Simulation in Produktion
und Logistik 2017
Sigrid Wenzel & Tim Peter (Hrsg.)
kassel university press, Kassel 2017

Data Farming im Kontext von Produktion und Logistik

Data Farming for Production and Logistics

Niclas Feldkamp, Sören Bergmann, Steffen Straßburger, TU Ilmenau, Ilmenau (Germany), niclas.feldkamp@tu-ilmenau.de, soeren.bergmann@tu-ilmenau.de, steffen.strassburger@tu-ilmenau.de

Thomas Schulze, Universität Magdeburg, Magdeburg (Germany), thomas.schulze@ovgu.de

Abstract: Simulation is an established methodology for planning and evaluating manufacturing and logistics systems. Usually simulations experts conduct experiments for a prior defined goal and by minimizing the number of simulation runs. In contrast to that, data farming describes an approach for using the simulation model as a data generator for broad scale experimentation with a broader coverage of system behaviour. This paper demonstrates how to apply data farming methodologies on simulation models in the context of production and logistics and how to analyse massive amounts of simulation data using data mining and visual analytics.

1 Einführung

Simulation ist ein etabliertes Werkzeug zur Planung und Steuerung komplexer Produktions- und Logistiksysteme. Insbesondere ist Simulation eine Schlüsselkomponente auch hinsichtlich der Herausforderungen im Kontext von Industrie 4.0 (Krückhans und Meier 2013). Der Ablauf traditioneller Simulationsstudien ist hierbei üblicherweise darauf ausgelegt, einen vorher definierten Projektumfang bzw. ein konkretes (Projekt-) Ziel zu erfüllen, z. B. die Optimierung eines Fertigungslayouts (Law 2003; Sargent 2013). Zur Zielerreichung wird daher in der praktischen Anwendung meist durch manuelles Experimentieren oder durch die Nutzung von simulationsbasierten Optimierungsverfahren nach guten Lösungen gesucht. In jedem Fall wird hierbei versucht, die Anzahl der durchzuführenden Experimente möglichst gering zu halten (Law 2014).

Mit steigender Rechenleistung und der allgemeinen Verfügbarkeit von Big-Data-Infrastrukturen sowie Fortschritten auf dem Gebiet des Data Mining erwächst hier allerdings eine weitere Anwendungsmöglichkeit für Simulationsmodelle: Das Durchführen einer sehr großen Bandbreite von Experimenten, um versteckte, vorher unbekannte und möglicherweise nützliche Wirkzusammenhänge aufzudecken.

Gerade in komplexen Systemen können möglicherweise Beziehungen, Probleme oder sogar Lösungen vorhanden sein, die über das definierte Projektziel einer klassischen Simulationsstudie hinausgehen, aber dennoch zur Entscheidungsunterstützung beitragen können. Das Vorhandensein solcher Zusammenhänge kann bereits durch einfache Sensitivitätsanalysen gezeigt werden (Rabe et al. 2008). Grundlage bildet die Methodik des *Data Farming* (Horne und Meyer 2010; Kallfass und Schlaak 2012). Hierbei wird das vorab validierte Simulationsmodell als Datengenerator genutzt, um mit Hilfe von intelligentem Experimentdesign ein möglichst großes oder gar vollständiges Spektrum an Modell- bzw. Systemverhalten (Wirkungsraum) abdecken zu können (Sanchez 2014; Sanchez und Wan 2009; Kleijnen et al. 2005). Auf der erzeugten Datenbasis können dann mit Hilfe von Data-Mining-Methoden interessante Zusammenhänge in den erzeugen Daten aufgedeckt werden (Fayyad et al. 1996). Ein Vorgehensmodell für eine solche Wissensentdeckung in Simulationsmodellen wurde in vorrausgegangen Publikationen bereits entwickelt (Feldkamp et al. 2015a).

Ziel dieses Beitrags ist die Vertiefung der Methodik von Data Farming und Wissensentdeckung speziell für Simulationsmodelle im Kontext von Produktion und Logistik. Aufbauend auf den bisherigen Veröffentlichungen werden die Besonderheiten von Produktion- und Logistiksystemen in das Zentrum der Betrachtung gestellt. Hierzu wird gezeigt, welche Fragestellungen und Probleme relevant sind und was die zuvor beschriebene Methodik leisten kann. Zudem wird dargestellt, wie dies konzeptionell und technisch umgesetzt werden kann.

