

Gewinnung und Aufbereitung von Simulationsdaten zu Vergleichszwecken

M. Ritzschke, T. Wiedemann

ritzschk@informatik.hu-berlin.de; wiedem@cs.tu-berlin.de

Humboldt-Universität Berlin, Unter den Linden 6, D-10099 Berlin; TU Berlin

Zusammenfassung

Entscheidend für den Gesamterfolg von Simulationsuntersuchungen ist letztlich die richtige Bewertung und Interpretation der Ergebnisse. Der Beitrag stellt für die modell- und auch simulatorübergreifende Auswertung von Simulationsergebnissen zwei mögliche Vorgehensweisen gegenüber. Insbesondere durch eine Anlehnung an die Data-Warehouse-Technologie kann in Zukunft eine wesentlich größere Bandbreite an Auswertungs- und Vergleichsmöglichkeiten erreicht werden.

1 Einführung

Bei der Durchführung von Simulationsstudien mit großem Experimentieranteil, z.B. im Rahmen der Systemauslegung und Planung von Fertigungssystemen, beim Vergleich von Simulatoren anhand der erzielten Simulationsergebnisse [1] oder bei der heuristischen Suche nach einer „besten“ Lösung unter gegebenen Randbedingungen müssen die anfallenden Rohdaten zu Vergleichszwecken aufbereitet werden. Dazu gibt es mehrere Möglichkeiten:

- Traditionell wird die simulatoreigene Standardausgabe in Form von Reports benutzt. Die in den Standardausgaben der Simulatoren implementierten statistischen Verfahren sind naturgemäß unterschiedlich stark ausgebaut. So kann das Simulationssystem Taylor eine Zustandsgrafik der Systemelemente oder bei stochastischen Eingangsgrößen eine automatische Konfidenzintervallberechnung generieren. Beide Ausgabemöglichkeiten sind bei anderen Simulatoren nur selten vorhanden. Zur Auswertung der Reports im Sinne eines Vergleichs sind die Ergebnisse in der Regel manuell weiter zu verarbeiten, da modell- oder simulatorübergreifende Vergleichsmöglichkeiten fehlen. Selbst bei einer Nutzung von Statistiksoftware oder Tabellenkalkulationssystemen und der Automatisierung von Handlungsabläufen mit Hilfe von Makros bleibt dies jedoch ein sehr aufwendiges Unterfangen (vgl. [3]).
- Eine zweite Möglichkeit besteht darin, nur die interessierenden Daten zusammengefaßt in ein externes File zu schreiben und mit Hilfe eines nutzerspezifischen Auswertungstools in der gewünschten Form aufzubereiten. Dieser Weg führt zu erheblichen Einsparungseffekten.

- Bei einer dritten Vorgehensweise, welche im Zusammenhang mit der verteilten Simulation und der Schaffung entsprechender Schnittstellen immer mehr in den Vordergrund tritt, wird die Auswertung vollständig vom Simulator getrennt. Dabei werden bei der Simulation erzeugte Trace-Files extern weiterverarbeitet. Solche Trace-Files können im Extremfall alle bei der Simulation auftretenden Ereignisse enthalten und daher sehr umfangreich sein. Sie erlauben dann aber beliebige Auswertungen, die weit über die Möglichkeiten der ersten beiden Auswertungsvarianten hinausgehen (Zeitreihendarstellungen, Trendfunktionen etc.). In [2] findet man eine ausführliche Darstellung verschiedener Varianten der Erzeugung von Trace-Files, insbesondere auch Überlegungen zur automatisierten Durchführung der erforderlichen Modellmodifikation.

Im Beitrag sollen die zweite und dritte Variante an Beispielen erläutert und die Vor- und Nachteile gegenübergestellt werden.

