bzw. Optimierung gestaltet sind. Für die eigentliche

statistische Ergebnisauswertung gibt es keine explizite Unterstützung, wobei einige Verfahren von ExpertFit auch für die Auswertung nützlich sind.

WITNESS bietet mit den Zusatzmodulen „Optimizer“ und „Scenario Manager“ die Berechnung von Konfidenzintervallen mit der t-Statistik an. Für eine weitere statistische Auswertung der Ergebnisse, wie z.B. bei Six Sigma Projekten, wird auf die Anbindung des Statistikwerkzeugs „Minitab“ verwiesen [Wall09].

Zusammenfassend bieten nur die Simulationswerkzeuge Arena und Plant Simulation sowie ansatzweise AutoMod mit AutoStat zumindest in Teilen eine Unterstützung des Anwenders bei der statistischen Auswertung. QUEST, Enterprise Dynamics und WITNESS ermöglichen die Berechnung von Konfidenzintervallen, allerdings ohne auf die notwendigen Voraussetzungen hinzuweisen. Dosimis-3, ProModel und SIMPROCESS verwiesen auf externe Zusatzmodule wie Stat::Fit, ExpertFit oder MS-Excel.

2.8.2 Statistikwerkzeuge

Die Diskussion der Statistikwerkzeuge betrifft die beiden leistungsfähigen und bekannte Analysesysteme SPSS und SAS, das kompaktere Werkzeug SIMSTAT sowie die verbreitete Tabellenkalkulation Excel.

SPSS [IBMC10] von IBM SPSS Inc. als ein verbreitetes und leistungsfähiges Statistikwerkzeug bietet eine Vielzahl von statistischen Prozeduren, eingebettet in eine komfortable Benutzungsumgebung. Grafische Darstellungen wie Liniendiagramme, Histogramme und Boxplots sind verfügbar. Parameterschätzungen, bspw. für Mittelwerte oder Mediane, werden einschließlich der zugehörigen Konfidenzintervalle angeboten.

Für den Vergleich von Mittelwerten nach einem einfaktoriellen ANOVA-Modell bietet SPSS in der Version 19 als post-hoc Statistik mehrere Verfahren. Unter der Prämisse gleicher Varianzen stehen vierzehn Verfahren, einschließlich der Bonferroni- und der Šidák-Ungleichung sowie den Verfahren von Tukey und Scheffé zur Verfügung. Bei der Annahme ungleicher Varianzen bietet SPSS die Verfahren T2-Tamhane, T3-Dunnett, Games-Howell und C-Dunnett an.

SPSS ist ein mächtiges Statistikwerkzeug, allerdings unterstützt es nur unzureichend bei der statistischen Aufbereitung von Daten aus Simulationen mit der Problematik der autokorrelierten und nicht normalverteilten Stichproben. Die Konfidenzintervalle nach der t-Statistik verlassen sich auf deren Robustheit, alternative Berechnungen für Konfidenzintervalle werden nicht angeboten. Die Verfahren der Zeitreihenanalyse in SPSS sind für die Ergebnisauswertung für Simulationen unkomfortabel bis ungeeignet. Bei den post-hoc Verfahren der ANOVA wird nirgends auf die, teilweise notwendige, Voraussetzung der Normalverteilung hingewiesen.

SIMSTAT [Prov05] ist ein ähnlich benutzerfreundliches Statistikwerkzeug wie SPSS, das sich auch bei der Benutzerführung und grafischen Aufbereitung an SPSS anlehnt. Simstat bietet in der Version 2.5 neben der üblichen deskriptiven Statistik zum Schätzen der Verteilungsparameter wie SPSS eine einfache ANOVA mit den post-hoc Tests von Scheffé, Tukeys HSD (honestly significant difference), Newman-Keuls und LSD (least significant difference) sowie eine große Anzahl von Verteilungsfreien Verfahren.

Als Defizit ist SIMSTAT trotz seiner Vielzahl an statistischen Verfahren, vergleichbar mit SPSS, nur begrenzt für die spezielle Problematik der statistische Auswertung von häufig nicht normalverteilten Simulationsdaten geeignet.

Das Statistikprogramm **SAS/STAT** [SASI10a] ist ein leistungsfähiges Analysewerkzeug, vergleichbar mit SPSS. Ebenso wie SPSS ist es für allgemeine statistische Aufgaben ausgelegt und kann bei der Auswertung von Simulationsergebnissen nur teilweise von Nutzen sein. SAS/STAT bietet paired-t-Tests und robuste Methoden für Datensätze mit Ausreißern. Konfidenzintervalle für Parameter und Quantile werden aber unter der Voraussetzung einer Normalverteilung geschätzt. Vergleichbar mit SPSS bietet SAS/STAT Prozeduren für eine ANOVA mit elf Verfahren der MMV als Post-hoc-Test, darunter Dunnett, Duncan, Tukey-Kramer, Scheffé, Gabriel und Hochberg, sowie die Bonferroni- und der Šidák-Ungleichungen [SASI10b]. Bei der Klassifikation dieser Verfahren wird dabei auf Hsu [Hsu96] referenziert.

Die Defizite bei SAS/STAT sind wie das Programm selbst mit denen von SPSS vergleichbar. Robuste Verfahren für autokorrelierte und nicht normalverteilte Beobachtungen sind nicht explizit verfügbar. Durch die Komplexität der statistischen Prozeduren leidet teilweise auch die Nachvollziehbarkeit und Plausibilität.

Das Tabellenkalkulationswerkzeug **Microsoft Excel** [MiCo11] bietet mit dem „Add-In“ „Analysefunktionen“ verschiedene statistische Verfahren (in der Version 2003). Es werden eine ein- und zweifaktorielle Varianzanalyse (ANOVA), eine Schätzung der Populationskenngrößen mit einem Konfidenzintervall auf Basis der t-Statistik für den Mittelwert, Histogramme sowie der Zweistichproben-F-Test und der Zwei-Stichproben-t-Test bei gleichen und ungleichen Varianzen angeboten.

Bei Excel sind die Defizite für den Einsatz zur Auswertung von Simulationsergebnissen offensichtlich. Die Berechnung der Tests und der Varianzanalyse erwartet unabhängige und normalverteilte Daten. MMV als Post-hoc Test werden nicht angeboten.

Keines der vier Statistikwerkzeuge ist explizit für die Auswertung von Simulationsdaten gestaltet, die angebotenen Verfahren sind häufig nicht robust bei

Abweichungen von den Voraussetzungen und geeignete Alternativen sind meist nicht verfügbar. Die Plausibilität der Verfahren ist für einen industriellen Anwender ebenfalls eher durchschnittlich. Eine Einbindung in den Ablauf der Simulationsmethoden ist nicht erkennbar.

2.9 Stand der Praxis

Praxisnahe Veröffentlichungen und anwendungsorientierte Projekte beschreiben selten eine korrekte statistische Auswertung. Teilweise sind auch Standardwerke zur Simulation im deutschsprachigen Raum in Bezug auf eine methodisch korrekte statistische Auswertung, auch bei nicht normalverteilten oder abhängigen Stichproben, nicht aussagekräftig.

Kuhn u.a. [KuWR93] als Herausgeber des „*Handbuchs Simulationsanwendungen in Produktion und Logistik*“ beschreiben nur kurz die Problematik der statistischen Auswertung, ohne auf konkrete Verfahren einzugehen.

Einige praxisnahe Veröffentlichungen zur Anwendung der Simulation wie bspw. jene von **Košťuariak und Gregor** [KoGr95] beschreiben u.a. detailliert die statistische Modellierung der Eingangsparameter, vernachlässigen aber die statistische Auswertung der Ergebnisse fast vollständig.

Der **ASIM-Leitfaden für Simulationsbenutzer in der Produktion und Logistik** [ASIM97] ist beim Thema statistische Auswertung zurückhaltend. Es werden zusätzlich zur Berechnung des Mittelwerts auch noch Minimal- und Maximalwerte sowie Streuung und Varianz für genauere Aussagen empfohlen. Konfidenzintervalle als Gütekriterium für Parameterschätzungen oder Verfahren für den Vergleich von alternativen Modellen werden nicht erwähnt.

Kuhn und Rabe [KuRa98] als Herausgeber der Fallbeispielsammlung „*Simulation in Produktion und Logistik*“ informieren oberflächlich über die Notwendigkeit der statistischen Auswertung und zeigen keinerlei Hilfestellung auf.

Küll und Stähly [KüSt99] beschreiben das Zusammenspiel zwischen der strategischen Versuchsplanung und der eher taktischen statistischen Auswertung einzelner Simulationsläufe. Verschiedene, teilweise anspruchsvolle Verfahren werden angesprochen, für eine operative Vorgehensweise wird jedoch auf die zitierte Literatur verwiesen.

Wichtige Hinweise für die Beschreibung von Anforderungen an eine verbesserte statistische Auswertung kommen u.a. von Studien und Anwenderbefragungen.

Die Studie und Anwenderbefragung von **Heitmann u. a.** [HHRW97] bei rund 400 deutschen Unternehmen zeigt noch erheblichen Verbesserungsbedarf bei der Experimentdefinition, Ergebnisauswertung und -darstellung.

Die Studien von **Mackulak u.a.** [MaCS94] in den USA sowie von **Hlupic** [Hlup99] in Europa, mit denen wichtige Erwartungen an eine industrielle Simulationsumgebung erhoben wurden, ergeben einen vergleichbaren Bedarf bei der „Output Statistics Analysis“ bzw. „Analysis of Results“, vor allem beim Testen von Hypothesen, der Ergebnisinterpretation und dem Berechnen von Konfidenzintervallen.

Vor diesem Hintergrund diskutieren **Sanchez u.a.** [SCHH94] die Anforderungen an zukünftige Entwicklungen für die Unterstützung der statistischen Ergebnisauswertung. Dabei wird sowohl die Problematik der Normalverteilung der Stichproben und die Plausibilität der Verfahren betont als auch die Integration in eine durchgängige Vorgehensweise gefordert.

Nikoukaran u.a. [NiHP99] präsentieren eine Kriterienliste für die Auswahl von Simulationssoftware, welche auch detaillierte Anforderungen an die statistische Auswertung, die Einbindung des Anwenders sowie an vor- und nachgelagerte Verfahren beinhaltet.

Tao und Guo [TaGo01] entwickeln für ihr Trainingswerkzeug „empirische Prinzipien für die statistische Analyse“ („Statistical Analysis Experimental Principles“) für eine verbesserte Vorgehensweise. Insbesondere das Prinzip 6-2 ist wesentlich und konkretisiert eine wichtige Anforderung:

„Use statistical procedures within user’s capabilities.“

Weitere relevante Prinzipien empfehlen visuelle Methoden und deren Bestätigung mit (statistischen) Test-Prozeduren.

Auch **Hollocks** [Holl01] Studie zum Verhalten der europäischen Simulationsanwender zeigt einen erheblichen und konkreten Bedarf an robusten, praktikablen und nachvollziehbaren statistischen Auswertungsverfahren, vor allem für „Nicht-Spezialisten“.

Der Stand der statistischen Auswertung von Simulationsergebnissen bei praxisrelevanten industrienahen Simulationsstudien lässt sich u.a. an Hand der veröffentlichten Handbücher, Richtlinien, Ergebnisse von Arbeitskreisen und anwendungsorientierten Veröffentlichungen beurteilen. Aus den evaluierten Veröffentlichungen ergibt sich ein ambivalentes Bild: Obwohl die richtige Vorgehensweise bekannt und u.a. in den oben beschriebenen Arbeiten dokumentiert ist, werden bei vielen Simulationsprojekten die Aufgaben der statistischen Auswertungen ignoriert und die Ergebnisse unkritisch in „Business Grafiken“ aufbereitet und dargestellt.

Eine grobe Analyse der Veröffentlichungen und Beiträge von **ASIM**-Symposien und -Jahrestagungen sowie internationalen praxisrelevanten Fachzeitschriften wie bspw. **Simulation** und **Industrial Engineering** zeigt, dass nur in seltenen

Fällen Simulationsergebnisse statistisch korrekt mit Konfidenzintervallen beschrieben oder die Ergebnisse des Vergleichs der simulierten Systemalternativen mit einer Varianzanalyse auf statistische Signifikanz bewertet werden.

2.10 Ableitung des Anforderungsprofils

Auf Grundlage der beschriebenen Motivation und Problemstellung, der Rahmenbedingungen (Kapitel 1), Ansprüche der Anwender sowie Defiziten aus dem Stand der Technik können **Anforderungen an ein integriertes Verfahren zur robusten statistischen Auswertung von Simulationsergebnissen** abgeleitet und dargestellt werden.

Diese Anforderungen lassen sich in die drei Bereiche

1. Robustheit,
2. Plausibilität und
3. Integration

gliedern, welche im Folgenden dargestellt werden. Diese abgeleiteten Anforderungen werden ausführlich beschrieben und dienen zur inhaltlichen Konzeption und Entwicklung des integrierten robusten Verfahrens in den nächsten drei Kapiteln.

2.10.1 Robustheit

2.10.1.1 Robuste statistische Verfahren

Es werden **robuste statistische Verfahren (Anforderung 1)** benötigt, die sich unempfindlich in Hinsicht auf die Erfüllung bzw. Verletzung ihrer Voraussetzungen verhalten. Diese statistische Robustheit bedeutet (nach Grammel [Gram91] und Huber [Hube04]) für die Verfahren:

- ein sinnvolles Verhalten bei den angenommenen Voraussetzungen und Modellannahmen,
- kleine Abweichungen von den Modellannahmen können die Eigenschaften nur gering beeinflussen und
- größere Abweichungen verursachen keine katastrophalen Ergebnisse.

Die Kriterien für die Robustheit eines Tests oder Schätzverfahrens sind sowohl die Validitätsrobustheit als auch die Effizienzrobustheit und werden in Kapitel 3.6.3 erläutert. Eine genaue und detaillierte Ausführung des statistischen Robustheitsbegriffes ist bei **Büning** [Büni91:31] zu finden.

2.10.1.2 Verfahren ohne Voraussetzung einer Normalverteilung

Es müssen statistische Methoden in dem integrierten Verfahren bereitgestellt werden, welche auch **ohne die Voraussetzung von normalverteilten Stichproben** korrekte Ergebnisse liefern (**Anforderung 2**). Die Modellannahme einer Normalverteilung ist ein Ausschlusskriterium für diese statistischen Verfahren. Der Verweis auf den zentralen Grenzwertsatz zum Erreichen einer Normalverteilung der Beobachtungen ist leider nicht ausreichend.

2.10.2 Plausibilität

2.10.2.1 Einfache und plausible Verfahren

Die Verfahren müssen **einfach** und für den Anwender aus der Praxis **nachvollziehbar und plausibel (Anforderung 3)** sein. Anspruchsvolle Methoden der Zeitreihenanalyse, die auf Spektralanalysen und sonstige Verfahren im Frequenzbereich beruhen, sind hier aufgrund ihrer Komplexität und „Abstraktheit“ nicht geeignet. Dasselbe gilt für andere, für einen Praktiker „anspruchsvolle“, Verfahren zur Effizienzsteigerung wie bspw. „Bootstrapping“ [Ross97] oder „Jackknifing“ Methoden [KlvG92].

2.10.2.2 Entscheidungskompetenz beim Anwender

Dem Anwender soll kein automatisiertes Verfahren oder ein „Expertensystem“ bereitgestellt werden. Die eigentliche Entscheidungskompetenz soll beim Anwender liegen (**Anforderung 4**), vor allem auch um die Akzeptanz der Verfahren und getroffenen Entscheidungen zu sichern. In Verbindung mit plausiblen Methoden soll der Anwender selbst diese Entscheidungen treffen, sei es bei einer Klassifikation der Beobachtungen als auch bei der Auswahl und dem Einsatz der Verfahren. Dies beinhaltet auch die Möglichkeit einer Sensitivitätsanalyse, um damit die Auswirkungen alternativer Anwenderentscheidungen vergleichen zu können.

2.10.3 Integration

2.10.3.1 Integrierter modularer Aufbau aus Einzelverfahren

Ein integrativer Aufbau des Verfahrens aus modularen Einzelmethoden (**Anforderung 5**) ist eine wesentliche Anforderung an das Gesamtverfahren. Bei Bedarf und modifizierten Anforderungen der Anwender oder einer Weiterentwicklung der statistischen Verfahren kann das integrierte Verfahren mit geringem Aufwand angepasst werden. Einzelne Verfahren können verbessert, ergänzt oder ersetzt werden.

2.10.3.2 Integration in eine durchgängige Methodenkette bei Simulationsprojekten

Das integrierte Verfahren muss sich nahtlos in die durchgängige Methodenkette und Vorgehensweise bei der Durchführung von Simulationsprojekten einpassen (**Anforderung 6**). Dabei ist die in Kapitel 1 beschriebene Vorgehensweise nach der VDI Richtlinie 3633 (Abb. 1.2) die gültige Referenz.

2.10.4 Anforderungsprofil

Die obigen sechs inhaltlichen Anforderungen werden noch von generellen und allgemeinen Anforderungen an ein betrieblich nutzbares Verfahren wie Anwendbarkeit („Usability“) und Anpassbarkeit (Flexibilität) begleitet. Die Abb. 2.3 fasst das Anforderungsprofil an das integrierte Verfahren zusammen.

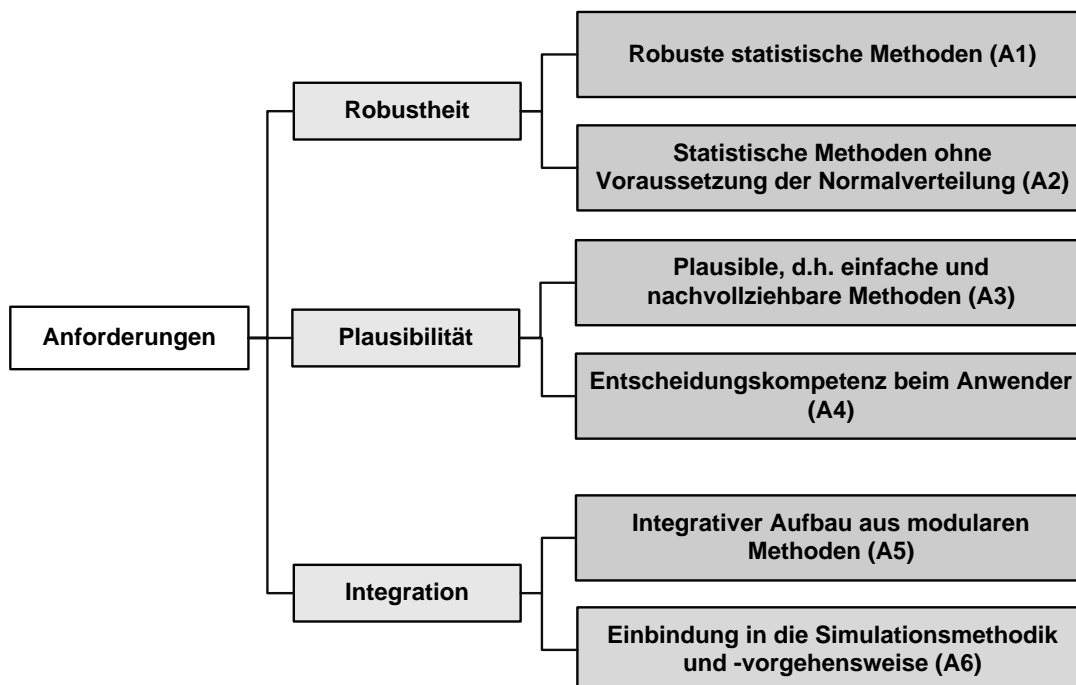


Abbildung 2.4: Anforderungsprofil an das integrierte Verfahren.

2.11 Zusammenfassung der Bewertung des Stands der Technik

Abschließend wird der Stand der Technik zusammenfassend bewertet. Bei der Betrachtung der Verfahren hinsichtlich der

- Auswertungsmethodiken,
- Auswertungswerkzeuge,
- Simulationssysteme und
- Statistikprogramme

ergeben sich wesentliche Defizite und Kritikpunkte bei den drei Aspekten der Robustheit, Plausibilität und Integration, welche damit, wie oben beschrieben, als relevante Anforderungen an ein neu zu entwickelndes Verfahren gelten. Tabelle 2.1 zeigt dazu eine Zusammenfassung der Bewertung der untersuchten Verfahren.

Mit dieser Zusammenfassung der Bewertung wird aufgezeigt, dass trotz eines großen und dringenden Bedarfs für die statistische Auswertung von Simulationsergebnissen derzeit kein robustes, plausibles und integriertes Verfahren verfügbar ist.

	1) Robustheit							
	A) robust	B) nicht normal	2) Plausibilität	A) Plausibel	B) Entsch. B. Anw.	3) Integration	A) Integr. Aufbau.	B) Sim-Methodik
Auswertungsmethodiken								
LAW (07)	o	—		o	o		o	+
NELSON (01)	o	—		o	o		o	o
KLEIJNEN (87)	+	+		—	o		o	o
BANKS/CARSON/NELSON/NICOL (09)	o	—		o	—		o	o
ALEXOPOULOS/SEILA/KIM (01/02)	o	—		o	—		o	o
GOLDSMAN/NELSON (98)	o	—		o	o		o	o
LIEBL (95)	o	—		+	o		o	o
FISHMAN (01)	o	—		—	o		o	o
NAKAYAMA (08)	—	—		o	o		—	o
ROSS (02)	o	—		o	—		o	o
VDI (97)	o	+		o	+		o	+
WENZEL et. al. (08)	—	—		o	o		o	+
Auswertungswerkzeuge								
SCHEIFELE (86)	+	—		+	o		+	o
HADDOCK (87)	—	—		o	—		o	o
MELlichAMP/YOUNG (89)	—	—		o	—		o	—
PIERREVAL/DESLANDRES (91)	—	—		o	o		o	o
MOLLAMUSTAFAOGLU (93)	—	—		—	o		o	o
TAO/NELSON (97)	—	—		—	o		o	o
LE-MIN YIN (98)	—	—		—	o		o	o
HOAD/ROBINSON/DAVIES (07)	—	—		o	—		o	o
Simulationssysteme								
ARENA	o	—		o	o		o	+
AUTOMOD	o	—		o	o		o	o
PLANT SIMULATION	o	—		o	o		o	o
QUEST	o	—		o	—		—	o
DOSIMIS	—	—		o	—		—	—
ENTERPRISE DYNAMICS	o	—		o	—		—	o
PROMODELL	—	—		o	—		—	o
SIMPROCESS	—	—		o	—		—	o
WITNESS	—	—		o	—		—	o
Statistikprogramme								
SPSS	o	o		o	+		—	—
SIMSTAT	o	o		o	+		—	—
SAS/STAT	o	o		o	+		—	—
MS-EXCEL	o	—		o	+		—	—

Legende: + größtenteils erfüllt, o teilweise erfüllt, - nicht erfüllt

Tabelle 2.1: Zusammenfassung der Bewertung der untersuchten Verfahren.

3 Konzeption des integrierten Verfahrens

3.1 Integriertes Vorgehensmodell

In den Kapiteln 1 und 2 wurde der Bedarf an einem integrierten Verfahren für die robuste statistische Auswertung von Simulationsdaten hergeleitet. Die derzeit eingesetzten statistischen Methoden, Vorgehensweisen und Verfahren haben wesentliche Defizite im praktischen Einsatz: Es wurde aufgezeigt, dass diese existierenden Verfahren für die beiden Ziele der Auswertung

- Schätzung von Leistungskennzahlen eines Modell mit Qualitätsangabe und
- Vergleich von Systemalternativen auf signifikante Unterschiede

erheblichen Verbesserungsbedarf haben.

Als häufige und typische Mängel der bekannten Vorgehensweisen lassen sich die folgenden Eigenschaften zusammenfassen: Die statistischen Methoden und Verfahren

- benötigen die Modellannahme der Normalverteilung und Unabhängigkeit,
- sind nicht robust,
- sind nicht einfach und plausibel sowie
- binden den Anwender nicht in die Entscheidungsfindung mit ein.

Des Weiteren fehlt meist eine Integration in die durchgängige Methodik der Simulationsprojekte. Wegen dieser vorhandenen Defizite werden die Verfahren in praktischen Anwendungen meistens nicht eingesetzt.

Das in diesem Kapitel entwickelte Konzept des **integrierten Vorgehensmodells**⁹ beschreibt die geeignete strukturierte Vorgehensweise der statistischen Auswertung einschließlich der einzelnen Teilaufgaben. Mit diesem Vorgehensmodell werden die beiden in Kapitel 1 beschriebenen Ziele der statistischen Auswertung erreicht und die in Kapitel 2 abgeleiteten Anforderungen an die einzelnen Verfahren sowie deren Integration in ein durchgängiges Verfahren erfüllt.

Dabei dient dieses **integrierte Vorgehensmodell** mit den Anforderungen aus Kapitel 2 in den folgenden Kapiteln 4 und 5 zur **Auswahl und Integration der Verfahren**. Die praktische industrielle Anwendung des integrierten Verfahrens

⁹ Ein Vorgehensmodell ist als Modell eine Vereinfachung der „Vorgehensweise“.

wird im Kapitel 6 beschrieben. Die dabei gesammelten Erfahrungen dienen zur kritischen Evaluation.

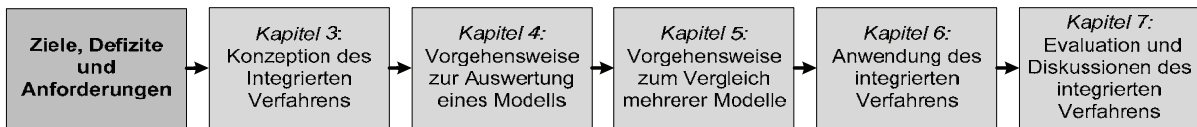


Abbildung 3.1: Weiterer Aufbau der Arbeit.

Dazu wird im folgenden Kapitel die Konzeption des integrierten Verfahrens mit Hilfe der abgeleiteten Anforderungen und notwendigen Funktionalitäten gestaltet. Der integrative Charakter des Verfahrens wird mit drei Eigenschaften dargestellt:

- Integration in die Simulationsmethodik,
- Einbindung des Anwenders sowie
- Auswahl und Zusammenführung der statistischen Methoden und Prozeduren in ein integriertes Verfahren.

Das Kapitel beschreibt darauf aufbauend das Konzept des Auswertungsprozesses und ist die Grundlage der detaillierten Spezifikation des integrierten robusten Verfahrens in den beiden folgenden Kapiteln.

3.2 Trennung von Modell und Experimentierumgebung

Bei der Einbindung der statistischen Auswertung in eine durchgängige Simulationsmethodik ist das Architektur-Prinzip der Trennung von Modell und Experimentierumgebung hilfreich.

Im Rahmen der “Objectives-driven Methodology” von **Zeigler** [Zeig84] wird diese konzeptionelle Aufteilung in Modell und Experimentierumgebung (engl. „Experimental Frame“) in Abb. 3.2 skizziert dargestellt.

Die statistische Auswertung einschließlich der operativen Datensammlung aus der Beobachtung der endogenen und exogenen Variablen ist mit der Einbindung in die Simulationsmethodik ein wichtiger Bestandteil der Experimentierumgebung und nicht des eigentlichen Simulationsmodells. In diesem Prinzip sind auch Parameter des Simulationsexperiments wie die Simulationsdauer und die Startwerte für die Zufallszahlengeneratoren in der Experimentierumgebung beschrieben.

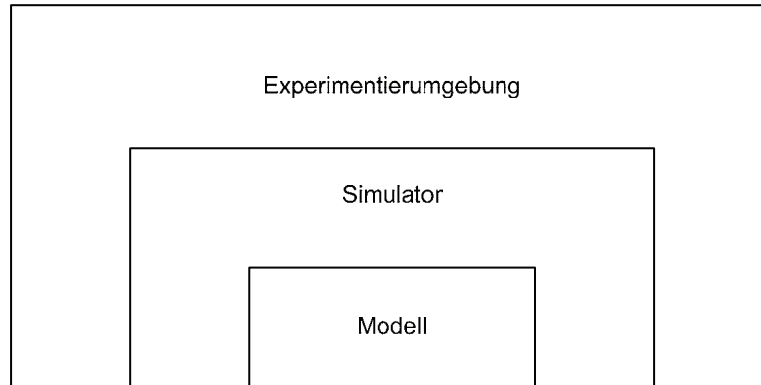


Abbildung 3.2: Modell und Experimentierumgebung.

Diese Positionierung der statistische Auswertung im Bereich der Experimentierumgebung ist als Rahmenbedingung für die Konzeption des integrierten Verfahrens von Bedeutung.

3.3 Einbindung in die Simulationsmethodik

Eine wesentliche Anforderung aus Kapitel 2 an das integrierte Verfahren ist die Einbindung der statistischen Auswertung in die Simulationsmethodik, d.h. in die durchgängige iterative Vorgehensweise bei der Durchführung einer Simulationsstudie, wie im Ablauf in Abb. 1.2 (ähnlich bei Bullinger u.a [BDAS86], in der VDI Richtlinie 3633 Blatt 1 [VDIF00] sowie auch bei Wenzel u.a. [WWCP08]) dargestellt.

Die statistische Auswertung ist keine vollständig in sich geschlossene Aufgabe innerhalb eines Simulationsprojekts, sondern interagiert mit den vor- und nachgestellten Schritten und Verfahren. Dabei müssen die Aufgaben, Verfahren und Ergebnisse der statistischen Auswertung iterativ in den Gesamtprozess eingebunden werden und sind Voraussetzung und Unterstützung für andere wesentliche Teilaktivitäten eines Simulationsprojekts. Die Verbindung der statistischen Auswertung mit den beschriebenen Teilaufgaben konkretisiert und spezifiziert damit die Anforderung aus Kapitel 2.

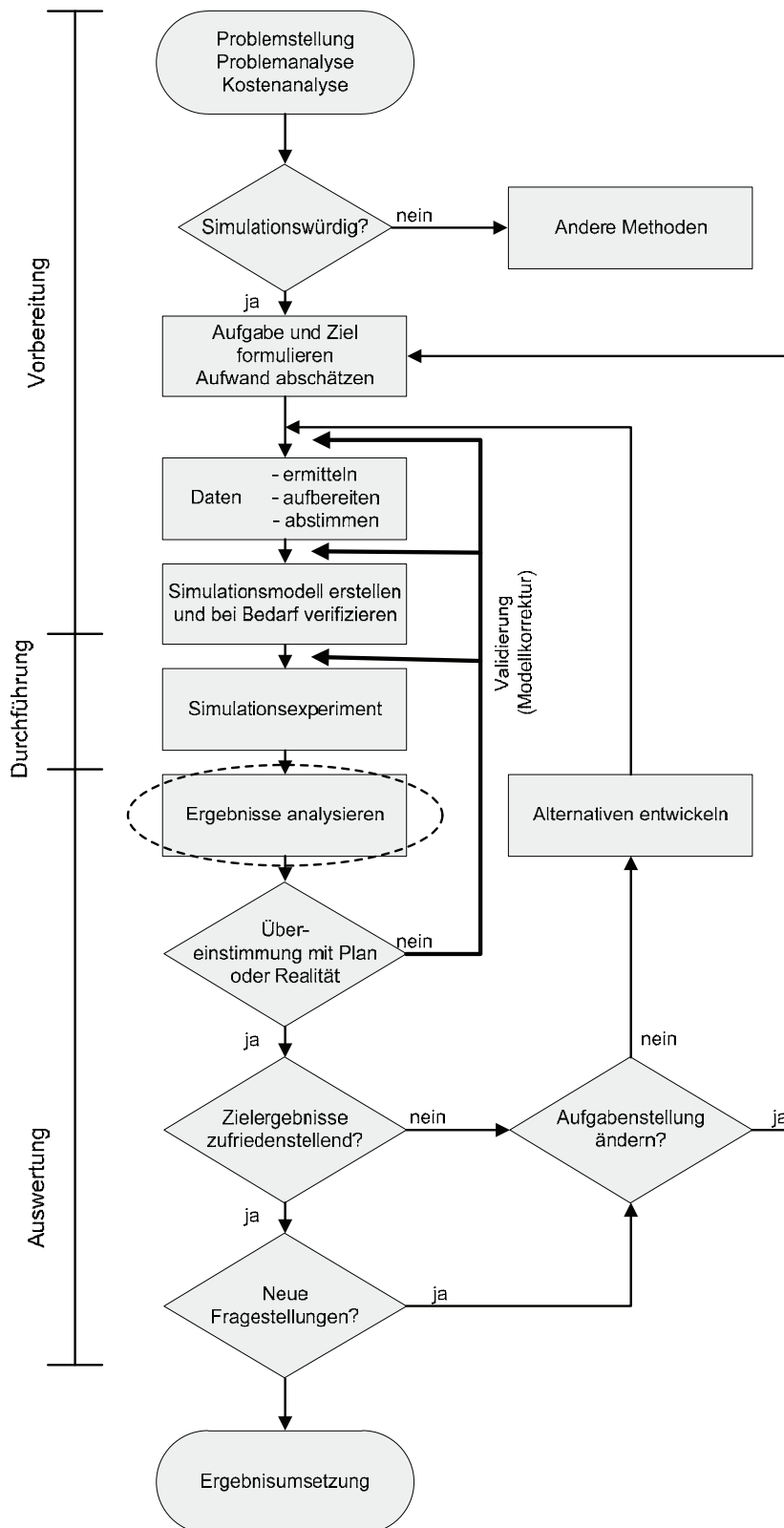


Abbildung 3.3: Einbindung der Auswertung in die Simulationemethodik nach VDI Richtlinie 3633 Blatt 1 [VDIF00].

3.3.1 Modellierung der Eingangsgrößen

Die stochastischen Eigenschaften des Simulationsmodells werden durch eine entsprechende Modellierung der stochastischen Eingangsgrößen dargestellt. Dabei werden auf Basis von (Pseudo-) Zufallszahlen die „passenden“ Verteilungen der Eingangsdaten bzw. Modellparameter erzeugt. **Law** beschreibt dies detailliert in seinem Standardwerk „*Simulation Modeling and Analysis*“ [Law07] und legt seinem Buch das entsprechende Programm „*ExpertFit*“ bei.

Die Abbildung der Eingangsgrößen gilt auch für eine hierarchische Modellierung unter Benutzung von Partialmodellen. Dabei dient die statistische Auswertung des Verhaltens von Partialmodellen bei diesen hierarchisch strukturierten Modellierungskonzepten zur Beurteilung und Modellierung des stochastischen Verhaltens dieser Eingangsgrößen. Somit können bei Bedarf komplexe Teilmodelle vereinfacht als Quellen oder Übertragungsfunktionen dargestellt werden und somit die gesamte Modellkomplexität reduziert werden. Diese Modellreduktion wirkt sich positiv auf die Simulationslaufzeiten aus, ohne dabei notwendigerweise die Ergebnisqualität erkennbar zu reduzieren.

3.3.2 Verifikation eines Modells

Die statistische Auswertung unterstützt die Verifikation des Simulationsmodells, indem über die Schätzung entsprechender Kennzahlen der Ergebnisse die interne Korrektheit des ganzen Simulationsmodells sowie von Teilmodellen bestätigt werden kann. Die Kontrolle der vorgegebenen Parameter der stochastischen Eingangsgrößen, d.h. der Modellparameter in den erzeugten Verteilungen dieser Eingangsdaten, ist ein weiterer Aspekt der vollständigen Verifikation eines Simulationsmodells.

3.3.3 Validierung der Simulation

Die Validierung vergleicht das Verhalten des Simulationsmodells im Rahmen der Fragestellung der Simulationsstudie mit dem Verhalten des abgebildeten realen Systems. Um dieses Verhalten des Simulationsmodells interpretieren und mit dem realen System vergleichen zu können, sind die Schätzungen der Kennzahlen mit ihrer Varianz bzw. dem Konfidenzintervall und damit einem Indikator für die Qualität der Schätzung unabdingbar.

Als eine wichtige Maßnahmen zur Validierung des Modells kann, bei der Verfügbarkeit von Daten des Realsystems, der direkte Vergleich der Verteilungsformen der Beobachtungen aus beiden Systemen sowie der Vergleich der jeweiligen Kennzahlen dienen. Damit können die statistischen Verfahren zum Vergleich von zwei Simulationsalternativen auch als Validierungsmethode zum Vergleich des Verhaltens von Modell und Realsystem dienen.

Eine gute Übersicht über die Methoden der Verifikation und Validierung findet sich bei **Law** [Law07] sowie bei **Rabe, Spiekermann** und **Wenzel** [RSW08] aus der ASIM Arbeitsgruppe Validierung.

3.3.4 Versuchsplanung

Die Versuchsplanung (engl. „Design of Experiments“, „DoE“) hängt direkt mit der statistischen Auswertung zusammen. Die Versuchsplanung definiert, welche Einstellungen der unabhängigen Variablen in welcher Kombination untersucht, d.h. simuliert werden sollen. Daraus ergibt sich mit den jeweiligen Verfahren der Versuchsplanung (siehe u.a. VDI-Richtlinie 3633 [VDIF97]) die Anzahl der Systemvarianten bzw. Szenarien. Die eigentliche Anzahl und ggf. Dauer der Simulationsläufe wird nun durch den Stichprobenbedarf der statistischen Verfahren und die erwartete Qualität der resultierenden Aussagen bestimmt.