Der Aufbau des Beitrags ist hierbei wie folgt: Nach der Einleitung folgt ein kurzer Forschungsüberblick zum Thema Data Farming und daran anknüpfender, relevanter Themengebiete. Danach werden die speziellen Herausforderungen und Fragestellungen eines Simulationsmodells im Kontext von Produktion und Logistik herausgearbeitet sowie Anknüpfungspunkte und Möglichkeiten der hier beschriebenen Methodik erarbeitet. Hieraus wird eine Vorgehensweise entwickelt und mit Hilfe einer Fallstudie veranschaulicht und validiert. Ein Kapitel mit Fazit und Ausblick schließt den Beitrag ab.

2 Grundlagen

Data Farming beschreibt eine Methodik für die Verwendung eines Simulationsmodells als Datengenerator mit Hilfe von effizientem Experimentdesign und High Performance Computing mit dem Ziel, das möglichst vollständige Spektrum von Ergebnisdaten zu erhalten und somit den Informationsgewinn zu verbessern (Sanchez 2007; Elmegreen et al. 2014; Horne und Meyer 2005; Horne und Meyer 2010). Ähnlich wie ein Farmer, der sein Land möglichst effizient kultiviert und damit versucht, seinen Ernteertrag zu maximieren, drückt die „Farming“-Metapher aus, dass hierbei der Datenertrag durch das Simulationsmodell maximiert wird (Sanchez 2014). Neue Ansätze für die Gestaltung von Simulationsexperimenten erlauben die umfassende Abbildung möglicher Wertekombinationen von Eingabeparametern bei gleichzeitig vertretbaren Datenmengen (Kleijnen et al. 2005; Sanchez und Wan 2009). Data Farming wurde ursprünglich für militärische Gefechtssimulationen entwickelt, da hierbei sowohl die Modelle an sich als auch die daran gestellten Fragestellungen eine hohe Komplexität aufweisen, der klassische Simulationsstudien nicht mehr Rechnung tragen konnten (Horne und Meyer 2010). Aufbauend auf dem Data-Farming-Konzept wurde eine Methodik für das Auffinden von verstecken, potenziell

nützlichen Wirkzusammenhängen in Ergebnisdaten von nicht-militärischen Simulationsmodellen entwickelt (Feldkamp et al. 2015a; Feldkamp et al. 2016b). Hierbei werden sehr große Mengen von Simulationsdaten mit Hilfe von Data Farming erzeugt und mittels Data Mining verarbeitet. Die eigentliche Analyse der generierten Simulationsergebnisdaten sowie deren Beziehungen zu den Eingangsparametern (Faktoren) ist auf interaktiver, visueller Analyse aufgebaut. Visualisierung ist im allgemeinen ein wichtiges Werkzeug, wenn eine Interpretation von Daten gefordert ist (Thomas und Cook 2005). Übliche Visualisierungstechniken, die bei Simulationsstudien eingesetzt werden, umfassen Animationen, Time-Plots oder Graphen bestimmter Kenngrößen z B. in einem Konfidenzintervall über replizierte Simulationsläufe (Law 2014). Die Forschungsdisziplin *Visual Analytics* geht allerdings noch einen Schritt weiter und strebt nach der konsequenten Verzahnung von Datenanalyse und -Visualisierung, deren Verbindungsglied die menschliche Fähigkeit zur Schlussfolgerung darstellt. Gefördert wird dies durch ein hohes Maß an Interaktivität (Keim et al. 2008; Thomas und Cook 2005). Dieses Konzept wurde in den Ansatz zur Wissensentdeckung in Simulationsdaten integriert (Feldkamp et al. 2015b; Feldkamp et al. 2016a) und in Industriefallstudien erprobt (Feldkamp et al. 2016b). Abbildung 1 zeigt das schematische Vorgehensmodell für die Wissensentdeckung in Simulationsdaten. Durch Methoden des Data Farming wird eine große Datenmenge erzeugt, um das Systemverhalten in Gänze abzubilden. Die eigentliche Wissensentdeckung erfolgt dann durch visuell gestützte Analysen, insbesondere der Beziehungen von Eingangs- zu Ausgangsdaten. Dazu werden die Daten vorher durch Data-Mining-Algorithmen verarbeitet, beispielsweise durch eine Klassifizierung mittels Clustering-Verfahren.