2 Excelbasierte Auswertung und Vergleich von Simulationsergebnissen

2.1 Datenbeschaffung

Von den im Rahmen einer Simulationsstudie interessierenden Kenngrößen erfassen die Simulatoren in der Regel standardmäßig nur einen Teil, andere müssen durch spezielle Anweisungen, zusätzliche Modellelemente oder gesonderte Berechnungen ermittelt werden. Beispielsweise werden bei der Simulation typischer fertigungsorganisatorischer Problemstellungen (vgl. Referenzmodelle in [1]) die Kenngrößen *Maschinen- und Arbeitsplatzauslastung*, *mittlere* und *maximale Pufferfüllung*, *Durchsatz* von allen betrachteten Simulatoren bereitgestellt, nicht aber *Zustandswahrscheinlichkeiten*, *Durchlaufzeiten*, *mittlerer Bestand* und *Bestandsentwicklungen* im System (oder in Systembereichen).

Die zur Auswertung bestimmten Rohdaten werden, ergänzt durch einen Kopfteil mit Verwaltungsangaben (Anzahl der Simulationsläufe, Art der Daten, Datum etc.), in eine Textdatei ausgegeben,

2.2 Datenauswertung mit SIDAT1

Mittels des in Visual Basic programmierten Excel-Tools SIDAT 1 erfolgt die Aufbereitung und Auswertung der Rohdaten in 3 Schritten (Bild 1):

Datenkonvertierung

Obwohl die Datenausgabe in ein Textfile vom Nutzer speziell programmiert werden muß, weichen die Ausgabeformate trotzdem noch in Details voneinander ab, so daß eine

Konvertierung in ein einheitliches internes Datenformat erforderlich ist. Bei Aufruf der Konvertierungsmodule werden weitere benötigte Informationen abgefragt, wie

- Anzahl, Bezeichnung und Pfad zu den Rohdatenfiles
- Anzahl der auszuwertenden Simulationsläufe
- gewünschtes Konfidenzniveau für die Konfidenzintervallberechnung
- Diagrammbeschriftungen

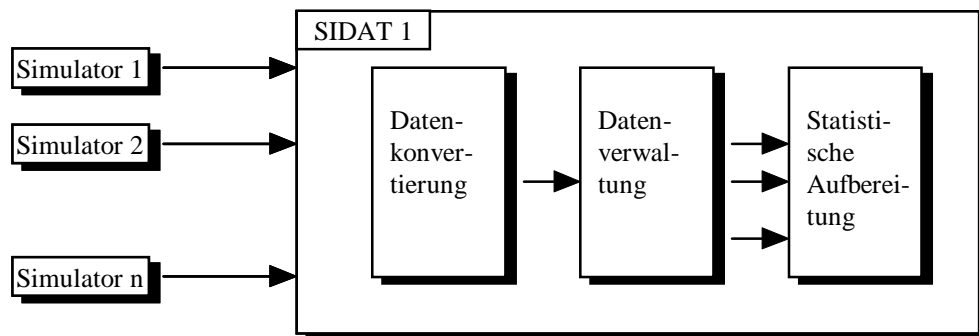


Bild 1: Datenauswertung mit Excel-Tool SIDAT

Datenverwaltung

Die konvertierten Daten und die Zusatzinformationen werden in speziellen Projektdateien gespeichert, so daß sie auch nachträglich für weitere Auswertungen zur Verfügung stehen.

Statistische Auswertung und Vergleich

Werden deterministische Prozeßabläufe simuliert oder stochastische Prozesse mit Hilfe eines langen Simulationslaufes erfaßt, können die Daten sofort in Tabellen und Diagrammen dargestellt werden (Einzelwertdarstellung). Für den Fall von Wiederholungsläufen mit unterschiedlichen Zufallszahlenreihen sind spezielle Module zur Mittelwertbildung und Konfidenzintervallberechnung sowie darauf basierender grafischer Darstellung implementiert.