Sowohl **Kleijnen** [Klei98] als auch **VDI Richtlinie 3633 Blatt 3** [VDIF95] bieten eine geeignete Einführung in die Versuchsplanung mit Verweisen auf die weiterführende Literatur.

3.3.5 Optimierung

Die korrekte statistische Auswertung der Simulationsergebnisse ist die prinzipielle Grundlage jeglicher Ansätze zur Optimierung von stochastisch ereignisdiskret modellierten Systemen. Erst die Schätzung der Kennzahlen mit Konfidenzintervallen und die Verfahren zum Vergleich von Systemvarianten ergeben signifikante Unterschiede (oder deren Abwesenheit) und sind damit die Basis für sämtliche Optimierungsstrategien und -verfahren.

Als Einstieg in die Optimierung von simulierten Modellen ist **Law** [Law07] und **Andradottir** [Andr98] geeignet.

3.4 Einbindung des Anwenders

Ein wesentlicher Bestandteil des Konzepts des integrierten Verfahrens ist die Einbindung des Anwenders in den Auswertungsprozess und die notwendigen Entscheidungen. Es soll kein autonomes regelbasiertes Expertensystem entwickelt werden, sondern der Anwender soll den Auswertungsprozess verstehen und steuern können. Diese Integration des Anwenders beinhaltet die drei wesentlichen und aufeinander aufbauenden Aspekte

1. Visualisierung von Daten,
2. Entscheidungskompetenz beim Anwender und
3. Möglichkeit der Sensitivitätsanalyse,

welche erheblich zur Plausibilität und Sicherung der Akzeptanz des gesamten integrierten Verfahrens beitragen.

3.4.1 Visualisierung

Die Visualisierung bzw. grafische Aufbereitung der Daten und Informationen in Diagrammen ist eine wesentliche Voraussetzung und Methode zur Einbindung des Anwenders. Erst die visuelle Aufbereitung ermöglicht es, Eigenschaften der Zeitreihen und der Verteilungen der Beobachtungen erkennen zu können.

Speziell bei der qualitativen Klassifizierung der Stichproben ist die Visualisierung ein wichtiger Zwischenschritt und dient vor allem auch zur Plausibilitätsprüfung der Ergebnisse der statistischen Verfahren¹⁰.

3.4.2 Entscheidungskompetenz

Das integrierte Verfahren ist kein autonomes Expertensystem, der Anwender trifft die relevanten Entscheidungen, insbesondere bei einer qualitativen Klassifikation der Beobachtungen. Diese Entscheidungskompetenz beim Anwender fördert die Plausibilität der einzelnen Verfahren und ist ein wesentliches Element der Akzeptanz des integrierten Verfahrens und seiner Ergebnisse durch den Anwender.

Speziell bei einer qualitativen Klassifikation ist die Einbindung und Entscheidung durch den Anwender anstatt einer „Automatisierung“ des Verfahrens durch entsprechende Regeln, evtl. mit Unterstützung der Fuzzy-Logik, einfacher und robuster. Die Entscheidung durch den Benutzer ist in Verbindung mit der Notwendigkeit der Sensitivitätsanalyse sinnvoll.

3.4.3 Sensitivitätsanalyse

Eine Sensitivitätsanalyse (auch Empfindlichkeitsanalyse) untersucht die Auswirkung der Variation der Eingangsgrößen oder Annahmen auf das Ergebnis. Besonders in Verbindung mit der Entscheidungskompetenz des Anwenders wird die Notwendigkeit einer Sensitivitätsanalyse betont.

Die Untersuchung der Empfindlichkeit der Auswertungsergebnisse (Größe von Konfidenzintervallen und signifikante Unterschiede von Systemalternativen) von der Klassifikation der Unabhängigkeit, Anzahl und Verteilung der Stichproben durch den Anwender ist ein wesentlicher Bestandteil des zu entwickelnden integrierten Verfahrens.

¹⁰ Siehe hierzu auch Huff [Huff54] zur Interpretation von statistischen Ergebnissen.

3.5 Konzeption des integrierten Auswertungsprozesses mit Teilaufgaben

Die konzeptionelle Struktur und kausale Abfolge des integrierten Auswertungsprozesses ergibt sich aus den Zielen, den einzelnen Teilaufgaben und deren Voraussetzungen und Abhängigkeiten. Wie in Kapitel 2.4 beschrieben, lässt sich das Konzept des integrierten Auswertungsprozesses aus der Strukturierung nach den beiden Zielen und Auswertungsabläufen

- **Schätzung von Leistungskennzahlen mit Qualitätsangabe** eines Modells und
- **Vergleich von Systemalternativen** auf signifikante Unterschiede

entwickeln. Dabei ist das Schätzen der Leistungskennzahlen auch eine notwendige Vorarbeit für den Vergleich der Systemalternativen.

3.5.1 Schätzung der Leistungskennzahl eines Modells mit Qualitätsangabe

Um für ein Simulationsmodell eine statistisch korrekte Schätzung von Leistungskennzahlen mit Qualitätsangabe durchführen zu können, müssen die entsprechenden Teilaufgaben durch geeignete Verfahren erfüllt werden.

Damit ergeben sich fünf wichtige und relevante Teilaufgaben für den integrierten Auswertungsprozess eines Simulationsmodells mit der Schätzung der Kennzahlen:

- **Bestimmung der stationären Phase:** Wenn die Simulationsstudie vor allem das stationäre Verhalten des System betrifft und vorhandene Einschwingvorgänge dabei stören, muss diese transiente Phase der Zeitreihe der Beobachtungen identifiziert und entsprechend ignoriert oder abgeschnitten werden.
- **Unabhängige Stichproben bei Simulationen ohne festes Ende:** Falls bei Simulationen ohne festes Ende die Beobachtungen aus einem langen Simulationslauf, einer Replikation, entnommen werden sollen, muss die vorhandene Autokorrelation und Autokovarianz entweder exakt bestimmt oder einfach beseitigt werden.
- **Grafische Darstellung von Verteilungen und Zeitreihen:** Um den Anwender in den Auswertungsprozess einbinden zu können, müssen geeignete grafische Darstellungsformen der Verteilungsfunktionen und Zeitreihen bereitgestellt werden.
- **Schätzung von Kennzahlen der Verteilung:** Die Schätzung von relevanten Kennzahlen, (Lage-) Parametern und Momenten der Verteilung der Beobachtungen dient sowohl zur Schätzung der Leistungskennzahl selbst als auch zur Beurteilung der Verteilung der Stichproben.

- **Berechnung von Konfidenzintervallen für Mittelwertschätzungen:**
Um die Kennzahlen mit Konfidenzintervallen als Qualitäts- bzw. Güteangabe schätzen zu können, benötigt man unabhängige Beobachtungen und eine Aussage über die Form ihrer Verteilung oder speziell über die Form der Abweichung von der Normalverteilung.

Die Ziele und Abläufe des Auswertungsprozesses mit den fünf Teilaufgaben und ihren Relationen sind in Abb. 3.4 schematisch dargestellt.

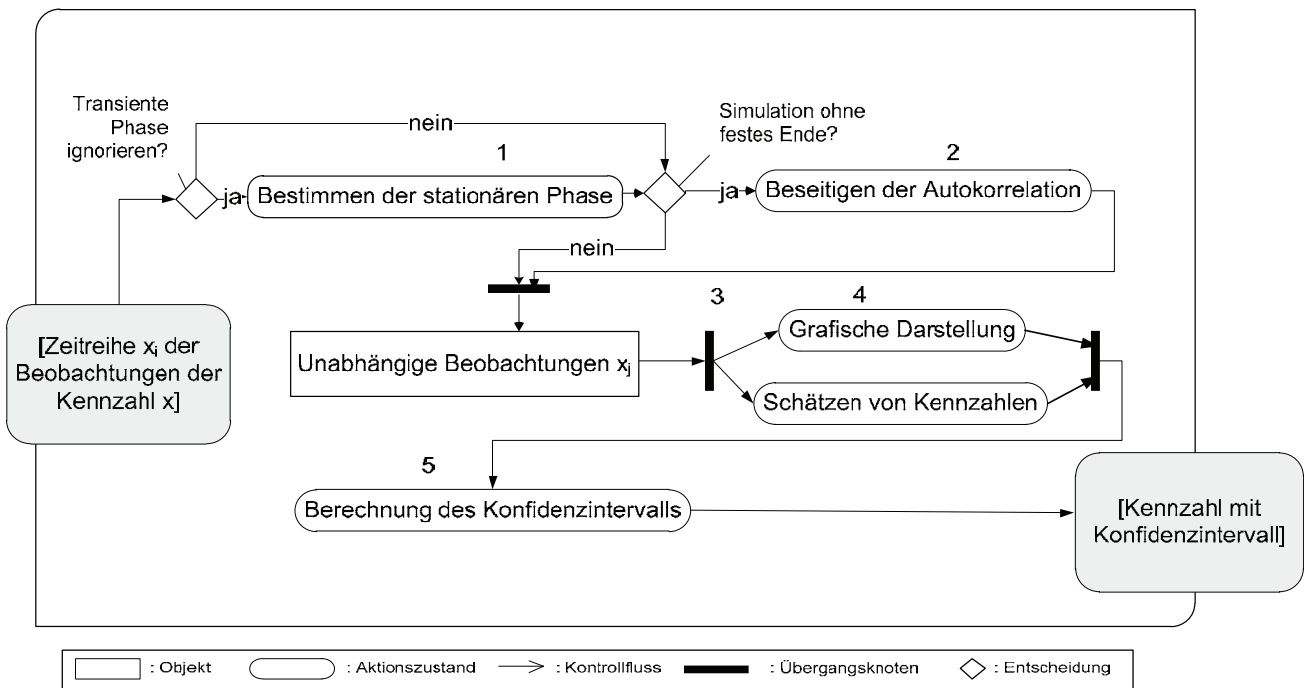


Abbildung 3.4: UML Diagramm der Abläufe und Teilaufgaben des Auswertungsprozesses eines Modells.

Im Folgenden werden die einzelnen Teilaufgaben mit ihren Anforderungen und Rahmenbedingungen beschrieben.

3.5.1.1 Bestimmung der stationären Phase

Je nach Ziel der Simulationsstudie und der entsprechenden Modellierung sollen die untersuchten Systeme entweder

- mit und einschließlich einer transienten Einschwingphase untersucht und simuliert oder
- nur das stationäre Systemverhalten betrachtet werden, da eine transiente Anfangsphase die Messung der Kennzahlen verfälscht.

Im zweiten Fall beim Fokus auf das stationäre Verhalten entsteht die Aufgabe, den Übergang von der transienten in die stationäre Phase der Zeitreihe zu bestimmen und die Beobachtungen aus der transienten Phase zu ignorieren bzw. abzuschneiden.

Die Verfahren zur Bestimmung der stationären Phase unterteilen sich in zwei Prinzipien:

- Direkte, heuristische Verfahren zur Bestimmung der stationären Phase, welche einen bestimmten Zeitpunkt vorschlagen und
- indirekte, statistische Tests, welche Aussagen machen, ob eine Verzerrung durch die transiente Phase an einem konkreten Zeitpunkt signifikant ist oder nicht.

Zusätzlich sollte die stationäre Phase mittels einer grafischen Aufbereitung der Zeitreihe vom Simulationsanwender direkt geschätzt werden. Eine Glättung der Zeitreihe durch Aggregation von einzelnen Werten bzw. aus verschiedenen Replikationen wird empfohlen.

Konsequenterweise ist in der praktischen Anwendung eine Kombination der direkten und indirekten Verfahren mit der grafischen Aufbereitung sinnvoll. Die erhaltenen Ergebnisse werden auch durch eine grafische Darstellung der Zeitreihe plausibel und nachvollziehbar.

3.5.1.2 Beseitigen der Autokorrelation

Wenn die Aufgabenstellung eine Simulation ohne festes Ende ermöglicht, ergibt sich die Option, die Beobachtungen aus einem langen Simulationslauf als Stichproben zu nutzen. Allerdings sind diese Stichproben autokorreliert¹¹ und nicht unabhängig. Es werden einfache, plausible und robuste Verfahren zur Beseitigung der Autokorrelation benötigt, da sonst durch eine Unterschätzung der Varianz ein zu geringes Konfidenzintervall berechnet wird.

Kompliziertere Verfahren auf Basis der Methode der Spektralanalyse oder einer Analyse der Varianzstruktur sind relativ komplex, für einen Praktiker schwer nachvollziehbar, wenig plausibel und damit ungeeignet.

Prinzipiell erfüllt das Verfahren der „Batch Means“ diese Anforderungen gut, allerdings ohne die etwas „anspruchsvollen“ Varianten dieser Methode, wie u.a. von Fishman [Fish96] beschrieben. Der Zielkonflikt zwischen der Effizienz anspruchsvoller Batch Means Verfahren und der Einfachheit und Robustheit der

¹¹ Die Autokorrelation der Beobachtungen aus einem Warteschlangensystem ist proportional zu dessen Verkehrsdichte ρ .

simplen Batch Means wird eindeutig zu Gunsten der Unkompliziertheit und Plausibilität des einfachen Verfahrens entschieden.

3.5.1.3 Grafische Darstellung

Eine mehrfach auftretende Teilaufgabe ist die Visualisierung und grafische Darstellung von Teil- und Endergebnissen des Auswertungsprozesses in Diagrammen. Es werden Darstellungen für

- Zeitreihen,
- Histogramme,
- Dichtefunktionen und
- Box-Plots

benötigt, um den Anwender sinnvoll in den Auswertungsprozess einzubinden. Dabei ist es notwendig, diese Visualisierung von Daten aus mehreren Szenarien in einer vergleichenden grafischen Darstellung zu präsentieren. Bei der Darstellung von Zeitreihen kann eine Glättung durch Zusammenfassen von mehreren Zeitpunkten sinnvoll sein.

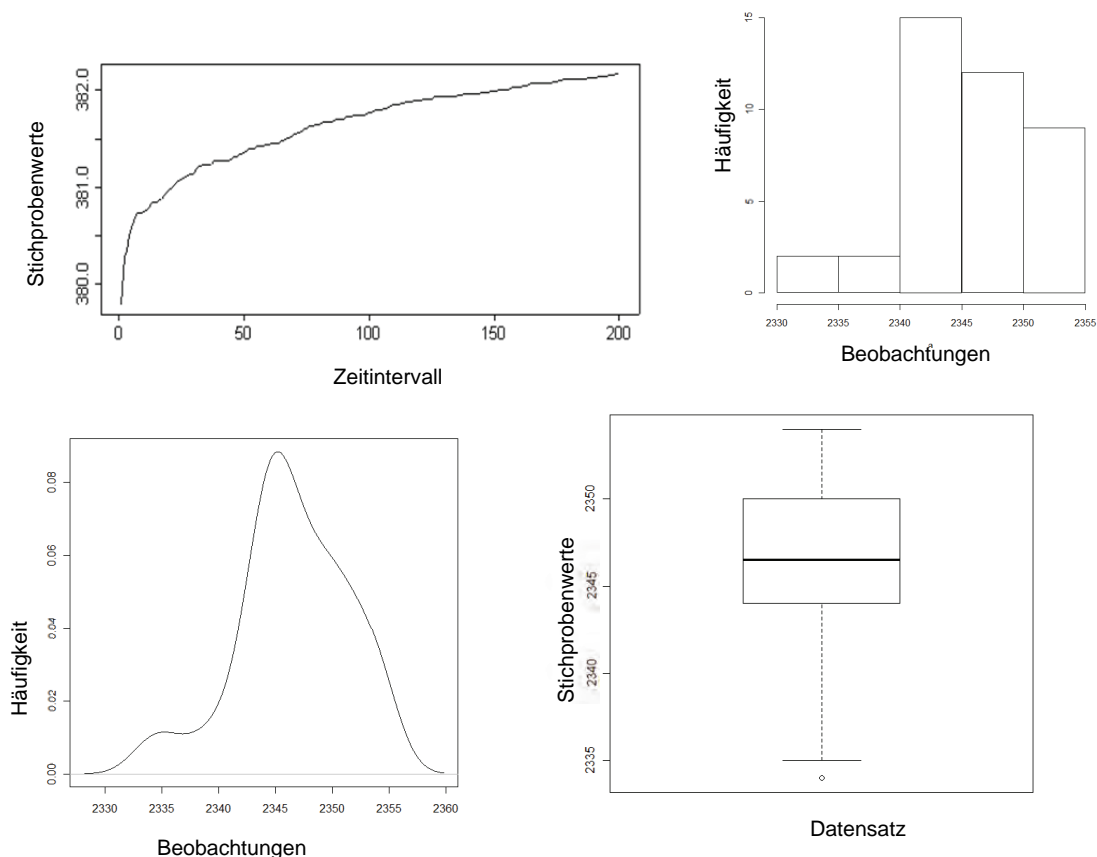


Abbildung 3.5: Grafische Darstellungsformen zur Unterstützung der Auswertung (Zeitreihe, Histogramm, Dichtefunktion, Box-Plot).

Dabei werden auch Box-Plots als ein grafisches Verfahren der Explorativen Datenanalyse (EDA) (siehe Hartung [Hart02:825]) eingesetzt. Anspruchsvollere Darstellungen der grafischen EDA (siehe duToit, Steyn und Stumpf [dTSS86]) sind nicht notwendig.

3.5.1.4 Schätzung von Kennzahlen der Verteilung

Die Schätzung von Kennzahlen ist eine Punktschätzung im statistischen Sinne. Das Schätzen von Kennzahlen der Verteilung reduziert sich nicht nur auf die Aufgabe der Bestimmung des arithmetischen Mittels \bar{x} als bester Schätzer für die eigentliche Leistungskennzahl, sondern beinhaltet auch das Schätzen der Werte für Median, Modus und Varianz der Verteilung der Beobachtungen. Für eine Beurteilung der Verteilung für eine Klassifizierung und damit einer korrekten Qualitätsangabe der Mittelwertschätzung sind auch weitere Lage- und Formparameter der Stichprobenverteilung zu bestimmen, um die Art der Abweichung von der Modellannahme der Normalverteilung richtig beurteilen zu können.

3.5.1.5 Konfidenzintervalle als Qualitätsangabe

Als Qualitätsangabe, Güteindikator bzw. Maßzahl für den Fehler für eine Mittelwertschätzung ist das Konfidenzintervall¹² bzw. der Vertrauensbereich bei praktischen Anwendungen üblich und am meisten verbreitet. Wie in Kapitel 2 beschrieben sind alternative Qualitätsangaben, wie der von **Schmeiser** [Schm01], [Schm04] vorgeschlagene geschätzte Standardfehler eher ungewöhnlich und werden praktisch nicht benutzt.

Üblicherweise basiert die Berechnung der Konfidenzintervalle auf der t-Statistik, welche allerdings die „klassischen Voraussetzungen“ benötigt. Da jedoch speziell die Normalverteilung der Beobachtungen bei Simulationsergebnissen nicht angenommen werden darf, sind alternative robuste, angepasste und ggf. verteilungsfreie Verfahren zur Berechnung eines Konfidenzintervalls notwendig.

3.5.2 Vergleich von Systemalternativen

Bei Simulationsprojekten geht es häufig nicht nur um die Bestimmung von Leistungskennzahlen eines Systems. Es sollen alternative Systeme bzw. Simulationsmodelle auf Grundlage dieser Leistungskennzahlen verglichen und die

¹² Der Begriff *Konfidenzintervall* oder *Vertrauensbereich* wurde nach Sachs und Hedderich [SaHe06] von J. Neyman und E.S. Pearson eingeführt.

signifikanten Unterschiede, falls vorhanden, identifiziert werden. Erwartet werden Aussagen wie

- „System A ist bei Merkmal Y besser als System B“ und
- „System B ist bei Merkmal Y ungefähr gleich System C“.

Diese Überprüfung von Systemalternativen auf signifikante Unterschiede ist eine klassische Aufgabenstellung der Statistik und nicht nur bei der Auswertung von Simulationsergebnissen notwendig. Häufig untersucht man, ob und wie sich welche Faktoren bei Experimenten auf eine Ausgangsgröße auswirken.

Bei der Auswertung von Simulationsergebnissen gibt es allerdings wie in Kapitel 1 beschrieben die Besonderheit, dass die Beobachtungen

- nicht als normalverteilt und
- bei Systemalternativen ohne Varianzgleichheit (Homoskedastizität)

angenommen werden müssen. Durch die Kontrolle der Simulation kann allerdings eine identische Anzahl von Beobachtung für jede simulierte Alternative vorausgesetzt werden.

Der einfache Fall des Vergleichs von genau zwei Alternativen mit der Nullhypothese gleicher Mittelwerte

$$H_0 : \mu_1 = \mu_2 \quad (3.1)$$

mit unbekanntem und ungleichen Varianzen ist in der Literatur (u.a. bei Hartung [Hart02:510]) auch als klassisches „**Behrens-Fisher-Problem**“ mit der approximativen Lösung von Welch bekannt.

3.5.2.1 Modell der Varianzanalyse

Um generell mehrere Systemalternativen auf signifikante Unterschiede zu untersuchen, ist das Konzept und die Grundidee der einfachen **Varianzanalyse** oder ANOVA (Analysis of Variance) nach R. A. Fisher gut geeignet.

Für alle k Systemalternativen werden jeweils $n_i (i = 1, \dots, k)$ unabhängige Beobachtungen $y_{ij} (j = 1, \dots, n_i)$ der Ausprägung eines Merkmals Y benötigt. Damit ergibt sich das Modell der Varianzanalyse mit k Effekten bzw. Systemalternativen als

$$y_{ij} = \mu + \alpha_i + e_{ij} \quad (3.2)$$

mit

- μ : Mittelwert der Gesamtstichprobe,
- α_i Effekt der i-ten Faktorstufe, d.h. Systemalternative,
- e_{ij} : Störvariable, die als unabhängig und normalverteilt mit Erwartungswert 0 und gleicher Varianz angenommen wird.

Die Varianzanalyse zeigt eventuell vorhandene Unterschiede der Systemvarianten auf und verwirft ggf. die Nullhypothese H_0 der identischen Mittelwerte der Systemalternativen.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (3.3)$$

Allerdings benutzt die klassische Varianzanalyse den F-Test, welcher (nach Sachs und Hedderich [SaHe06:368]) empfindlich für Abweichungen von der Normalverteilung ist, und testet nur, ob die Mittelwerte signifikant verschieden sind.

Um die eigentlichen Unterschiede zwischen den Systemvarianten und vor allem deren Richtung zu identifizieren, werden die Verfahren der **Mehrfachen Mittelwertvergleiche** als Post-Hoc-Tests der Varianzanalyse eingesetzt (vgl. Hartung [Hart02:614]), vor allem auch, um der Alphafehler-Kumulierung (Alpha-Inflation) der mehrfachen paarweisen Vergleiche entgegen zu wirken.

Alternativ zu den Verfahren der Mehrfachen oder Multiplen Mittelwertvergleiche (engl. Multiple Comparison Procedures, MCP) sind auch die Ranking and Selection Verfahren für eine Bestimmung des Richtungsunterschieds der Mittelwerte geeignet. **Swisher u.a.** [SJY03] geben in ihrem Artikel eine gute Übersicht über die Entwicklung, die Eignung und den Zusammenhang der Verfahren der Multiplen Mittelwertvergleiche sowie Ranking and Selection bei der Auswertung von Simulationsergebnissen.

Um nun bei der Simulationsauswertung die Fragestellung nach „signifikanten Unterschieden“ der untersuchten Alternativen zu beantworten, ohne diese Signifikanz in absoluten Werten selbst festlegen zu müssen bzw. zu wollen, sind die Verfahren der Mehrfachen Mittelwertvergleiche vorteilhafter als die entsprechenden Ranking and Selection-Verfahren. Der nachfolgende Abschnitt erläutert die prinzipiellen Unterschiede der beiden Verfahren.

3.5.2.2 Gegenüberstellung von Multiplen Mittelwertvergleichen und Ranking & Selection Verfahren

Die Tabelle 3.1 beschreibt die Unterschiede der beiden Verfahrensarten hinsichtlich des Konzepts, der Ergebnisse und der jeweiligen Vor- und Nachteile bei der Auswertung von Simulationsdaten zum Vergleich von Systemalternativen.

Verfahren	Multiple Mittelwertvergleiche	Ranking & Selection Verfahren
Konzept	Ein vorhandener signifikanter Unterschied wird durch die Eingangsgrößen bestimmt.	Der signifikante Unterschied d^* wird vorgegeben, es muss der Stichprobenumfang angepasst, d.h. vergrößert werden.
Eingangsgrößen	<ul style="list-style-type: none"> • Fehlerwahrscheinlichkeit • Anzahl der Systemalternativen • Art der Stichprobenverteilung • Anzahl der Stichproben 	<ul style="list-style-type: none"> • Fehlerwahrscheinlichkeit • Anzahl der Systemalternativen • Art der Stichprobenverteilung • Anfängliche Anzahl der Stichproben • Signifikanter Unterschied d^*
Ergebnisse	Reihenfolge Signifikante Unterschiede	Reihenfolge Ggf. „bestes System“
Vorteile	Der Anwender bekommt eine Aussage über signifikante Unterschiede.	Durch die Vorgabe von d^* bestimmt der Anwender, was für ihn ein signifikanter Unterschied ist.
Nachteile	Ggf. finden sich keine signifikanten Aussagen bei großen Varianzen, trotz vorhandener Unterschiede. Evtl. werden Aussagen bei höheren Stichprobenumfängen möglich.	Der Stichprobenumfang ist nicht festgelegt. Ggf. hoher Simulationsaufwand, um die notwendigen Stichprobenumfänge bereitzustellen.

Tabelle 3.1: Vergleich der Verfahren der Multiplen Mittelwertvergleiche mit Ranking & Selection Verfahren.

Für die robuste integrierte Auswertung der Simulationsergebnisse und die Beantwortung der Frage nach den signifikanten Unterschieden haben die MMV-Verfahren zwei wesentlich Vorteile:

- Ein signifikanter Unterschied wird ggf. erkannt und muss nicht vom Anwender vorgegeben werden.
- Die Anzahl der Stichproben wird von Anwender vorgegeben und kontrolliert.

Bei der Betrachtung von „vielen“¹³ Systemvarianten bietet sich allerdings auch eine Kombination der beiden Verfahren an: Mit Ranking & Selection-Verfahren und einem geringen, vorgegebenen signifikanten Unterschied d^* wird eine Vorauswahl („Screening“) getroffen; diese ausgewählten Alternativen werden dann mit den MMV-Verfahren auf signifikante Unterschiede, deren Richtung und damit auf eine Reihenfolge hin untersucht.

3.6 Auswahl und Integration der Verfahren

Mit dem Konzept des integrierten Auswertungsprozesses mit seinen Teilaufgaben sowie den in Kapitel 2 abgeleiteten Anforderungen lassen sich die geeigneten Verfahren identifizieren, auswählen und zum Gesamtverfahren zur robusten statistischen Auswertung integrieren. Für eine effektive Selektion der Verfahrenskandidaten sind die drei Anforderungen

- 1. Robustheit**
- 2. Plausibilität**
- 3. Verwendbarkeit ohne Voraussetzung der Normalverteilung**

als Schlüsselkriterien der Auswahl geeignet. Speziell die ersten beiden Schlüsselkriterien werden im Folgenden erläutert.

3.6.1 Robustheit

Für die Definition des statistischen Begriffs der „Robustheit“ werden die bei Büning [Büni91:31] und Grammel [Gram91:16] beschriebenen Konzepte benutzt.

Für einen statistischen Test oder eine Schätzung muss zuerst ein Modell festgelegt werden, mit dessen Modellannahmen Aussagen über die Eigenschaften des Bereichs- oder Punktschätzers bzw. des Testverfahrens definiert werden können. Robuste statistische Verfahren werden

- bei geringen Abweichungen von den Modellannahmen nur gering beeinflusst und
- erweisen sich bei größeren Abweichungen als nicht überproportional schlecht.

Die beiden quantifizier- und messbaren Kriterien für das Auswahlkriterium Robustheit sind:

¹³ „Viele“ Systemvarianten bedeutet in diesem Zusammenhang mehr als 10 ($k > 10$), vgl. auch Kleijnen [Klei87].

- **Validitätsrobustheit:**
Die Validitätsrobustheit wird durch die Differenz zwischen dem nominalen (unter Modellannahmen) und dem tatsächlichen (bei Abweichungen von einigen Modellannahmen) Test- oder Konfidenzniveau bestimmt.
- **Effizienzrobustheit:**
Die Effizienzrobustheit beurteilt die Veränderung der Güte oder Trennschärfe eines Verfahrens, wenn die Modellannahmen nicht erfüllt sind.

Beide Kriterien werden im Rahmen von Monte-Carlo-Studien zur Beurteilung und Auswahl genutzt.

3.6.2 Einfachheit und Plausibilität

Speziell das Kriterium der Plausibilität und damit der Nachvollziehbarkeit für einen Simulationsanwender ohne vertiefte mathematische Kenntnisse schließt eine Reihe von mathematisch anspruchsvollen oder durch vorausgesetzte Modellannahmen entsprechend komplexe Verfahren aus. Als Konsequenz werden damit Vorgehensweisen und statistische Verfahren nicht weiter betrachtet, welche auf

- den Resampling-Konzepten wie „Bootstrapping“ oder „Jackknifing“ (siehe u.a. [Klei87:77], [Chen06], [ChCu09] und [Rinn08:489]),
- einer Transformation in den Frequenzbereichen und einer harmonischen Analyse (siehe u.a. [Law07:272] oder [Fish96:353]),
- der Methode der Standardisierten Zeitreihen (siehe u.a. [Law 07:526] oder
- einer logarithmischen Transformation und Rücktransformation (siehe u.a. [Klei87:18])

basieren.

Wenn ein Anwender auf Grundlage der Auswertungsverfahren Entscheidungen treffen soll, muss die entsprechende Nachvollziehbarkeit vorhanden sein und dürfen die Verfahren nicht als „Black-Box“ Verfahren angewendet werden.

Einfachheit, Plausibilität und Nachvollziehbarkeit sind relative und subjektive Kriterien, welche nicht direkt gemessen werden können, sondern im Rahmen von praktischen Tests von Anwendern mit industriellem Hintergrund bewertet werden.

3.6.3 Vorgehensweise zur Auswahl und Integration der Verfahren

Mit den oben beschriebenen Anforderungen und den Rahmenbedingungen der Simulationsmethodik ergibt sich ein vierstufiges Vorgehen zur Auswahl und Integration der Verfahren:

1. Literaturanalyse für die Identifikation von Kandidaten für Verfahren, welche die Anforderungen aus Kapitel 2 und vor allem die Schlüsselkriterien erfüllen.
2. Monte-Carlo-Studien zur Evaluation und Auswahl eines Teils der Verfahren, welche in der Literatur nur unvollständig für die Simulationsauswertung evaluiert wurden.
3. Überprüfung auf Plausibilität, Praxistauglichkeit und Anwenderakzeptanz der Verfahren im Rahmen von Simulationsprojekten für praktische industrielle Anwendungen.
4. Integration in ein robustes Gesamtverfahren.

Die Teilergebnisse der ersten beiden Schritte dieses Auswahlprozesses, die Literaturanalyse und die Monte-Carlo-Studien wurden ausführlich im Rahmen des ISSTAS Projekts [IAO98] dokumentiert. Die Monte-Carlo Studien fokussieren sich auf die Verfahren zur Bestimmung der stationären Phase und den Vergleich von Alternativen.

Die Überprüfung auf Plausibilität, Praxistauglichkeit und Anwenderakzeptanz wurde in zwei Projekten in intensiver Zusammenarbeit mit Anwendern aus der industriellen Praxis durchgeführt. Dabei handelt es sich um die Simulationsstudien einer Textilproduktion und einer Motorenfertigung, welche in den Arbeiten von **Stallkamp** [Stal94] und **Rassmann** [Rass97] beschrieben sind.

3.6.4 Monte-Carlo-Verfahren

Als Monte-Carlo-Verfahren oder -Methode¹⁴ werden mathematische Verfahren auf der Grundlage einer großen Anzahl von Zufallsexperimenten bezeichnet. Daher auch die Bezeichnung „Monte-Carlo“ nach dem Stadtteil in Monaco mit dem Sitz der bekannten Spielbank. Die Monte-Carlo-Methode verwendet dabei Prinzipien der Wahrscheinlichkeitsrechnung und Statistik, um geeignete, komplexe Probleme, welche analytisch schwierig sind, zumindest annäherungsweise numerisch zu lösen. Dies gilt und funktioniert auch für Aufgaben, welche gar keinen probabilistischen Hintergrund haben.

¹⁴ Metropolis und Ulam [MeUl49] vom Los Alamos Laboratory verwenden 1949 zum ersten Mal den Begriff „Monte Carlo Methode“, um dieses Verfahren zu beschreiben.

Nach **Fishman** [Fish96:1] bieten die Monte-Carlo Methoden näherungsweise Lösungen für eine Vielzahl von mathematischen Aufgaben durch statistische Stichprobenexperimente mit Hilfe eines Rechners.

Eine erweiterte Definition wird von **Kleijnen** [Klei08:2] benutzt: Alle rechnerbasierten Verfahren mit dem Einsatz von (Pseudo-) Zufallszahlen sind Monte-Carlo-Verfahren. Damit wird auch die stochastische ereignisdiskrete Simulation im Prinzip als Monte-Carlo-Verfahren interpretiert.

3.6.5 Die Monte-Carlo-Studien zur Evaluation der Verfahren

In Anlehnung an die Studien zur Untersuchung der Verfahren der MMV von **Tamhane** [Tamh79] und **Dunnnett** [Dunn80], [Dunn82] wurden Monte-Carlo-Studien zur Bewertung und Auswahl von Verfahren für die Aufgaben

- der MMV und
- zur Bestimmung der stationären Phase

durchgeführt.

Hierbei wurden bei der Evaluierung der MMV-Verfahren in Erweiterung der obigen Untersuchungen im Wesentlichen auch das Verhalten bei schiefen Verteilungen und ungleichen Varianzen bewertet.

Da bei Monte-Carlo-Verfahren die Ergebnisqualität direkt von der Qualität der Erzeugung der (Pseudo-) Zufallszahlen sowie der abgeleiteten Verteilungen abhängt, wurden die beiden Monte-Carlo-Studien auf Arbeitsplatzrechnern mit Verwendung der NAG Softwareunterprogramm-bibliothek [NAG06] gerechnet. Damit ist eine numerisch und stochastisch korrekte Generierung der (Pseudo-) Zufallszahlen sowie der verwendeten Verteilungen gewährleistet.

Die Ergebnisse der Monte-Carlo-Studien werden hier zusammengefasst vorgestellt, eine detaillierte Ergebnisdarstellung ist in dem ISSTAS Projektbericht [IAO98] beschrieben.

3.6.5.1 Die Monte-Carlo-Studie zur Evaluation der Verfahren der Mehrfachen Mittelwertvergleiche

Die auf Grundlage der Literaturstudie gewählten Verfahren der MMV werden mit einer Monte-Carlo-Studie untersucht. Die Eignung der Verfahren wird dabei nach zwei Kriterien beurteilt:

Validitätsrobustheit: Einhaltung des vorgegebenen (nominalen) multiplen Signifikanzniveaus.

Effizienzrobustheit: Die Trennschärfe, Güte oder Power eines Verfahrens. Wie gut erkennt das Verfahren, dass die Nullhypothese zu verwerfen ist?

Die Nullhypothesen besagen, dass sich die Mittelwerte nicht unterscheiden. Das Problem wird paarweise betrachtet und man erhält mehrere Nullhypothesen.

$$H_{0i,j} : \mu_i = \mu_j \quad (3.4)$$

Ein geeignetes Verfahren hält das multiple Signifikanzniveau ein und hat relativ kleine Konfidenzintervallbreiten. Das bedeutet auch, dass ein empfehlenswertes MMV-Verfahren mit geringer Wahrscheinlichkeit eine richtige Nullhypothese ablehnt und mit hoher Wahrscheinlichkeit die Alternativhypothese bestätigt, falls diese richtig ist.

Bei der Studie wurden jeweils 1000 Monte-Carlo-Läufe mit dem multiplen Gesamtniveau $\alpha_G = 0,05$ durchgeführt und fünf Systemalternativen bzw. Szenarien bei einem Stichprobenumfang von 20 betrachtet. Die Kriterien zur Beurteilung der MMV sind:

1. Wie verhält sich die geschätzte Wahrscheinlichkeit für einen multiplen Fehler 1. Art zu dem vorgegebenen multiplen Signifikanzniveau von $\alpha_G = 0.05$?
2. Wie klein sind die mittleren Breiten der Konfidenzintervalle?

Als Referenz wurde die Standardnormalverteilung $N(0,1)$ verwendet, symmetrische Verteilungen wurden durch eine Student-t-Verteilung sowie schiefe Verteilungen durch die Beta- und X^2 -Verteilungen dargestellt.

Als Ergebnis der Monte-Carlo-Studie ergab sich eine Abhängigkeit von der Verteilungsform:

- **Symmetrische Abweichungen:** 10% getrimmte Schätzer für Mittelwert und Varianz ermöglichen ein verbessertes Verhalten der untersuchten Verfahren.
- **Unsymmetrische und schiefe Verteilungen:** je nach Varianzunterschieden erfüllen der t-Test mit der Bonferroni-Ungleichung (Homoskedastizität) oder der Wilcoxon-Ranktest mit der Šidák-Ungleichung (Heteroskedastizität) am besten die beiden obigen Kriterien.