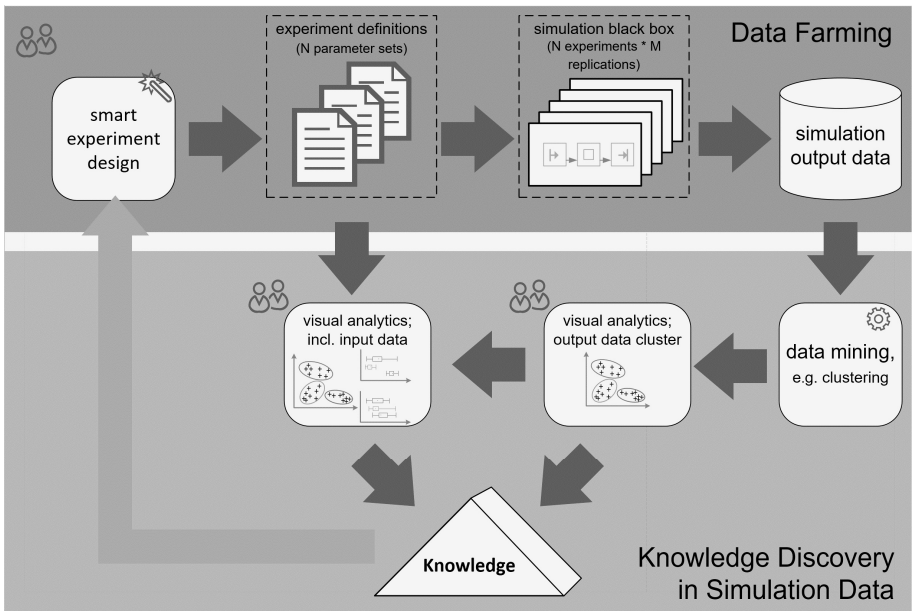


Abbildung 1: *Wissensentdeckung in Simulationsdaten (Feldkamp et al. 2015a, S. 5).*

3 Data Farming in Produktion und Logistik

Data Farming stellt die Untersuchung der Beziehungen und Zusammenhänge zwischen Eingabeparametern und Ergebnisdaten, sowie den Ergebnisdaten an sich in den Vordergrund. Diese Methodik wurde, wie oben bereits erwähnt, im militärischen Kontext entwickelt und hauptsächlich für komplexe, agentenbasierte Gefechtssimulation eingesetzt. Wie in unseren vorherigen, darauf aufbauenden Publikationen bereits gezeigt, lässt sich die Data-Farming-Methodik auch sehr gut in diskret ereignisgesteuerten Modellen anderer Anwendungsdomänen einsetzen. Abgeleitet aus den Erfahrungen bisheriger Data-Farming-Projekte und Publikationen haben wir Fragestellungen für den Anwendungskontext von Simulationsmodellen für Produktion und Logistik entwickelt. Typische Leitfragen, die hierbei mit Data Farming bearbeitet werden können, sind u. a.:

- Welche Eingabeparameter haben den größten Einfluss die Ergebnisdaten?
- Gibt es signifikante Wechselwirkungen zwischen den Eingabeparametern?
- Wie verteilen sich die Ergebnisdaten?
- Welche Auftragsportfolios/Produktmixe sind kritisch hinsichtlich bestimmter Zielgrößen?
- Welche Kombinationen von verschiedenen Eingabedaten (Konfigurationen) sind robust gegenüber Produktmix/-sequenz?
- Welche Konfigurationen sind robust gegenüber nicht beeinflussbaren Störgrößen?
- Welche Eingabeparameter haben einen (positiven oder negativen) Einfluss auf die Robustheit von Konfigurationen?
- Welche Ergebnisgrößen sind empfindlich gegenüber Zufallseinflüssen und Schwankungen?

Entsprechend dem obigen Vorgehensmodell zur Wissensentdeckung in Simulationsdaten werden im Folgenden beispielhaft einige wenige ausgewählte Leitfragen anhand einer Fallstudie bearbeitet.

4 Fallstudie

Für eine Demonstration der Vorgehensweise wird ein Modell einer Fertigungslinie in Siemens Plant Simulation implementiert. Dieses Modell wird mit dem Ziel entwickelt, leicht verständliche Prozesse darzustellen, aber dennoch eine ausreichende Komplexität für die Durchführung eines Data-Farming-Projektes aufzuweisen. Abbildung 2 zeigt ein 2D- und 3D-Layout dieses Modells.

In diesem Modell werden fünf verschiedene Produkttypen auf Werkstückträger geladen und auf einer Förderstrecke transportiert. Der anteilige Mix von Produkttypen kann variieren, wobei ein eingehender Auftrag in einem Eingangspuffer gelagert und ggf. umsortiert werden kann, bis die Freigabe erfolgt und der Auftrag auf einen freien Werkstückträger montiert werden kann. Produkte werden sowohl maschinell an Fertigungsstationen, als auch manuell auf Arbeitsplätzen bearbeitet. Abschließend werden bearbeitete Produkte einer Qualitätsprüfung unterzogen, bevor sie abgeladen werden und das System verlassen.