Auch bei Wiederholungsläufen kann die Einzelwertdarstellung von Nutzen sein. So enthält Bild 2 die ermittelten Produktionsraten von 30 Simulationsläufen mit dem Referenzmodell „Reihenfertigung“ [1] für verschiedene Pufferkapazitäten. Da in diesem Modell von zufälligen Bearbeitungszeiten ausgegangen wird, führt jeder Lauf erwartungsgemäß zu einem anderen Durchsatz, die Produktionsrate streut. Erst die Verbindung der zum jeweiligen Lauf gehörenden Ergebnisse miteinander machte deutlich (Bild 3), daß beim verwendeten Simulationsmodell keine Unabhängigkeit der Startzahlen der Zufallszahlengeneratoren bei den einzelnen Läufen vorhanden war. Bild 4 zeigt dagegen das Ergebnis eines echten (pseudo-)zufälligen Prozessablaufes auf Basis des korrigierten Simulationsmodells.

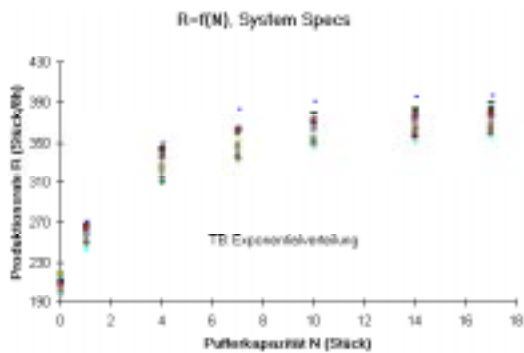


Bild 2: Punktdarstellung mit SIDAT

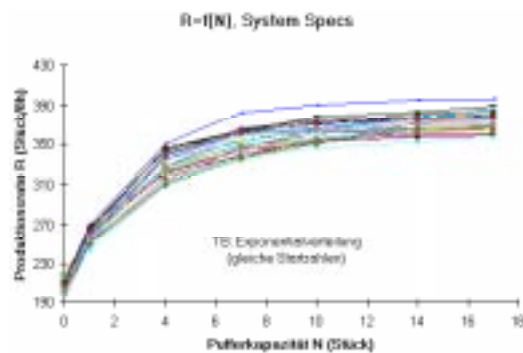


Bild 3: Liniendiagramm

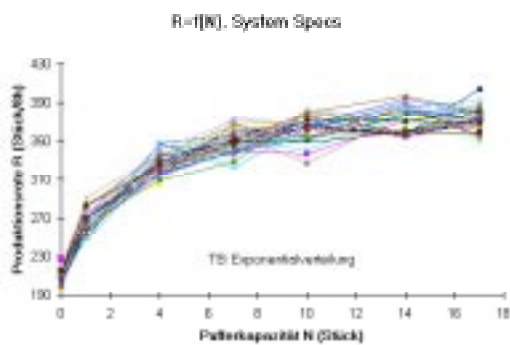


Bild 4: Korrekter Zufallsprozeß

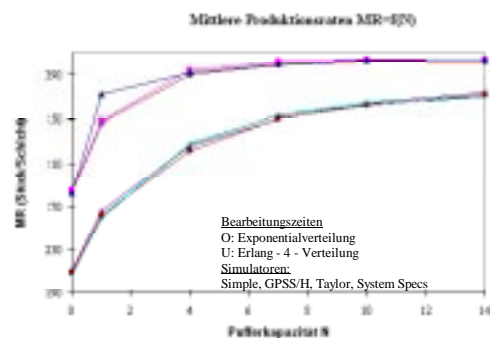


Bild 5: Überlagerungsdiagramm

Alle gewonnenen Diagramme können in einem letzten Schritt miteinander überlagert dargestellt werden. Das ermöglicht vielfältige Vergleiche z.B. auch die Gegenüberstellung von Ergebnissen, die mit unterschiedlichen Simulatoren gewonnen wurden (Bild 5).