Auf Basis der Ergebnisse dieser Monte-Carlo-Studie sowie der Arbeiten von **Tamhane** [Tamh79] und **Dunnnett** [Dunn80], [Dunn82] werden im folgenden Kapitel 5 die geeigneten MMV-Verfahren ausgewählt.

3.6.5.2 Die Monte-Carlo-Studie zur Evaluation der Verfahren zur Bestimmung der stationären Phase

Zur Bewertung der Verfahren zur Bestimmung der stationären Phase wird in der Monte-Carlo-Studie ein einfacher Warteschlangenprozess der Struktur

M/M/1 FIFO (in der Kendall-Notation) simuliert. Die Ankunftsrate ist exponentialverteilt mit dem Parameter λ , die Servicezeiten exponentialverteilt mit dem Parameter ω . Die zu untersuchenden Verfahren sollen den stationären Zustand, der bei Verkehrsdichte ρ mit

$$\rho = \lambda / \omega < 1 \quad (3.5)$$

erreichbar ist, richtig erkennen. Die betrachtete Variable ist die mittlere Wartezeit \bar{w} .

Bei den Tests als indirekte Verfahren zur Erkennung der Stationarität wurde das Signifikanzniveau auf $\alpha = 0,05$ festgelegt.

Im Rahmen der Studie wurden 100 Monte-Carlo Läufe durchgeführt und dabei jeweils 1000 Stichproben, d.h. aufeinander folgende Zeitpunkte, simuliert. Dabei wurden die direkten heuristischen Verfahren und die indirekten Tests evaluiert, in wie weit die unterschiedlichen Charakteristiken der stationären und transienten Phase zu erkennen sind. Die Ergebnisse wurden durch die grafisch aufbereiteten Zeitreihen verifiziert.

Die Monte-Carlo-Studie ergab die zuverlässigsten Ergebnisse für das heuristische Verfahren von Fischer [Fisc89] sowie für den indirekten Test von Vassiliacopoulos [Vass89]. Der indirekte Test alleine lieferte nicht immer anschauliche Ergebnisse, war allerdings im Rahmen der Monte-Carlo-Studie besser als die Test-Prozedur von Schruben, Singh und Thierney [ScST83][Klei87:110]. In der Gesamtbetrachtung erreichte eine Kombination von einem direkten heuristischen Verfahren mit dem indirekten Test von Vassiliacopoulos die besten Resultate beim Erkennen des Übergangs von der transienten in die stationäre Phase.

3.7 Konzept des Auswertungsprozesses

Zusammenfassend lässt sich das Konzept des integrierten robusten Verfahrens der statistischen Auswertung bei der Erfüllung von Rahmenbedingungen durch den Ablauf der Teilaufgaben mit der jeweiligen Auswahl der Verfahren darstellen. Es ist damit ein **mehrstufiges adaptives Verfahren**¹⁵, bei dem, abhängig von der Aufgabenstellung und einer ersten Beurteilung und Klassifikation der Beobachtungen, geeignete robuste Prozeduren ausgewählt und verwendet werden.

Damit ergibt sich der in Abb. 3.6 in einem UML Aktivitäten-Diagramm dargestellte Gesamtablauf der vier bis sieben Teilprozeduren:

¹⁵ Siehe auch Büning [Büni91:2].

- 1. Bestimmung der stationären Phase:** Falls bei einer Simulationsstudie nur das stationäre Verhalten interessant ist und eine zeitlich vorgelagerte transiente Einschwingphase ignoriert werden soll, ist eine Bestimmung des Übergangs von der transienten in die stationäre Phase notwendig. Die Beobachtungen aus der transienten Phase werden anschließend abgeschnitten bzw. ignoriert.
- 2. Eliminierung der Autokorrelation:** Sollen die Beobachtungen einem langen Simulationslauf entnommen werden, muss die Autokorrelation beseitigt werden, um die Unabhängigkeit der Stichproben zu gewährleisten.
- 3. Schätzung von Mittelwert, Median, Modus und Varianz sowie weiteren Parametern und Momenten der Verteilung:** Aus den Beobachtungen werden die obigen vier Kennzahlen und weitere Lageparameter der Verteilung geschätzt.
- 4. Grafische Darstellung der Verteilung der Beobachtungen:** Die Visualisierung der Verteilung der Stichproben ist eine wichtige Maßnahme, um eine Abweichung von der Normalverteilung zu erkennen. Weitere Diagramme sind bei der Darstellung der Zeitreihen und Konfidenzintervallen notwendig.
- 5. Klassifikation der Verteilung** mittels der geschätzten Verteilungsparameter und der grafischen Darstellung vor allem bezüglich ggf. vorhandener Abweichungen von der Modellannahme der Normalverteilung.
- 6. Berechnung eines Konfidenzintervalls für den Mittelwert mit geeigneten Verfahren:** Ein Konfidenzintervall ist der notwendige Güteindikator für die Punktschätzung einer Kennzahl, auch bei nicht normalverteilten Stichproben.
- 7. Mehrfache Mittelwertvergleiche zum Bestimmen signifikanter Unterschiede von Systemalternativen:** Um die signifikanten Unterschiede der Systemalternativen und deren Richtung bei verschiedener Varianz und Abweichungen von der Normalverteilung zu erkennen, werden die Prozeduren der mehrfache Mittelwertvergleiche verwendet.

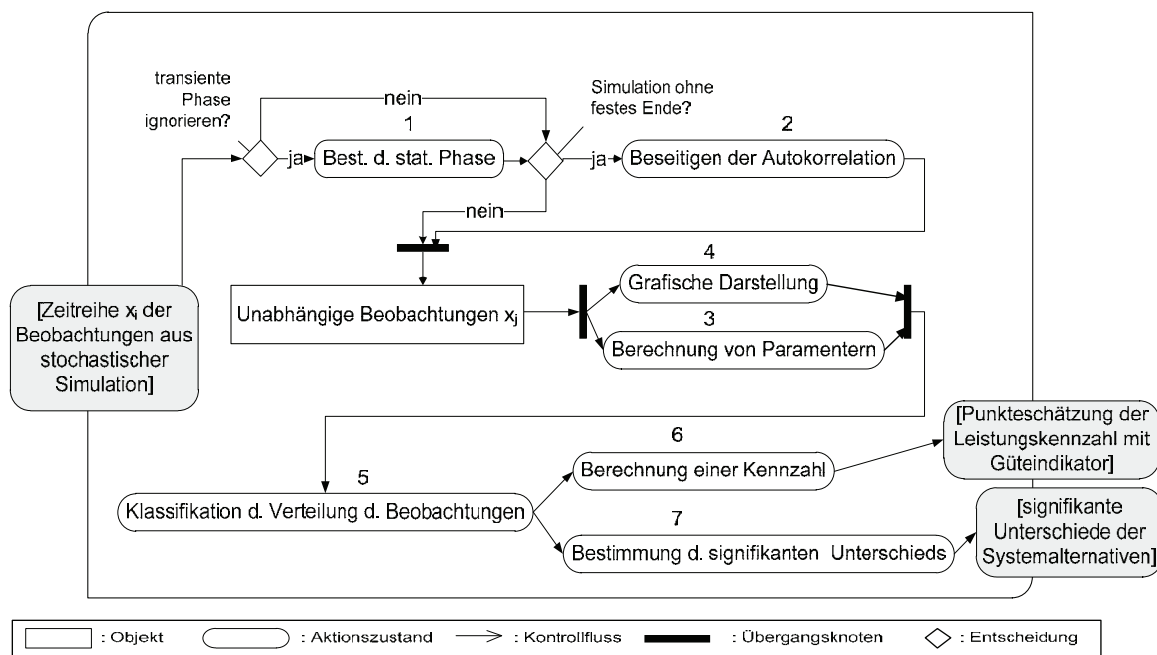


Abbildung 3.6: UML Diagramm des Gesamtablaufs der integrierten Vorgehensweise zur Auswertung.

Dieses Konzept des Auswertungsprozesses wird nun in den folgenden Kapiteln 4 und 5 als integriertes Verfahren mit einem detaillierten Vorgehensmodell beschrieben.

3.8 Sonstige Aufgaben bei der Auswertung

Beim überwiegenden Teil der praktischen Simulationsstudien ist der arithmetische Mittelwert als erwartungstreuer Schätzer der Leistungskennzahlen gesucht. Die vorgestellten Aufgaben bei der Auswertung konzentrieren sich daher auf eine univariate Schätzung und den Vergleich von Mittelwerten. Weitere Aufgaben oder Fragestellungen der statistischen Auswertung werden hier nicht vertieft.

Für die Schätzungen von alternativen Parametern, Kennzahlen und Momenten einer Verteilung wie Varianz, Quantile und Perzentile, Anteile oder Proportionen sei auf die aus der Schätzung des Mittelwerts abgeleiteten Verfahren, wie u.a. bei **Kleijnen** [Klei87:28] beschrieben, verwiesen.

Die Auswertung eines Simulationsmodells bezieht sich im Rahmen dieser Arbeit stets auf den univariaten Fall einer Kennzahl. Falls mehrere Kennzahlen bei den Alternativen in Form einer multivariaten Analyse gleichzeitig verglichen

werden sollen, bietet sich hierzu der Einsatz der Bonferroni-Ungleichung an¹⁶. Diese Ungleichung wird im Kapitel 5 bei dem Vergleich von mehreren Simulationsmodellen beschrieben. Weitere multivariate Analysemethoden sind in den Standardwerken bei **Backhaus u.a.** [BEPW93] sowie bei **Hartung** und **Elpelt** [HaEl99] beschrieben.

¹⁶ Vgl. auch die Empfehlung von Law [Law07:537].

4 Vorgehensweise zur Auswertung eines Modells

4.1 Zielsetzung und Rahmenbedingungen

In diesem Kapitel werden die in Kapitel 3 konzipierte Vorgehensweise zur Auswertung eines Modells konkretisiert sowie die Entscheidungsregeln und die statistischen Verfahren entwickelt und integriert.

Das Ziel bei der statistischen Auswertung der Simulationsdaten eines Modells ist die „bestmögliche“ Schätzung einer Kennzahl eines Simulationsmodells. Mit einer statistischen Sichtweise ist dies die Punktschätzung des Erwartungswerts bzw. Mittelwerts mit einem Konfidenzintervall als Qualitätsangabe. Dabei ist die Berechnung des Konfidenzintervalls abhängig von der Anzahl und Verteilung der Stichproben.

Typische und häufig untersuchte Kennzahlen bei Simulationsprojekten in der Produktion sind beispielsweise die Durchlaufzeit, der Durchsatz, die Wartezeit, die Länge von Warteschlangen oder die Auslastungen von Ressourcen [VDIF00].

Wie im vorigen Kapitel mit dem Konzept des integrierten Verfahrens dargestellt, gibt es vor der eigentlichen Punktschätzung des Mittelwerts je nach Gestaltung der Simulationsstudie ggf. noch die beiden vorbereitenden Aufgaben:

- Bestimmung der Stationarität, d.h. bei Anlaufprozessen der Zeitpunkt des Übergangs von der transienten in die stationäre Phase der Zeitreihe,
- Beseitigung der Autokorrelation, d.h. bei der Auswertung von Beobachtungen nur eines Simulationslaufs die Beseitigung der meist vorhandenen Autokorrelation dieser Stichproben.

Für diese eventuell notwendigen vorbereitenden Aufgaben und die eigentliche Punktschätzung mit Angabe des Konfidenzintervalls werden nun die den Anforderungen entsprechenden Regeln und Verfahren entwickelt und integriert.

Die dabei verwendeten Methoden werden im Anhang detailliert erläutert und beschrieben.

4.2 Bestimmung der stationären Phase

Aus der Kombination der Ergebnisse der Monte-Carlo-Studie aus Kapitel 3 mit den Anforderungen an Robustheit und Plausibilität aus Kapitel 2 lassen sich die geeigneten Methoden zur Bestimmung der stationären Phase ableiten.

Die Kombination und Integration des favorisierten direkten Verfahrens und des indirekten statistischen Tests aus der Monte-Carlo-Studie mit einem einfachen intuitiven und grafisch-interaktiven Verfahren aus der Literatur erfüllt am besten diese Anforderungen.

Damit ist für praktische Anwendungen eine Kombination der drei Prozeduren am besten geeignet:

- Das einfache direkte Verfahren „**Fischer**“ von **Fischer** [Fisc89] und
- das intuitive und grafisch-interaktive Verfahren „**Welch**“ nach **Welch** [Welc83], [Law07:509] sowie
- der indirekte Test auf Stationarität „**Vassiliacopoulos-Test**“ von **Vassiliacopoulos** [Vass89].

Der „Vassiliacopoulos-Test“ wird auch von **Ma und Kochhar** [MaKo93] in ihrer Vergleichsstudie empfohlen. Alle drei Methoden werden im Anhang detailliert erläutert.

Damit ergibt sich zum Ermitteln der stationären Phase die folgende integrative Vorgehensweise aus fünf wesentlichen Schritten:

1. Die Zeitreihe wird grafisch veranschaulicht. Die Wahl der Achseneinteilung ist von großer Bedeutung und daher sollten die Skalierungen variiert werden.
2. Der Benutzer lässt von dem direkten Verfahren „Fischer“ eine Abschneidegrenze berechnen.
3. Falls die empfohlene Abschneidegrenze nicht plausibel erscheint, wird mit dem „Welch“ Verfahren eine alternative Abschneidegrenze berechnet. Dabei werden mehrere, $n \geq 5$, Replikationen durchgeführt, die erhaltenen Zeitreihen punktweise gemittelt und die „gemittelte“ Zeitreihe betrachtet. Zusätzlich zur leichteren Bestimmung der Abschneidegrenze wird die Zeitreihe durch Mittelung w benachbarter Beobachtungen, d.h. durch einen gleitenden Durchschnitt mit Breite w , geglättet.
4. Das Testverfahren von Vassiliacopoulos sollte die gewählte, ggf. spätere, Abschneidegrenze überprüfen. Der Test liefert als Ergebnis die Aussage „stationäre Phase erreicht“ oder „stationäre Phase noch nicht erreicht“.
5. Falls der Test die Abschneidegrenze ablehnt, sollte der Benutzer eine andere, spätere, Grenze angeben und den Vassiliacopoulos-Test wiederholen.

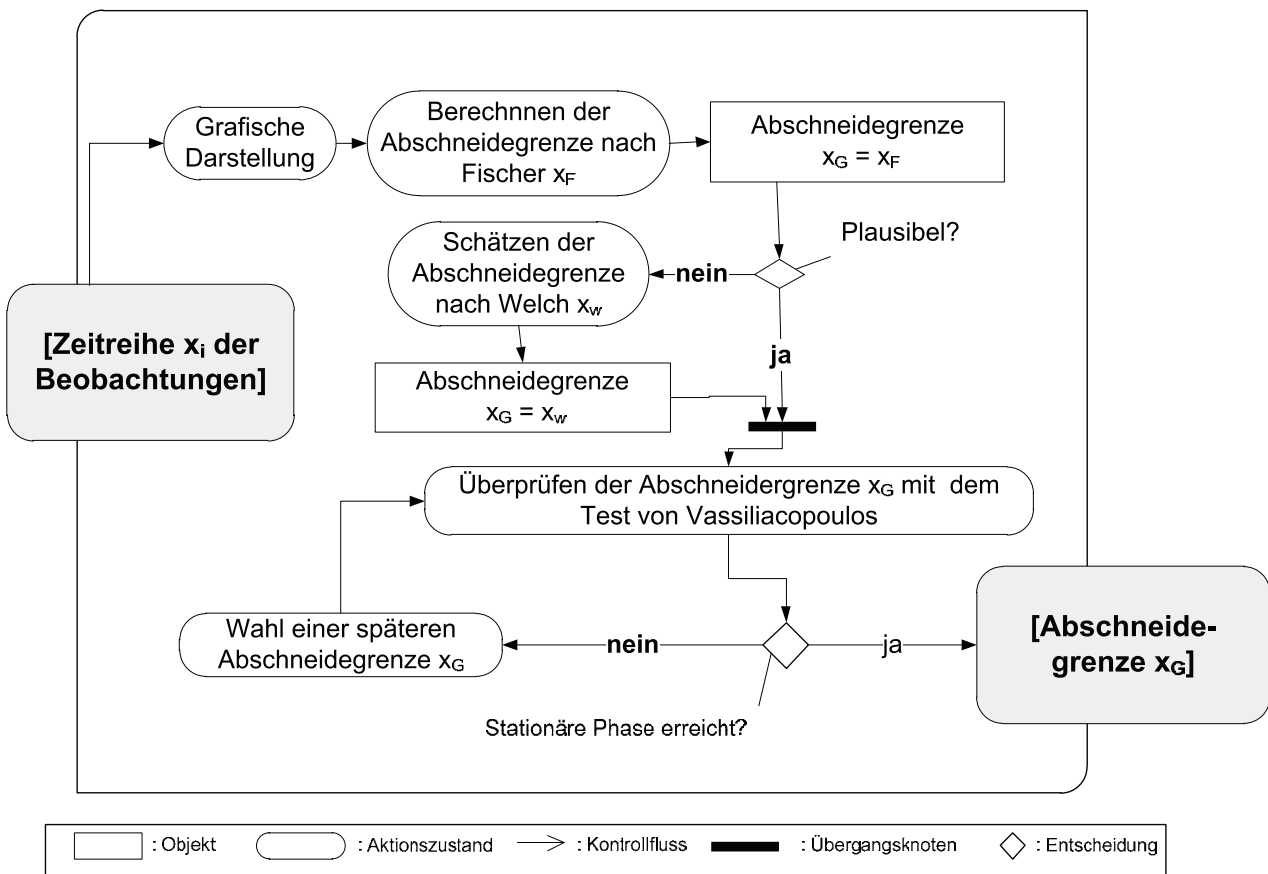


Abbildung 4.1: UML Diagramm der Vorgehensweise für die Bestimmung der stationären Phase.

Das Ende der transienten Übergangsphase ist für eine Simulation allein aus dem Verlauf einer einzigen Zeitreihe unter dem Gesichtspunkt eines hohen Signifikanzniveaus oft nur schwer zu ermitteln. Eine weitere Möglichkeit zur leichteren Bestimmung der Abschneidegrenze besteht darin, die Zeitreihe durch Mittelung benachbarter Beobachtungen, d.h. durch Batch Means (s.u.) anstatt eines gleitenden Durchschnitts, zu glätten.

Generell empfiehlt sich beim Bestimmen des Phasenübergangs ein sicherheitsorientiertes Vorgehen, um damit im Zweifelsfall einige bereits stationäre Beobachtungen zu ignorieren.

4.3 Eliminierung der Autokorrelation der Stichproben mit Batch Means

Bei der Nutzung der Stichproben aus nur einem Simulationslauf für die Schätzung einer Kennzahl muss die Autokorrelation der Stichproben beseitigt werden. Ansonsten wird durch eine Unterschätzung der Varianz ein zu kleines

Konfidenzintervall und damit eine zu hohe positive Qualität der Schätzung berechnet.

Wie in Kapitel 3 beschrieben ist für die Erfüllung der Anforderungen aus Kapitel 2 das einfache und plausible Verfahren der „**nichtüberlappenden Batch Means**“ am besten geeignet. Diese Batch Means berechnen sich durch eine Gruppierung der n fortlaufenden Stichproben $x_i, x_{i+1}, x_{i+2}, \dots, x_n$ in b so genannter Batches der Größe m mit $n = bm$. Es werden die Mittelwerte B_j der einzelnen Batches berechnet.

$$B_j = \frac{1}{m} \sum_{i=1}^m x_{(j-1)m+i}, j = 1, \dots, b \quad (4.1)$$

Da die Kovarianz von x_k und x_{k+1} mit wachsendem i normalerweise gegen Null geht, reduziert sich auch die Kovarianz von B_j und B_{j+1} für eine wachsende Batch Größe m . Damit können die Batch Means B_j für entsprechend große m näherungsweise als unabhängig betrachtet werden und anstatt der eigentlichen Stichproben x_i für die Punktschätzung mit Hilfe des arithmetischen Mittels (siehe unten) verwendet werden.

Bei dem Batch Mean Verfahren BM gilt die empirische Regel von **Schmeiser** [Schm82], dass die Anzahl der Batches b nicht weniger als zehn und nicht mehr als zwanzig betragen soll.

$$10 \leq b \leq 20 \quad (4.2)$$

Die Korrelationsfreiheit, die Unabhängigkeit und Zufälligkeit der Mittelwerte der einzelnen Batches bzw. Gruppen wird mit einem **Korrelogramm** überprüft und mit dem **Run-Test** getestet. Gegebenenfalls muss die Anzahl der Stichproben bzw. Batchgröße m in den b einzelnen Batches vergrößert und damit der Simulationslauf entsprechend verlängert werden. Abb. 4.2 beschreibt diese Vorgehensweise in einem UML Diagramm.

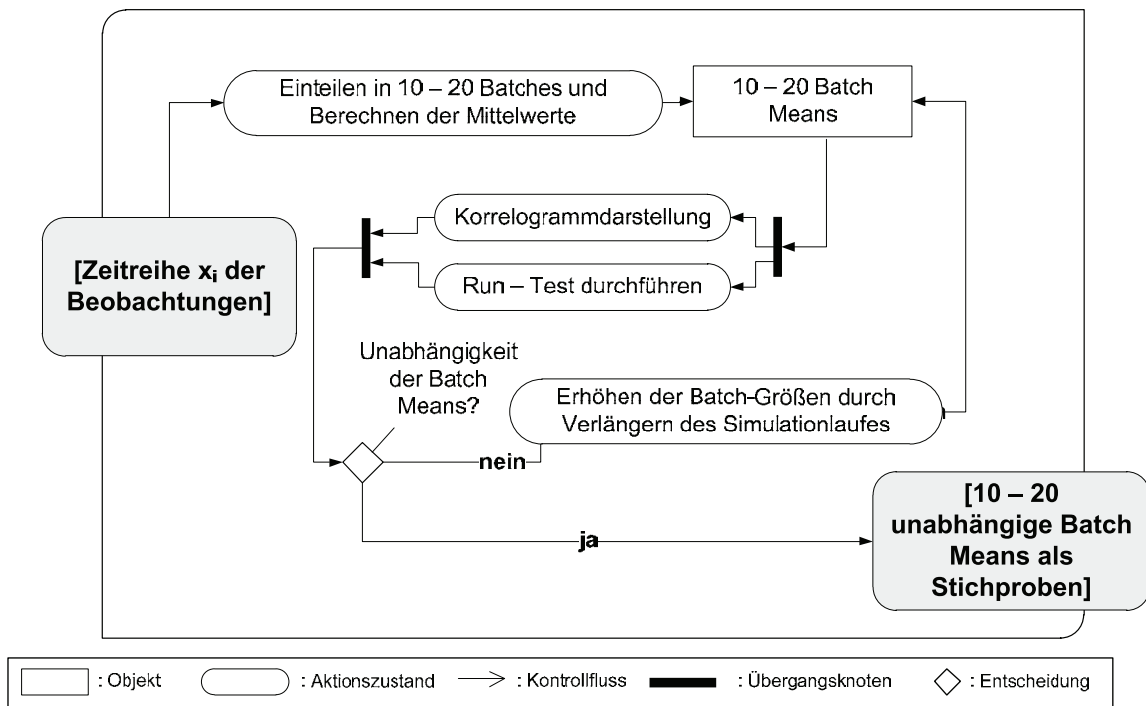


Abbildung 4.2: UML Diagramm der Vorgehensweise für die Beseitigung der Autokorrelation mit dem Batch Mean Verfahren.

4.4 Schätzung einer Kennzahl und Bewertung der Ergebnisqualität

Für die Punktschätzung einer Kennzahl ist das arithmetische Mittel $\bar{x}(n)$ der Stichproben x_i mit

$$\bar{x}(n) = \frac{\sum_{i=1}^n x_i}{n} \quad (4.3)$$

der erwartungstreue Punktschätzer für die Kennzahl der Grundgesamtheit. Das Konfidenzintervall zu dieser Punktschätzung wird bei Gültigkeit der „klassischen Voraussetzungen“ **NID** aus der t-Statistik berechnet.

Die Verfahren zur Berechnung des Konfidenzintervalls einer Mittelwertschätzung zur Bewertung der Ergebnisqualität hängen von der Form der erhaltenen Verteilung der Stichproben und dabei speziell vom Grad der Abweichung der Stichprobenverteilung von der Normalverteilung ab. Wenn sich die empirische Stichprobenverteilung stark von der Normalverteilung unterscheidet oder, wie häufig bei Simulationsergebnissen, stark abweichende Formen annimmt, so gilt natürlich die Annahme einer Normalverteilung nicht mehr.

4.4.1 Lageparameter der Verteilung

Um die Verteilung der Stichproben bzgl. der Abweichung von der Modellannahme der Normalverteilung beurteilen zu können, ist die Schätzung von weiteren Kennzahlen und Lageparametern notwendig. Die Verteilungsform lässt sich beurteilen, wenn die neben dem Mittelwert auch Modus x_{mod} (oder „Modalwert“ oder „häufigster Wert“ und damit der „maximale Wert der Dichtefunktion“) und Median \tilde{x} (der Medianwert teilt die sortierten Beobachtungen in zwei gleiche Hälften) sowie die zweiten bis vierten Momente Varianz σ^2 , Schiefe γ_1 und Exzess γ_2 bestimmt werden. Hierbei gilt für die Schiefe als das auf die dritte Potenz der Standardabweichung σ bezogene dritte zentrale Moment $\mu_3(X)$

$$\gamma_1(X) = \frac{\mu_3(X)}{\sigma^3(X)} = \frac{E((X - E(X))^3)}{\text{Var}(X)^{\frac{3}{2}}} \quad (4.4)$$

mit dem Erwartungswert $E(X)$ und der Varianz $\text{Var}(X)$. Der Exzess γ_2 , auch Kurtosis oder Wölbung genannt, ist als die normierte Form des vierten zentralen Moments $\mu_4(X)$

$$\gamma_2(X) = \frac{\mu_4(X)}{\sigma^4(X)} - 3 = \frac{E((X - E(X))^4)}{\text{Var}(X)^2} - 3 \quad (4.5)$$

definiert. Beide Momente sind bei einer standardisierten Normalverteilung $N(0,1)$ gleich null und damit geeignete Indikatoren für eine Abweichung von dieser Modellannahme.

4.4.2 Anzahl und Verteilung der Stichproben

Bei der Schätzung eines Mittelwerts einschließlich des Konfidenzintervalls hängt die Qualität der Schätzung von der Anzahl und Verteilung der Stichproben ab. Als pragmatische Empfehlung für die Größenordnung der Stichproben n ist in Übereinstimmung mit der Standardliteratur (Kleijnen [Klei87:24], [Klei08:80] und Law [Law07:504]) eine anfängliche Anzahl zwischen zehn und zwanzig ($10 \leq n \leq 20$) zu empfehlen. Da allerdings nach der t-Statistik die Konfidenzintervallbreite Br ungefähr umgekehrt proportional zur Quadratwurzel der Stichprobenanzahl ist, muss für eine höhere Genauigkeit ggf. die Anzahl der Stichproben n entsprechend vergrößert werden.

$$Br \sim 1/\sqrt{n} \quad (4.6)$$

Unter den Verfahren, welche zur Beurteilung einer Abweichung der Verteilung von der Modellannahme der Normalverteilung herangezogen werden können, ist vor allem der Chi-Quadrat-Anpassungstest (X^2 -Test) geeignet¹⁷. Ob die Stichproben als normalverteilt betrachtet werden können, sollte aber nicht allein auf Basis dieses Tests entschieden werden. Der Anwender sollte mittels der dritten und vierten Momente und vor allem mit der grafischen Darstellung der Verteilungsdichte entscheiden, ob die Abweichungen von der Normalverteilung toleriert werden können oder nicht. Diese etwas pragmatischere Vorgehensweise, anstatt nur auf der Basis von Testergebnissen zu entscheiden, kann und sollte, wegen der Robustheit der nachfolgend beschriebenen Verfahren durchaus angewandt werden. Der Anwender muss entscheiden, ob die Stichproben als normalverteilt betrachtet werden können, und falls nicht, in welcher Form und mit welchem Ausmaß diese Verteilung von einer Normalverteilung abweicht.

Um die Form der Stichprobenverteilung beurteilen zu können, kann es notwendig sein, die Anzahl der Stichproben zu vergrößern. Um eine erhaltene empirische Verteilung der Modellannahme der Normalverteilung anzunähern, kann der Zentrale Grenzwertsatz von Nutzen sein:

Wenn die beobachtete Variable ein Mittelwert ist oder aus dem additiven Zusammenwirken vieler Einflüsse resultiert, sollte sich die Verteilung der Stichproben mit Vergrößerung der Stichprobenanzahl nach dem zentralen Grenzwertsatz einer Normalverteilung annähern¹⁸.

Falls aber, wie in den meisten Fällen (u.a. nach [Klei87]), nur geringe bis mäßige Abweichungen von der Normalverteilung vorliegen, so kann eines der folgenden Verfahren verwendet werden.

4.4.3 Beurteilung und Klassifizierung der Verteilungsform

Zur Beurteilung und Klassifizierung der Verteilungsform der Beobachtungen in Hinsicht auf die Art der Abweichung von der Normalverteilung werden drei Methoden bzw. Kriterien empfohlen:

- Der X^2 -Test,

¹⁷ Kleijnen [Klei87:16], [Klei87:95], [Klei08:80] und Law [Law07:340] empfehlen als weitere oder alternative Anpassungstests bzw. „Goodness-of-Fit-Tests“ die Verfahren von Kolmogorow-Smirnoff und Anderson-Darling.

¹⁸ Alternativ kann mit einer In-Transformation der schiefen Verteilung die Abweichung von der Normalverteilung ebenfalls reduziert werden. Mit der Transformation wird auch eine nur im Bereich der positiven reellen Zahlen R^+ definierte Verteilung auf die ganzen reellen Zahlen R abgebildet. Allerdings ist die Rücktransformation des Konfidenzintervalls nicht trivial und deshalb für Praktiker nicht unbedingt geeignet.

- die Berechnung und das Verhältnis der dritten und vierten Momente der Verteilung, Schiefe γ_1 und Exzess γ_2 sowie
- die grafische Aufbereitung und Darstellung der Verteilung im Vergleich zu einer entsprechenden Normalverteilung als Referenz.

Die Integration des Anwenders bei der qualitativen Beurteilung der Verteilungsform respektive der Abweichung von der Normalverteilungsannahme ist wichtig, um die Plausibilität und Akzeptanz der folgenden Entscheidungen und Berechnungen zu sichern.

4.4.3.1 Grafische Darstellung der Verteilungsdichte

Wie in Kapitel 2 beschrieben ist die Einbindung des Anwenders eine wichtige Anforderung an das integrierte Verfahren. Eine geeignete Visualisierung der Daten und Zwischenergebnisse als Grundlage und Unterstützung für die notwendigen Entscheidungen des Anwenders ist unerlässlich, um die erforderliche Plausibilität zu sichern und damit die eigentliche Akzeptanz des Verfahrens zu gewährleisten.

Bei einer grafischen Darstellung der Verteilung der Stichproben zur Beurteilung und Klassifizierung durch den Simulationsanwender ist die Dichtefunktion besser geeignet als die eigentliche Verteilungsfunktion, insbesondere wenn es einen Vergleich mit einer Normalverteilung als Referenz betrifft.

Für die Schätzung und grafische Aufbereitung der Dichtefunktion der Verteilung der Beobachtungen eignen sich sowohl ein Histogramm als auch eine geglättete Darstellung mit Dichteschätzern. Bei der Visualisierung der Schätzung der empirischen Dichte sollte eine geeignete glättende Kernfunktion, wie z.B. der „Gauss-Kern“ oder der „Tukey’s biweight kernel estimator“ [Rinn08:495], [Hart82:872], [VDIF95], als Kern des Dichteschätzers verwendet werden.

Bei der Schätzung der Verteilungsdichte der Beobachtungen mit einem Histogramm kann die Wahl des Startwerts, die Klassenanzahl und Klassenbreite vom Anwender bestimmt werden. Speziell die Klassenbreite sollte variiert werden, da diese erheblichen Einfluss auf die Form und Darstellung des resultierenden Histogramms hat.

Bei beiden Darstellungsarten der Dichteschätzung der Stichproben ist es für die Klassifizierung der Verteilung hilfreich, eine entsprechende Normalverteilung

mit den gleichen Lageparametern Mittelwert und Varianz als Referenzverteilung in derselben Abbildung darzustellen¹⁹.

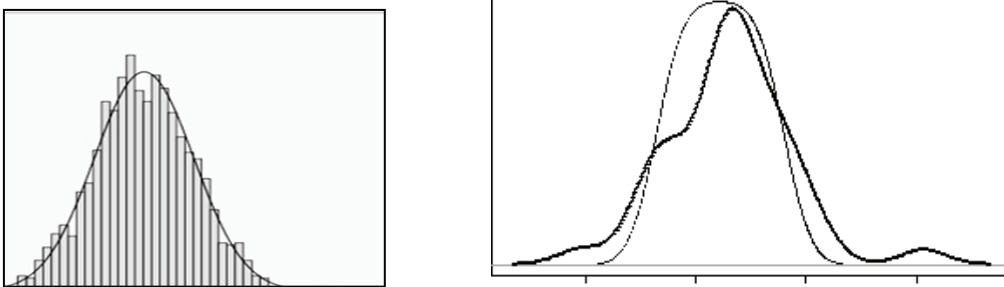


Abbildung 4.3: Histogramm und Dichteschätzung einer empirischen Verteilungsfunktion mit einer Normalverteilung als Referenz.

Zusätzlich zur Visualisierung der Dichtefunktion empfiehlt sich eine Darstellung von Informationen über die Verteilung in Box-Plots (auch Kasten-Zeichnungen). Speziell schiefe Abweichungen von einer Normalverteilung lassen sich mit Box-Plots sehr gut erkennen. Die explorative Darstellung mit Box-Plot zeigen die Symmetrie, Streuung und Ausreißer einer Verteilung (siehe u.a. [Rinn08: 49]). Der zweigeteilte Kasten besteht aus den Werten des Medians \tilde{x} sowie dem oberen $x_{0,75}$ und unteren Quartil $x_{0,25}$ der empirischen Verteilung.

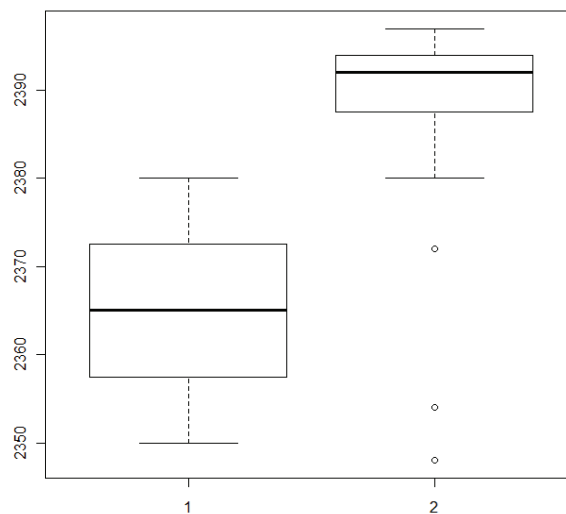


Abbildung 4.4: Box-Plots einer empirischen Verteilungsfunktion mit einer Normalverteilung als Referenz.

Das Ziel der grafischen Darstellung der Schätzung der Verteilungsdichte der Beobachtungen ist nicht eine möglichst gute Anpassung an die Normalverteilung

¹⁹ Eine entsprechende Vorgehensweise unterstützt auch das Programm „ExpertFit“ [Law07], welches zu Modellierung von stochastischen Eingangsgrößen (siehe Kap. 3.3) mit passenden Verteilungen dient.

lung als Referenz, sondern eine grafische Unterstützung des Anwenders bei der anschließenden Klassifikation der Abweichungen von dieser Modellannahme in gering, symmetrisch oder schief.

4.4.3.2 Indikatoren und Klassifikation der Verteilung der Beobachtungen

Da bei Abweichung von der Modellannahme der Normalverteilung alternative Verfahren für die Berechnung des Konfidenzintervalls notwendig sind, muss die Verteilung der Beobachtungen beurteilt und entsprechend klassifiziert werden. Die notwendigen Indikatoren für diese Klassifikation sind die Schätzungen von

- Mittelwert \bar{x} , Modus x_{mod} und Median \tilde{x} ,
- zweitem Moment der Varianz s^2 , drittem Moment der Schiefe γ_1 und viertem Moment des Exzess γ_2 ,
- die Beschreibung mit Box-Plots,
- die grafische Darstellung der Verteilungsdichte mit Histogramm und Dichteschätzung im Vergleich zu eine Normalverteilung als Referenz sowie
- das Ergebnis des X^2 -Tests.

Mit dem statistischen Test, der grafischen Darstellung und den Kennzahlen bzw. den Relationen der Kennzahlen lässt sich die Verteilung der Beobachtungen in drei Kategorien der Abweichungen von der Normalverteilung klassifizieren, um danach mit jeweils geeigneten Verfahren ein Konfidenzintervall als Güteindikator schätzen zu können.