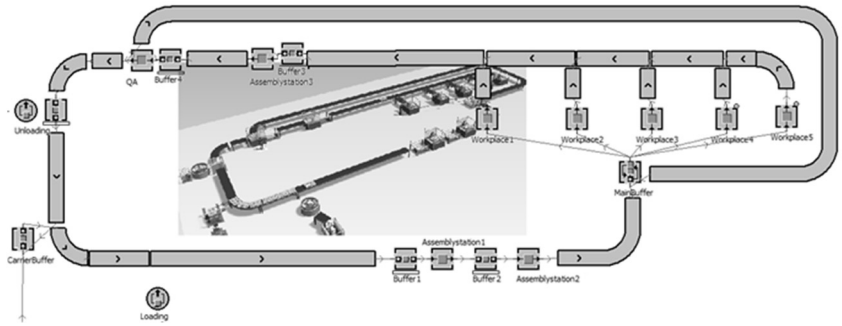


Abbildung 2: *2D- und 3D-Darstellung des Plant Simulation Modells.*

Der erste Schritt für ein Data-Farming-Projekt (nach Modellimplementierung und -validierung) besteht im Definieren von Eingangsparametern (Faktoren) und das Erstellen eines Experimentplans. Faktoren lassen sich in verschiedene Kategorien unterteilen, wie Entscheidungsfaktoren, die sich im realen System tatsächlich beeinflussen lassen und Störfaktoren, welche unkontrollierbar sind. Auch mehr als zwei Kategorien sind denkbar, die tatsächliche Ausgestaltung hängt hierbei vom jeweiligen Modell und Anwendungskontext ab. Im vorliegenden Modell wurde zwischen Entscheidungsfaktoren und Faktoren für den Produktmix unterschieden, d. h. jeweils ein Faktor für den prozentualen Anteil eines bestimmten Produkttyps (Tabelle 1).

Tabelle 1: *Übersicht über die Entscheidungsfaktoren des Modells.*

Name des Faktors	**Beschreibung**	**Wertebereich**
LoadingTime	Zeit für das Aufladen auf Werkstückträger	10 – 60 s
UnloadingTime	Zeit für das Abladen vom Werkstückträger	10 – 60 s
ArrivalTime	Zeitabstand für Auftragsfreigabe	100 – 300 s
ClearanceStrategy	Sortierstrategie für Aufträge (FIFO, Losgröße von 5/10/unbegrenzt)	1 - 4
BufferXCap	Puffergröße (ein Faktor je Puffer)	1 - 100
#Workplaces	Anzahl Arbeitsplätze	1 - 4
#Carriers	Anzahl Werkstückträger im System	1 - 100
WP_ProcTimeVar	Erlaubte Streuung der Bearbeitungszeit	100 – 300 s
QA_ProcTimeMean	Mittlere Prozesszeit der QA-Station	100 – 300 s

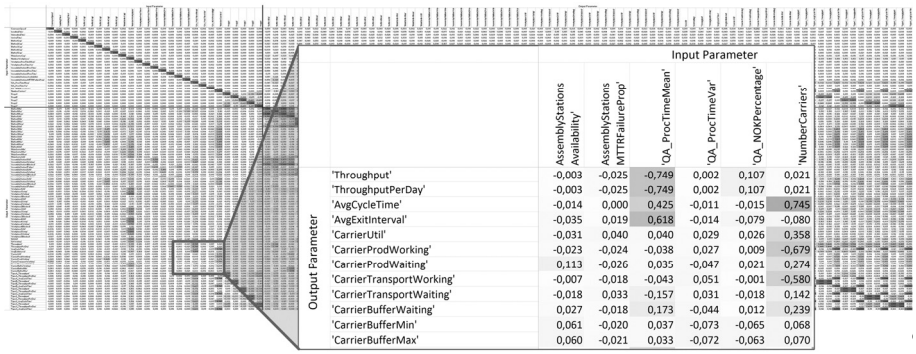
Für beide Faktorkategorien, Entscheidungsfaktoren und Produktmix, werden jeweils voneinander unabhängige Experimentpläne basierend auf der *Nearly-Orthogonal-Latin-Hypercube-Methode* (NOLH) erstellt und miteinander gekreuzt. Dieses Vorgehen hat den Vorteil, dass sich Effekte in Entscheidungsfaktoren jeweils über die komplette Bandbreite der Produktmixausprägungen untersuchen lassen. Ist ein bestimmter Effekt signifikant über alle Experimente, kann diese daraus abgeleitete Erkenntnis als robust gegenüber Schwankungen im Produktmix angesehen werden.

Robustheit ist mitunter ein sehr wichtiges, aber schwer zu bestimmendes Kriterium bei der Planung von Produktions- und Logistiksystemen. Hieraus resultierten für das finale Experimentdesign ca. 260.000 Simulationsläufe, welche auf einem Rechencluster mit zehn Maschinen parallel ausgeführt wurden.