3 Grundprinzipien einer vollständigen externen Simulationsauswertung

3.1 Datenbeschaffung

Bei einer vollständig externen Simulationsauswertung müssen alle relevanten Einzelereignisse ausgewertet werden. Der Zeitpunkt und die Art des Ereignisses kann wiederum über Tracefiles oder andere Standardausgaben der Simulationssysteme ermittelt werden. Praktische Versuche zeigten, daß selbst Standardtraceprotokolle kommerzieller Systeme nicht immer alle Informationen bereithalten. In solchen Fällen sind nutzerspezifische Ausgaben notwendig.

Basisparameter für die spätere Auswertung sind der Ereigniszeitpunkt, die Art des Ereignisses und die beteiligten Objekte (z.B. Maschinen und Produkte). Bei der Erfassung werden auch die Typen der beteiligten Objekte unterschieden, so daß spätere Auswertungen auch produktgruppenbezogen sein können. In den Standardausgaben von Simulationssystemen lassen sich derartige Zuordnungen und Gruppierungen meist nicht durchführen.

3.2 Datenaufbereitung unter Verwendung von Datenbanktechnologien

In Abhängigkeit von der Dauer und Anzahl der Simulationsläufe können einige zehn- bis hunderttausend Datensätze zu einem Gesamtexperiment anfallen, welche nur in einer Datenbank sinnvoll verwaltet werden können. Das Einlesen der Daten in die realisierte Analysedatenbank SimMiner [4] muß zur Zeit noch über Zwischenformate wie Textdateien oder Excellisten erfolgen, da die wenigsten Simulationssysteme direkt in eine Datenbank ausgeben können.

Bei der Aufbereitung und Analyse der Eingangsdaten sind analog zur Auswertung im Simulationssystem folgende Aufgaben zu realisieren:

- die datenbankgerechte Ablage und Indizierung nach wichtigen Suchkriterien,
- Konvertierung und mathematische, statistische Aufbereitung der Rohdaten,
- die Verdichtung von Einzelwerten zu Trendkurven und statistisch relevanten Aussagen,
- die Ermittlung von Korrelationen und Abhängigkeiten zwischen einzelnen Werten und Modellobjekten.

Verläßt man mit dieser Anforderungsdefinition den Simulationsbereich, stößt man unweigerlich auf die gegenwärtig stark expandierenden Methoden und Systeme des Data Warehousing, OLAP (Online-Analytical Processing) und Data Mining. Obwohl die Datenmengen bei der Simulationsauswertung um Größenordnungen geringer sind, orientiert der folgende Ansatz auf Grundprinzipien und Architekturen des Data Warehouse und Data Mining-Bereiches.

Im Sinne des Data-Warehouse-Konzeptes stellen die Eingangsdaten operative Daten dar. Zur Gewährleistung schneller Analysen müssen diese Daten verdichtet werden. Bei der Verwendung einer relationalen Datenbank zur Datenablage lassen sich mit den verfügbaren SQL-Gruppierungsbefehlen sehr leicht die entsprechenden Aggregationen als Wertetabellen erzeugen und in die Datenbank zurückschreiben.

3.3 Datenbankbasierte Resultatsanalyse

Bei der Auswertung der Daten sind prinzipiell alle Data-Warehouse-Konzepte anwendbar. So lassen sich aus dem gesamten Datenbestand an operativen und aggregierten Daten, welcher im Data-Warehouse-Bereich als hyperdimensionaler Würfel dargestellt wird (Bild 6), beliebig viele Subansichten generieren (Bild 7).