Geringe Abweichungen von der Normalverteilung:

Kriterien für eine geringe Abweichung von einer Normalverteilung sind:

- Die Werte für Mittelwert \bar{x} , Modus x_{mod} und Median \tilde{x} haben relativ geringe Abweichungen,
- die Werte von Schiefe γ_1 und Exzess γ_2 sind sehr klein,
- der Box-Plot ist mit einer Normalverteilung vergleichbar,
- der X^2 -Test bestätigt die Hypothese der Normalverteilung und
- die grafischen Darstellungen der Verteilungsdichte sind mit einer Normalverteilung vergleichbar.

Auch wenn der X^2 -Test die Annahme der Normalverteilung ablehnt, sollte der Anwender die Verteilung der Beobachtungen bei der Erfüllung der qualitativen Kriterien entsprechend klassifizieren.

Symmetrische Abweichungen von der Normalverteilung:

Die Kriterien für eine symmetrische Abweichung von einer Normalverteilung sind:

- Die Werte für Mittelwert \bar{x} , Modus x_{mod} und Median \tilde{x} haben relativ geringe Abweichungen,
- der Wert der Schiefe γ_1 ist klein, aber der Betrag des Exzess γ_2 ist deutlich größer Null,
- der Box-Plot ist symmetrisch,
- der X^2 -Test verwirft die Hypothese der Normalverteilung und
- die grafischen Darstellungen der Verteilungsdichte zeigen deutliche symmetrische Abweichungen von einer Normalverteilung als Referenz.

Bei der Einordnung als Verteilung mit symmetrischen Abweichungen zum Modell der Normalverteilung sollte vor allem die grafische Darstellung als wesentliches Kriterium dienen.

Schiefe Abweichung von der Normalverteilung:

Als Kriterien für eine schiefe Abweichung von einer Normalverteilung gelten:

- Die Werte von Modus x_{mod} und/oder Median \tilde{x} haben deutliche Abweichungen vom Mittelwert \bar{x} ,
- der Betrag des Wert der Schiefe γ_1 ist deutlich größer Null,
- der Box-Plot zeigt eine schiefe Verteilung,
- der X^2 -Test verwirft die Hypothese der Normalverteilung und
- die grafischen Darstellungen der Verteilungsdichte zeigen deutliche schiefe Abweichungen zu einer Normalverteilung als Referenz.

Bei Simulationsstudien von Systemen, welche auch als Warteschlangenmodelle abgebildet werden können, treten häufig rechtschiefe Verteilungen mit einem positiven dritten Moment und einem Modus und Median kleiner als dem Mittelwert auf.

Rechtsschiefe Verteilung: $\gamma_1 > 0$ und $x_{\text{mod}} < \bar{x}$, $\tilde{x} < \bar{x}$ (4.7)

Die Klassifikation der Verteilungsform und die Indikatoren der Beurteilung sind in der folgenden Tabelle 4.1 zusammengefasst.

Beurteilung der Abweichung	Indikatoren
Geringe Abweichung	<ul style="list-style-type: none"> • $x_{\text{mod}} \sim \tilde{x} \sim \bar{x}$ • $\gamma_1 \sim 0, \gamma_2 \sim 0$ • Box-Plot $\sim (N)$ • X^2-Test bestätigt H_0 • Histogramm & Dichte $\sim (N)$
Symmetrische Abweichung	<ul style="list-style-type: none"> • $x_{\text{mod}} \sim \tilde{x} \sim \bar{x}$ • $\gamma_1 \sim 0, \gamma_2 > 0$ • Box-Plot symmetrisch • X^2-Test lehnt H_0 ab • Histogramm & Dichte symmetrisch $\neq (N)$
Schiefe Abweichung	<ul style="list-style-type: none"> • $x_{\text{mod}} \neq \bar{x}, \tilde{x} \neq \bar{x}$ • $\gamma_1 > 0$ • Box-Plot unsymmetrisch • X^2-Test lehnt H_0 ab • Histogramm & Dichte schief $\neq (N)$

Tabelle 4.1: Klassifikation der Abweichung von der Normalverteilung.

Bei extremen Formen der Abweichung von der Modellannahme der Normalverteilung wie Mehrwipfligkeit oder extremer Schiefe hilft ggf. eine deutliche Vergrößerung der Stichprobenanzahl nach dem zentralen Grenzwertsatz.

Die für die jeweiligen drei Klassifikationen geeigneten und empfohlenen Verfahren zur Berechnung des Konfidenzintervalls werden im Folgenden dargestellt.

Eine Sensitivitätsuntersuchung der Entscheidung, d.h. wie verändert sich das berechnete Konfidenzintervall (und damit die Güte der Schätzung der Kennzahlen) bei jeweils unterschiedlicher Klassifikation der Verteilungsform, trägt erheblich zur Plausibilität und damit zur Anwenderakzeptanz bei.

4.4.4 Geringe Abweichung von der Normalverteilung:

Bei geringen Abweichungen von der Normalverteilung kann man sich auf die Robustheit des t-Tests verlassen. Robust bedeutet, dass dieser Test nicht empfindlich auf Abweichungen von der Normalverteilung reagiert. Wenn eine „große“ Anzahl n von Beobachtungen vorliegt, ist die t-Statistik robuster als bei nur wenigen Stichproben. Mehr als 30 Stichproben gelten in Anlehnung an **Kleijnen** [Klei87:24], [Klei08:80] als „große“ Anzahl von Beobachtungen (siehe auch 5.3.2).

Bei der Anwendung der t-Statistik unter klassischen Voraussetzungen gilt

$$t_{n-1} = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{s_x / \sqrt{n}} \quad (4.8)$$

mit dem beobachteten empirischen Mittelwert \bar{x} , dem „wahren“ Mittelwert μ und der Schätzung der Varianz $s_{\bar{x}}^2$ der n Beobachtungen. Die Zufallsvariable t_{n-1} besitzt eine t-Verteilung mit $n-1$ Freiheitsgraden. Damit kann die Wahrscheinlichkeit, dass t_{n-1} einen bestimmten Wert überschreitet, berechnet werden. Um ein Konfidenzintervall für die Sicherheitswahrscheinlichkeit von $100(1-\alpha)$ Prozent zu berechnen, wird mit t_{n-1} der Wert bezeichnet, für den gilt:

$$\Pr(t_{n-1} \geq t_{n-1}^{\alpha}) = \alpha \quad (4.9)$$

Aus Gleichung 4.8 und 4.9 folgt:

$$\Pr\left(-t_{n-1}^{\alpha/2} < \frac{\bar{x} - \mu}{s_{\bar{x}}} < t_{n-1}^{\alpha/2}\right) = 1 - \alpha \quad (4.10)$$

gilt mit einer Sicherheit von $100(1-\alpha)$ Prozent, dass der wahre Mittelwert μ im Konfidenzintervall

$$\left[\bar{x} - t_{n-1}^{\alpha/2} s_{\bar{x}}, \bar{x} + t_{n-1}^{\alpha/2} s_{\bar{x}}\right] \quad (4.11)$$

liegt. Die Intervallbreite Breite vergrößert sich bei zunehmender Sicherheit $(1-\alpha)$ und wird kleiner mit einer zunehmenden Zahl von Beobachtungen nach (4.6).

4.4.5 Symmetrische Abweichungen von der Normalverteilung

Wenn die Verteilungen der Beobachtungen zwar von der Normalverteilung abweichen, aber dennoch symmetrisch sind sollten verteilungsfreie Verfahren wie der Vorzeichen- bzw. Zeichentest (oder der Vorzeichenrangtest nach Wilcoxon) eingesetzt werden, (vgl. hierzu auch [Gram91:58] und [Klei87:24]) um ein entsprechendes Konfidenzintervall zu berechnen.

Die verteilungsfreien Verfahren haben keine speziellen Voraussetzungen bezüglich der Verteilung, außer der Symmetrie. Die Verfahren können somit auf

die ganze Gruppe von symmetrischen Verteilungen, d. h. wenn Median und Mittelwert ungefähr gleich sind, angewandt werden²⁰.

Bei einer symmetrischen Stichprobenverteilung und einer „großen“ Stichprobenanzahl ($n > 30$) kann die Binominalverteilung durch die Standardnormalverteilung approximiert werden. Danach berechnet sich bei $n > 30$ ein $1-\alpha$ Konfidenzintervall für den Mittelwert μ bei nach der Größe geordneten Beobachtungen wie folgt: Mit den Rangzahlen

$$r = \left\lceil \frac{n}{2} - \frac{z^{\alpha/2} \sqrt{n}}{2} \right\rceil \quad (4.12)$$

und

$$s = \left\lceil \frac{n}{2} + \frac{z^{\alpha/2} \sqrt{n}}{2} \right\rceil \quad (4.13)$$

wird das Konfidenzintervall für den Mittelwert μ bestimmt.

$$x_{(r)} \leq \mu \leq x_{(s)} \quad (4.14)$$

Dabei ist $z^{\alpha/2}$ das (in Tabellen dokumentierte oder berechenbare) $\alpha/2$ -Quantil der standardisierten $N(0,1)$ Normalverteilung.

Entsprechend dieser Vorgehensweise lassen sich bei Bedarf auch Konfidenzintervalle für andere Quantile der Verteilung berechnen²¹.

4.4.6 Schiefe Abweichungen von der Normalverteilung

Wenn die Beobachtungen deutlich nicht normalverteilt sind, sondern, wie häufig bei Simulationsdaten, eine rechtsschiefe Verteilung besitzen, liegt eine deutliche Abweichung von der Modellannahme der Normalverteilung vor. Damit sollte der Anwender sich nicht mehr auf die Robustheit der t-Statistik verlassen, sondern eine speziell für diesen Fall modifizierte t-Statistik verwenden.

Bei einer qualitativen Klassifikation der Verteilungen als „stark schief“ sollte als zusätzliche Maßnahme die modifizierte t-Statistik auch mit robusten Schätzern kombiniert werden.

²⁰ Wenn der Median anstatt des Mittelwerts betrachtet wird, können die verteilungsfreien Verfahren auch auf unsymmetrische Verteilungen angewendet werden.

²¹ Siehe exemplarisch bei Hartung [Hart02:263]).

4.4.6.1 Modifizierte t-Statistik nach Johnson

Speziell für Beobachtungen mit schiefen Verteilungen modifiziert **Johnson** die t-Statistik (nach **Kleijnen** [Klei87:22]), indem er diese Schiefe als drittes Moment der beobachteten Verteilung berücksichtigt. Johnsons t-Statistik definiert \tilde{t} als Erweiterung von (4.3) wie folgt:

$$\tilde{t}_{n-1} = \left[(\bar{x} - \mu) + \frac{\hat{\mu}_3}{6s^2n} + \frac{\hat{\mu}_3}{3s^2} (\bar{x} - \mu)^2 \right] \left(\frac{s^2}{n} \right)^{-\frac{1}{2}} \quad (4.15)$$

Dabei ist $\hat{\mu}_3$ die Schätzung der Schiefe und s^2 die Schätzung der Varianz der erhaltenen Verteilung.

Für die Berechnung eines Konfidenzintervall empfiehlt Johnson [John78:538] den quadratischen Term mit $(\bar{x} - \mu)^2$ zu vernachlässigen. Damit reduziert sich (4.15) auf

$$\tilde{t}_{n-1} = \left[(\bar{x} - \mu) + \frac{\hat{\mu}_3}{6s^2n} \right] \left(\frac{s^2}{n} \right)^{-\frac{1}{2}} \quad (4.16)$$

und damit ein Konfidenzintervall mit der modifizierten t-Statistik nach Johnson auf

$$\left[\bar{x} + \frac{\hat{\mu}_3}{6s^2n} - t_{n-1}^{\alpha/2} s_{\bar{x}}, \bar{x} + \frac{\hat{\mu}_3}{6s^2n} + t_{n-1}^{\alpha/2} s_{\bar{x}} \right] \quad (4.17)$$

Diese Johnson t-Statistik \tilde{t} ist bei schiefen Abweichungen von der Normalverteilung zur Schätzung eines Konfidenzintervalls geeignet.

4.4.6.2 Robuste Schätzer

Bei extrem schiefen Verteilungen der Stichproben können die obigen Schätzer für Mittelwert, Varianz und Schiefe für die statistischen Momente durch robuste Schätzer ersetzt werden. Die Robustifizierung der Schätzer findet durch „Trimmen“ oder „Winsorisieren“ statt. Bei einem mit $\beta = 0,1$ getrimmten bzw. winsorisierten Schätzer werden bei den nach der Größe geordneten Stichproben jeweils die obersten und die untersten 10% ignoriert bzw. beim Winsorisieren durch den nächsten benachbarten Wert ersetzt [Büni91:54], [Hart02:881]. Typischerweise verwendet man Trimm- bzw. Winsor-Werte von $0,05 \leq \beta \leq 0,2$, d.h. 10% bis 40% der extremen Werte werden ignoriert bzw. ersetzt [SaHe06:70].

4.4.6.3 Sensitivitätsanalyse und weitere Beobachtungen

Bei der Berechnung des Konfidenzintervalls ist eine Sensitivitätsanalyse mit einer alternativen Entscheidung bei der Klassifikation der Verteilung in Schritt 5 sinnvoll²². Dabei sind die beiden alternativen Klassifikationen der Verteilung der Beobachtungen und die jeweils resultierenden Konfidenzintervalle interessant und relevant, um bei Bedarf entsprechend abgesicherte und konservative Güteaussagen für die Kennzahlenschätzung treffen zu können.

Da die Breite des berechneten Konfidenzintervalls (bei geringen und schiefen Abweichungen) nach Relation 4.6 ungefähr umgekehrt proportional zur Quadratwurzel der Anzahl der Beobachtungen ist, verringern weitere zusätzliche Beobachtungen entsprechend die Größe des Konfidenzintervalls.

Diese weiteren Stichproben reduzieren bei einer additiven Eigenschaft der Beobachtungen nach dem zentralen Grenzwertsatz auch die Abweichung von der Normalverteilung, wodurch ggf. auch eine neue Klassifikation als „geringe Abweichung“ und der Einsatz der t-Statistik möglich ist.

Je nach Rechenintensität des Simulationsmodells ist eine Vergrößerung der Stichprobenanzahl generell sinnvoll und empfehlenswert, da sich die Aussagequalität der Ergebnisse damit stetig verbessert.

4.5 Integration der Vorgehensweise zur Auswertung eines Modells

Als Detaillierung des in Kapitel 3.7 konzipierten Auswertungsprozesses lässt sich die Vorgehensweise zur statistischen Auswertung eines Simulationsmodells mit einem Ablauf in acht Hauptschritten beschreiben. Dabei sind die ersten beiden Schritte bzw. Teilprozeduren nur bei einer entsprechenden Gestaltung der Simulationsstudie notwendig.

1. **Bestimmung der stationären Phase durch grafische Darstellung, direkte Verfahren und Tests:**

Wenn bei einer Simulationsstudie nur das stationäre Verhalten interessant ist, muss die vorgelagerte transiente Phase erkannt und abgeschnitten werden. Dazu wird die folgende Prozedur empfohlen:

- a. Grafische Darstellung der Zeitreihe mit einer Variation der Achsenskalierung.
- b. Berechnen der Abschneidegrenze mit dem direkten Verfahren „Fischer“.

²² Auch Kleijnen [Klei87:27] empfiehlt, im Zweifelsfall ebenfalls alternative Berechnungen zu betrachten: „*In case of doubt apply several procedures*“.

- c. Alternativ, falls die Abschneidegrenze mit dem „Fischer“ Verfahren nicht plausibel erscheint, ist ein Mitteln über mehrere Replikationen und ein Glätten durch Zusammenfassen von einzelnen Werten, wie nach der Methode von „Welch“, zu empfehlen .
 - d. Überprüfen der gewählten Abschneidegrenze mit dem Testverfahren von Vassiliacopoulos. Der Test liefert als Ergebnis die Aussage „stationäre Phase erreicht“ oder „stationäre Phase noch nicht erreicht“.
 - e. Wenn der Test von Vassiliacopoulos die Abschneidegrenze ablehnt, sollte er mit einer „späteren“ Abschneidegrenze wiederholt werden.
2. **Eliminierung der Autokorrelation durch „Batch Means“:**
 Wenn die Beobachtungen einem einzigen langen Simulationslauf entnommen werden, muss mit dem Verfahren der Batch Means die vorhandene Autokorrelation beseitigt werden.
- a. Einteilung der Zeitreihe in 10 bis 20 Batches.
 - b. Überprüfen der „Batch Means“ auf Unabhängigkeit mit einem Korrelogramm und dem Run-Test.
 - c. Erhöhen der Batch-Größen durch ein Verlängern des Simulationslaufs, bis die Unabhängigkeit der Batch Means erreicht ist.
3. **Schätzung von Mittelwert, Median, Modus sowie den 2., 3. und 4. Momenten Varianz, Schiefe und Exzess der Verteilung der Beobachtungen:**
 Aus den Beobachtungen werden die obigen sechs Kennzahlen sowie die Lage- und Formparameter der Verteilung geschätzt. Diese Kennzahlen dienen zur Klassifikation der Verteilung bezüglich vorhandener Abweichungen von der Modellannahme der Normalverteilung.
4. **Grafische Darstellung der Verteilungsdichte der Beobachtungen:**
 Die Visualisierung der Verteilungsdichte der Stichproben ist eine wichtige Maßnahme, um die Form der Abweichung von der Normalverteilung zu erkennen. Dazu sind Dichteschätzer mit geeigneter Kernfunktion (z.B. „Gauss“ oder „Tukey’s biweight kernel“), ein Histogramm, bei dem Klassenanzahl und -breite variiert werden sowie Box-Plots am besten geeignet.
5. **Klassifizierung der Verteilung:**
 Die empirische Verteilung wird nach der Abweichung von der Modellannahme der Normalverteilung klassifiziert. Als Kriterien dazu dienen:

- Der X^2 -Test
 - Die grafische Darstellung aus der Verteilung wie in 4) mit Dichteschätzer, Histogramm und Box-Plots
 - Die Beurteilung von Parametern und Momenten aus 3)
6. **Berechnung des Konfidenzintervalls für den Mittelwert gemäß der Klassifikation mit geeigneten Verfahren:**
- Bei geringen Abweichungen: t-Statistik
 - Bei symmetrischen Abweichungen: Nichtparametrisches Konfidenzintervall mit dem Vorzeichen-Test
 - Bei schiefen Abweichungen: Johnsons t-Statistik
 - Bei sehr schiefen oder extremen Verteilungen: Trimmen oder Winsorisieren der Stichproben und Erhöhen der Anzahl der Beobachtungen.
7. **Sensitivitätsanalyse:**
Bei der Berechnung des Konfidenzintervalls ist eine Sensitivitätsanalyse mit einer alternativen Klassifikation und damit einem anderen Test sinnvoll, um die Auswirkungen der Entscheidung auf die Breite des Konfidenzintervalls zu erkennen.
8. **Weitere Beobachtungen:**
Weitere zusätzliche Beobachtungen verringern nach Relation 4.6 ungefähr umgekehrt proportional zur Wurzel ihrer Anzahl die Größe des Konfidenzintervalls und reduzieren bei einem additiven Merkmal der Stichprobe nach dem zentralen Grenzwertsatz auch die Abweichung von der Normalverteilung.

Ein iteratives Vorgehen mit einem Rücksprung von der Erzeugung weiterer Stichproben zu einem Einstieg bei Teilprozedur eins bis drei ist häufig angebracht, um die Konfidenzintervallbreite zu reduzieren und damit die Güte der Schätzung der Kennzahl eines Simulationsmodells zu verbessern.

Als Zusammenfassung der Vorgehensweise zeigt die folgende Abbildung 4.4 ein UML-Diagramm des Gesamtablaufs zur Auswertung eines Simulationsmodells.

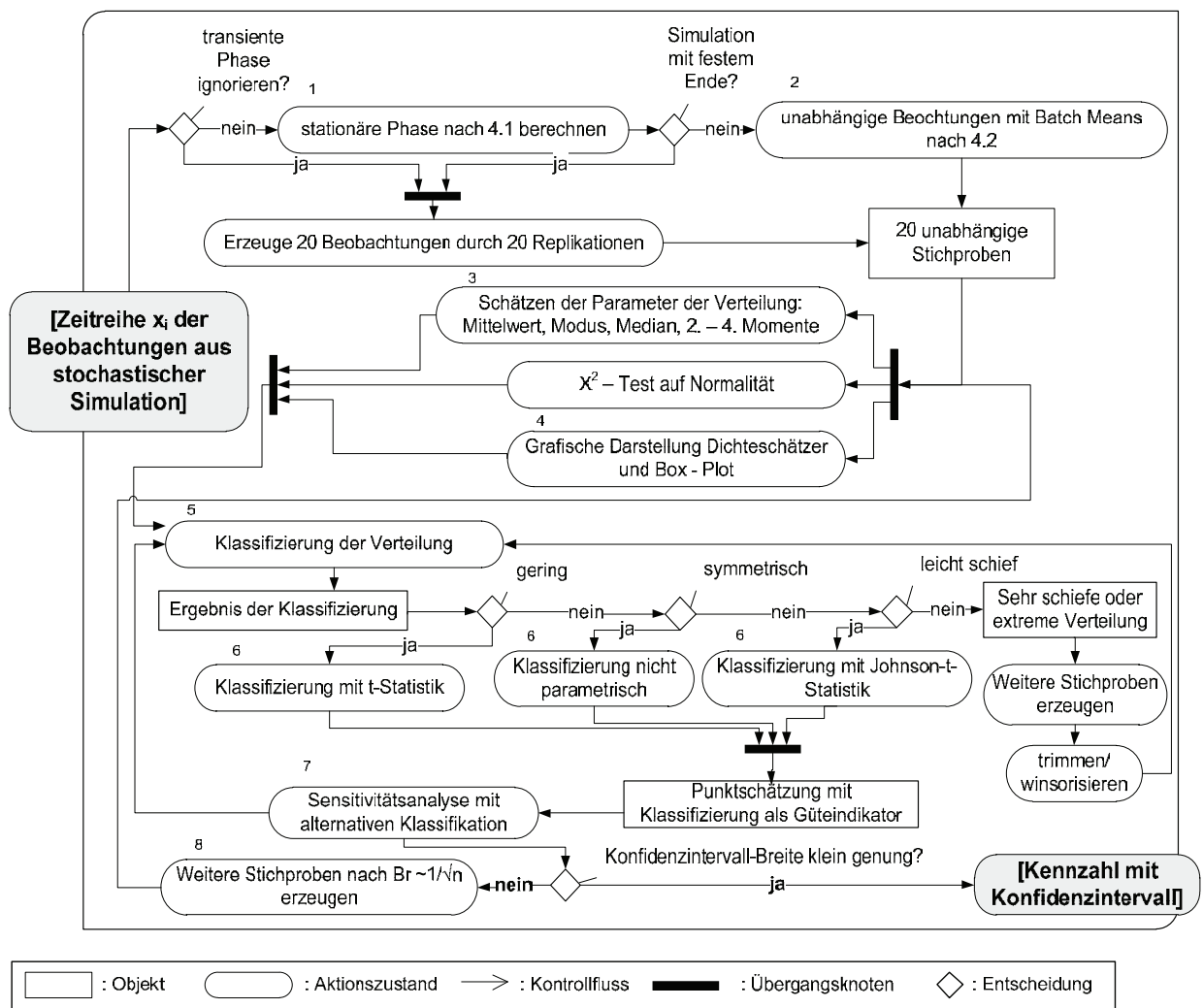


Abbildung 4.4: UML-Diagramm zur Vorgehensweise zur Auswertung eines Modells.

4.6 Zusammenfassung

Die in diesem Kapitel dargestellte Vorgehensweise beschreibt ein neues integriertes Verfahren zur statistischen Auswertung eines Simulationsmodells. Das Verfahren erfüllt die Anforderungen an die Robustheit und Plausibilität. Es ist kein automatisiertes Verfahren, der Anwender wird explizit durch grafische Darstellungen, Entscheidungen bei der Klassifizierung von Verteilungen und der folgenden Sensitivitätsanalyse eingebunden.

Wie bereits in Kapitel 3 konzipiert ist die Erweiterbarkeit für den multivariaten Fall mit Hilfe der Bonferroni-Ungleichung gegeben.

5 Vorgehensweise zum Vergleich von mehreren Simulationsmodellen

5.1 Zielsetzung und Rahmenbedingungen

In diesem Kapitel werden die im Kapitel 3 konzipierte Vorgehensweise zum Vergleich von mehreren Simulationsmodellen dargestellt sowie die Entscheidungsregeln und die statistischen Verfahren entwickelt und integriert. Die dabei verwendeten Methoden werden im Anhang detailliert erläutert und beschrieben.

Die vergleichende Bewertung von alternativen Systemen und Modellen ist eine typische und häufige Aufgabenstellung für Simulationsstudien. Die unterschiedlichen Simulationsmodelle, auch als Systemalternativen oder -varianten bezeichnet, können sich dabei sowohl in der Struktur der Modelle als auch in der Parametrisierung unterscheiden. Bei diesem Vergleich der Ergebnisse der Simulation von mehreren Systemalternativen sollen signifikante Unterschiede und deren Richtung bei einer bestimmten Leistungsgröße bzw. Kennzahl aufgezeigt werden. „Signifikant“ bedeutet im Rahmen dieser Arbeit „statistische Signifikanz“, d.h. mit einer definierten Wahrscheinlichkeit $1-\alpha$ unterscheiden sich die aus den Simulationsergebnissen geschätzten Leistungsgrößen der alternativen Systeme.

Ein typisches Ziel der Simulationsuntersuchung mit mehreren Systemvarianten ist es, Aussagen über Leistungsgrößen (hier die Durchlaufzeit) treffen zu können wie:

„System A hat mit 95% Wahrscheinlichkeit eine höhere Durchlaufzeit als System B.“

„System B unterscheidet sich mit 95% Wahrscheinlichkeit bei der Durchlaufzeit nicht von System C.“

Bei Aussagen über signifikante Unterschiede beim Vergleich von Simulationsmodellen wird die Wahrscheinlichkeit meist in der Form $(1-\alpha) \cdot 100$ in Prozent beschrieben.

Um nun eine Fragestellung nach „signifikanten Unterschieden“ beantworten zu können, ohne diese Signifikanz in absoluten Werten selbst festlegen zu müssen bzw. zu wollen, sind, wie in Kapitel 3.4.3.1 erläutert, die Verfahren der Mehrfachen Mittelwertvergleiche (MMV) für praktische Anwendungen geeignet²³.

²³ Das ergänzende Konzept der Ranking & Selection Verfahren, bei dem der Anwender einen signifikanten Unterschied vorgibt, wurde in Kap. 3 kurz vorgestellt.

5.2 Auswahl der geeigneten Verfahren der Mehrfachen Mittelwertvergleiche

Zur Auswahl der geeigneten MMV-Verfahren wurden die durch eine Literaturstudie identifizierten Verfahren mit den in Kapitel 3 beschriebenen Monte-Carlo Studien auf Ihre Eignung untersucht. Grundsätzlich müssen die MMV-Verfahren die Anforderungen aus Kapitel 2 nach Robustheit und Plausibilität erfüllen. Basierend auf dem Konzept des Auswertungsprozesses aus Kapitel 3 werden nun die geeigneten Verfahren ausgewählt und integriert.

Die Auswahl der MMV-Verfahren erfolgt mit der Klassifizierung der Beobachtungen der Systemvarianten nach den unterschiedlichen Ausprägung der drei wesentlichen Kriterien:

- Anzahl der Stichproben
- Vorhandene Varianzunterschiede
- Art der Abweichung von der Normalverteilung

Bei der Evaluation der MMV-Verfahren durch die Monte-Carlo-Studie in Kapitel 3.6.5.1 für die Eignung zum Vergleich von Simulationsalternativen wurde vor allem auf die beiden Leistungsmerkmale

- Einhaltung des (multiplen) Signifikanzniveaus und
- Trennschärfe der Verfahren

Wert gelegt.

5.2.1 Anzahl der Stichproben

Zur Beurteilung der Anzahl der Stichproben und der Klassifikation in „klein, mittel“ und „groß“ gibt es entsprechende Empfehlungen aus der Literatur.

Als „**große** Anzahl von Stichproben“ darf in Übereinstimmung mit den Empfehlungen und der Klassifizierung von **Law** [Law07:504] und **Kleijnen** [Klei87:24] [Klei08:80] eine pragmatische Größenordnung größer als 30 bzw. 100 interpretiert werden ($n > 30$, $n > 100$). Entsprechend werden zehn und weniger ($n \leq 10$) als eine **kleine** Anzahl Stichproben bezeichnet. Daraus abgeleitet können Stichprobenumfänge zwischen zehn und 30 als „mittelgroß“ klassifiziert werden.

Qualitative Beurteilung	Anzahl der Stichproben n
Groß	$n > 30$, ggf. $n > 50-100$
Mittel	$10 < n \leq 30$
Klein	$n \leq 10$

Tabelle 5.1: Qualitative Klassifikation der Stichproben.

Als häufig in der Literatur erwähnte Faustregel ist eine initiale Stichprobenanzahl von $n = 20$ sinnvoll.

5.2.2 Varianzunterschiede

Beim Vergleich von Simulationsmodellen kann man prinzipiell bereits *a priori* von unterschiedlichen Varianzen, auch Heteroskedastizität genannt, ausgehen (vgl. [Klei08:88]). Zur eigentlichen Beurteilung und Klassifizierung der Varianzunterschiede durch den Anwender empfiehlt es sich, nicht nur die absoluten Beträge zu betrachten, sondern vor allem auch mit den relativen Verhältnissen zu argumentieren.

Die Homoskedastizität bzw. die gegenteilige Heteroskedastizität der Stichprobenverteilungen lässt sich auch mit den Unterschieden der jeweiligen empirischen Variationskoeffizienten

$$v = \frac{s}{\bar{x}} \quad (5.1)$$

beschreiben. Der empirische Variationskoeffizient als Quotient der empirischen Standardabweichung durch den arithmetischen Mittelwert wird als Prozentsatz angegeben. Allerdings wird der Variationskoeffizient als dimensionslose Kennzahl „unbrauchbar“, wenn der arithmetische Mittelwert gegen Null geht.

Ergänzend zum Variationskoeffizient sind auch die Verfahren der Dispersionsvergleichstest verfügbar, um eine eventuelle Hypothese der Homoskedastizität zu bestätigen. Der einfache und plausible **Cochran-Test** zur Prüfung der Gleichheit mehrerer Varianzen ist nach Sachs und Hedderich [SaHe06:421] auch bei leicht schiefen Abweichungen von der Normalverteilung geeignet. Der Cochran-Test arbeitet mit der Nullhypothese H_0 der Varianzgleichheit.

Falls der Cochran-Test bei einem entsprechend sinnvollen Signifikanzniveau α die Nullhypothese der Varianzgleichheit bei den Stichproben aus den Systemvarianten nicht bestätigt, sollte man von entsprechend großen Varianzunterschieden ausgehen.

Der Fall der geringen Varianzunterschiede sollte dann angenommen werden, wenn der Cochran-Test die Nullhypothese der Homoskedastizität nur knapp ablehnt und die Variationskoeffizienten ähnlich sind.

Qualitative Beurteilung der Varianzunterschiede	Indikatoren
Große Varianzunterschiede	<ul style="list-style-type: none"> • Variationskoeffizienten v_i unterschiedlich, • Grafische Verteilungsformen und Boxplots zeigen unterschiedliche Varianzen bzw. Quartile • Cochran-Test lehnt H_0 ab
Geringe Varianzunterschiede	<ul style="list-style-type: none"> • Variationskoeffizienten v_i ähnlich, • Grafische Verteilungsformen und Boxplots zeigen geringe unterschiedliche Varianzen bzw. Quartile • Cochran-Test lehnt H_0 ab
Keine Varianzunterschiede	<ul style="list-style-type: none"> • Variationskoeffizienten v_i sehr ähnlich, • Grafische Verteilungsformen und Boxplots zeigen keine unterschiedliche Varianzen bzw. Quartile • Cochran-Test bestätigt H_0

Tabelle 5.2: Qualitative Klassifikation der Stichproben.

Allerdings erhält man prinzipiell beim Einsatz des Cochran-Tests eine mehrstufige Testprozedur mit entsprechenden Auswirkungen auf das Gesamtniveau α (vgl. Büning [Büni91:143]) der Entscheidungssicherheit.

5.2.3 Abweichung von der Normalverteilung

Um die Abweichungen von der Normalverteilung beurteilen zu können, ist die bereits in Kapitel 4.1 und 4.5 beschriebene Vorgehensweise zu empfehlen. Die entsprechende Beurteilung der empirischen Verteilungsform anhand der drei Kriterien

- Entscheidung des X^2 -Test,
- grafische Darstellung der empirischen Verteilungsdichten und der Box-Plots im Vergleich mit den entsprechenden Normalverteilungen und
- Beurteilung von Form- und Lageparametern und dritten und vierten Momenten

kann damit auch aus dem Vorgehen zur Bestimmung der Konfidenzintervallbreite übernommen werden. Die Tabelle 4.1 gilt entsprechend für die Klassifikation zur Auswahl der MMV.

5.3 Empfohlene Verfahren der Multiplen Mittelwertvergleiche

Auf Basis der jeweiligen Ausprägungen der drei Kriterien

- Anzahl der Stichproben,
- vorhandene Varianzunterschiede sowie
- Art der Abweichung von der Normalverteilung

lassen sich mit den Ergebnissen der Monte-Carlo-Studien die für die jeweilige Situation geeigneten MMV-Verfahren empfehlen. Die Auswahl der Verfahren wird dazu dreidimensional in einem Würfel mit den obigen Kriterien als Achsen in Abb. 5.1 dargestellt.

Bei einer entsprechenden Konstellation mit der Erfüllung der jeweiligen Kriterien sind folgende sieben Verfahren geeignet und werden empfohlen²⁴:

1 **Tu: Tukey-Test**

(Tukey) ([Hart02:616] und [SaHe06:436]) Der Tukey-Test benutzt die Verteilung der Studentisierten Extremwerte (auch Studentisierte Spannweite bzw. engl. Studentized Range).

2 **GH: GH-Test**

(Games-Howell) ([Dunn82], [Tamh79] und [HoTa87:180ff]) Der GH-Test nach Games und Howell basiert auf dem Tukey-Test, allerdings wird die Zahl der Freiheitsgrade mit der Methode von Welch geschätzt.

3 **C: C-Test**

(Dunnnett-C) ([Dunn80] und [HoTa87:180ff]) Der C-Test von Dunnnett basiert wie der GH-Test auf dem Tukey-Test, beim C-Test werden die Freiheitsgrade nach der Methode von Cochran approximiert.

4 **T3: T3-Test - (Dunnnett-T3)** ([Dunn80] und [HoTa87:180ff]) Der T3-Test von Dunnnett basiert ebenfalls auf dem Tukey-Test, benutzt aber das

²⁴ Die Abkürzungen sind konsistent zur referenzierten Literatur, vor allem Hochberg und Tamhane [HoTa87], Hsu [Hsu96], Miller [Mill81] und Büning [Büni91] gewählt. In Klammern sind für die ersten fünf Verfahren die Bezeichnungen im SPSS Menu "Einfaktorielle ANOVA: Post-Hoc-Mehrfachvergleiche" angegeben.

Studentisierte Maximalmodul (engl. Studentized-Maximum-Modulus) und die Šidák-Ungleichung.

- 5 **t+B: t-Test mit der Bonferroni-Ungleichung** - (Bonferroni) ([Law07:552], [Klei87:42], [Klei87:100] sowie [SaHe06:123] und [Mill81:67]) Die Bonferroni-Ungleichung in Kombination mit dem Zweistichproben-t-Test ist ein relativ einfaches und plausibles Standardverfahren für multiple Mittelwertvergleiche. Bei häufig vorhandenen Varianzunterschieden wird die Zahl der Freiheitsgrade (wie beim GH-Test) mit der Methode von Welch geschätzt. Mit der Bonferroni-Ungleichung wird die individuelle Fehlerrate α bei k Vergleichen einfach durch α/k ersetzt.
- 6 **W+S: Wilcoxon-Test mit der Šidák-Ungleichung** - Der Wilcoxon-Test in Verbindung mit der Šidák-Ungleichung ist ein robustes verteilungsfreies Verfahren. Bei der Šidák-Ungleichung wird die Fehlerrate α bei k Vergleichen durch $1-(1-\alpha)^{1/k}$ ersetzt.
- 7 **T10: getrimmte Schätzer für Mittelwert und Varianz** - ([Hart02:616], [SaHe06:70] und [Büni91:92]) Die um jeweils 10% der extremen Werte getrimmten Schätzer ($\beta = 0,1$) für Mittelwert und Varianz dienen zur „Robustifizierung“ in der Kombination mit einem der vier ersten Verfahren Tu, GH, C oder T3.

Diese sieben Verfahren sind im Anhang detailliert dargestellt und erläutert. Die folgende Abb. 5.1 zeigt die würfelförmige Darstellung der Empfehlung der Verfahren bei der jeweiligen Ausprägung der obigen drei Kriterien.