Die Simulationsergebnisse wurden in CSV-Dateien geschrieben und nach Durchführung aller Experimente aggregiert. Für die Analyse und Visualisierung der Daten wurden hauptsächlich MATLAB und R verwendet. Im Folgenden sollen nun beispielhaft drei typische Leitfragen (siehe Abschnitt 3) besprochen werden:

Frage 1: Welche Eingabeparameter haben den größten Einfluss auf die Ergebnisdaten?

Um ein allgemeines Verständnis und eine erste Einschätzung für die Beziehungen zwischen Faktoren und Ergebnisdaten im Modell zu erhalten, ist das Erstellen einer Korrelationsmatrix über alle Parameter hilfreich. Diese wird in Abbildung 3 gezeigt. Die farbliche Kodierung zeigt hierbei jeweils positive oder negative Korrelationsbeziehungen zwischen zwei Parametern an.



Input Parameter

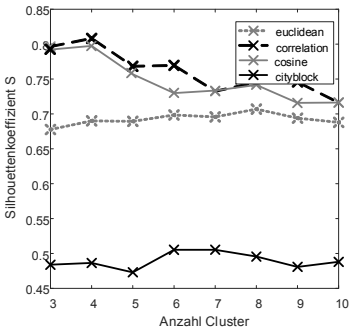
Output Parameter	AssemblyStations Availability'	AssemblyStations MTTRFailureProp'	'QA_ProcTimeMean'	'QA_ProcTimeVar'	'QA_NOKPercentage'	'NumberCarriers'
'Throughput'	-0,003	-0,025	-0,749	0,002	0,107	0,021
'ThroughputPerDay'	-0,003	-0,025	-0,749	0,002	0,107	0,021
'AvgCycleTime'	-0,014	0,000	0,425	-0,011	-0,015	0,745
'AvgExitInterval'	-0,035	0,019	0,618	-0,014	-0,079	-0,080
'CarrierUtil'	-0,031	0,040	0,040	0,029	0,026	0,358
'CarrierProdWorking'	-0,023	-0,024	-0,038	0,027	0,009	-0,679
'CarrierProdWaiting'	0,113	-0,026	0,035	-0,047	0,021	0,274
'CarrierTransportWorking'	-0,007	-0,018	-0,043	0,051	-0,001	-0,580
'CarrierTransportWaiting'	-0,018	0,033	-0,157	0,031	-0,018	0,142
'CarrierBufferWaiting'	0,027	-0,018	0,173	-0,044	0,012	0,239
'CarrierBufferMin'	0,061	-0,020	0,037	-0,073	-0,065	0,068
'CarrierBufferMax'	0,060	-0,021	0,033	-0,072	-0,063	0,070

Abbildung 3: *Korrelationsmatrix über alle Faktoren- und Ergebnisparameter.*

Der schwarze Trennstrich markiert die Aufteilung zwischen Faktoren und Ergebnisparametern. Faktoren sollten untereinander generell unkorreliert sein, da sonst eventuelle Effekte nicht verzerrungsfrei analysiert werden können. Eine gute Experimentdesignmethode sollte dies gewährleisten. Lediglich die Faktoren für den Produktmix korrelieren miteinander, da diese per Definition in Summe immer 100% ergeben müssen. Betrachtet man die Korrelationen zwischen Faktoren und Ergebnisdaten, werden einige Einflussbeziehungen offenkundig. Zwar ist eine Korrelation eigentlich ungerichtet, in einem Simulationsmodell kann aber ein kausaler Zusammenhang zwischen Faktoren und Ergebnisdaten unterstellt werden. So hat im Beispielmodell etwa der Faktor Prozesszeit der Quality-Assurance-Station (QA) einen stark negativen Einfluss auf den Durchsatz. Für eine erste Einschätzung des Modellverhaltens ist eine Korrelationsmatrix gut geeignet. Allerdings ist zu beachten, dass insbesondere Interaktionseffekte, das heißt zwei oder mehr Faktoren, die erst in einer bestimmten Kombination einen signifikanten Effekt auf einen Ergebnisparameter ausüben, hierbei nicht ersichtlich werden.

Frage 2: Wie verteilen sich die Ergebnisdaten?

Um mögliche Strukturen in den Ergebnisdaten zu entdecken und diese zu klassifizieren, wird ein Clustering-Algorithmus verwendet. Dieser gruppiert die Datensätze der Simulationsexperimente in Ähnlichkeitsgruppen (Cluster) anhand bestimmter Dimensionen ein. Werden die Simulationsexperimente anhand von Ergebnisparametern gruppiert, lassen sich die Experimente folglich in verschiedene Performancegruppen einteilen. Als Maßzahl für die Güte eines Clusterings lässt sich der sogenannte Silhouettenkoeffizient heranziehen. Abbildung 4 zeigt diesen Wert jeweils für verschiedene Clusteranzahlen und Distanzmaße. Das Distanzmaß ist eine algorithmische Berechnungsvorschrift für die Bewertung der Ähnlichkeit von Datenpunkten.