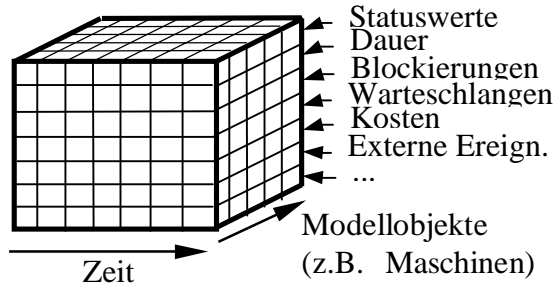


Bild 6: Basisdatenwürfel

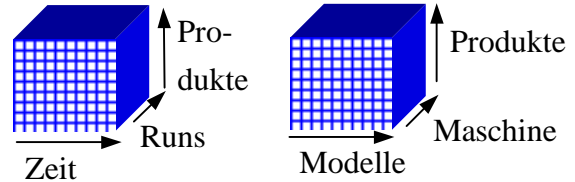


Bild 7: Subansichten

Zusätzlich zur schichtweisen Extraktion der Daten kann auch eine interaktive Verfeinerung des jeweiligen Detaillierungsgrades erfolgen. Diese im Data-Mining als Drilldown bezeichnete Technik ist bei der Simulationsauswertung besonders für den Übergang von allgemeinen Statistiken, wie z.B. mittlerer Warteschlangenlängen, zu konkreten zeitlichen oder stationsbezogenen Einzelverläufen interessant.

Allgemein erlaubt das Datenmodell einen Drilldown von allen Modellen über einzelne Simulationsläufe bis hinunter zu bestimmten Typen oder Einzelinstanzen von Objekten des Modells (Bild 8), welche sowohl stationäre Bedienstationen oder temporäre Objekte, wie z.B. Produkte, darstellen können.

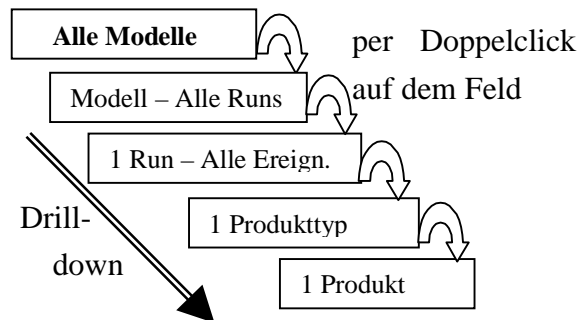


Bild 8: Drilldownebenen

Die Daten jeder Ebene lassen sich unter Nutzung

der Datenbankstandardfunktionen als gruppierte Listen oder auch als Grafiken ausgeben.

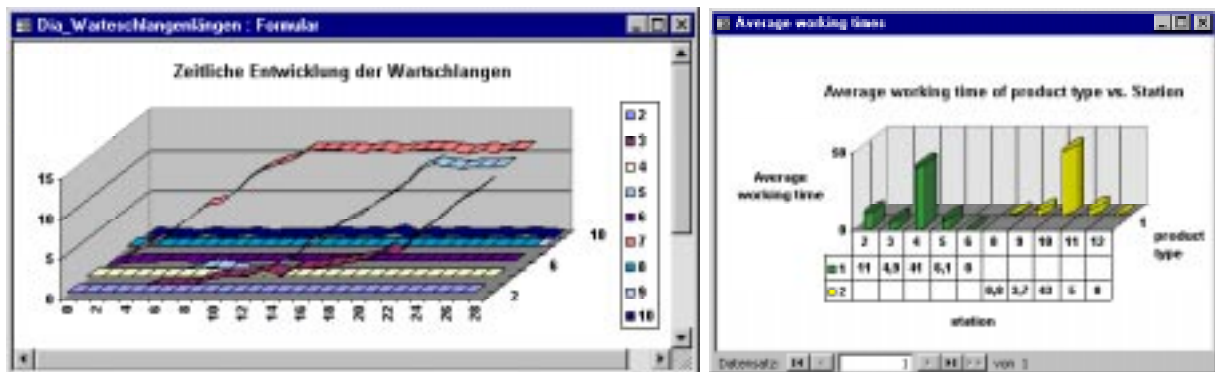


Bild 9: Beispiele zur datenbankbasierten Ergebnispräsentation

4 Bewertung der vorgestellten Vorgehensweisen

Verglichen mit dem traditionellen, meist auf einzelnen Ausdrucken bestehenden Vergleich von Simulationsergebnissen und -modellen ergeben die beiden betrachteten Vorgehensweisen folgende Vorteile:

- Nach dem Einlesen der Rohdaten können alle Vergleiche auf einer einheitlichen Basis vorgenommen werden. Alle Ergebnisse können unabhängig von den konkreten Simulationssystemen in gleich strukturierten Diagrammen oder Listen dargestellt werden.
- Im Beispiel der verwendeten Microsoft-Office-Produkte Excel und Access können auch Anwender ohne größere Programmierkenntnisse mit Hilfe von Assistenten weitere Ergebnispräsentationen oder weiterführende Auswertungen generieren.

Neben diesen allgemeinen Vorteilen sind auch einige Unterschiede bei der Anwendung relevant:

- Im vorgestellten Analysetool SIDAT werden hauptsächlich die Standardausgaben der Simulationssysteme für die Analyse verwendet. Dies erzeugt eine geringe Menge an zu verarbeitenden Rohdaten und erlaubt dadurch sehr schnelle und übersichtliche Modell- und Ergebnisvergleiche.
- Bei der vollständig externen, datenbankbasierten Auswertung kommt es durch das Einlesen der umfangreichen Ereignisinformationen aus den Simulations-Tracefiles zu einem relativ großen Datenvolumen. Als Nachteil ergeben sich dadurch höhere Hardwareanforderungen und im Vergleich zu SIDAT längere Laufzeiten bei der Rohdatenübernahme und -aufbereitung. Kompensiert wird dies jedoch durch die Verwendung identischer Kalkulationsvorschriften zur Erzeugung aller Statistiken und die Verfügbarkeit aller zugrundeliegenden Rohdaten, wodurch tiefergehende Detailanalysen des Verhaltens einzelner Modellobjekte (vgl. Bild 9) möglich sind.

Falls eine Auswertung auf dem Niveau allgemeiner Statistiken ausreichend ist und die notwendigen Rohdaten in vergleichbarer Form von allen verwendeten Systemen geliefert werden, können mit einem Tool wie SIDAT sehr einfach und schnell Vergleiche generiert werden. Bei hohen Anforderungen an die Vergleichbarkeit der Ergebnisse oder einer Weiterverwendung der Simulationsergebnisse in einem Fertigungsleitstand, welcher konkrete Daten zu einzelnen Aufträgen benötigt, ist der datenbankbasierte Ansatz günstiger und aussagekräftiger. Aufgrund der Ähnlichkeit beider Vorgehensweisen ist auch eine Kombination beider Systeme denkbar, welche wie SIDAT auf bereits vorliegenden Statistiken und zusätzlich auch auf selbst kalkulierten Werten arbeiten kann.

Literatur

- [1] M. Ritzschke. Vergleich von Simulationsergebnissen verschiedener diskreter Simulatoren anhand fertigungsorganisatorischer Referenzmodelle, *Tagungsband zum 10. Symposium Simulationstechnik Journal*, Dresden, 1996, S.89-94
- [2] J. Tolujew, G. Neumann. Variable Tracefiles für nutzerorientierte Analysen in der Materialflußsimulation, *Proceedings der Tagung „Simulation und Visualisierung '98“*, SCS-Europe BVBA, Ghent, Belgium, 1998, S. 322-333
- [3] M. Ritzschke. Entwicklung von Benchmarktests zum Vergleich diskreter Simulatoren, *Tagungsband zum 9. Symposium Simulationstechnik*, Stuttgart, 1994, S. 217-222
- [4] T. Wiedemann. Data Mining im Simulationsdatenschwungel - eine neue Technologie mit dem M/SQL-Ansatz, *Proceedings der Tagung „Simulation und Visualisierung '98“*, SCS-Europe BVBA, Ghent, Belgium, 1998, S. 334-341