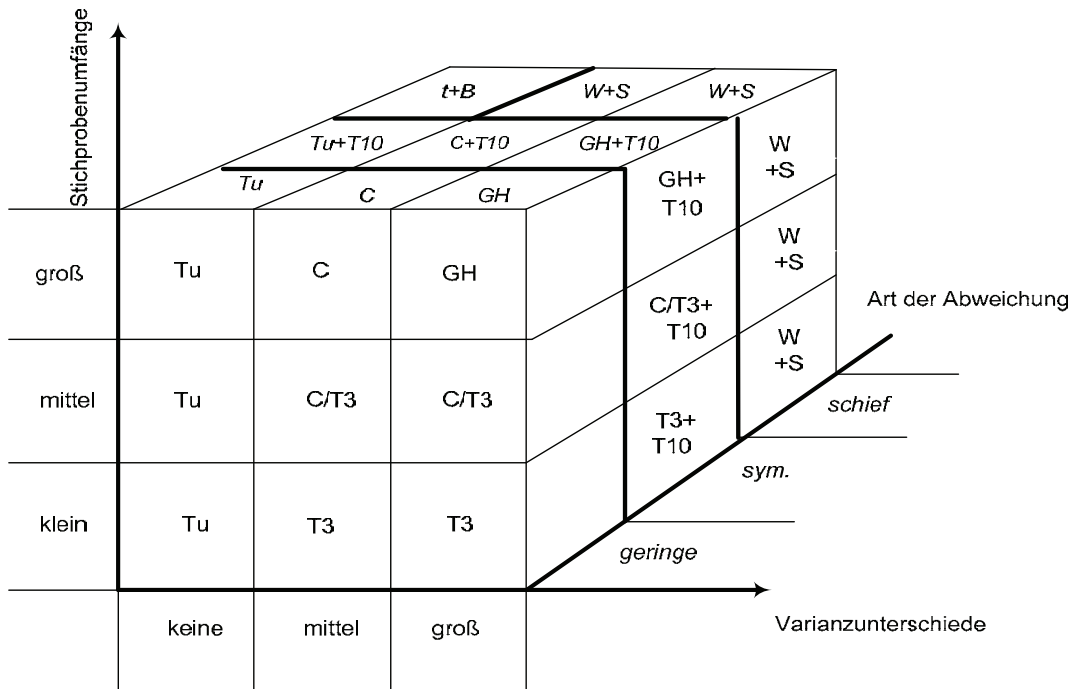


Abbildung 5.1: Geeignete Verfahren der mehrfachen Mittelwertvergleiche in Abhängigkeit von den drei Kriterien.

Diese dreidimensionale Darstellung kann wie in Abb. 5.2 in drei Ansichten je nach Art der Abweichung von der Normalverteilung aufgelöst werden.

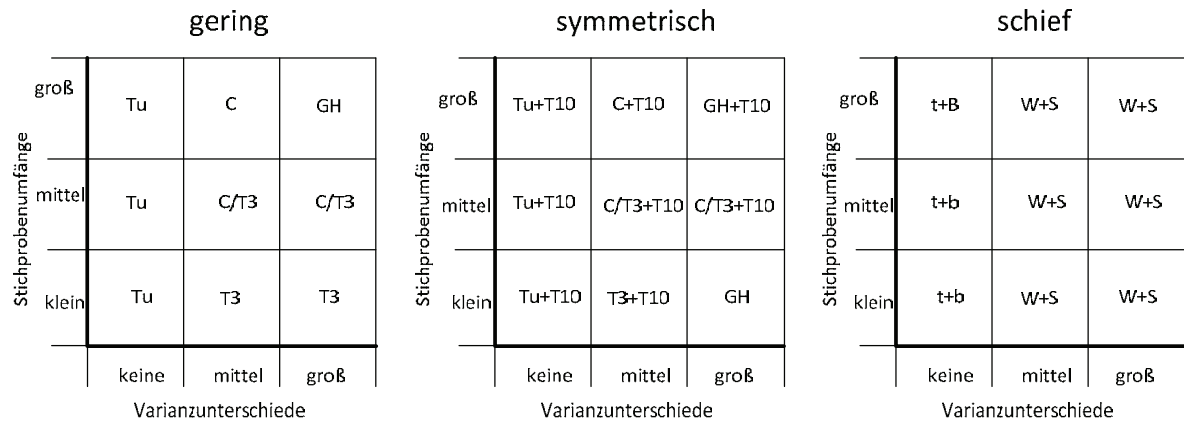


Abbildung 5.2: Geeignete Verfahren der mehrfachen Mittelwertvergleiche aufgeteilt nach Art der Abweichung von der Normalverteilung.

Die MMV-Verfahren liefern mit der vorgegebenen Wahrscheinlichkeit Aussagen über vorhandene Unterschiede bei den Leistungskennzahlen der Systemalternativen oder erkennen keinen signifikanten Unterschied, wie z.B.

$$\text{System A} > \text{System B} \sim \text{System C}$$

Wie bei der Berechnung eines Konfidenzintervalls in Kapitel 4 steigt auch bei den mehrfachen Systemvergleichen die Trennschärfe der Verfahren mit der Anzahl der Stichproben.

5.3.1 Vereinfachte Darstellung und Auswahl

Eine vereinfachte Darstellung und Auswahl ergibt sich, wenn der Würfel durch das Entfallen des Kriteriums der Stichprobenanzahl auf zwei Dimensionen reduziert wird und nur die obere „horizontale Scheibe“ betrachtet wird.²⁵

Unter der Voraussetzung, dass bei Simulationsstudien mit nicht zu komplexen Modellen oft eine „große Anzahl“ von Stichproben ($n \geq 30$) zur Verfügung steht bzw. bei Bedarf unter entsprechendem Einsatz von Rechenzeit erzeugt werden kann, ergibt sich eine einfachere Darstellung des Würfels als 3x3 Quadrat. Die Abbildung 5.2 stellt diese Vereinfachung der Auswahl mit der Reduktion auf die beiden Kriterien „Varianzunterschiede“ und „Abweichungen von der Normalverteilung“ dar.

Art der Abweichung	schief	t + B	W + S	W + S
	sym.	Tu + T10	C + T10	GH + T10
	gering	Tu	C	GH
		keine	mittel	groß
		Varianzunterschiede		

Abbildung 5.3: Vereinfachte Auswahl geeigneter Verfahren der mehrfachen Mittelwertvergleiche.

Diese Auswahl der geeigneten Verfahren ist mit der Nutzung des Kriteriums „Abweichungen von der Normalverteilung“ plausibel nachzuvollziehen:

²⁵ Im Rahmen von frühen Veröffentlichungen der ersten Vorergebnisse in der VDI-Richtlinie 3633 [VDIF95] und in einem Vorlesungsskript [WaWa96] wurde die Vorgehensweise stark vereinfacht und die Auswahl auf die beiden Tests „Tukey“ und „Wilcoxon“ reduziert.

- Bei geringen Abweichungen ist der Tukey-Test auch mit einer modifizierten Berechnung der Freiheitsgrade als C- oder GH-Test robust genug, um empfohlen zu werden.
- Bei symmetrischen Abweichungen werden die drei Tests Tu, C und GH mit den symmetrisch getrimmten Schätzern noch zusätzlich robustifiziert.
- Bei schiefen Abweichungen ist der verteilungsfreie Wilcoxon-Test mit der Šidák-Ungleichung am besten geeignet. Beim Sonderfall schiefer Verteilungen ohne Varianzunterschiede ist allerdings der Zweistichproben-Test mit der Bonferroni-Ungleichung trennschärfer und damit effizienter.

Als weitere Empfehlung sollte bei allen Verfahren der Mehrfachen Mittelwertvergleiche ein gesamtes Signifikanzniveau von mindestens $\alpha = 0,05$ gewählt und nicht zu viele Systemalternativen betrachtet werden. Ggf. bietet sich eine Vorauswahl mit einem der Ranking & Selection Verfahren an.

5.3.2 Sensitivitätsanalyse und weitere Beobachtungen

In Analogie zur Vorgehensweise bei der robusten Berechnung eines Konfidenzintervalls in Kapitel 4 ist auch bei den Verfahren der MMV eine Sensitivitätsanalyse mit einer alternativen Entscheidung bei der Klassifikation der Verteilungen sinnvoll. Eine Untersuchung, ob signifikante Unterschiede bestätigt werden oder verloren gehen, falls ein „im Würfel benachbartes“ Verfahren verwendet wird, steigert die Akzeptanz und die Plausibilität der Vorgehensweise erheblich.

Falls diese Sensitivitätsuntersuchung allerdings stark variierende Aussagen über vorhandene Unterschiede der Systemvarianten ergibt, sollten für diese Systemvarianten weitere Stichproben aus den Simulationsläufen erzeugt werden. Dadurch kann eine entsprechend große Anzahl von Beobachtungen, siehe 5.2.1, notwendig werden.

5.4 Integration der Vorgehensweise zum Vergleich mehrerer Simulationsmodelle

Die in Kapitel 3.7 im Konzept des integrierten Verfahrens beschriebene Methode der MMV wird im Folgenden noch in acht wesentlichen Schritten dargestellt. Die Schritte 1 und 2 mit der Schätzung der Lage- und Formparameter der Verteilungen der Beobachtungen aus den Systemalternativen, die grafische Darstellung der Verteilungsdichten sowie Visualisierung als Box-Plots kann, falls bereits vorhanden, direkt aus den Methoden zur Auswertung eines Modells in Kapitel 4 übernommen werden.

1. Schätzung von Mittelwert, Median, Modus sowie den 2., 3. und 4. Momenten Varianz, Schiefe und Exzess der Verteilungen der Beobachtungen aus den Systemalternativen:

Für die Systemalternativen werden die obigen sechs Kennzahlen sowie die Lage- und Formparameter der empirischen Verteilungen geschätzt. Diese Kennzahlen dienen zur Klassifikation der Verteilungen bezüglich vorhandener Abweichungen von der Modellannahme der Normalverteilung.

2. Grafische Darstellung der Verteilungsdichten der Beobachtungen aus den Systemalternativen:

Die Visualisierung der Verteilungsdichten der Stichproben der Systemalternativen ist eine wichtige Maßnahme, um die Form der Abweichung von der Normalverteilung zu erkennen. Dazu sind Dichteschätzer mit geeigneter Kernfunktion (z.B. „Gauss“ oder „Tukey's biweight“), ein Histogramm, bei dem Klassenanzahl und -breite variiert werden sowie Box-Plots am besten geeignet.

3. Klassifizierung nach Anzahl der Beobachtungen:

Eine qualitative Klassifikation der Stichprobenanzahl mit den Grenzen zehn und 30 ist notwendig, um geeignete Verfahren auswählen zu können.

4. Klassifizierung der Verteilung der Systemalternativen nach der Art der Abweichung:

Die empirische Verteilung wird nach der Abweichung von der Modellannahme der Normalverteilung klassifiziert. Als Kriterien dazu dienen:

- Der X^2 -Test
- Die Visualisierung mit Box-Plots
- Die grafische Darstellung der Verteilungsdichte im Vergleich zur Dichte einer entsprechenden Normalverteilung
- Die Beurteilung von Parametern und Momenten

5. Klassifizierung der Verteilungen der Systemalternativen nach Varianzunterschieden:

Die empirischen Verteilungen werden nach Varianzunterschieden qualitativ klassifiziert. Als Kriterien dazu dienen:

- Die Variationskoeffizienten v_i
- Die grafischen Verteilungsformen und Boxplots
- Der Cochran-Test

6. Auswahl des geeigneten Verfahrens der MMV:

Ein geeignetes Verfahren der MMV wird gemäß der obigen Klassifikation nach den drei Kriterien

- Stichprobenanzahl,
- Varianzunterschiede und
- Abweichungen von der Normalverteilung

aus der würfelförmigen Darstellung ausgewählt. Mit einer vorgegebenen Irrtumswahrscheinlichkeit (meist kleiner oder gleich 5 %) werden vorhandene signifikante Richtungsunterschiede erkannt. Es ergeben sich Aussagen über Unterschiede der Kennzahlen der Systemalternativen wie z.B. System A > System B ~ System C.

7. Sensitivitätsanalyse:

Bei der Berechnung des signifikanten Richtungsunterschieds ist eine Sensitivitätsanalyse mit einer alternativen Klassifikation und damit einem „benachbarten“ Verfahren sinnvoll, um die Auswirkungen der Klassifikationsentscheidung auf das Identifizieren eines Richtungsunterschieds zu erkennen.

8. Weitere Stichproben:

Weitere zusätzliche Stichproben erhöhen die Trennschärfe der Verfahren und ermöglichen das Erkennen von vorhandenen Unterschieden. Allerdings wird gemäß der Anzahl der Systemalternativen ein entsprechend erhöhter Simulationsaufwand notwendig.

Vergleichbar mit der Vorgehensweise für die Auswertung eines Simulationsmodells in Kapitel 4 ist auch hier ein iteratives Vorgehen mit einem Rücksprung von der Erzeugung weiterer Stichproben zu einem Einstieg bei Teilprozedur **eins** bis **drei** oft notwendig, um mit einer vergrößerter Stichprobenanzahl eventuell vorhandene Richtungsunterschiede der Systemalternativen signifikant erkennen zu können.

Allerdings ist der entsprechend hohe Rechenaufwand der Simulationsmodelle meist eine praktische Einschränkung.

Die Vorgehensweise der Auswahl der geeigneten Verfahren der MMV zeigt die folgende Abbildung 5.3 in einem UML-Diagramm:

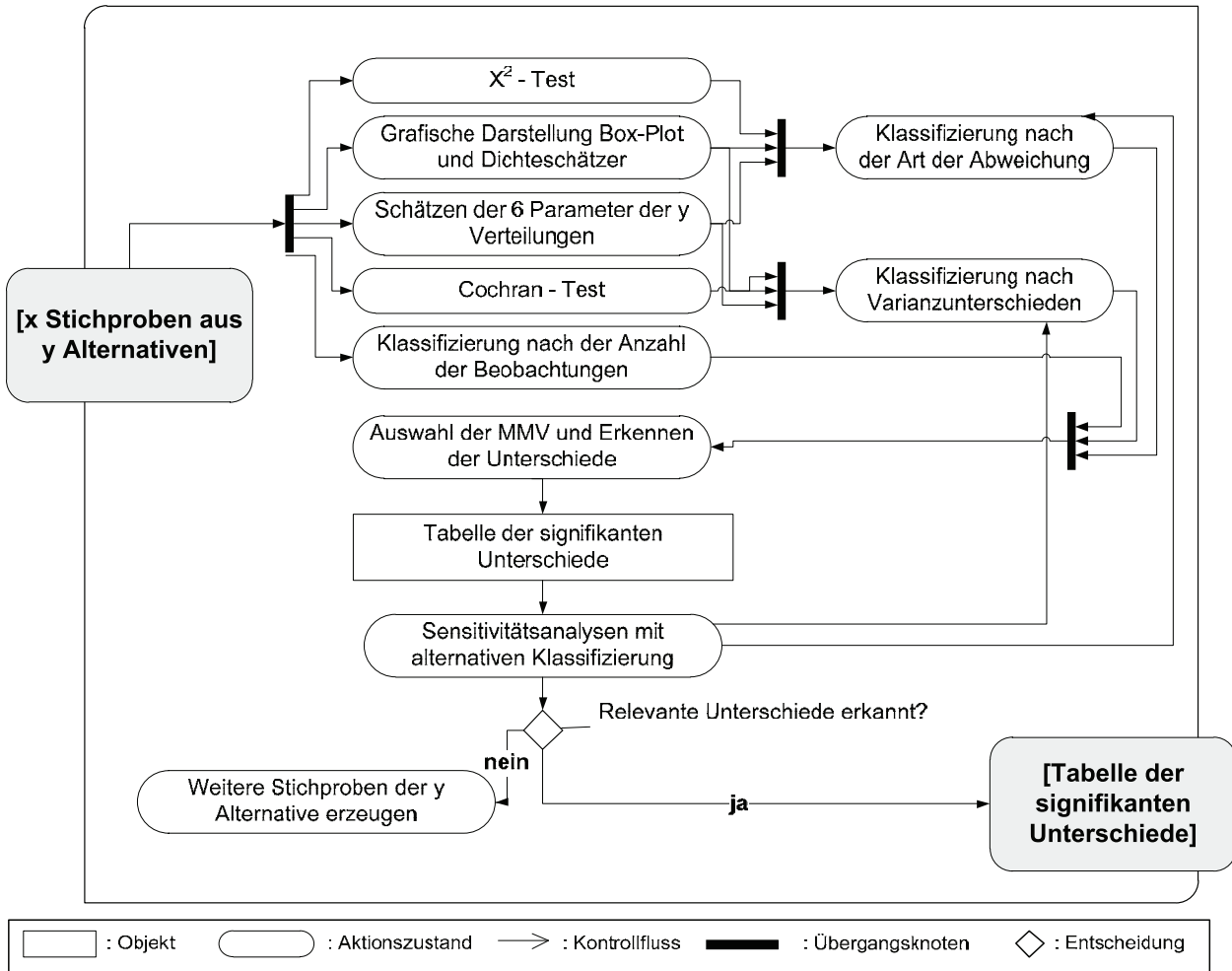


Abbildung 5.4: UML Diagramm der Vorgehensweise zum Vergleich von mehreren Simulationsalternativen.

5.5 Zusammenfassung

Die in diesem Kapitel dargestellte Vorgehensweise beschreibt ein neues integriertes Verfahren zum Vergleich von mehreren Simulationsmodellen. Die zweistufige adaptive Vorgehensweise mit der Auswahl der geeigneten MMV ergibt ein plausibles, robustes und iteratives Verfahren, um bei Simulationsprojekten die Systemalternativen auf signifikante Richtungsunterschiede untersuchen zu können.

Auch hier ist bei Bedarf eine Erweiterung auf einen multivariaten Anwendungsfall mit der Bonferroni-Ungleichung möglich.

6 Anwendung des integrierten Verfahrens

6.1 Anwendung des integrierten Verfahrens im industriellen Umfeld

Das in dieser Arbeit beschriebene integrierte Verfahren wurde bei den Simulationsstudien im Rahmen von zwei industriellen Projekten eingesetzt. Die beiden kooperativen Forschungs- und Technologieentwicklungsprojekte²⁶ „ManuBuild“ [Manu11] und „CeC-made-shoe“ [CeC05] wurden im 6. Rahmenprogramm der Europäischen Kommission im Bereich der Produktionsforschung gefördert.

Die industrielle Anwendung des integrierten Verfahrens soll die Erfüllung der Anforderungen überprüfen, die praktische Anwendbarkeit bestätigen und vor allem zeigen, dass die als Ziel der Arbeit erwarteten Ergebnisse einer statistisch korrekten Auswertung von Simulationen erreichbar sind.

Dazu werden zwei industrielle Simulationsstudien benutzt, welche unterschiedliche Produktionssysteme aus der Bau- und Schuhindustrie abbilden, um damit sowohl die praktische Anwendbarkeit in diesen Bereichen als auch eine Übertragbarkeit zu bestätigen. Die beiden Anwendungsszenarien einer Produktion von Fertigbauteilen sowie einer flexiblen Schuhproduktion werden im Folgenden vorgestellt.

6.2 Beschreibung der Anwendungsszenarien

Das europäische Forschungs- und Technologieentwicklungsprojekt „**ManuBuild**“ wurde von der Europäischen Kommission im 6. Rahmenprogramm unter der Projektnummer 515825 unterstützt. „ManuBuild – Open Building Manufacturing“ ist ein industriegeführtes kooperatives Forschungsprojekt über industrialisiertes Bauen mit 25 Partnern aus zehn europäischen Ländern [Manu11]. Die Projekt-Vision soll einem Kunden ermöglichen, qualitativ hochwertig errichtete Gebäude mit einem erheblichen Grad an Designflexibilität zu relativ niedrigen Kosten zu erwerben. Ein wichtiges Entwicklungsziel des Projekts sind innovative Produktionsmethoden und Technologien für ein „Open Building Manufacturing“.

Ein wesentlicher Bestandteil dieser innovativen Produktionsmethoden der Bauindustrie ist das Konzept und der Prototyp einer flexiblen und mobilen Fabrik für Fertigbauteile, wie es von dem spanischen Bauunternehmen Dragados entwickelt wurde.

²⁶ Engl. „RTD Project“.

Das zweite europäische Forschungs- und Technologieentwicklungsprojekt „**CeC-made-shoe**“ der Schuhindustrie wurde ebenfalls von der europäischen Kommission im 6. Rahmenprogramm mit der Projektnummer 507378 gefördert. Das Projekt „Comfort, Environmental and Custom made Shoe“ sollte die europäische Schuhindustrie von einem produktzentrierten Ansatz zu einem humanorientierten Industriemodell führen. Dabei wurde ein Ansatz mit den drei Aspekten Komfort (Comfort), Umwelt (Environment) und Individualität (Custom) als Projektname und Aufgabenstellung benutzt [CeC05].

Das Projekt nutzte die „Integrierte Produktionsfabrik“ in Vigevano als Modellfabrik für neue flexible Produktionskonzepte. Dabei sollen unter anderem Produktionsmethoden für „konventionelle“ und „Bio-Schuhe“ verglichen werden [CeC06].

6.3 Simulations- und Auswertungsumgebung

Die Anwendung des integrierten Verfahrens nutzt eine einfach verfügbare Hard und Softwareumgebung und ist prinzipiell unabhängig von der eingesetzten Simulationssoftware.

Die Software-Umgebung der beiden Anwendungen auf Grundlage eines „normalen“ intel-basierten PCs mit Microsoft Windows XP besteht im Wesentlichen aus den folgenden Programmen:

- Stochastische ereignisdiskrete Simulation mit Dosimis3 [SDZ07]
- Dateiaufbereitung mit Microsoft Excel [MiCo11]
- Statistische Auswertung und Darstellung mit R [RPro11]

Kleinere Filter- und Aufbereitungsprozeduren wurden mit dem aus einer UNIX Betriebssystem-Umgebung bekannten Werkzeug „awk“ bzw. „gawk“ [FSF09] durchgeführt.

Bei der Auswertung der Dosimis3 Simulationen wurden sowohl die Statistikausgaben in der *.slg Datei verwendet als auch direkt auf das Ereignisprotokoll der Trace-Dateien *.tra zugegriffen. Die relevanten Daten wurden mit gawk in neue Dateien extrahiert. Für die Visualisierung der aufbereiteten Daten in Form von Zeitreihen, Verteilungen, Verteilungsdichten und Box-Plots wurde R verwendet. R ist (nach Sachs und Hedderich [SaHe06:633]) *„in erster Linie eine Programmiersprache und Programmierumgebung für die statistische Analyse von Daten. R kann einerseits elementare mathematische Rechenoperationen ausführen, berechnet andererseits aber auch anspruchsvolle komplexe statistische Funktionen“*.

6.4 Simulation einer Produktion für Fertigbauteile

6.4.1 Aufgabenstellung

Im Rahmen des ManuBuild Projekts sollte mit Hilfe einer Simulationsstudie das innovative Konzept des Prototyps einer flexiblen und mobilen Fabrik für Fertigbauteile, vor allem im Gegensatz zu traditionellen stationären Fertigungskonzepten, untersucht und bewertet werden. Die Idee der mobilen Fabrik basiert auf Überlegungen des spanischen Bauunternehmens „Dragados“²⁷, dass ab einer gewissen Größe des Bauprojekts eine lokale Produktion der verwendeten Fertigbauteile vor Ort erhebliche logistische Vorteile bedeuten kann. Abb. 6. 1 zeigt eine Skizze der flexiblen und mobilen Fabrik mit der Verwendung von auf LKW verladbaren Standardcontainern.

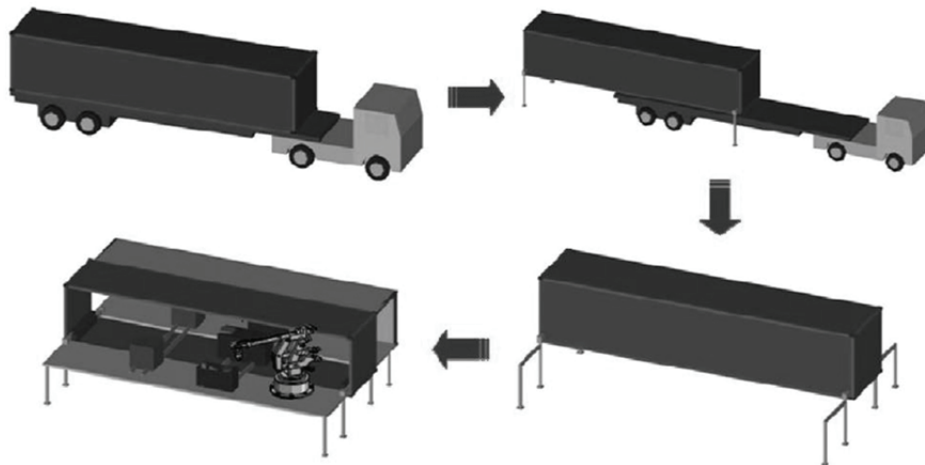


Abbildung 6.1: Konzept der mobilen Fabrik (aus [Manu08]).

Die flexible automatisierte Montage einer „Service Core“ Komponente ist ein wesentlicher Bestandteil des Konzepts der mobilen Fabrik. Die Bewertung dieser automatisierten Option im Vergleich zu einem traditionellen manuellen Zusammenbau ist die Aufgabenstellung für die Simulationsstudie. Der „Service Core“ besteht aus einem metallischen Rahmen, an welchem die notwendigen Leitungen für Küchen, Badezimmer oder Versorgungsbereiche angeschlossen werden. Dabei enthält der „Service Core“ bereits einen Großteil der Leitungen und Anschlüsse für Wasser und Heizung, Lüftung und Klima (engl. kurz „HVAC“), welche sonst individuell vor Ort installiert werden müssen. Abb. 6.2

²⁷ Das Bauunternehmen Dragados ist ein Teil des spanischen Baukonzerns „Grupo ACS“ (Actividades de Construcción y Servicios, SA) aus Madrid. Informationen über Dragados und ACS gibt es unter www.acsgroup.com .

zeigt eine CAD-Skizze des „Service Core“, dessen Montage in der mobilen Fabrik simuliert werden soll:

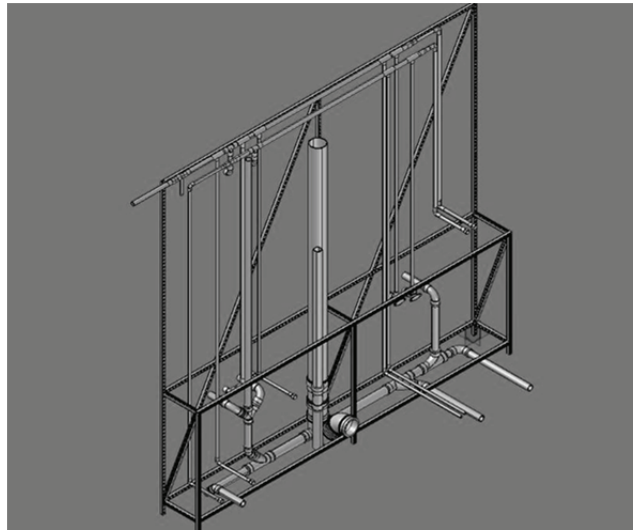


Abbildung 6.2: CAD-Skizze des Service Cores (aus [Manu08]).

Dass eine automatisierte Montage im Hinblick auf die Durchlaufzeit deutliche Vorteile im Vergleich zu einer manuellen Installation besitzt ist offensichtlich. Die wesentliche industrielle Fragestellung an die Simulation galt einer genaueren Quantifizierung der Herstell- bzw. Durchlaufzeit der automatisierten Montage, vor allem hinsichtlich der für eine Automatisierung notwendigen Investitionen.

6.4.2 Beschreibung der Anwendung des integrierten Verfahrens

Das Ziel der Anwendung des integrierten Verfahrens war die Bestimmung der Durchlaufzeit bei der automatisierten Montage der „Service Core“ Komponente. Dazu war die Betrachtung eines stationären Systemzustands notwendig, um die Aussagen nicht durch ein Anlaufverhalten zu beeinflussen.

Für die Simulationsstudie in Zusammenarbeit mit Dragados wurde mit dem Simulationswerkzeug Dosimis3 die automatisierte Systemvariante der Montage der „Service Core“ Komponente in der geplanten mobilen Fabrik modelliert und simuliert. Das Simulationsmodell besteht aus den fünf Sektoren

- Materialeingangslager,
- Vorbereitung,
- Transport Montage,
- Montage sowie
- Ausgangslager

und ist in Abb. 6.3 dargestellt. Zusätzlich zu den Sektoren sind in dem Modell noch Steuerungs- und Messmodule abgebildet, die nur der Simulation dienen und denen keine realen Komponenten entsprechen.

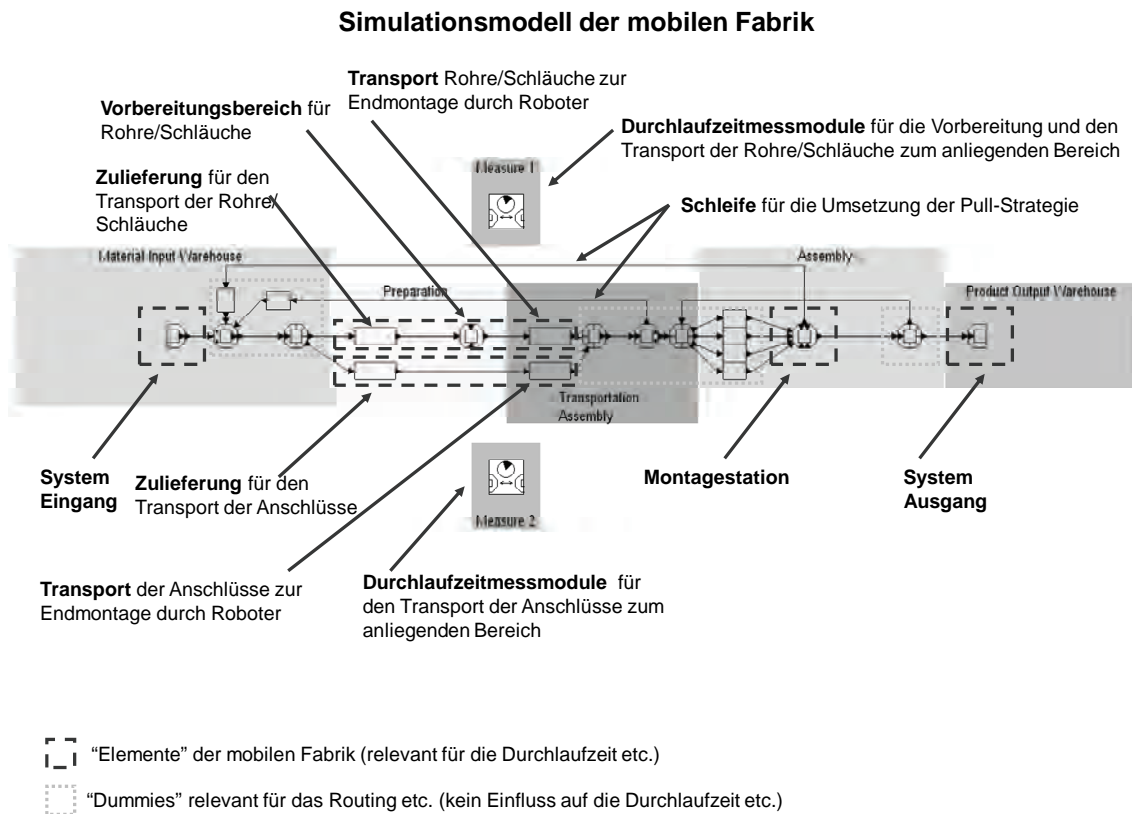


Abbildung 6.3: Simulationsmodell der Montage der „Service Core“ Komponente in der mobilen Fabrik (aus [Manu08]).

Das Simulationsmodell wurde leicht idealtypisch dargestellt, unterschiedliche Handhabungszeiten der verschieden langen Rohre wurden nicht abgebildet. Die exponentialverteilten Ausfall- oder Wartungszeiten wurden für eine durchschnittliche Verfügbarkeit von 97% berechnet. Das Modell wurde verifiziert und soweit wie möglich validiert. Zur Unterstützung der Validierung des Simulationsmodells wurde sowohl ein rein deterministisches Modell ohne jegliche Zufallseinflüsse als auch ein Modell mit stochastischen Einflüssen der Transport-, Ausfall- und Wartungszeiten modelliert. Für die Schätzung von Durchsatz und Durchlaufzeit sollte eine achtstündige Produktionsschicht ohne die Anlaufphase simuliert werden.

Da die automatisierte Montage der mobilen Fabrik stets mit einem leeren und verfügbaren System beginnt, sind evtl. vorhandene transiente Anlaufeffekte beim diesem Simulationsmodell nicht relevant bzw. systemimmanent.

Es werden zwanzig Stichproben der durchschnittlichen Durchlaufzeiten (in Minuten) aus den zwanzig unabhängigen Replikationen benutzt, um die Kennzah-

len Mittelwert, Modus und Median zu schätzen, die zweiten bis vierten statistische Momente zu berechnen und die Verteilungsdichte und Box-Plots darzustellen. Zusammen mit dem X^2 -Test dient dies zur Klassifizierung der Abweichung von der Normalverteilung.

Tabelle 6.1 zeigt diese Kennzahlen der 20 Stichprobenverteilungen und das Testergebnis.

Kennzahl (Durchlaufzeit in Minuten)	Systemalternative „automatisch“
Mittelwert \bar{x}	396,85
Modus x_{mod}	379,27
Median \tilde{x}	381,31
Varianz s^2	549,00
Schiefe γ_1	4,68
Exzess γ_2	-0,74
X^2 -Test p-Wert	1,069e-07

Tabelle 6.1: Stationäre Eigenschaften der Zeitreihe der Durchlaufzeiten (bei 20 Stichproben).

Die geschätzten Dichten der Stichproben der durchschnittlichen Durchlaufzeiten und die zugehörigen Box-Plots der Verteilungen der simulierten Systemalternativen „automatisch“ werden in Abb. 6.4 gezeigt.

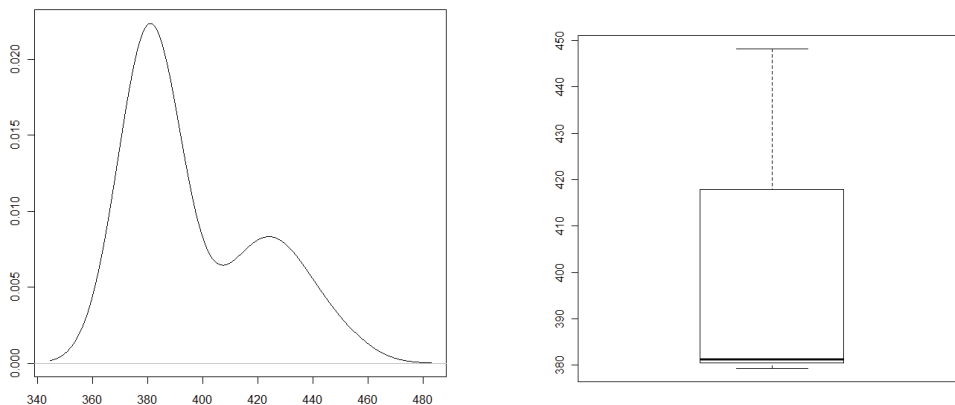


Abbildung 6.4: Dichten und Box-Plots Verteilung der Durchlaufzeiten der simulierten Systemalternative (bei 20 Stichproben).

Mit diesen Kennzahlen aus Tab. 6.1 und den Darstellungen aus Abb. 6.4 lässt sich die Stichprobenverteilung nun bezüglich der Abweichung von der Modellannahme der Normalverteilung als (rechts-) „schief“ klassifizieren. Damit wird nach Kapitel 4.4.6 die Johnson t-Statistik \tilde{t} (siehe Formel 4.15) als geeignetes Verfahren zur Berechnung des Konfidenzintervalls ausgewählt. Allerdings ergibt

der additive Korrekturterm von Johnson $\frac{\hat{\mu}_3}{6s^2n}$ durch die relativ hohe Varianz s^2 im Nenner mit $7,104 \cdot 10^{-5}$ einen vernachlässigbar geringen Wert und damit das „schiefe“ Konfidenzintervall nach Johnson praktisch denselben Bereich wie die „normale“ Berechnung.

Das 95% Konfidenzintervall für den Mittelwert 396,85 ergibt sich mit den Formeln 4.15 und 4.17 zu [385,88 – 407,81].

Im Rahmen einer Sensitivitätsanalyse wird auch das Konfidenzintervall als Resultat der „inkorrekten“ Klassifizierungen „symmetrisch“ mit $z^{0,025} = 1,96$ und damit nach den Formeln 4.12 und 4.13 die beiden Ränge $r = 6$ und $s = 15$ berechnet. Die beiden Konfidenzintervalle sind in Tabelle 6.3 dargestellt.

Klassifizierung der Verteilungsform der Beobachtungen	95% Konfidenzintervall für den Mittelwert 396,85
Schief und Normal	[385,88 – 407,81]
(Symmetrisch)	([380,49 – 415,79])

Tabelle 6.2: Mittelwerte der durchschnittlichen Durchlaufzeit mit Konfidenzintervall auf Basis der Verteilungsform bei 20 Stichproben (alternative inkorrekte Klassifizierung in Klammern).