Strukturierung	Silhouettenkoeff
stark	0,75 < S ≤ 1,00
mittel	0,50 < S ≤ 0,75
schwach	0,25 < S ≤ 0,50
keine Struktur	0,00 < S ≤ 0,25

Abbildung 4: *Silhouettenkoeffizient für verschiedene Clusteranzahlen und Distanzmaße.*

Das beste Clustering erhält man folglich durch die Wahl von vier Clustern und dem „Correlation"- Distanzmaß. Das daraus resultierende, finale Clustering ist in Abbildung 5 dargestellt.

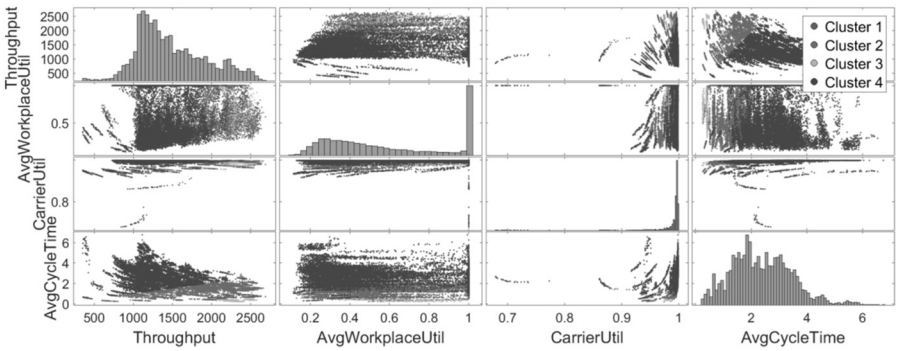


Abbildung 5: *Matrixplot für relevante Ergebnisparameter.*

Die hierfür herangezogenen Ergebnisparameter sind Durchsatz, durchschnittliche Auslastung der Arbeitsplätze und Werkstückträger sowie die durchschnittliche

Durchlaufzeit. Dargestellt sind jeweils 2D-Streudiagramme in sämtlichen Kombinationen über alle vier Pararemter, gefärbt nach der jeweiligen Clusterzuordnung. Diagonal finden sich zusätzlich jeweils Histogramme, um den allgemeinen Wertebereich und die Häufigkeit der Parameterwerte einschätzen zu können. Cluster 3 beinhaltet die Simulationsexperimente mit der besten Performance, d h. hoher Durchsatz, hohe Auslastung von Arbeitsplätzen und Wertstückträgern sowie eine geringe Durchlaufzeit. Im nächsten Schritt sollen nun analysiert werden, wie sich die Parameterwerte der einzelnen Faktoren in den jeweiligen Clustern, insbesondere Cluster 3 verteilen.

Frage 3: Gibt es signifikante Wechselwirkungen zwischen den Faktoren?

Wenn ein bestimmter Faktorwert in einem Cluster dominant ist, kann man davon ausgehen, dass dieser Wert maßgeblich für die Clusterzuordnung ist und somit einen Einfluss auf die korrespondieren Ergebnisgrößen hat. Auf der anderen Seite hat ein Faktor keinen Einfluss, wenn er innerhalb eines Clusters nahezu gleich verteilt ist. Sind mehrere Faktorwerte dominant, ist davon auszugehen, dass die bestimmte Kombination dieser Werte maßgeblich ist und möglicherweise ein Interaktionseffekt besteht. Abbildung 6 (oben) zeigt sogenannte Spinnennetzdiagramme für ausgewählte Faktoren jeweils für die vier Cluster. Die gepunkteten Linien zeigen hierbei jeweils das obere und untere Quartil der Daten an, die durchgezogene Linie den Median. Je enger die Linien bei einander liegen, desto dominanter ist ein bestimmter Faktorwert innerhalb des Clusters. Der Zielcluster 3 hat jeweils in allen fünf Dimensionen sehr dominante Faktorausprägungen. Für eine bessere Ablesbarkeit der konkreten Werte wurden für Cluster 3 zusätzlich je ein Box-Whisker-Plot erstellt (Abb. 6 unten).