Als weiterer Aspekt der Sensitivitätsanalyse werden die Stichproben auf 40 verdoppelt, um den Einfluss der Anzahl der Beobachtungen auf die Verteilungsform und die Konfidenzintervallbreite zu untersuchen.

Es werden nun 40 Stichproben der durchschnittlichen Durchlaufzeiten aus den 40 unabhängigen Replikationen benutzt, um – wie oben –, die relevanten Kennzahlen zu schätzen, die Verteilungsdichte und Box-Plots darzustellen. Zusammen mit dem χ^2 -Test dient dies zur Klassifizierung der Abweichung von der Normalverteilung.

Die Tabelle zeigt diese Kennzahlen der Stichprobenverteilung und das Testergebnis.

Kennzahl (Durchlaufzeit in Minuten)	Systemalternative „automatisch“
Mittelwert \bar{x}	395,88
Modus x_{mod}	380,73
Median \tilde{x}	386,13
Varianz s^2	424,22

Schiefe γ_1	1,32
Exzess γ_2	0,77
χ^2 -Test p-Wert	4,926e-06

Tabelle 6.3: Stationäre Eigenschaften der Zeitreihe der Durchlaufzeiten (bei 40 Stichproben).

Die geschätzten Dichten der 40 Stichproben der durchschnittlichen Durchlaufzeiten und die zugehörigen Box-Plots der Verteilungen der simulierten Systemalternativen „automatisch“ werden in Abb. 6.5 gezeigt.

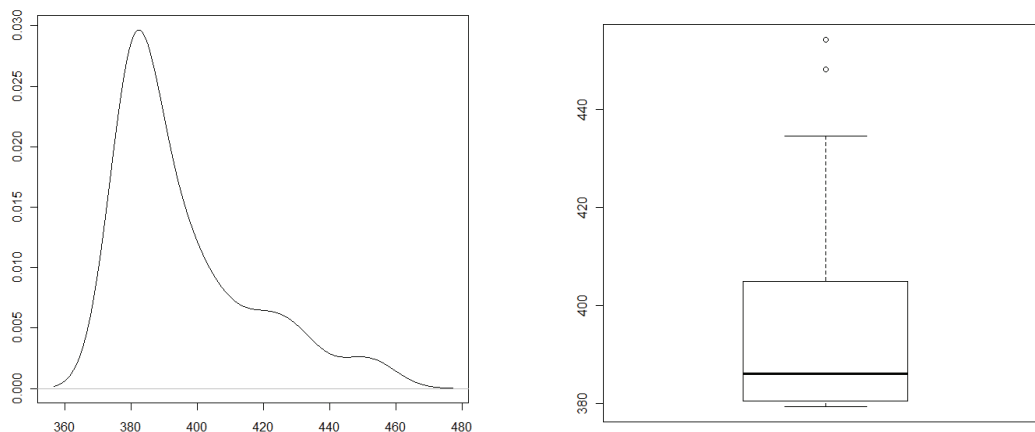


Abbildung 6.5: Dichten und Box-Plots Verteilung der Durchlaufzeiten der simulierten Systemalternative (bei 40 Stichproben).

Mit diesen Kennzahlen aus Tab. 6.3 und den Darstellungen aus Abb. 6.5 lässt sich die Stichprobenverteilung nun bezüglich der Abweichung von der Modellannahme der Normalverteilung als (links-) „schief“ klassifizieren. Allerdings ist auch bei 40 Beobachtungen der Korrekturterm von Johnson mit $1,2964 \cdot 10^5$ vernachlässigbar klein, so dass praktischerweise die normale t-Statistik verwendet wird. Das 95% Konfidenzintervall für den Mittelwert 395,88 ergibt sich damit zu [389,30 – 402,45].

Im Rahmen einer Sensitivitätsanalyse wird auch hier das Konfidenzintervall als Resultat der „symmetrischen“ Klassifizierung mit $z^{0,025} = 1,96$ und damit den Rängen $r = 14$ und $s = 27$ berechnet und in der Tabelle 6.4 in Klammern dargestellt.

Die Klassifizierung der Verteilungsform bleibt unverändert „schief“. Tabelle 6.4 zeigt die Konfidenzintervalle bei verdoppeltem Stichprobenumfang. Erwartungsgemäß sind die Konfidenzintervallbreiten bei einer verdoppelten Stichprobenanzahl nach Relation 4.6 ungefähr um den Faktor $\sqrt{2}$ kleiner.

	Mittelwert 395,88 mit 95% Konfidenzintervall
Schief und Normal	[389,30 – 402,45]
(Symmetrisch)	([380,85 – 396,93])

Tabelle 6.4: Mittelwerte der durchschnittlichen Durchlaufzeit mit Konfidenzintervall auf Basis der Verteilungsform bei 40 Stichproben (alternative inkorrekte Klassifizierung in Klammern).

Damit liefert das integrierte Verfahren eine Schätzung der Leistungskennzahl der Durchlaufzeit mit einer statistisch korrekten Berechnung des Konfidenzintervalls als Güteindikator der Schätzung.

Dieses Konfidenzintervall als Güteindikator der Mittelwertschätzung betont für den Praktiker die stochastische Natur der Simulationsergebnisse. Es gibt keinen „richtigen“ Mittelwert, sondern nur eine korrekte Schätzung auf Basis der vorliegenden Beobachtungen mit einem der Irrtumswahrscheinlichkeit entsprechenden Konfidenzintervall.

6.4.3 Bewertung

Mit der Anwendung des integrierten Verfahrens zur Auswertung der Simulationsergebnisse der Montageprozesse der „mobilen Fabrik“ im Rahmen des ManuBuild Projekts konnte die Auswertungsqualität und damit die Bewertung des innovativen Konzepts deutlich verbessert werden. Damit wurde auch die Grundlage für bessere simulationsbasierte Entscheidungen für den Einsatz des mobilen Produktionskonzepts geschaffen.

Als wesentliche Vorteile der Anwendung des integrierten Verfahrens ergeben sich:

- Eine Sensibilisierung der industriellen Anwender für die Problematik und Notwendigkeit der statistischen Simulationsauswertung sowie der Grenzen der Aussagefähigkeit der Ergebnisse aus Simulationsstudien, insbesondere bei der Schätzung von Leistungskennzahlen,
- eine hohe Nachvollziehbarkeit und Plausibilität des integrierten Verfahrens bei Anwendern ohne vertiefte Statistikenkenntnisse sowie
- eine gesteigerte Akzeptanz der Simulationsergebnisse inklusive ihrer Güte bzw. Unschärfe der Schätzung der Kennzahlen.

Durch die Verwendung des integrierten und robusten Verfahrens wurde eine bessere Auswertungs- und Entscheidungsqualität von Simulationsstudien er-

reicht. Der Einsatz der einzelnen Methoden und Prozeduren sowie die damit erreichten Zwischenergebnisse waren stets nachvollziehbar und zielorientiert.

Der Nachteil des erhöhten Aufwands bei der Simulationsdurchführung durch die Erzeugung der Anzahl der notwendigen unabhängigen Stichproben sowie der Aufwand für die statistische Auswertung mit dem integrierten Verfahren wurde für die wesentliche Verbesserung der Aussagequalität und damit auch der nachgelagerten Entscheidungen als untergeordnet betrachtet.

6.5 Simulation einer Schuhproduktion

6.5.1 Aufgabenstellung

Beim Projekt CeC-made-shoes wurde das integrierte Auswertungsverfahren verwendet, um mit Hilfe einer Simulationsstudie die Schuhproduktion in der Modellfabrik in Vigevano zu verbessern. Abb. 6.6 zeigt einen Ausschnitt aus der integrierten Schuhfabrik in Vigevano.



Abbildung 6.6: Ausschnitt aus der integrierten Schuhfabrik in Vigevano (aus [CeC06]).

Die dargestellte industrielle Anwendung dient zum Vergleich von Produktionssystemalternativen bei der Produktion von „Bio-Schuhen“, in diesem Fall einem vorhandenen Ist-Zustand im Vergleich zu drei alternativen Szenarien. Dabei ist der Durchsatz, d.h. die Anzahl der produzierten Schuhpaare pro Arbeitsschicht, die relevante Kennzahl. Der Auftrags- und Modell-Mix, d.h. die Unterschiede in den Arten der produzierten Schuhe, wie klassisch oder sportlich, wurde bei der Anwendung des integrierten Auswertungsverfahrens nicht weiterbetrachtet, um

eine multivariate Aufgabenstellung zu vermeiden. Für multivariate Aussagen kann das integrierte Auswertungsverfahren wie in Kapitel 5.5 beschrieben jederzeit mit der Bonferroni-Ungleichung erweitert werden.

6.5.2 Beschreibung der Anwendung des integrierten Verfahrens

Das Ziel der Anwendung des integrierten Verfahrens war die Bestimmung des Durchsatzes an Schuhpaaren bei der vorhandenen Konfiguration der Produktionsanlage als auch bei den drei alternativen Lösungen. Die entsprechenden Systemmodelle „Ist“ und „Szenarien 1-3“ wurden in Dosimis3 modelliert. Das verifizierte und an der realen Fabrik validierte Simulationsmodell bestand in Dosimis3 aus den fünf Teilmodellen

- Zuschneidebereich,
- Vor-Nähen und Nähbereich,
- Vorbereitungs- und Montagebereich,
- Materiallager sowie
- Kommissionierung (als rein virtueller Bereich, für den es keine reales Vorbild gab)

und ist in Abb. 6.7 dargestellt. Bei diesem Simulationsmodell wird der Auftrags- und Modellmix über eine Microsoft Excel-Datei an Dosimis3 übergeben.

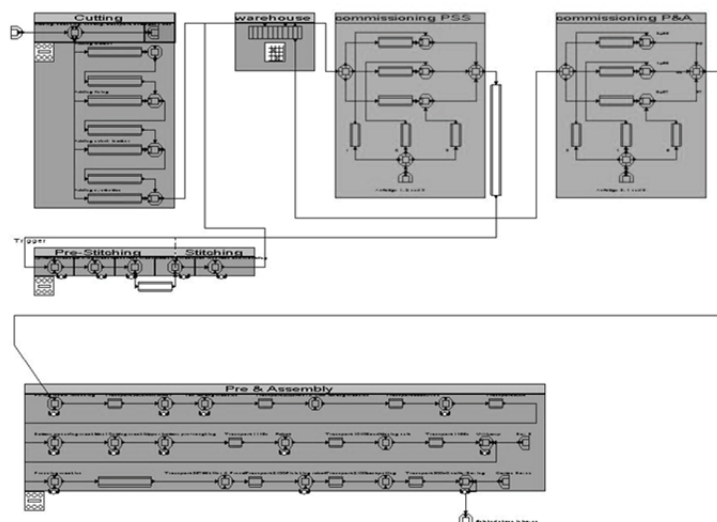


Abbildung 6.7: Simulationsmodell der integrierten Schuhfabrik in Vigevano (aus [CeC06]).

Um die evtl. vorhandenen signifikanten Unterschiede bei der Leistungskennzahl „Durchsatz“ der vier Systemalternativen bestimmen zu können, wurden von allen vier Simulationsmodellen 40 unabhängige Stichproben benötigt. Die Stichproben wurden durch jeweils 40 unabhängige Replikationen mit neuen unab-

hängigen Zufallszahlenstartwerten erzeugt. Tabelle 6.5 zeigt die Kennzahlen der Verteilung der Stichproben, Abb. 6.8 die vier zugehörigen geschätzten Verteilungsdichten und Box-Plots bei 40 Stichproben.

Kennzahl	Ist	Szenario 1	Szenario 2	Szenario 3
Mittelwert \bar{x}	2353,70	2346,55	2392,08	2388,52
Modus x_{mod}	2353	2345	2396	2393
Median \tilde{x}	2353	2346.5	2396	2392
Varianz s^2	15,96	23,33	129,51	100,31
Varianzkoeffizient v	0,0017	0,0021	0,0048	0,0042
Schiefe γ_1	-0,26	-0,51	-3,39	-2,83
Exzess γ_2	-0,39	0,26	11,41	8,04
X^2 -Test p-Wert	0,0073	0,54	< 2,2e-16	2,519e-09

Tabelle 6.5: Kennzahlen der vier Systemalternativen aus vierzig Stichproben.

Bei der Verteilungsform erkennt der X^2 -Test nur Szenario 1 als normalverteilt und liefert bei den anderen drei Systemalternativen sehr geringe p-Werte.

Der Cochran-Test der vier Varianzen berechnet als Testgröße 0,4813 und lehnt die Nullhypothese der Varianzgleichheit mit dem p-Wert von 4,838e-05 ab.

Damit lassen sich die Verteilungen der Stichproben klassifizieren, um ein geeignetes MMV-Verfahren auszuwählen:

Mit der Kategorisierung der vierzig Beobachtungen als große Stichprobenanzahl mit großen Varianzunterschieden und schiefen Verteilungen ist „W+S“, der Wilcoxon-Test mit der Šidák-Ungleichung, am besten geeignet.

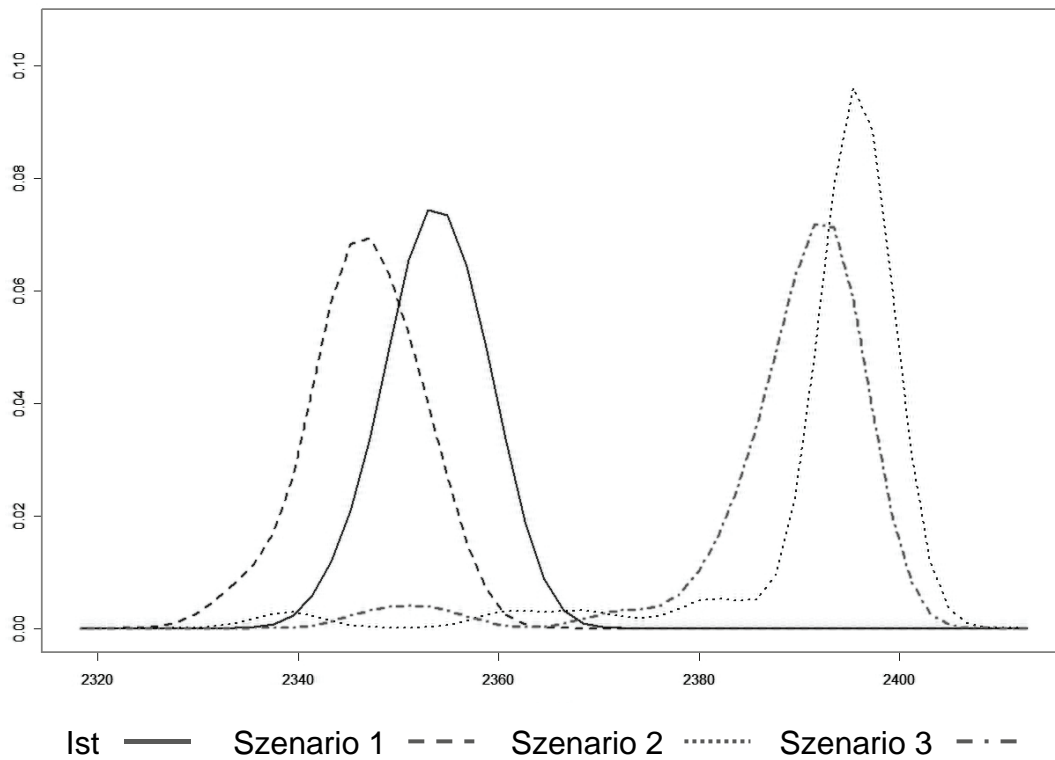


Abbildung 6.8: Geschätzte Dichten der Verteilung des Durchsatzes der vier Systemalternativen bei vierzig Stichproben.

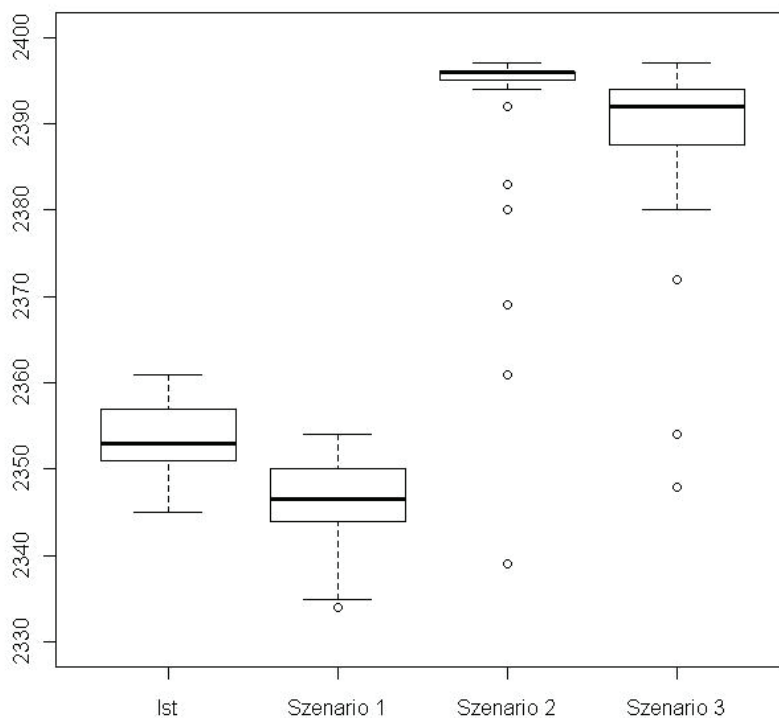


Abbildung 6.9: Geschätzte Dichten und Box-Plots der Verteilung des Durchsatzes der vier Systemalternativen bei vierzig Stichproben.

Für den Vergleich der vier Systemalternativen benötigt man mit dem Binomialkoeffizient $\binom{4}{2} = 6$ paarweise Vergleiche. Der Wilcoxon-Test mit der Šidák-Ungleichung erkennt beim Gesamtsignifikanzniveau von $\alpha_G = 0,05$ und damit dem individuellen Signifikanzniveau von $\alpha_k = 0,008512$ **alle** Unterschiede bei der Kennzahl Durchsatz zwischen den Systemalternativen, alle p-Werte sind deutlich kleiner als α_k .

Wilcoxon/Šidák (p-Werte)	Szenario 1	Szenario 2	Szenario 3
Ist	9,328e-9	1,964e-13	6,079e-13
Szenario 1	-	1,539e-13	4,698e-14
Szenario 2	-	-	6,204e-7

Tabelle 6.6: Testergebnisse der sechs Mittelwertvergleiche: p-Werte des Wilcoxon Tests.

Damit lassen sich die vier Systemalternativen nach den signifikanten Unterschieden zum Niveau $\alpha_G = 0,05$ in der Kennzahl Durchsatz wie folgt eindeutig sortieren:

$$\text{Szenario 1} < \text{Ist} < \text{Szenario 3} < \text{Szenario 2}$$

Eine im Rahmen einer Sensitivitätsanalyse durchgeführte alternative Beurteilung der Unterschiede mit dem Zweistichproben-t-Test und der Bonferroni-Ungleichung, trotz der deutlich unterschiedlichen Varianzen, ergibt ein teilweise anderes Ergebnis. Mit demselben Gesamtsignifikanzniveau von $\alpha_G = 0,05$, was ein individuelles Signifikanzniveau von $\alpha_k = 0,008333$ bedeutet, wird kein statistisch signifikanter Unterschied zwischen Szenario 2 und Szenario 3 erkannt (p-Wert $> \alpha_k$).

t-Test/Bonferroni (p-Werte)	Szenario 1	Szenario 2	Szenario 3
Ist	3,597e-10	<2,2e-16	<2,2e-16
Szenario 1	-	<2,2e-16	<2,2e-16
Szenario 2	-	-	0,1427

Tabelle 6.7: Testergebnisse der sechs Mittelwertvergleiche: p-Werte des Zweistichproben-t-Tests.

Im Gegensatz zum Zweistichproben-t-Test kann der ausgewählte Wilcoxon-Test einen signifikanten Unterschied zwischen den beiden Systemalternativen Szenario 2 und Szenario 3 identifizieren.

Damit hilft das integrierte Verfahren dem Anwender, die statistisch signifikanten Unterschiede zwischen den vier Systemalternativen mit einem definierten

Güteniveau von 95% zu erkennen und auf dieser Grundlage entsprechend abgesicherte Entscheidungen zu treffen.

6.5.3 Bewertung

Durch den Einsatz des integrierten Verfahrens zur Auswertung der Simulationen der integrierten Schuhfabrik konnten unterschiedlich Alternativen zur Neugestaltung des Produktionssystems untersucht und signifikante Unterschiede erkannt werden.

Damit wurde die Entscheidungssicherheit bei den Verbesserungsmaßnahmen zu dem vorgegebene Qualitätsniveau ($\alpha_G = 0,05$) der Signifikanz der Simulationsergebnisse eindeutig verbessert und die Akzeptanz der Ergebnisse und Empfehlungen deutlich gesteigert. Die Vorgehensweise erwies sich als strukturiert und zielführend.

Die wesentlichen Vorteile der Anwendung des integrierten Verfahrens bei der Simulation der Schuhproduktion waren:

- Verbesserte Entscheidungsqualität mit dem Bewusstsein und der Quantifizierung der Qualität bzw. auch Unschärfe der Simulationsergebnisse,
- hohe Akzeptanz, Plausibilität und Praxisnähe des integrierten Verfahrens durch die Nachvollziehbarkeit der Methoden als auch deren Auswahl,
- Verständnis für Notwendigkeit und Vorgehensweise bei der statistischen Auswertung als auch bei der Durchführung der gesamten Simulationsstudie.

Durch die Verwendung des integrierten Verfahrens wurden notwendige Entscheidungen auf Grundlage der Simulationsdaten statistisch korrekt hergeleitet und mit Güteindikatoren hinterlegt.

Der Nachteil des erhöhten Aufwands, insbesondere der Rechenzeit, jedoch kaum der Personalstunden, wurde durch die deutlich verbesserte Ergebnis- und Entscheidungsqualität aufgehoben und sogar umgekehrt.

6.6 Zusammenfassung

Die Anwendung bei zwei industriellen Simulationsstudien von Produktionssystemen in der Bau- und Schuhindustrie zeigt, dass durch den praktischen Einsatz die Erreichung der Zielsetzung der Entwicklung des integrierten Verfahrens bestätigt werden kann. Vorhandene Defizite bei der Auswertung von Simulationsergebnissen sind durch die Vorgehensweise beseitigt und die Auswertungs- und Ergebnisqualität verbessert.

Das integrierte Verfahren erweist sich als strukturiert und zielorientiert. Die Nachvollziehbarkeit der Methoden sichert die Plausibilität und Akzeptanz vor allem bei industriellen Anwendern ohne vertiefte Statistikkennntnisse. Die beiden Anwendungen zeigen auch die Notwendigkeit des evtl. vorhandenen Mehraufwands bei der Simulationsdurchführung zur Erzeugung der Anzahl der benötigten Stichproben.

7 Evaluation und Diskussion des integrierten Verfahrens

7.1 Evaluation

Die praktische industrielle Anwendung des integrierten Verfahrens bei den beiden Simulationsstudien zeigt, dass die Vorgehensweise und Methoden für eine robuste und statistisch korrekte Auswertung der Simulationen von Produktionssystemen geeignet sind. Im Einzelnen sollen die Erfahrungen aus der konkreten Anwendung des entwickelten Verfahrens für Simulationsprojekte in der Bau- und Schuhindustrie mittels der Anforderungen aus Kapitel 2.10 überprüft werden.

7.1.1 Robustheit

Die beiden Anforderungen des Einsatzes von robusten statistischen Verfahren und der statistischen Prozeduren ohne die Voraussetzung der Normalverteilung wurden bei der Anwendung des integrierten Verfahrens erfüllt. Statistische Verfahren, welche, wie z.B. der Tukey-Test T_u , die Normalverteilung voraussetzen, wurden nur bei entsprechender Robustheit, bei geringen Abweichungen von der Modellannahme der Normalverteilung sowie nach expliziter Robustifizierung eingesetzt.

Die beiden relevanten Kriterien der statistischen Robustheit nach Kapitel 3.6.1 der Validität, d.h. der Gültigkeit, und der Effizienz wurden von dem integrierten Verfahren in der praktischen Anwendung erfüllt.

7.1.2 Plausibilität

Sowohl die einzelnen Methoden als auch das integrierte Verfahren selbst und die erhaltenen Ergebnisse waren für die Anwender stets nachvollziehbar und plausibel. Durch die Einbeziehung des Anwenders in die Klassifikation der Verteilung der Stichproben und der resultierenden Zwischenergebnisse und Entscheidungen wurde die Akzeptanz gesichert. Die Sensitivitätsbetrachtungen bei einer alternativen Klassifikation und der Vergrößerung der Stichprobenanzahl steigern die Plausibilität und damit das Vertrauen in das robuste integrierte Verfahren.

Das Vermeiden von komplexen statistischen Prozeduren und die relative Einfachheit der einzelnen Methoden, ggf. auch mit Kompromissen bei der statistischen Effizienz, haben sich bei den beiden praktischen industriellen Anwendungen bewährt.

7.1.3 Integration

Die beiden integrativen Aspekte des entwickelten Verfahrens, die Integration aus modularen statistischen Methoden sowie die Einbindung in die durchgängige Gesamtvorgehensweise bei der Durchführung von Simulationsprojekten, haben in der praktischen Anwendung ihre Vorteile aufgezeigt.

Der integrative Aufbau des Auswertungsverfahrens aus einzelnen Methoden und Prozeduren hat sich bei der Anwendung bewährt, ohne dass eine Anpassung oder Ergänzung notwendig war. Insbesondere die obige Anforderung der Plausibilität und Nachvollziehbarkeit wurde durch den integrierten modularen Aufbau wesentlich unterstützt.

Der Vorteil der Integration in die durchgängige Methodenkette und Vorgehensweise bei der Durchführung von Simulationsprojekten wurde bei beiden praktischen Anwendungen bestätigt. Die in Kapitel 1 referenzierte Vorgehensweise nach VDI Richtlinie 3633 hat sich bei beiden Anwendungen bewährt.

7.1.4 Allgemeine Anforderungen

Die beiden allgemeinen Anforderungen der Anwendbarkeit und Anpassbarkeit des robusten integrierten Verfahrens wurden bei der Anwendung auf unterschiedliche Weise validiert.

Durch die Eignung und Leistungsfähigkeit des integrierten Verfahrens wurde die Anwendbarkeit des Verfahrens direkt und explizit bestätigt. Der direkte Nutzen der Anwendung des entwickelten Verfahrens wurde durch die verbesserte Ergebnis- und Entscheidungsqualität dokumentiert.

Die Anpassbarkeit des entwickelten Verfahrens wurde bei beiden Anwendungen nur indirekt evaluiert, da kein konkreter Bedarf für eine Modifikation vorlag. In Rahmen der Diskussion der Ergebnisse mit den industriellen Anwendern wurde eine angepasste Darstellung der Ergebnisse, besonders beim Vergleich der Systemalternativen der Schuhproduktion, erwogen und zurückgestellt. Eine prinzipielle Anpassbarkeit ist durch den modularen integrativen Aufbau inhärent begründet.

7.1.5 Zusammenfassung der Evaluation

Die beiden industriellen Anwendungen haben gezeigt, dass das integrierte Verfahren die statistische Auswertung von Simulationsergebnissen in der Produktion verbessert. Eine Zusammenfassung der Erfüllung der Anforderungen zeigt Tabelle 7.1.

Anforderungen	Anwendung „Manubuild“	Anwendung „CeC made Shoe“
Robuste statistische Methoden A1	+	+
Ohne Normalverteilung als Modell A2	+	+
Plausible Methoden A3	+	+
Entscheidungskompetenz beim Anwender A4	+	+
Integrativer Aufbau aus modularen Methoden A5	+	+
Integration in die Simulationsmethodik A6	+	+
Anwendbarkeit	+	+
Anpassbarkeit	n	o

Legende: + größtenteils erfüllt, o teilweise erfüllt, - größtenteils nicht erfüllt, n nicht benötigt

Tabelle 7.1: Zusammenfassung der relativen Erfüllung der Anforderungen beider industriellen Anwendungen.

Das Ergebnis der Arbeit mit dem integrierten Verfahren zur robusten statistischen Auswertung von Simulationsergebnissen in der Produktion wurde bei beiden praktischen Anwendungen positiv evaluiert. Im Vergleich zu den seitherigen Vorgehensweisen bei der Auswertung von Simulationsstudien wurde die Ergebnis- und Entscheidungsqualität signifikant verbessert.

7.2 Diskussion

Bei der Bewertung des integrierten Verfahrens wurde gezeigt, dass die entwickelte Vorgehensweise zielführend, anwendbar und von erheblichem Nutzen ist. Zusätzlich zur Erfüllung der Anforderungen werden im Folgenden noch weitere, für das entwickelte Verfahren relevante, Gesichtspunkte betrachtet und diskutiert.

7.2.1 Aufwand

Die beiden Anwendungen des Verfahrens haben gezeigt, dass mit einer statistisch korrekten Auswertung der Simulationsergebnisse ein leicht erhöhter Aufwand verbunden ist. Dieser höhere Aufwand, welcher häufig im Wesentlichen in Rechenzeit für zusätzliche Stichproben aus weiteren Replikationen investiert wird, ist relativ klein im Vergleich zum Gesamtaufwand des Anwenders für die Durchführung einer Simulationsstudie. Der eigentliche zusätzliche Aufwand ist

gering, da vor allem Simulations- und Rechenzeit und weniger direkter Zeitaufwand des Anwenders benötigt wird.

Um eine gute Relation von Aufwand und Nutzen zu erzielen, dürfen die Anforderungen an die statistische Qualität der Ergebnisse nicht unrealistisch hoch angesetzt werden. Die Größe der Signifikanzbereiche B_r hat nach der Relation 4.6 $B_r \sim 1/\sqrt{n}$ erheblichen Einfluss auf die Anzahl der notwendigen Stichproben n . Insbesondere bei Produktionssystemen mit einer großen inhärenten Varianz der beobachteten Kennzahlen können evtl. ambitionierte statistische Güteindikatoren, d.h. enge Konfidenzintervalle und kleine Signifikanzniveaus entweder mit erheblichen Simulationsaufwand oder nur unter Einsatz von zusätzlichen, relativ komplexen, varianzreduzierenden Methoden wie „gemeinsamen Zufallszahlen“ (Common Random Numbers) [Law07:578] erreicht werden.

7.2.2 Effizienz versus Plausibilität und Robustheit

Sowohl bei der Auswahl der Methoden als auch bei der praktischen Anwendung des integrierten Verfahrens wurde die Plausibilität und Robustheit der statistischen Methoden und Prozeduren höher gewertet als die Effizienz. Allerdings stehen die drei Eigenschaften in einem Zielkonflikt.

Die Robustheit und Effizienz einer Prozedur sind messbare statistische Funktionen, während die Plausibilität als relative Eigenschaft stark vom Kenntnissstand des Anwenders abhängt. Damit können bei entsprechenden statistischen Kenntnissen und bei Bedarf auch effizientere und damit komplexere Verfahren wie u.a. in Kap 3.6.2. beschrieben, für den Anwender plausibel und nachvollziehbar sein.

Andererseits kann eine geringere Effizienz des Gesamtverfahrens meist durch eine erhöhte Anzahl von Stichproben und damit durch einen höheren Simulationsaufwand (ohne direkte Beteiligung des Anwenders) wieder ausgeglichen werden.

Die Robustheit der Verfahren sollte bei einer Anpassung oder Erweiterung des Verfahrens nicht durch effizienzsteigernde Maßnahmen kompromittiert werden. Im Gegenteil: Effizienzverluste durch eine Robustifizierung der Prozeduren, wie in Kap. 4.4.6.2 und bei Büning [Büni91:53] beschrieben, werden bewusst akzeptiert und durch zusätzliche Stichproben, d.h. weitere Replikationen, wieder ausgeglichen.

7.2.3 Anwendungsbereich und Übertragbarkeit

Das integrierte Verfahren eignet sich nicht nur für die statistische Auswertung von Produktionssimulationen, sondern generell für die Auswertung der Simulationen mit stochastischen und ereignisdiskreten Modellen. Dies beinhaltet

Simulationen von Modellen anderer Anwendungsbereiche wie bspw. Prozesse und Abläufe innerhalb anderer Unternehmensbereiche sowie von Dienstleistungen. Auch Simulationen von logistischen Abläufen und unternehmensübergreifende Kooperationen können und sollten, wenn dabei stochastische und ereignisdiskrete Modelle verwendet werden, mit dem entwickelten Verfahren ausgewertet werden.

Eine Übertragbarkeit ist teilweise auch für andere stochastische Modelle gegeben. Die Module des integrierten Verfahrens, welche sich nicht auf Zeitreihen, sondern auf unabhängige Beobachtungen beziehen, sind auch für Daten und Ergebnisse aus den Simulationen von statischen, d.h. nicht zeit-dynamischen stochastischen Modellen, wie bspw. der Risikoanalyse, geeignet.

Die Anwendung des Verfahrens ist auch für die Auswertung der Simulation von Kombinationen stochastischer ereignisdiskreter Modellen mit kontinuierlichen oder mit statischen Modellen sinnvoll. Diese kombinierten Modelle werden bei der hierarchischen Simulation komplexer Produktionssysteme verwendet.

7.2.4 Erweiterbarkeit

Die bereits bei den allgemeinen Anforderungen dokumentierte Eigenschaft der Anpassbarkeit gewinnt an Bedeutung, wenn das integrierte robuste Verfahren zukünftig für weitere Einsatzgebiete erweitert werden soll. Diese Erweiterbarkeit dient dazu, weitere statistische Methoden und Prozeduren für zusätzliche Anwendungsfälle einzubinden. Aus der praktischen Anwendung des Verfahrens ergeben sich mehrere Erweiterungsoptionen, welche bei der Konzeption der Verfahren in Kap. 3 teilweise schon erwähnt wurden.

- **Verfahren für multivariate Modelle und Simulationen:**

Bei multivariaten Leistungskennzahlen des Produktionssystems kann wie in den Kap. 4.6 und 5.5 beschrieben die Bonferroni-Ungleichung verwendet werden. Allerdings kann diese Vorgehensweise zu „ungewöhnlichen“ Signifikanzniveaus führen. Alternative multivariate Verfahren der Varianzanalyse werden bei Hartung und Elpelt [HaEl99:692ff] beschrieben. Eine multivariate Erweiterung des robusten integrierten Verfahrens erscheint sinnvoll und nur bedingt aufwändig.

- **Verfahren für Simulationen mit vielen Eingangsfaktoren:**

Für Simulationen mit vielen Eingangsfaktoren sollte nach Kleijnen [Klei87:320ff] mit einem „Aussieben“ (Screening) der relevanten Faktoren die Komplexität des Modells und damit die Anzahl der Simulationsexperimente soweit wie

sinnvoll reduziert werden. Danach muss mit den Methoden der Versuchsplanung eine möglichst effiziente, d.h. geringe Anzahl von notwendigen Simulationsversuchen entworfen werden. Ergänzend dazu können, wie in Kap. 3.5 erwähnt, mit einem stufenweisen Vorgehen zur Identifikation der relevanten Eingangsfaktoren mit den R&S-Verfahren und einem geringen, vorgegebenem signifikanten Unterschied d^* eine Reihenfolge festgelegt und eine Vorauswahl durchgeführt werden, bevor mit den MMV-Verfahren die eigentlichen signifikanten Unterschiede bestimmt werden. Eine Erweiterung des Verfahrens für viele Eingangsfaktoren benötigt damit sowohl die Einbindung der Verfahren der Versuchsplanung als auch die Integration der R&S-Verfahren und generiert damit eine gewisse Komplexität.

- **Varianzreduzierende Verfahren:**

Insbesondere bei Simulationsmodellen mit einer starken inhärenten Varianz können varianzreduzierende Verfahren sinnvoll und notwendig sein. Law [Law07:572] gibt zu den bekanntesten Verfahren, wie „gemeinsamen Zufallszahlen“ (Common Random Numbers) oder „gegensätzliches Merkmal“ (Antithetic Variate), einen guten Überblick. Allerdings geht bei der Verwendung dieser Verfahren die Unabhängigkeit der einzelnen Stichproben und Beobachtungen verloren und wird durch eine gewollte (positive oder negative) Korrelation ersetzt. Damit ist aber eine der wesentlichen Voraussetzungen der statistischen Prozeduren des entwickelten integrierten robusten Verfahrens verletzt. Eine mögliche Erweiterung mit der Nutzung von varianzreduzierenden Verfahren erfordert eine erneute Bewertung und Auswahl der verwendeten statistischen Prozeduren, was einen großen Aufwand und eine deutlich höhere Komplexität des integrierten Verfahrens impliziert.

All diese Erweiterungsoptionen sollten den Anforderungen der Einfachheit und Plausibilität entsprechen sowie die Kenntnisse und Fähigkeiten eines industriellen Anwenders nicht überfordern.