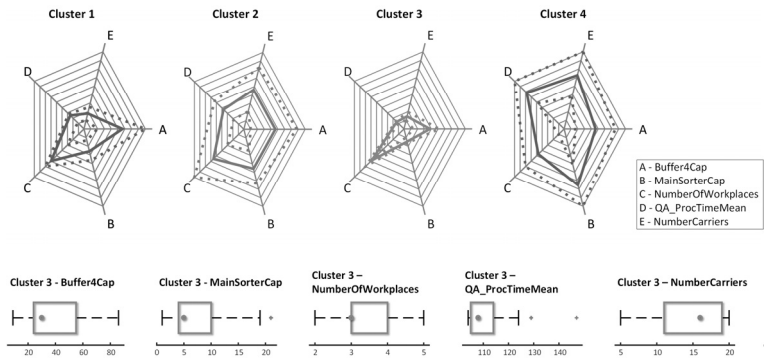


Abbildung 6: *Spinnennetzdiagramme und Box-Whisker-Plots für ausgesuchte Inputparameter.*

Um eine Zuordnung zu Cluster 3 und damit die beste Systemperformance im vorliegenden Beispielmodell zu erhalten, sollten somit u. a. drei Arbeitsplätze bereitstehen, die Prozesszeit der QA-Station zwischen 110 und 120 Sekunden und die Anzahl der Werkstückträger bei etwa 15 liegen. Weiter ist zu erwähnen, dass die gewonnenen Erkenntnisse durch das gekreuzte Experimentdesign für alle Variationen des Produktmixes gültig sind und somit als robust betrachtet werden können. Die konkreten Parameterwerte sollten jedoch durch weitere Experimente validiert werden.

5 Fazit und Ausblick

Der Beitrag diskutiert die Einsatzmöglichkeiten des Ansatzes zur Wissensentdeckung in Simulationsdaten speziell im Kontext von Produktions- und Logistiksystemen. Hierzu werden die Besonderheiten solcher Systeme betrachtet, insbesondere wurden typische, durch den Ansatz bearbeitbare Fragestellungen aufgezeigt. Zu Verdeutlichung werden beispielhaft einige typische Fragestellungen in Form einer Fallstudie aufgegriffen und bearbeitet. Es kann gezeigt werden, dass Wirkzusammenhänge aufgedeckt und potenziell nützliches Wissen generiert werden kann. Da trotzdem per definitionem vom Vorhandensein valider Modelle ausgegangen wird, empfiehlt es sich, die Ergebnisse, wie bei allen simulationsbasierter Methoden, kritisch auf Plausibilität zu prüfen. Gerade komplexe, teilautomatisierte Methoden wie die des Data Mining und Visual Analytics können im Extremfall Effekte aufdecken, die ausschließlich auf den enthaltenen Modellfehlern basieren. Weiterer Entwicklungsbedarf besteht bei den genutzten Werkzeugen, um auch Nicht-Experten die Nutzung des Ansatzes zu ermöglichen. Hierbei sind sowohl die Integration der benötigten Methoden in die Simulatoren als auch die interaktive Visualisierung großer Datenmengen eine Herausforderung für die nahe Zukunft. Des Weiteren bestehen diverse Forschungsbedarfe, z. B. bzgl. des Experimentdesigns der Produktmixe unter Beachtung der Qualitätskriterien, wie der Orthogonalität, Ausgeglichenheit, sowie der Anzahl der benötigten Experimente. Weitere Forschungen zu den Themen Robustheit oder dynamisches Experimentdesign sind ebenfalls denkbar.

Literatur

Elmegreen, B.G.; Sanchez, S.M.; Szalay, A.S.: The Future of Computerized Decision Making. In: Tolk, A.; Diallo, S.D.; Ryzhov, I.O.; Yilmaz, L.; Buckley, S.; Miller, J.A. (Hrsg.): Proceedings of the 2014 Winter Simulation Conference, Savannah GA. Piscataway, N.J.: IEEE Inc, 2014, S. 943–949.

Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P.: From Data Mining to Knowledge Discovery in Databases. AI Magazine 17 (1996), S. 37–54.

Feldkamp, N.; Bergmann, S.; Strassburger, S.: Knowledge Discovery in Manufacturing Simulations. In: Proceedings of the 3rd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation. New York, NY, USA: ACM, 2015a, S. 3–12.

Feldkamp, N.; Bergmann, S.; Strassburger, S.: Visual Analytics of Manufacturing Simulation Data. In: Yilmaz, L.; Chan, W.K.V.; Moon, I.; Roeder, T.M.; Macal, C.; Rossetti, M.D. (Hrsg.): Proceedings of the 2015b Winter Simulation Conference, Huntington Beach. Piscataway, N.J.: IEEE Inc, 2015b, S. 779–790.

Feldkamp, N.; Bergmann, S.; Strassburger, S.: Innovative Analyse- und Visualisierungsmethoden für Simulationsdaten. In: Nissen, V.; Stelzer, D.; Straßburger, S.; Fischer, D. (Hrsg.): Multikonferenz Wirtschaftsinformatik (MKWI) 2016. Ilmenau: TU Ilmenau Universitätsbibliothek 2016, S. 1737–1748.

Feldkamp, N.; Bergmann, S.; Strassburger, S.; Schulze, T.: Knowledge Discovery in Simulation Data: A Case Study of a Gold Mining Facility. In: Roeder, T.M.; Frazier, P.I.; Szechtman, R.; Zhou, E.; Huschka, T.; Chick, S.E. (Hrsg.): Proceedings of the 2016 Winter Simulation Conference, Washington, DC, USA. Piscataway, N.J.: IEEE Inc, 2016b, S. 1607–1618.

Horne, G.E.; Meyer, T.: Data farming and defense applications. In: Armstrong, R.; McNamara, J.; Pinelli, T.E. (Hrsg.): MODSIM World Conference and Expo, Hampton, VA, USA. Hampton, VA: Langley Research Center, 2010, S. 74–82.

Horne, G.E.; Meyer, T.E.: Data Farming: Discovering Surprise. In: Kuhl, M.E.; Steiger, N.M.; Armstrong, F.B.; Joines, J.A. (Hrsg.): Proceedings of the 2005 Winter Simulation Conference, Orlando, FL. USA. Piscataway, N.J.: IEEE Inc, 2005, S. 1082–1087.

Kallfass, D.; Schlaak, T.: NATO MSG-088 Case Study Results to demonstrate the Benefit of using Data Farming for Military Decision support. In: Laroque, C.; Himmelspach, R.; Pasupathy, R.; Rose, O.; Uhrmacher, A. (Hrsg.): Proceedings of the 2012 Winter Simulation Conference (WSC 2012), Berlin, Germany. Piscataway, N.J.: IEEE Inc, 2012, S. 1–12.

Keim, D.A.; Mansmann, F.; Schneidewind, J.; Thomas, J.; Ziegler, H.: Visual Analytics: Scope and Challenges. In: Simoff, S.; Boehlen, M.H.; Mazeika, A. (Hrsg.): Visual Data Mining: Theory, Techniques and Tools for Visual Analytics: Springer 2008,

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

Krückhans, B.; Meier, H.: Industrie 4.0 – Handlungsfelder der Digitalen Fabrik zur Optimierung der Ressourceneffizienz in der Produktion. In: Dangelmaier, W.; Laroque, C.; Klaas, A. (Hrsg.): Proceeding der 15. ASIM Fachtagung Simulation in Produktion und Logistik 2013, Paderborn. Paderborn: HNI-Verlagsschriftenreihe, 2013, S. 31–40.

Law, A.M.: How to conduct a successful simulation study. In: Chick, S.; Sanchez, P.J.; Ferrin, D.; Morrice, D.J. (Hrsg.): Proceedings of the 2003 Winter Simulation Conference, New Orleans, LA, USA. Piscataway, N.J.: IEEE Inc, 2003, S. 66–70.

Law, A.M.: Simulation Modeling and Analysis. New York, N.Y.: McGraw Hill Book Co 2014.

Rabe, M.; Spieckermann, S.; Wenzel, S.: Verifikation und Validierung für die Simulation in Produktion und Logistik: Vorgehensmodelle und Techniken. Berlin, Heidelberg: Springer 2008.

Sanchez, S.M.: Work Smarter, Not Harder: Guidelines for Designing Simulation Experiments. In: Henderson, S.G.; Biller, B.; Hsieh, M.-H.; Shortle, J.; Tew, J.D.; Barton, R.R. (Hrsg.): Proceedings of the 2007 Winter Simulation Conference. Piscataway, N.J.: IEEE, 2007, S. 84–94.

Sanchez, S.M.: Simulation Experiments: Better Data, Not Just Big Data. In: Tolk, A.; Diallo, S.D.; Ryzhov, I.O.; Yilmaz, L.; Buckley, S.; Miller, J.A. (Hrsg.): Proceedings of the 2014 Winter Simulation Conference, Savannah GA. Piscataway, N.J.: IEEE Inc, 2014, S. 805–816.

Sanchez, S.M.; Wan, H.: Better than a petaflop: The power of efficient experimental design. In: Rossetti, M.D.; Hill, R.R.; Johansson, B.; Dunkin A.; Ingalls, R.G. (Hrsg.): Proceedings of the 2009 Winter Simulation Conference (WSC 2009), Austin, TX. Piscataway, N.J.: IEEE Inc, 2009, S. 60–74.

Sargent, R.G.: Verification and validation of simulation models. Journal of Simulation 7 (2013) 1, S. 12–24.

Thomas, J.J.; Cook, K.A.: Illuminating the Path: Research and Development Agenda for Visual Analytics. Los Alamitos, California: IEEE Computer Society 2005.