8 Ausblick

Das entwickelte integrierte Verfahren bietet eine wichtige und wesentliche Unterstützung für industrielle Anwender beim Einsatz der Simulationstechnik für die Gestaltung von Produktionssystemen. Bei der Entwicklung des integrierten Verfahrens wurde besonderer Wert sowohl auf die Einfachheit, Nachvollziehbarkeit und Plausibilität der Vorgehensweise als auch der einzelnen statistischen Prozeduren und Methoden gelegt.

Der integrative Aufbau des Verfahrens erlaubt eine zukünftige Weiterentwicklung, um neuen oder veränderten Anforderungen entsprechen zu können. Speziell die Einbindung und Rolle des Anwenders bei der Beurteilung von Beobachtungen und Klassifikation von Zwischenergebnissen kann bei variierenden statistischen Vorkenntnissen durchaus unterschiedlich anspruchsvoll gestaltet werden. Bei entsprechendem statistischem Verständnis des Anwenders können einzelne einfachere statistische Prozeduren durch effizientere Alternativen ergänzt oder ersetzt werden.

Darüber hinaus bieten sich als anschließende und komplementäre Aufgaben für zukünftige Forschungsprojekte drei wesentliche Themenbereiche an:

- 1 Die Kombination der Versuchsplanung mit der statistischen Auswertung:** Die Verbindung der „strategischen“ Methoden [Klei08:9] der Metamodelle und Versuchsplanung mit dem in der vorliegenden Arbeit entwickelten „taktischen“ Verfahren der statistischen Auswertung der stochastischen Simulationsexperimente eröffnet ein großes Potenzial, um den gesamten Simulationsaufwand signifikant zu reduzieren. Dabei werden sowohl die „richtigen“ Parameterkombinationen für die Versuche ausgewählt und damit die Anzahl der zu simulierenden Systemalternativen reduziert als auch die Simulationsexperimente mit dem integrierten robusten Verfahren richtig ausgewertet. Sowohl der Aufwand für den Simulationsanwender als auch der Rechenaufwand für die eigentlichen Simulationsversuche kann bei gleicher oder verbesserter Entscheidungsqualität erheblich reduziert werden.
- 2 Die Verbindung der stochastischen Modellierung der Eingangsdaten mit dem integrierten robusten Verfahren der statistischen Auswertung:** Für eine Steigerung der Effizienz eines Simulationsprojekts ist eine Verbindung der stochastischen Modellierung der Eingangsdaten mit der statistischen Auswertung interessant, weil sich dadurch eine Sensitivitätsanalyse für den Einfluss des Modellierungsaufwands auf die Ergebnisqualität durchführen lässt. Eine erhöhte Genauigkeit der Modellierung der stochastischen Eingangsdaten mit einer empirischen oder theoretischen Verteilungsfunktion [Law07:279] kann, bis zu einer gewissen Grenze, bei vielen Produktionssystemen zu einer verbesserten Modellvalidität und Er-

gebnisqualität führen. Eine Aufwand-Nutzen-Relation wird dabei durch das entwickelte integrierte Verfahren mit den geschätzten Konfidenzintervallbreiten sowie den erkannten signifikanten Unterschieden der Systemalternativen ermöglicht. Eine systematische Betrachtung dieser Input-Output-Relation beinhaltet weiterführende Forschungsaufgaben.

3 Die Einbeziehung varianzreduzierender Methoden ohne erhebliche Nachteile bei der Einfachheit und Plausibilität des Gesamtverfahrens:

Die bekannten varianzreduzierenden Methoden der „gemeinsamen Zufallszahlen“ (Common Random Numbers) der „gegensätzlichen Merkmale“ (Antithetic Variate) [Law07:572] verletzen die Voraussetzung der Unabhängigkeit der Beobachtungen, da die generierenden Replikationen bewusst korreliert werden. Um bei Produktionssystemen und den abgeleiteten Simulationsmodellen mit einer erheblichen systemimmanenten Varianz dennoch relativ kleine Konfidenzintervalle berechnen und signifikante Unterschiede zwischen Systemalternativen erkennen zu können, sind Erweiterungen des entwickelten integrierten robusten Verfahrens sinnvoll. Dies beinhaltet ggf. weitere Monte-Carlo-Studien über die Gesamteffizienz des integrierten Verfahrens als auch Untersuchungen in Bezug auf Einfachheit, Nachvollziehbarkeit und Plausibilität aus Anwendersicht.

Abschließend gibt es weiterhin Forschungsbedarf für die Fragen nach einer durchgängigen und anwenderfreundlichen methodischen Unterstützung für die gesamte Durchführung von Simulationsexperimenten mit Produktionssystemen. Dabei sollten vorhandene Partiallösungen, wie in Kapitel 1.4 beschrieben, in eine anwenderorientierte, methodisch richtige und konsistente Gesamtverfahrensweise für die Simulation von Produktionssystemen integriert werden.

9 Zusammenfassung

Ein erfolgreiches Unternehmen muss seine Produkte und Dienstleistungen schnell und kostengünstig auf internationalen Märkten anbieten sowie flexibel auf veränderte Anforderungen reagieren können. Die Simulationstechnik als eine Schlüsseltechnologie zur frühen und schnellen Planung, Entwicklung und Bewertung von Produkten, Dienstleistungen und Produktionsprozessen ist dabei von entscheidender Bedeutung. Allerdings sind bei der aufwändigen und komplexen statistischen Auswertung der Ergebnisse von ereignisdiskreten stochastischen Simulationen noch große Defizite vorhanden. Daher muss diese Aufgabe für industrielle Anwender methodisch unterstützt, verbessert und vereinfacht werden.

Das Ziel der vorliegenden Arbeit war es, ein integriertes Verfahren zur robusten statistischen Auswertung von Simulationsergebnissen in der Produktion zu entwickeln, das Simulationsanwender ohne vertiefte Statistikkenntnisse bei der korrekten Auswertung ihrer Simulationsergebnisse unterstützt und damit besser abgesicherte Entscheidungen ermöglicht.

Um dieses Ziel zu erreichen, wurde auf Grundlage der Simulationstechnik im industriellen Einsatz der Stand von Forschung, kommerziellen Werkzeugen und Praxis analysiert. Aus den erkannten Defiziten in den drei wesentlichen Bereichen Robustheit, Plausibilität und Integration wurde ein entsprechendes Anforderungsprofil für das zu entwickelnde integrierte und robuste Verfahren zur statistischen Auswertung der Simulationsergebnisse in der Produktion abgeleitet.

Bei der Konzeption des integrierten Vorgehensmodells wurde anhand des Anforderungsprofils sowohl die Einbindung in die Simulationsmethodik und des Anwenders als auch der strukturierte Auswertungsprozess gestaltet. Dazu wurden, teilweise mit Unterstützung von Monte-Carlo-Studien, die geeigneten statistische Methoden und Prozeduren ausgewählt und integriert.

Auf Grundlage dieses Konzepts wurde für statistische Auswertung eines Simulationsmodells das detaillierte integrierte Verfahren entwickelt. Der Anwender wird eingebunden und dabei unterstützt, Kennzahlen einer Produktionssimulation mit einem Konfidenzintervall korrekt schätzen und damit auch die Qualität der Schätzung richtig beurteilen zu können.

Für den Vergleich der Ergebnisse von mehreren Simulationsmodellen hinsichtlich signifikanter Unterschiede wurden auf Basis einer Klassifikation der Beobachtungen die geeigneten Verfahren der Multiplen Mittelwertvergleiche in eine robuste Vorgehensweise integriert.

Das entwickelte integrierte und mehrstufig adaptive Verfahren wurde bei zwei industriellen Simulationsstudien in der Bau- und Schuhindustrie angewendet.

Die Evaluation des entwickelten Verfahrens bei der Simulation von unterschiedlichen Produktionssystemen, einer mobiler Fabrik für Fertigbauteile sowie einer flexiblen Schuhproduktion, bestätigte den Nutzen, die Anwendbarkeit sowie die Übertragbarkeit. Dabei wurde auch die Erreichung der Zielsetzung und die Erfüllung des Anforderungsprofils überprüft und bestätigt.

Eine abschließende Diskussion des entwickelten integrierten robusten Verfahrens zeigt ein großes Potenzial bei der Übertragbarkeit für die Auswertung anderer Anwendungsbereiche von stochastischen Simulationen sowie die Möglichkeiten für weitere zukünftige Erweiterungen.

10 Abstract

Successful companies must be able to offer their products and services on international markets in a fast and cost efficient way as well as react flexible on changing requirements. Being a key technology for the early and fast planning, development and evaluation of products, services and production processes, the technique of simulation has a decisive impact on this. However, there are large deficits regarding the complex and time consuming statistical evaluation of the results of stochastic discrete-event simulations. Therefore, this task has to be improved, simplified and supported by a methodology for industrial users.

The objective of this thesis is to develop an integrated method for the robust statistical analysis of simulation results in manufacturing. This method will support simulation users without detailed statistical knowledge with the correct analysis of simulation results and enable more reliable decisions.

The current state of research, commercial tools and practical applications were analyzed on basis of the industrial situation. Based on the identified deficits in the three relevant areas of robustness, plausibility and integration, the corresponding profile of requirements was defined to support the development of the integrated and robust method for the statistical analysis of simulation results in manufacturing.

For the concept of the integrated procedure model, requirements were used to design the integration of the user and the overall simulation methodology as well as the structured process of statistical analysis. Appropriate statistical methods and procedures were selected, partly with the support of Monte-Carlo studies, and integrated.

Based on this concept of the integrated procedure model, the detailed integrated method for the statistical analysis of a single simulation model was developed. The user is integrated and supported to estimate the performance indicators of a manufacturing simulation correctly, including the assessment of the estimation's quality.

In order to compare the results from multiple simulation models for significant differences, appropriate procedures of multiple comparisons of means were integrated based on a classification of the distribution of observations.

The developed integrated and multistage adaptive method for the robust statistical analysis of simulation results was applied and validated for two industrial simulation studies of manufacturing processes in the construction and shoe industry. The evaluation of the simulation of two different manufacturing systems, a mobile factory for pre-manufactured components and a flexible shoe production system confirmed the benefit, usability and transferability of the new

integrated method. The achievement of the objectives and the fulfillment of the requirements were also verified and confirmed.

The final discussion of the integrated method for the robust statistical analysis of simulation results showed a large potential in the transferability for the analysis of stochastic simulations of other areas of application as well as the opportunity for future extensions.

11 Literaturverzeichnis

- [Abdi07] H. Abdi: *Bonferroni and Šidák corrections for multiple comparisons*. In: N.J. Salkind (Ed.): *Encyclopedia of Measurement and Statistics*. Thousand Oaks. Sage. S. 103-107. 2007.
- [Alex07] C. Alexopoulos: *Statistical Analysis of Simulation Output: State of the Art*. In: S.G. Henderson, B. Biller; M.-H. Hsieh, J. Shortle, J.D. Tew, und R.R. Barton (Hrsg.): *Proceedings of the 200 Winter Simulation Conference*. 2007.
- [AlGo03] C. Alexopoulos und D. Goldman: *To batch or not to batch*. In: S. Chick, P.J. Sánchez, D. Ferrin und D.J. Morrice (Hrsg.): *Proceedings of the 2003 Winter Simulation Conference*. 2003.
- [AlKi02] C. Alexopoulos und S.-H. Kim: *Output Data Analysis for Simulations*. In: E. Yücesan, C.-H. Chen, J.L. Snowdon und J.M. Charnes (Hrsg.): *Proceedings of the 2002 Winter Simulation Conference*. 2002.
- [AlSe98] C. Alexopoulos und A.F. Seila: *Output Data Analysis*. In: J. Banks (Hrsg.): *Handbook of Simulation*. New York u.a. John Wiley & Sons. 1998.
- [Andr98] S. Andradóttir: *Simulation Optimization*. In: J. Banks (Hrsg.): *Handbook of Simulation*. New York u.a. John Wiley & Sons. 1998.
- [AnAr01] S. Andradóttir und N.T. Argon: *Variance estimation using replicated batch means*. In: B. A. Peters, J. S. Smith, D.J. Medeiros und M.W. Rohrer (Hrsg.): *Proceedings of the 2001 Winter Simulation Conference*. 2001.
- [ASIM97] ASIM-Fachgruppe 4.5.6: *Leitfaden für Simulationsbenutzer in der Produktion und Logistik*. ASIM-Mitteilungen Nr.58, 1997.
- [Bank96] J. Banks: *Output Analysis Capabilities of Simulation Software*. *Simulation* 66:1. S. 23-30, 1996.
- [Bank98] J. Banks (Hrsg.): *Handbook of Simulation*. New York u.a. John Wiley & Sons. 1998.
- [Bank00] J. Banks: *Getting started with AutoMod*. AutoSimulations. Bountiful, Utah. 2000. (sowie <http://www.automod.de/akademische-seite.html> am 24.8.2010).
- [BCNN10] J. Banks, J.S. Carson II, B.L. Nelson und D.M. Nicol: *Discret-Event System Simulation*. 5th Ed. Upper saddle River, New Jersey. Pearson. 2010.
- [BDAS86] H.-J. Bullinger, K. Dungs, E.-D. Ammer, U.A. Seidel und B. Weller: *Systematische Montageplanung*. München. Hanser. 1986.
- [Bech54] R.E. Bechhofer: *A single sample multiple decision procedure for ranking means of normal populations with known variances*. *Annals of Mathematical Statistics*. 25. S. 16-39. 1954.

- [BeSG95] R.E. Bechhofer, T.S. Santner und D.M. Goldsman: *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York u.a. John Wiley & Sons. 1995.
- [Bopp01] R. Bopp: Ein Verfahren zur Restwertgenerierung in der Konstruktion. Heimsheim. Jost Jetter Verlag. 2001.
- [Büni91] H. Büning: *Robuste und adaptive Tests*. Berlin, New York. De Gruyter. 1991.
- [BuWa97] H.-J. Bullinger und F. Wagner: *Petri-net models for logistic planning*. In: A. Kusiak und M. Bielli (Hrsg.) *Designing Innovations in Industrial Logistics Modelling*, 1997.
- [Caci08] CACI Products Company: *User's Manual SIMPROCESS Release 4*. 2008 http://simprocess.com/products/simprocess_manuals.html am 24.8.2010.
- [Cars96] J.S. Carson: *AutoStat: Output statistical analysis for AutoMod users*. In: J.M. Charnes, D.J. Morrice, D.T. Brunner and J.J. Swain (Hrsg.): *Proceedings of the Winter Simulation Conference*. 1996.
- [CeC05] CeC made shoe Consortium: *CeC made shoe Newsletter*. 1. 2005. <http://www.cec-made-shoe.com/> am 7.1. 2011.
- [CeC06] CeC made shoe Consortium: *Deliverable D2-5.1-6 Production Test Case Report*. CeC made shoe Projekt. 2006.
- [Chen07] E.J. Chen: *Indifference-zone subset selection procedures: using sample means to improve efficiency*. In: S.G. Henderson, B. Biller; M.-H. Hsieh, J. Shortle, J.D. Tew, und R.R. Barton (Hrsg.): *Proceedings of the 200 Winter Simulation Conference*. 2007.
- [Chen06] R. Cheng: *Resampling methods*. In: S.G. Hendersen und B.L. Nelson (Eds.): *Handbooks in Operations Research and Management Science. Vol. 13 Simulation*. North Holland, Elsevier. Amsterdam. 2006.
- [ChCu09] R. Cheng und C. Currie: *Resampling methods of analysis in simulation studies*. In: M.D. Rosseti, R.R. Hill, B. Johansson; A. Dunkin und R.G. Ingalls. (Hrsg.): *Proceedings of the 2009 Winter Simulation Conference*. 2009.
- [ChGM97] C. Chien, D. Goldsman und B. Melamed. *Large-Sample Results for Batch Means*. *Management Science*. Vol.43. No.9. 1997.
- [ChLi95] L. Chen: *Testing the Mean of Skewed Distributions*. *Journal of the American Statistical Association*. Vol. 90, Nr. 430. S. 767-772. 1995.
- [DELM08] DELMIA Corp: *QUEST© D5R19SP3 Hilfemenu*. Auburn Hills, MI. 2008.
- [DIN05] DIN Deutsches Institut für Normung (Hrsg.): *DIN EN ISO 9000 Qualitätsmanagementsysteme Grundlagen und Begriffe. ISO 9000:2005*. Beuth. Berlin. 2005.
- [Dött95] S. Döttling: *Simulationsgestützte hierarchische Produktionsregelung am Beispiel der Textilveredelung*. Renningen-Malmheim. expert. 1995.

- [DuDa75] E.J. Dudewicz und S.R. Dalal: *Allocation of Observations in Ranking and Selections with Unequal Variances*. Sankhya B37: 28-78. 1975.
- [Dunn80] C.W. Dunnett: *Pairwise multiple comparisons in the unequal variance case*. JASA Vol.75 Nr. 372, Applications Section, S. 796-800. 1980.
- [Dunn82] C.W. Dunnett: *Robust multiple Comparisons*. In: Communications in Statistics, A 10 (20), Theory + Methods, S. 2057-2071, 1982.
- [dTSS86] S.H.C. duToit; A.G.W. Steyn und R.H. Stumpf: *Graphical Exploratory Data Analysis*. New York u.a. Springer. 1986.
- [Fisc88] J. Fischer: *Die statistische Auswertung von Simulationsdaten*. In: K. Feldmann und B. Schmidt (Hrsg.): *Simulation in der Fertigungstechnik*. Berlin, Springer-Verlag, 1988.
- [Fisc89] J. Fischer: *Verfahren zur Bestimmung der transienten Phase*. Internes Arbeitspapier. Fraunhofer-Institut für Arbeitswirtschaft und Organisation – IAO. Stuttgart. 1989.
- [Fish78] G.S. Fishman: *Principles of Discrete Event Simulation*. New York. John Wiley. 1978.
- [Fish96] G.S. Fishman: *Monte Carlo: concepts, algorithms, and applications*. New York u.a. Springer. 1996.
- [Fish01] G.S. Fishman: *Discrete Event Simulation: Modeling, Programming, and Analysis*. New York u.a. Springer. 2001.
- [FSF09] Free Software Foundation: *Gawk-GNU Project*. <http://www.gnu.org/software/gawk/> am 11.1. 2011.
- [GaBM96] M.A. Gallagher, K.W. Bauer und P.S. Maybeck: *Initial Data Truncation for Univariate Output of Discrete-event Simulation Using the Kalman Filter*. Management Science. 42. S. 559-575. 1996.
- [Gram91] J. Grammel: *Anwendung multipler Mittelwertvergleiche auf die Ergebnisse von ereignisdiskreten Simulationen*. Diplomarbeit. Ludwig-Maximilians-Universität München, 1991.
- [Gram93] J. Grammel: *Bemerkungen zu den Monte Carlo Studien*. Internes Arbeitspapier. Fraunhofer-Institut für Arbeitswirtschaft und Organisation – IAO. 1993.
- [GoNe98] D. Goldsmans und B.L. Nelson: *Comparing Systems via Simulation*. In: J. Banks (Hrsg.): *Handbook of Simulation*. New York u.a. John Wiley & Sons. 1998.
- [GoNS91] D. Goldsmans, B.L. Nelson und B. Schmeiser: *Methods for selecting the best system*. In: B.L. Nelson, W.D. Kelton und G.M.Clark (Hrsg.): *Proceedings of the 1991 Winter Simulation Conference*. 1991.
- [Gupt65] S.S. Gupta: *On multiple decision (ranking and selection) rules*. Technometrics. 7. S. 225-245. 1965.

- [Hadd87] J. Haddock: *An expert system framework based on a simulation generator*. Simulation 48:2 S. 45-53. 1987.
- [Hart02] J. Hartung: *Statistik*. 13. Auflage. München, Wien. Oldenbourg. 2002.
- [HaEl99] J. Hartung und B. Elpelt: *Multivariate Statistik*. 6. Auflage. München, Wien. Oldenbourg. 1999.
- [HeNe06] S.G. Hendersen und B.L. Nelson (Eds.): *Handbooks in Operations Research and Management Science. Vol. 13 Simulation*. North Holland, Elsevier. Amsterdam. 2006.
- [HHRW97] K. Heitmann, A. Hirschberg, E. Rauh und J. Wunderlich: *Zusammenfassung der Studie: Einsatz der Simulationstechnik*. iwv, FAPS. München, Erlangen. 1997.
- [Hlup99] V. Hlupic: *Simulation Software: User's Requirements*. Computers & Industrial Engineering. 37. S. 185-188. 1999.
- [HoRD09] K. Hoard, S. Robinson und R. Davies: *Automating the Analysis of Simulation Output Data*. <http://www2.warwick.ac.uk/fac/soc/wbs/projects/autosimoo/> am 11.12.2009.
- [HoTa87] Y. Hochberg und A.C. Tamhane: *Multiple Comparison Procedures*. New York u.a. John Wiley & Sons. 1987.
- [Holl01] B.W. Hollocks: *Discrete Event Simulation: An Inquiry to User Practice*. Simulation Practice and Theory. 8. S. 451-471. 2001.
- [Hsu96] J.C. Hsu: *Multiple Comparisons*. Boca Raton u.a. Chapman & Hall/CRC. 1996.
- [Hube04] P.J. Huber: *Robust Statistics*. New York u.a. John Wiley & Sons. 2004.
- [Huff54] D. Huff: *How to lie with Statistics*. New York. Norton. 1954.
- [IAO98] Fraunhofer Institut für Arbeitswirtschaft und Organisation - IAO: *ISSTAS – Integriertes System zur Unterstützung der statistischen Auswertung von Simulationsdaten*. Projektabschlussbericht. Stuttgart. 1998
- [IBMC10] IBM Corporation: *SPSS Statistics*. <http://www.spss.com> am 13.4.2010.
- [Inco04] Incontrol Enterprise Dynamics: *Enterprise Dynamics Tutorial ED6*. Maarssen. 2004. <http://www.incontrolsim.com/index.php/ed-tutorials.html> am 24.10.2010.
- [IoIE11] Institute of Industrial Engineers: *Industrial Engineer Magazine*. www.iienet2.org/Landing.aspx?id=973 am 15.4.2011.

- [ISO01] ISO: *ISO TC 69/SC 6 Practical guide to ISO 5725-2:1994 — Accuracy (trueness and precision) of measurement methods and results — Part 2 : Basic method for the determination of repeatability and reproducibility of a standard measurement method*. 2001.
http://www.stat.ucl.ac.be/cours/stat3320/Doc_2004_2005/GUIDE_ISO.pdf
 am 14.11.2011.
- [John78] N.J. Johnson: *Modified t-test and confidence intervals for asymmetrical populations*. Journal of the American Statistical Association. 73. S. 536-544. 1978.
- [KeSS98] W.D. Kelton, R.P. Sadowski und D.A. Sadowski: *Simulation with Arena*. WCB McGraw-Hill. Boston u.a. 1998.
- [KiNe06] S.-H. Kim und B.L. Nelson: *Selecting the Best System*. In: S.G. Hendersen und B.L. Nelson (Eds.): *Handbooks in Operations Research and Management Science. Vol. 13 Simulation*. North Holland, Elsevier. Amsterdam. 2006.
- [KiNe07] S.-H. Kim und B.L. Nelson: *Recent advances in ranking and selection*. In: S.G. Henderson, B. Biller; M.-H. Hsieh, J. Shortle, J.D. Tew, und R.R. Barton (Hrsg.): *Proceedings of the 2007 Winter Simulation Conference*. 2007.
- [KHPI01] V. Kachitvichyanukul, J.O. Henriksen, C.D. Pedgen, R.G. Ingalls und B. Schmeiser: *Simulation Environments for the new Millennium (Panel)*. In: B. A. Peters, J. S. Smith, D.J. Medeiros und M.W. Rohrer (Hrsg.): *Proceedings of the 2001 Winter Simulation Conference*. 2001.
- [Klei87] J.P.C. Kleijnen: *Statistical Tools for Simulation Practitioners*, New York, Marcel Dekker. 1987.
- [Klei98] J.P.C. Kleijnen: *Experimental Design for Sensitivity Analysis, Optimisation, and Validation of Simulation Models*. In: J. Banks (Hrsg.): *Handbook of Simulation*. New York u.a. John Wiley & Sons. 1998.
- [Klei08] J.P.C. Kleijnen: *Design and Analysis of Simulation Experiments*. New York. Springer. 2008.
- [KlvG92] J.P.C. Kleijnen und W. van Groenendal: *Simulation: A Statistical Perspective*, Chichester u.a, John Wiley & Sons, 1992.
- [KoGr95] J. Košturiak und M. Gregor: *Simulation von Produktionssystemen*. Wien New York, Springer, 1995.
- [KuRa98] A. Kuhn und M. Rabe (Hrsg.): *Simulation in Produktion und Logistik: Fallbeispielsammlung*, Berlin, Springer, 1998.
- [KuRW93] A. Kuhn, A. Reinhardt und H.-P. Wiendahl (Hrsg.): *Handbuch Simulationsanwendungen in Produktion und Logistik*, Wiesbaden, Vieweg, 1993.
- [KüSt99] R. Küll und P. Stähly: *Zur Planung und effizienten Abwicklung von Simulationsexperimenten*. In: J. Bietmann, W. Hummeltenberg, B. Schmidt, P. Stähly und T. Witte: *Simulation als betriebliche Entscheidungshilfe*. Heidelberg. Physica. 1999.

- [LaKe00] A.M. Law und W.D. Kelton: *Simulation Modeling and Analysis*. 3rd Ed., New York, McGraw-Hill. 2000.
- [Law07] A.M. Law: *Simulation Modeling and Analysis*. 4th Ed., New York, McGraw-Hill. 2007.
- [LaWi07] E.K. Lada und R.R. Wilson: *SBatch: A spaced batch means procedure for Simulation Analysis*. In: S.G: Henderson, B. Biller; M.-H. Hsieh, J. Shortle, J.D. Tew, und R.R. Barton (Hrsg.): *Proceedings of the 200 Winter Simulation Conference*. 2007.
- [Lieb95] F. Liebl: *Simulation*. 2. Auflage. München, Wien. Oldenbourg. 1995.
- [LiHa02] J.R. Linton und C.M. Harmonosky: *A comparison of selective initialization bias elimination methods*. In: E. Yücesan, C.-H. Chen, J.L. Snowdon und J.M. Charnes (Hrsg.) *Proceedings of the 2002 Winter Simulation Conference*. 2002.
- [MaKo93] X. Ma und A.K. Kochhar. *A Comparison Study of Two Tests for Detecting Initialization Bias in Simulation Output*. *Simulation*. 62. S. 94-101. 1993.
- [MACS94] G.T. Mackulac, J.K. Cochran und P.A. Savory: *Ascertaining important Features for industrial Simulation Environments*. *Simulation*. 63:4. S. 211-221. 1994.
- [MaIn04] P.S. Mahajan und R.G. Ingalls: *Evaluation of Methods used to detect warm-up Period in steady state Simulation*. In: R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters (Hrsg.): *Proceedings of the 2004 Winter Simulation Conference*. 2004.
- [Manu08] ManuBuild Consortium: *Deliverable 4.4-4 Flexible, mobile and autonomous factory prototype*. ManuBuild Project. 2008.
- [Manu11] ManuBuild Consortium: *Public Website*. ManuBuild Project. 2011. <http://www.manubuild.org/> am 5.1.2011.
- [MePa89] J.M. Mellichamp und Y.H. Park: *A Statistical Expert System For Simulation Analysis*. *Simulation*. 52:4. S. 134-139, 1989.
- [MeUl49] N. Metropolis und S. Ulam: *The Monte Carlo Method*. *Journal of the American Statistical Association*, Vol. 44, No. 247, S. 335-341. 1949.
- [MiCo11] Microsoft Corporation: *Microsoft Excel 2010*. <http://office.microsoft.com/de-de/excel/> am 11.1.2011.
- [Mill81] R.G. Miller: *Simultaneous Statistical Inference*. 2nd Ed. New York u.a. Springer. 1981.
- [Mitt95] J. Mittelstraß u.a. (Hrsg.): *Enzyklopädie Philosophie und Wissenschaftstheorie*. Band 2. Stuttgart. Metzler. 1995.

- [MoOG93] L. Mollamustafaoglu, G. Gurkan, und A.Y. Ozge: *Object-Oriented Design of Output Analysis Tool for Simulation Languages*. Simulation 60:1. S. 6-16. 1993.
- [MTYX10] A.C. Mokashi, J.J. Tejada, S. Yousefi, T. Xu, J.R. Wilson, A. Tafazzoli und N.M. Steiger: *Performance Comparison of MSER-5 and N-SKART on the Simulation Start-Up Problem*. In: B. Johansson, S. Jain, J. Montoya-Torres, J. Hukan und E. Yücesan (Hrsg.): *Proceedings of the 2010 Winter Simulation Conference*. 2010.
- [NAG06] The Numerical Algorithms Group Limited: *NAG Library Manual*. 2006. <http://www.nag.co.uk> am 2.12. 2008.
- [Naka08] M.K. Nakayama: *Statistical Analysis of Simulation Output*. In: S.J. Mason, R.R. Hill; L. Mönch, O. Rose, T. Jefferson und J.W. Fowler (Hrsg.): *Proceedings of the 2008 Winter Simulation Conference*. 2008.
- [Nels01] B.L. Nelson: *Statistical Analysis of Simulation Results*. In: G. Salvendy (Hrsg.) *Handbook of Industrial Engineering*. 3rd Ed. Wiley, 2001.
- [NiHP99] J. Nikoukaran, V. Hlupic und R.J. Paul: *A hierarchical framework for evaluating simulation software*. In: *Simulation Practice and Theory*. 7. S. 219-231. 1999.
- [PaSc10] R. Pasupathy und B. Schmeiser: *The initial ransient in steady-state point Estimation: Contexts, a Bibliography, the MSE Criterion, and the MSER Statistic*. In: B. Johansson, S. Jain, J. Montoya-Torres, J. Hukan und E. Yücesan (Hrsg.): *Proceedings of the 2010 Winter Simulation Conference*. 2010.
- [ProM10] ProModel Corporation: *Modelling and Simulation Software Products*. Utah. 2010. <http://www.promodel.com/products/> am 24.10. 2010.
- [Prov05] Provalis Research: *Simstat v2.5 Statistical Analysis and Simulation Software*. Montreal. 2005. <http://www.provalisresearch.com/simstat/simstw.html> am 25.1.2008.
- [RaKN88] V. Ramachandran, D.L. Kimbler und G. Naadimuthu: *Expert postprocessor for simulation output analysis*. *Computers Industrial Engineering*. 15.1-4. S. 98-103. 1988.
- [RaSW08] M. Rabe, S. Spiekermann und S. Wenzel: *Verifikation und Validierung für die Simulation in Produktion und Logistik*. Berlin u.a. Springer. 2008.
- [Rass97] T. Raßmann: *Materialflußuntersuchung einer Motorenproduktion durch ein ereignisdiskretes, stochastisches Simulationsmodell*. Studienarbeit. Universität Stuttgart IAT. 1997.
- [Rey83] W.J.J. Rey: *Introduction to Robust and Quasi-Robust Statistical Methods*. Berlin u.a. Springer. 1983.
- [Rinn08] H. Rinne: *Taschenbuch der Statistik*. 4. Aufl. Frankfurt. Harry Deutsch. 2008.

- [Rinn78] Y. Rinott: *On two-stage selection procedures and related probability inequalities*. Communications in Statistics A7. S. 799-811. 1978.
- [Robi02] S. Robinson: *A statistical Process Control Approach for estimating the Warm-up Period*. In: E. Yücesan, C.-H. Chen, J.L. Snowdon und J.M. Charnes (Hrsg.) Proceedings of the 2002 Winter Simulation Conference. 2002.
- [Ross02] S.M. Ross: *Simulation*. 3rd Ed. San Diego u.a. Academic Press. 2002.
- [RPro11] R Project: *The R Project for Statistical Computing*. <http://www.r-project.org/am> 11.1.2011.
- [SaHe06] L. Sachs und J. Hedderich: *Angewandte Statistik*. 12. Aufl. Berlin u.a. Springer. 2006.
- [SASI10a] SAS Institute: *SAS/STAT(R) 9.22 User's Guide*. Cary, NC. 2010. <http://support.sas.com/documentation> am 25.8.2010.
- [SASI10b] SAS Institute: *SAS/STAT(R) 9.22 User's Guide*. Cary, NC. 2010. http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#/documentation/cdl/en/statug/63347/HTML/default/statug_glm_a0000000867.htm am 25.8.2010.
- [SCHH94] P.J. Sanchez, F. Chance, K. Healy, J. Henriksen, W.D. Kelton und S. Vincent: *Simulation Statistical Software: An introspective Appraisal*. In: Proceedings of the 1994 Winter Simulation Conference. 1994.
- [Sche86] M. Scheifele: *A Postprocessor for the Statistical Analysis of the Simulation Output*. In: G.C. Vansteenkiste, E.J.H. Herkhoffs und L. Dekker (Hrsg.): Proceedings of the 2nd European Simulation Congress. Antwerpen. 1986.
- [Schm82] B. Schmeiser: *Batch Size Effects in the Analysis of Simulation Output*. Operations Research. 30. S. 556-568. 1982.
- [Schm01] B. Schmeiser: *Some myths and common errors in simulation experiments*. In: B. A. Peters, J.S. Smith, D.J. Medeiros und M.W. Rohrer (Hrsg.): Proceedings of the 2001 Winter Simulation Conference. 2001.
- [Schm04] B. Schmeiser: *Simulation Output Analysis: a tutorial based on one research thread*. In: R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters (Hrsg.): Proceedings of the 2004 Winter Simulation Conference. 2004.
- [Schru82] L.W. Schruben: *Detecting initialization bias in simulation output*. Operations Research. 30. S. 569-590. 1982.
- [ScST83] L.W. Schruben, H. Singh und L. Thierney. *Optimal tests for initialization bias in simulation output*. Operations Research. Vol. 31 (6). S.1167-1178. 1983.
- [SDZ07] SimulationsDienstleistungsZentrum GmbH: *DOSIMIS-3 für MS-WINDOWS – Benutzerhandbuch*. Dortmund. 2007.

- [ScYe02] B. Schmeiser und Y. Yeh: *On choosing a single criterion for confidence interval procedures*. In: E. Yücesan, C.-H. Chen, J.L. Snowdon und J.M. Charnes (Hrsg.) Proceedings of the 2002 Winter Simulation Conference. 2002.
- [Sher98] M. Sherman: *Data-based Choice of Batch Size for Simulation Output Analysis*. Simulation. July 1998.
- [SiCo11] Simulation Councils Inc: *SIMULATION*. <http://sim.sagepub.com> am 14.4.2011.
- [Siem08] Siemens Product Lifecycle Management Software II GmbH: *Technomatix Plant Simulation 9. Version 9.0.0. Hilfemenü*. 2008.
- [Siem10] Siemens Product Lifecycle Management Software Inc.: *Plant Simulation*. 2010.
http://www.plm.automation.siemens.com/en_us/products/tecnomatix/plant_design/plant_simulation.shtml am 24.8.2010.
- [SLWA02] N.M. Steiger, E.K. Lada, J.R. Wilson, C. Alexopoulos, D. Goldsman und F. Zouaoui: *ASAP2: An improved batch means procedure for simulation output analysis*. In: E. Yücesan, C.-H. Chen, J.L. Snowdon und J.M. Charnes (Hrsg.) Proceedings of the 2002 Winter Simulation Conference. 2002.
- [Spat06] D. Spath: *Neue FuE-Konzepte in der Digitalen Produktion*. In: FtK 2006 – Fertigungstechnisches Kolloquium. Stuttgarter Impulse. Stuttgart. FtK. S. 543-555. 2006.
- [Spat08] D. Spath: *Digitale Produktion*. 3. Forum Innovationscluster Digitale Produktion. IAO Stuttgart. 2008.
- [SpLR07] D. Spath, J. Lentjes und M. Richter: *Neue Ansätze für die Integration von Konstruktion und Planung im Rahmen der digitalen Produktion*. ZWF. 102. S. 73-77. 2007.
- [SpLe06] D. Spath und J. Lentjes: *An innovation cluster to cope with the challenges of digital manufacturing*. In: Manufacturing engineering in time of information society: 1st jubilee scientific conference. Gdansk University of Technology, Faculty of Mechanical Engineering. Danzig. 2006.
- [SpLe08] D. Spath und J. Lentjes: *Digitale Produktion. Von der Entwicklungsabsicherung zum Steuerungsinstrument*. Industrie Management. 24. 2. S. 27-30. 2008.
- [SpWS10] D. Spath, F. Wagner und M. Schubert: *Lean Development – Schlanke und effiziente Produktentwicklung*. Tagungsband. Stuttgart. Fraunhofer IAO. 2010.
- [Stal94] J. Stallkamp: *Struktur- und Parametervereinfachung bei Simulationsstudien mit ereignisdiskreten, stochastischen Modellen*. Studienarbeit. Universität Stuttgart IAT. 1994.

- [Stol78] M.R. Stoline: *Tables of the Studentized Augmented Range and Applications to Problems of Multiple Comparison*. JASA Vol.73 Nr 363, Theory + Methods, S. 656-660. 1978.
- [StWi00] N.M. Steiger und J.R. Wilson. *Experimental performance evaluation of batch means procedures for simulation output analysis*. In: J.A. Joines, R.R. Barton, K. Kang und P.A. Fishwick (Hrsg.): *Proceedings of the 2000 Winter Simulation Conference*. 2000.
- [SJY03] J.R. Swisher, S.H. Jacobson und E. Yücesan: *Discrete-event simulation optimization using ranking, selection, and multiple comparison procedures: A survey*. ACM Transactions on Modeling and Computer Simulation (TOMACS), 13:2: S. 134-154. 2003.
- [TaGu01] Y.-H. Tao und S.-M. Guo: *The Design of a web-based training System for Simulation Analysis*. In: B. A. Peters, J. S. Smith, D.J. Medeiros und M.W. Rohrer (Hrsg.): *Proceedings of the 2001 Winter Simulation Conference*. 2001.
- [TaHu88] R. Taylor und R.D. Hurrion: *An expert advisor for simulation experimental design and analysis*. AI and Simulation. S. 238-244. 1988.
- [Tamh79] A.C. Tamhane: *A comparison of procedures for multiple comparisons of means with unequal variances*. JASA Vol.74 Nr 366, Theory + Methods, S. 471-480. 1979.
- [TaNe97] Y.-H. Tao und B.L. Nelson: *Computer-assisted simulation analysis*. IIE Transactions (29) S. 221-231. 1997.
- [Vass89] G. Vassilacopoulos: *Testing for initialization bias in simulation output*. Simulation. 52 (4). S. 151-153. 1989.
- [VDIF93] VDI-Gesellschaft Fördertechnik Materialfluß Logistik (Hrsg.): *Richtlinie VDI 3633, Blatt 1: Simulation von Logistik-, Materialfluß- und Produktionssystemen: Grundlagen*. Düsseldorf, 1993.
- [VDIF00] VDI-Gesellschaft Fördertechnik Materialfluß Logistik (Hrsg.): *Richtlinie VDI 3633, Blatt 1: Simulation von Logistik-, Materialfluß- und Produktionssystemen: Grundlagen*. Entwurf. Düsseldorf, 2000.
- [VDIF95] VDI-Gesellschaft Fördertechnik Materialfluß Logistik I (Hrsg.): *Richtlinie VDI 3633, Blatt 3: Simulation von Logistik-, Materialfluß- und Produktionssystemen: Experimentplanung und -auswertung*. Düsseldorf, 1993.
- [VDIF96] VDI-Gesellschaft Fördertechnik Materialfluß Logistik I (Hrsg.): *Richtlinie VDI 3633, Entwurf: Simulation von Logistik-, Materialfluß- und Produktionssystemen: Begriffsdefinitionen*. Düsseldorf, 1996.
- [VDIF09] VDI-Gesellschaft Fördertechnik Materialfluß Logistik I (Hrsg.): *Richtlinie VDI 4499, Blatt 1: Digitale Fabrik – Grundlagen*. Düsseldorf, 2009.

- [VDIE95] VDI-Gesellschaft Entwicklung Konstruktion Vertrieb (Hrsg.): *Simulation in der Praxis – Neue Produkte effizient entwickeln*. Düsseldorf, VDI, 1995.
- [VWBZ09] S. Vajna, C. Weber, H. Bley und K. Zeman: *CAX für Ingenieure*. 2. Aufl. Berlin u.a. Springer. 2009.
- [Wagn10] F. Wagner: *Einführung in die Simulationstechnik*. Unterlagen zur Vorlesung an der Universität Stuttgart. <http://www.iat.uni-stuttgart.de/lehre/lehveranstaltungen/simulationstechnik.html> am 25.8.2010.
- [WaGr91] F. Wagner und J. Grammel: *ISSTAS*. Unveröffentlichtes Arbeitspapier. Fraunhofer-Institut für Arbeitswirtschaft und Organisation – IAO. Stuttgart. 1991.
- [WaWa94] F. Wagner und J. Warschat: *IAOSSAS - Ein Werkzeug zur methodisch korrekten statistischen Auswertung von Simulationsergebnissen*. In: G. Kampe und M. Zeitz (Hrsg.): *Simulationstechnik 9. Symposium in Stuttgart Oktober 1994*. Fortschritte in der Simulationstechnik Band 9. Braunschweig, Vieweg, 1994.
- [Wall09] A.P. Waller: *Witness for Six Sigma*. Lanner Group LTD, Reddich, Worcestershire. 2009.
http://www.lanner.com/en/media/witness/pdfs/six_sigma_witness_2009.pdf am 24.8.2010.
- [WaWa96] J. Warschat und F. Wagner: *Einführung in die Simulationstechnik*. Vorlesungsskript. Stuttgart. Universität Stuttgart Institut für Arbeitswissenschaft und Technologiemanagement. 1996.
- [WhCS00] K.P. White, M.J. Cobb und S.C. Spratt: *A comparison of five steady-state truncation heuristics for simulation*. In: J.A. Joines, R.R. Barton, K. Kang und P.A. Fishwick (Hrsg.) *Proceedings of the 2000 Winter Simulation Conference*. 2000.
- [WSC10] *Winter Simulation Conference*. <http://www.wintersim.org/> am 25.8. 2010.
- [Welc83] P.D. Welch: *The statistical Analysis of Simulation Results*. In: S.S. Lavenberg (Hrsg.): *The computer Performance Modeling Handbook*. New York. Academic Press. 1983.
- [WeNo00] S. Wenzel und B. Noche: *Simulationsinstrumente in Produktion und Logistik - eine Marktübersicht*. In: K. Mertins und M. Rabe: *The new simulation in production and logistics. Prospects, views and attitudes: 9. ASIM-Fachtagung Simulation in Produktion und Logistik, Berlin, 8.-9. März 2000*. Stuttgart: Fraunhofer IRB Verlag, S. 423-432. 2000.
- [West03] E. Westkämper: *Die digitale Fabrik*. In H.-J. Bullinger (Hrsg.) u.a.: *Neue Organisationsformen im Unternehmen: ein Handbuch für das moderne Management*. Berlin u.a. Springer. 2003.

- [West09] E. Westkämper: *Digitale Fabrik – Leitthema in der Produktionstechnik*. Editorial. wt Werkstattstechnik online. 99. S.91. 2009.
- [West10] E. Westkämper: *Potentiale der digitalen Produktion*. Editorial. wt Werkstattstechnik online. 100. S.119. 2010.
- [WWCP08] S. Wenzel, M. Weiß, S. Collisi-Böhmer, H. Pitsch und O. Rose: *Qualitätskriterien für die Simulation in Produktion und Logistik*. Berlin u.a. Springer. 2008.
- [YeSc00] Y. Yeh und B. Schmeiser: *Simulation Output Analysis via Dynamic Batch Means*. In: J.A. Joines, R.R. Barton, K. Kang und P.A. Fishwick (Hrsg.): *Proceedings of the 2000 Winter Simulation Conference*. 2000.
- [YeSc01] Y. Yeh und B. Schmeiser: *On the MSE robustness of batching estimators*. In: B. A. Peters, J. S. Smith, D.J. Medeiros und M.W. Rohrer (Hrsg.): *Proceedings of the 2001 Winter Simulation Conference*. 2001.
- [Yin99] L.-M. Yin: *Analytical Tool Suite for Stochastic Simulation Models*. <http://www.nmsr.labmed.edu/~yin> am 28.7.1999.
- [Zeig84] B.P. Zeigler: *Multifaceted Modelling and Discrete Event Simulation*. London u.a. Academic Press. 1984.

12 Anhang

12.1 Methoden

In der Arbeit wurden die nachfolgend beschriebenen Methoden und Prozeduren verwendet:

Korrelogramm

Ein Korrelogramm beschreibt die Autokorrelation einer Zeitreihe, indem die Korrelationskoeffizienten in Abhängigkeit von der Zeitdifferenz dargestellt werden. Nach Rinne [Rinn08:738] ist ein Korrelogramm „*die grafische Darstellung der Autokorrelationsfunktion über die Zeitdifferenz*“.

Batch Means Methode

Die Batch-Means-Methode zur Beseitigung der Autokorrelation einer Zeitreihe ist ein Standardverfahren und wird u.a. bei Law [Law07:520] beschrieben. In Kapitel 4.3 dieser Arbeit wird das einfache Verfahren der nicht überlappenden Batch Means erläutert.

Dichteschätzer

Die Schätzung und grafische Darstellung der Verteilungsdichte einer empirischen Verteilung erfolgt mit Dichteschätzern, auch als Kerndichteschätzern bei Rinne [Rinn08:494] bezeichnet. Im Gegensatz zu einem Histogramm wird dabei die Dichte der Verteilungsfunktion stetig geschätzt.

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „density“ und dem Kern „gaussian“ verwendet.

Bonferroni Ungleichung

Die Bonferroni Ungleichung der Wahrscheinlichkeitsrechnung beschreibt die gesamte Wahrscheinlichkeit von mehreren sich nicht ausschließenden Ereignissen nach Rinne [Rinn08:182] in der Form

$$Pr(\bigcap_{i=1}^n A_i) \geq 1 - \sum_{i=1}^n Pr(A_i) \quad (12.1)$$

oder nach Kleijnen [Klei78:101] mit der individuellen und gesamten Fehlerrate

α_k und α_G

$$\alpha_G \leq \sum_{k=1}^K \alpha_k \quad (12.2)$$

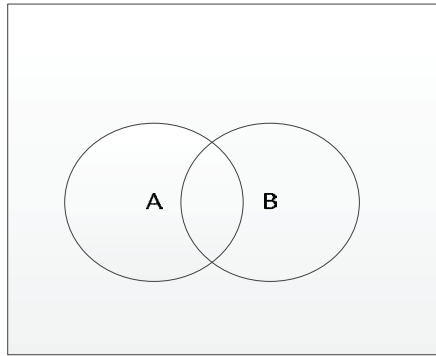


Abbildung 12.1: Venn-Diagramm zur Bonferroni-Ungleichung für „vereinbare“ Ereignisse.

Für den Fall $n = 2$ lässt sich die Ungleichung nach [SaHe06:114] und [Klei78:100] mit einem Venn-Diagramm der Wahrscheinlichkeiten wie in Abb. 12.1 erläutern.

Šidák-Ungleichung

Die Šidák-Ungleichung beschreibt nach Hochberg und Tamhane [HoTa87:95] und Rinne [Rinn08:182] etwas „schärfer“ als die Bonferroni-Ungleichung die Gesamtwahrscheinlichkeit von mehreren *unabhängigen* Ereignissen, indem die Summe durch das Produkt der Einzelwahrscheinlichkeiten ersetzt wird.

$$Pr(\cap_{i=1}^n A_i) \geq \prod_{i=1}^n Pr(A_i) \quad (12.3)$$

Diese Ungleichung kann analog zur Bonferroni-Ungleichung nach Hochberg und Tamhane [HoTa87:95] in einer Form mit der individuellen und gesamten Fehlerrate α_k und α_G dargestellt werden.

$$\alpha_k = 1 - (1 - \alpha_G)^{1/k^*} \text{ mit } k^* = \binom{k}{2} \quad (12.4)$$

Für viele praktische Anwendungen kann die Šidák-Ungleichung durch die bekanntere, einfachere und etwas plausiblere Bonferroni-Ungleichung ersetzt werden, da für die typischen Fehlerraten zwischen 0,1 und 0,01 die Ergebnisse sehr ähnlich sind.

Chi-Quadrat-Test (X^2 -Test)

Der X^2 -Test (Chi-Quadrat-Test) ist ein Anpassungstest, der in dieser Arbeit als Anpassungstest an die Normalverteilung verwendet wird (Kap. 4). Die Standardlehrbücher von Hartung [Hart02:182], Sachs und Hedderich [SaHe06:335] sowie das Taschenbuch von Rinne [Rinn08:576] beschreiben den Test ausführlich.

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „pearsons.test“ aus dem package „nortest“ verwendet.

Welch-Methode

Das intuitive und grafisch-interaktive Verfahren von Welch [Welc83] zur Bestimmung der stationären Phase ist bei Law [Law07:509ff] detailliert erläutert. Welchs Prozedur nutzt unabhängige Replikationen des Simulationsmodells und besteht aus vier Schritten.

1. Es werden n ($n \geq 5$) Replikationen mit jeweils m Beobachtungen durchgeführt. Dabei ist Y_{ji} die i -te Beobachtung der j -ten Replikation ($j = 1, 2, \dots, n; i = 1, 2, \dots, m$).
2. Der Prozess wird mit $\bar{Y}_i = \sum_{j=1}^n Y_{ji}/n$ ($i = 1, 2, \dots, m$) über die n Replikationen gemittelt und hat den selben Erwartungswert $E(\bar{Y}_i) = E(Y_i)$ und nur noch die Varianz $Var(\bar{Y}_i) = Var(Y_i)/n$.
3. Um die „hochfrequenten“ Schwingungen zu glätten, wird ein „gleitender Durchschnitt“ $\bar{Y}_i(w)$ mit der Breite w und $w \leq \lfloor m/4 \rfloor$, so dass

$$\bar{Y}_i(w) = \begin{cases} \frac{\sum_{s=-w}^w \bar{Y}_{i+s}}{2w+1} & \text{für } i = w+1, \dots, m-w \\ \frac{\sum_{s=-(i-1)}^{i-1} \bar{Y}_{i+s}}{2i-1} & \text{für } i = 1, \dots, w \end{cases} \quad (12.5)$$

gilt.

4. Es wird $\bar{Y}_i(w)$ für $i = 1, 2, \dots, m-w$ grafisch dargestellt und die Länge l der transienten Phase so gewählt, dass l der Wert jenseits des Zeitindex i ist, wo die Zeitreihe $\bar{Y}_1(w), \bar{Y}_2(w), \dots$ zu konvergieren scheint.

Die Welch-Methode ist auch in Abb. 12.2 als UML-Diagramm dargestellt.

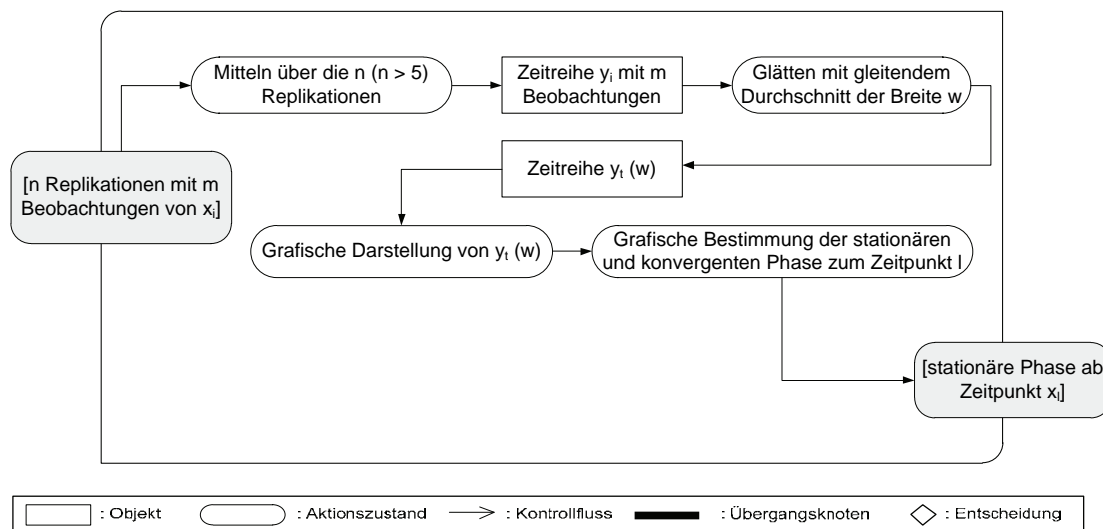


Abbildung 12.2 : UML-Diagramm der Welch-Methode zur Bestimmung der stationären Phase.

Law [Law07:511] empfiehlt fünf oder zehn Replikationen mit möglichst vielen Beobachtungen zu simulieren. Die Breite w des „gleitenden Durchschnitts“ sollte bewusst (wie bei einem Histogramm) variiert werden, um einen möglichst „geglätteten“ Verlauf zu erreichen. Falls in Schritt 3 kein geeigneter Wert für w gefunden werden kann, können fünf bis zehn zusätzliche Replikationen hilfreich sein.

Fischer-Methode

Die Fischer-Methode [Fisc89] als heuristische Prozedur zum Erkennen der stationären Phase ist in dem (unveröffentlichten) ISSTAS Projektbericht [IAO98:40] dokumentiert. Die iterative Methode besteht aus mehreren Schritten und ist in dem UML-Diagramm 12.3 dargestellt.

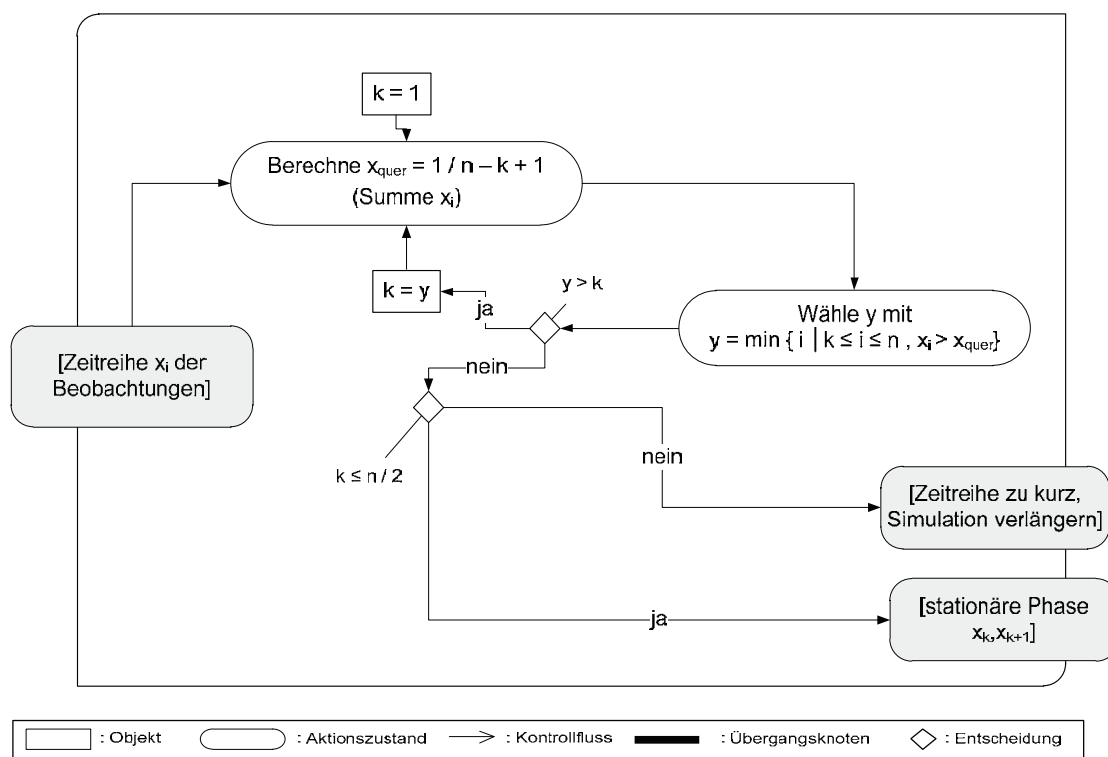


Abbildung 12.3: UML-Diagramm der Fischer-Prozedur zur Bestimmung der stationären Phase, vor allem für ein Anlaufen aus einem leeren Zustand.

Die zu untersuchende Zeitreihe x_i beinhaltet n Beobachtungen.

1. Der Startpunkt $k= 1$ wird festgelegt, d.h. man beginnt beim ersten Wert der Zeitreihe.
2. Das arithmetische Mittel \bar{x} wird aus den Werten, die zu Zeitpunkten später als k beobachtet wurden, berechnet:

$$\bar{x} = 1/(n - k + 1) \sum_{i=k}^n x_i \quad (12.6)$$

3. y wird aus den Beobachtungen ermittelt, welche in das arithmetische Mittel bei (2) eingehen, die größer als dieses Mittel sind.

$$y = \min\{i | k \leq i \leq n, x_i \geq \bar{x}\} \quad (12.7)$$

4. Ist der Zeitpunkt y der zu der bei (3) ermittelten Beobachtung größer als k , dann setze k gleich diesem Zeitpunkt und gehe zu (2). D.h., ist nicht schon der erste Wert, der in das arithmetische Mittel eingeht, größer als dieses arithmetische Mittel, dann beginnt man zu einem neuen Zeitpunkt noch einmal bei (2). Gilt obiges nicht, gehe zu (5).
5. Entweder endet die erkannte transiente Phase zum Zeitpunkt $k-1$, falls k kleiner oder gleich $n/2$ ist oder die betrachtete Zeitreihe war zu kurz, falls k größer als $n/2$ gilt.

Diese heuristische Prozedur ist speziell für das Anfahren aus leerem Zustand geeignet.

Run-Test

Der Run-Test (auch Runs-Test oder Wald-Wolfowitz-Test) ist ein parameterfreier Test auf Unabhängigkeit und Zufälligkeit der Stichproben einer Zeitreihe und bei Rinne [Rinn08:563] beschrieben. Die Nullhypothese ist dabei die Unabhängigkeit der Beobachtungen. Beim Run-Test einer Zeitreihe werden alle Beobachtungen, die größer als der Median sind, mit „+“, Beobachtungen, die kleiner sind, mit „-“ gekennzeichnet.

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „runs.test“ aus dem package „lawstat“ verwendet.

Cochran-Test

Der Cochran-Test bzw. der Cochran C-Test ist ein einfacher und bekannter Test für Varianzgleichheit, auch als Homoskedastizität bezeichnet. Nach Sachs und Hedderich [SaHe06:420] kann er auch bei leicht schiefen Verteilungen eingesetzt werden. Die Testprüfgröße C wird nach 12.8 aus der größten Varianz, geteilt durch die Summe aller k Varianzen, berechnet und mit den tabellierten Signifikanzschranken verglichen.

$$C = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_k^2} \quad (12.8)$$

Der Cochran C-Test wird auch im ISO Standard 5725 [ISO01:8] beschrieben.

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „cochran.test“ aus dem package „outlier“ verwendet.

Test von Vassiliacopoulos

Der Test von Vassiliacopoulos [Vass89] beruht auf einem verteilungsfreien Konzept mit einer standardisierten Reihenfolge von linearen Kombinationen der Ränge der Simulationsbeobachtungen. Da der Artikel von Vassiliacopoulos nicht überall verfügbar ist, wird die Test-Prozedur hier (wie im ISSTAS Projektbericht [IAO98:42]) ausführlicher beschrieben.

Es werden die Ränge der Beobachtungen berechnet und durch Aufsummieren sowie Abzug des Erwartungswerts eine Testgröße gebildet, welche den Erwartungswert Null hat und nach der Division durch die Varianz sich bei verschiedenen Zeitpunkten wie eine „Brownsche Brücke“ verhält. Damit konstruiert Vassiliacopoulos einen Test, dessen Verteilung approximativ bekannt ist.

Mit den Rängen der Beobachtungen des stochastischen Prozesses $\{x_k\}$ wird ein neuer stochastischer Prozess $\{U_N(k)\}$, bezogen auf die Anzahl der n Ränge, wie folgt definiert:

$$U_N(k) = \begin{cases} 2W_k - k \cdot (N+1), & k = 1, 2, \dots, N \\ 0, & k = 0 \end{cases} \quad (12.9)$$

mit

$$W_k = \sum_{j=1}^k r_j \quad (12.10)$$

und r_j als Rang der j-ten Beobachtung x_j .

Im Weiteren wird die Größe

$$C_N = \max_{1 \leq k \leq N} |U_N(k)| \quad (12.11)$$

benötigt. Die Bandbreite des stochastischen Prozesses $\{U_N(k)\}$ wird bzgl. der Standardabweichung normiert und die Simulationszeit auf das Einheitsintervall skaliert. Damit erhält man den Prozess

$$B_t = \sqrt{\frac{3}{N^2 \cdot (N+1)}} \cdot U_N(N \cdot t) \quad \text{mit } 0 \leq t \leq 1 \quad (12.12)$$

dessen Grenzverteilung für $N \rightarrow \infty$ bekannt ist. Die Grenzverteilung für (12.11) mit der Normierung in (12.12) stellt sich wie folgt dar:

$$\Pr\left(\max_{0 \leq t \leq 1} |B_t| \leq z\right) = 1 + 2 \sum_{j=1}^{\infty} (-1)^j \cdot \exp(-2j^2 z^2) \quad (12.13)$$

Damit ist auch die Grenzverteilung von C_N bekannt und für deren Signifikanzwahrscheinlichkeit unter der Realisierung c von C_N gilt approximativ:

$$\hat{\alpha} \cong 2 \exp\left\{\frac{-6c^2}{N^3 + N^2}\right\} \quad (12.14)$$

Die Testprozedur läuft dann in den folgenden drei Schritten ab:

1. Ermittle die Ränge r_k der Beobachtungen und berechne $\{U_N(k); k = 1, 2, \dots, N\}$.
2. Bestimme $c = \max_k |U_N(k)|$ und das zugehörige Signifikanzniveau $\hat{\alpha}$.
3. Lehne die Nullhypothese „keine transiente Phase“ ab, falls $\hat{\alpha}$ kleiner als ein vorgegebenes Signifikanzniveau α ist.

Das UML-Diagramm der Testprozedur von Vassiliacopoulos ist in Abb. 12.4 dargestellt.

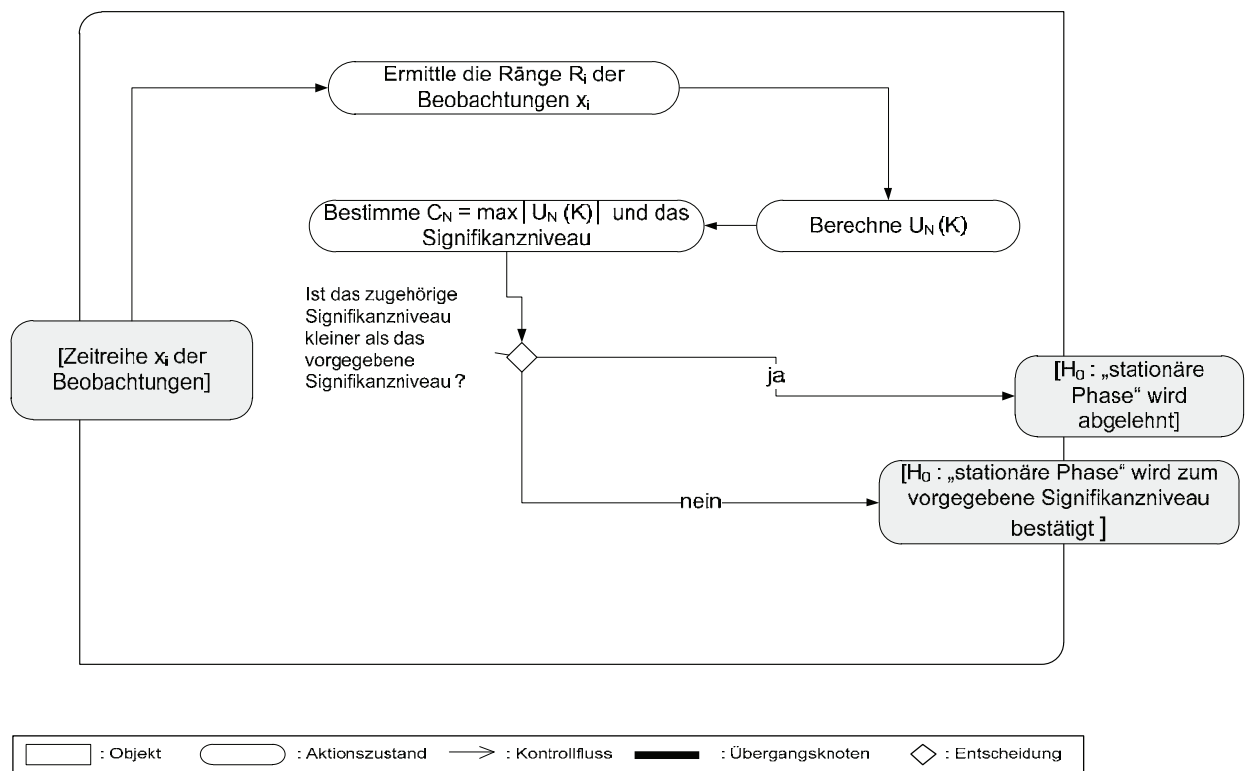


Abbildung 12.4: UML-Diagramm des Vassiliacopoulos-Tests zur Bestimmung der stationären Phase.

t-Test und t-Statistik

Der t-Test und die t-Statistik sind ein bekanntes Standardverfahren der Statistik. Die Variante des Zweistichproben-t-Tests ist bei Rinne [Rinn08:527], Hartung [Hart02:510] sowie Sachs und Hedderich [SaHe06:269] und auch in Verbindung mit der Bonferroni-Ungleichung bei Miller [Mill82:67] beschrieben. Beim „Behrens-Fischer-Problem“ der unbekanntenen und ungleichen Varianzen werden die Freiheitsgrade nach der Methode von Welch (Formel 12.17) approximiert.

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „t.test“ verwendet.

Johnsons t-Statistik

Um die t-Statistik auch bei schiefen Verteilungen benutzen zu können, hat Johnson [John78] die Differenz des geschätzten und wahren Mittelwerts $\bar{x} - \mu$ nach Chen [ChLi95] durch „die ersten Terme der inversen Cornish-Fisher Erweiterung“ ersetzt, um damit die Quantile einer schiefen Verteilung mit Hilfe der ersten drei Momente abschätzen zu können.

$$\tilde{t}_{n-1} = \left[(\bar{x} - \mu) + \frac{\hat{\mu}_3}{6s^2n} + \frac{\hat{\mu}_3}{3s^2} (\bar{x} - \mu)^2 \right] \left(\frac{s^2}{n} \right)^{-\frac{1}{2}} \quad (12.15)$$

Bei einer praktischen Anwendung wird die Johnson-t-Statistik ohne den quadratischen Term verwendet

$$\tilde{t}_{n-1} = \left[(\bar{x} - \mu) + \frac{\hat{\mu}_3}{6s^2n} \right] \left(\frac{s^2}{n} \right)^{-\frac{1}{2}} \quad (12.16)$$

und die t-Statistik „nur“ mit dem additiven Teil $\hat{\mu}_3/6s^2n$ ergänzt.

Tukey-Test und Tukey-Prozedur

Der Tukey-Test und die Tukey-Prozedur sind bekannte Standardverfahren der Statistik und bei Hartung [Hart02:616] sowie bei Sachs und Hedderich [SaHe06:436] beschrieben. Dieser auch als Tukey-Kramer Verfahren oder Tukey HSD (honestly significant differences) bezeichnete Test basiert auf der Verteilung der Studentisierten Extremwerte bzw. Studentisierten Spannweite (Studentized Range bzw. SR-Verteilung).

In dieser Arbeit wurde die Softwareumgebung R mit den Funktionen „TukeyHSD“ und „TK.test“ aus dem R-Package „DTK“ verwendet.

Games-Howell-Test

Der GH-Test von Games und Howell basiert auf dem Tukey-Test und erlaubt Varianzheterogenität. Der Test ist bei Hochberg und Tamhane [HoTa87:180]

und Sachs und Hedderich [SaHe06:439] detailliert beschrieben. Als Modifikation zum Tukey-Test wird die Zahl der Freiheitsgrade ν mit der Methode von Welch nach Formel 12.17 (zur ganzen Zahl gerundet) geschätzt wird.

$$\nu = \frac{\left[\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j} \right]^2}{\frac{s_i^4}{[n_i^2(n_i-1)]} + \frac{s_j^4}{[n_j^2(n_j-1)]}} \quad (12.17)$$

Während beim Tukey-Test die SR-Verteilung benutzt wird, werden beim GH-Test die Schranken der SAR-Verteilung („Studentized Augmented Range“) verwendet [SaHe06:439]. Der GH-Test ist bei großen Varianzunterschieden besser geeignet als der Dunnett-C-Test.

In dieser Arbeit wurde das Statistiksoftwarepaket SPSS mit der Funktion „Games-Howell“ in dem Menü „ANOVA: Analyze: General Linear Model: Univariate: PostHoc: verwendet. Falls die Statistiksoftware SPSS nicht verfügbar ist, kann der Test auch bei einer größeren Anzahl von Stichproben durch den Dunnett-C Test ersetzt oder „manuell“ mit der Tabelle der SAR-Schranken $q'_{\nu,k,1-\alpha}$ von Sachs und Hedderich [SaHe06:439] (aus Stoline [Sto178]) nach der dortigen Formel (12.18) berechnet werden.

$$\bar{x}_i - \bar{x}_j \mp \frac{q'_{\nu,k,1-\alpha}}{\sqrt{2}} \cdot \sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}} \quad (1 \leq i < j \leq k) \quad (12.18)$$

Dunnett-C-Test

Der Dunnett-C-Test [Dunn80] ist analog zum GH-Test, benutzt aber anstatt der Welch-Methode 10.5 die Prozedur von Cochran (als näherungsweise Lösung des „Behrens-Fischer-Problems“) zur Bestimmung der Freiheitsgrade. Der Dunnett-C-Test ist bei Sachs und Hedderich [SaHe06:440] beschrieben. Hochberg und Tamhane [HoTa87:180] weisen darauf hin, dass sich der Dunnett-C-Test und der GH-Test für wachsende Freiheitsgrade einander annähern.

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „DTK.test“ aus dem „R-Package „DTK“ verwendet.

Dunnett-T3 Test

Der Dunnett-T3-Test [Dunn80] basiert ebenfalls auf dem Tukey-Test, benutzt aber die „Studentized Maximum-Modulus“-Verteilung auf Basis der Kimball-Ungleichung (welche etwas „schärfer“ als die Šidák-Ungleichung ist) [HoTa87:189]. Nach Dunnett [Dunn80] ist der T3-Test vor allem bei kleinen

Freiheitsgraden aus wenigen Stichproben weniger konservativ als der C-Test und der GH-Test.

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „DTK.test“ aus dem „R-Package „DTK“ verwendet.

Wilcoxon-Rangsummentest

Der Wilcoxon-Rangsummentest ist nach Rinne [Rinn08:543] ein verteilungsfreier, d.h. nichtparametrischer Rangtest auf Lagealternativen bzw. nach Hartung [Hart02:513ff] ein Test für Lokationsvergleiche der Verteilungen. Es wird dabei gegen die Nullhypothese der gleichen Verteilungen getestet. Der Wilcoxon-Rangsummentest ist über die Beziehung

$$W = U + \frac{n_x(n_x+1)}{2} \quad (12.19)$$

äquivalent zum U-Test von Mann-Whitney. Der U-Test kann als das parameterfreie Gegenstück zum t-Test betrachtet werden [SaHe06:391]

In dieser Arbeit wurde die Softwareumgebung R mit der Funktion „wilcoxon.test“ verwendet.

12.2 Simulationsergebnisse

In der Arbeit wurden in Kapitel 6 bei der Anwendung des integrierten Verfahrens folgende Datensätze verwendet.

Eine Simulation der Produktion von Fertigbauteilen „Manubuild“ mit der Kennzahl **Durchschnittliche Durchlaufzeiten** (in Minuten). Die 20 bzw. 40 Replikationen des Simulationsmodells ergaben die folgenden Beobachtungen:

20 Stichproben	40 Stichproben
380,54	380,54
434,67	434,67
379,81	379,81
380,06	380,06
428,53	428,53
380,49	380,49
409,69	409,69
415,79	415,79
380,01	380,01
381,38	381,38
380,42	380,42
428,06	428,06
380,63	380,63
381,24	381,24
379,27	379,27
448,25	448,25
383,14	383,14
383,85	383,85
420,02	420,02
381,07	381,07
	390,08
	380,85
	380,15
	380,43
	406,75

	454,42
	379,77
	402,95
	393,06
	403,20
	421,58
	391,42
	394,46
	380,21
	393,59
	397,81
	396,93
	388,40
	380,69
	381,44

Eine Simulation der Schuhproduktion „CEC-made-shoe“ mit der Kennzahl **Durchsatz pro Woche**. Die vier alternative Simulationsmodelle (Szenarien) wurden mit jeweils 40 Replikationen simuliert und ergaben die folgenden Beobachtungen:

Ist	Szenario1	Szenario2	Szenario3
2345	2351	2369	2354
2358	2346	2395	2393
2358	2345	2397	2387
2353	2345	2383	2393
2353	2344	2396	2382
2357	2351	2396	2397
2356	2352	2396	2390
2354	2340	2397	2390
2360	2347	2397	2391
2350	2344	2396	2380
2358	2342	2361	2392
2351	2350	2395	2393

2351	2344	2397	2385
2353	2347	2396	2392
2359	2350	2395	2394
2347	2348	2395	2372
2361	2345	2395	2394
2351	2344	2396	2395
2354	2347	2396	2388
2352	2347	2394	2396
2360	2345	2396	2395
2356	2335	2395	2385
2353	2349	2397	2394
2353	2354	2339	2394
2352	2351	2394	2348
2353	2350	2392	2390
2356	2345	2396	2395
2358	2344	2395	2394
2358	2334	2396	2393
2345	2345	2396	2388
2354	2344	2380	2385
2351	2345	2396	2390
2347	2354	2396	2393
2357	2349	2396	2394
2351	2354	2396	2387
2353	2342	2396	2392
2356	2338	2397	2389
2353	2348	2396	2389
2352	2354	2395	2395
2349	2353	2397	2393